

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Государственное образовательное учреждение высшего профессионального образования
УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Н. Г. Ярушкина
Т. В. Афанасьева
И. Г. Перфильева

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Учебное пособие

Ульяновск
2010

УДК 004.8 (075)

ББК 32.813 я7

Я 94

Рецензенты профессор кафедры «Информационные технологии» УлГУ
профессор, д-р техн. наук И. В. Семушин,
кафедра «Телекоммуникационные технологии и сети» УлГУ.

*Утверждено редакционно-издательским советом
Ульяновского государственного технического университета
в качестве учебного пособия*

Ярушкина, Н. Г.

Я 94 Интеллектуальный анализ временных рядов : учебное пособие /
Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. – Ульяновск :
УлГТУ, 2010. – 320 с.

ISBN 978-5-9795-0618-0

Содержание пособия включает изложение основных подходов направления интеллектуального анализа временных рядов (Time Series Data Mining), которое в настоящее время объединяет статистические, нейросетевые и нечеткие модели и технологии анализа временных рядов. Описание моделей и технологий базируется на современном обзоре отечественных и зарубежных источников, системном подходе и сопровождается примерами и контрольными вопросами. Рассматриваются авторские модели временных рядов на основе анализа нечетких тенденций, позволяющие прогнозировать не только числовые значения временных рядов, но и направление их изменения, а также генерировать краткое описание поведения временного ряда в лингвистической форме.

Пособие предназначено для поддержки дисциплин «Интеллектуальные информационные технологии», «Информационные технологии», «Информатика», дневной, вечерней, заочной и дистанционной форм обучения, а также для магистрантов и аспирантов, специализирующихся в области интеллектуальной обработки данных.

УДК 004.8 (075)

ББК 32.813 я7

ISBN 978-5-9795-0618-0

© Ярушкина Н. Г., Афанасьева Т. В.,
Перфильева И. Г., 2010
© Оформление. УлГТУ, 2010

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	6
СПИСОК ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ	11
ГЛАВА 1. ВВЕДЕНИЕ В НЕЧЕТКИЕ ВРЕМЕННЫЕ РЯДЫ	13
Введение	13
1.1. Сферы применения нечетких экспертных оценок	14
Контрольные вопросы.....	20
1.2. Виды экспертных оценок	20
Контрольные вопросы.....	26
1.3. Определение нечетких временных рядов	27
Контрольные вопросы.....	31
1.4. Нечеткие временные ряды в системах поддержки принятия управленческих решений.....	31
Контрольные вопросы.....	34
1.5. Нечеткий временной ряд в системах автоматизации проектирования ..	35
Контрольные вопросы.....	40
1.6. Нечеткие временные ряды в задачах экспертной деятельности	41
Контрольные вопросы.....	44
Выводы	44
Библиографический список.....	45
ГЛАВА 2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ В МОДЕЛИРОВАНИИ НЕЧЕТКИХ ВРЕМЕННЫХ РЯДОВ	58
Введение	58
2.1. Нечеткие множества.....	59
Контрольные вопросы.....	64
2.2. Лингвистические переменные	65
Контрольные вопросы.....	67
2.3. Нечеткая логика.....	67
Контрольные вопросы.....	70
2.4. Нечеткие модели и системы.....	71
Контрольные вопросы.....	77
2.5. Обзор практических применений нечетких моделей	77
Контрольные вопросы.....	81
2.6. Современные тенденции развития теории нечетких множеств нечеткой логики и нечетких систем.....	82
Контрольные вопросы.....	95
Выводы	96
Библиографический список.....	98

ГЛАВА 3. ОБЗОР НАПРАВЛЕНИЙ И ПОДХОДОВ В МОДЕЛИРОВАНИИ ВРЕМЕННЫХ РЯДОВ	100
Введение	100
3.1. Статистический подход к моделированию временных рядов	101
Контрольные вопросы	111
Библиографический список	112
3.2. Нейросетевой подход к моделированию временных рядов	113
Контрольные вопросы	122
Библиографический список	122
3.3. Нечеткий подход к моделированию временных рядов	123
Контрольные вопросы	143
Библиографический список	143
3.4. Гибридные модели временных рядов	145
Контрольные вопросы	152
Библиографический список	152
3.5. Интеллектуальный анализ баз данных ВР	154
Контрольные вопросы	161
Выводы	162
Библиографический список	165

ГЛАВА 4. ОСНОВЫ ТЕОРИИ НЕЧЕТКИХ ТЕНДЕНЦИЙ НЕЧЕТКИХ ВРЕМЕННЫХ РЯДОВ	168
Введение	168
4.1. Концептуальная модель ACL-шкалы для генерации нечетких оценок	170
Контрольные вопросы	193
4.2. Классификация нечетких тенденций	194
Контрольные вопросы	202
4.3. Модель элементарной тенденции нечеткого временного ряда	202
Контрольные вопросы	209
4.4. FT-декомпозиция нечеткого временного ряда в базисе элементарных тенденций	209
Контрольные вопросы	210
4.5. Структурно-лингвистическая модель нечеткого временного ряда в базисе элементарной тенденции	210
Контрольные вопросы	217
Библиографический список	217

ГЛАВА 5. СТРУКТУРНО-ЛИНГВИСТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ НЕЧЕТКИХ ТЕНДЕНЦИЙ ВРЕМЕННЫХ РЯДОВ	219
Введение	219
5.1. Структурно-лингвистический подход при анализе временных рядов ..	219
Контрольные вопросы	224
5.2. Сегментация НВР в терминах нечетких тенденций	225
Контрольные вопросы	233

5.3. Кластеризация НВР в терминах нечетких тенденций	234
Контрольные вопросы.....	236
5.4. Частотный анализ	236
Контрольные вопросы.....	238
5.5. Классификация НВР в терминах нечетких тенденций.....	239
Контрольные вопросы.....	246
5.6. Извлечение правил	246
Контрольные вопросы.....	249
5.7. Моделирование ВР на основе нечетких тенденций	250
Контрольные вопросы.....	270
5.8. Поиск аномалий во временных рядах	270
Контрольные вопросы.....	277
5.9. Резюмирование нечеткого временного ряда	278
Контрольные вопросы.....	279
5.10. Принципы построения интеллектуальной системы анализа нечетких тенденций временных рядов.....	279
Контрольные вопросы.....	285
Выводы	285
Библиографический список.....	286
 ГЛАВА 6. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ НЕЧЕТКОГО МОДЕЛИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ И АНАЛИЗА НЕЧЕТКИХ ТЕНДЕНЦИЙ	
Введение	289
6.1. Программная система моделирования временных рядов на основе элементарных нечетких тенденций FuzzyTend	289
6.2. Вычислительный эксперимент	300
Контрольные вопросы.....	313
Выводы	313
Библиографический список.....	314
 ЗАКЛЮЧЕНИЕ	 316
 ГЛОССАРИЙ.....	 317

*...вычисления со словами находятся на авансцене,
тогда как обычные вычисления проходят как бы
за занавесом и скрыты от взгляда пользователя...*

Л. Заде

ВВЕДЕНИЕ

В связи с ростом хранимых, упорядоченных во времени, данных о характеристиках объектов, процессов и систем в промышленности, экономике, медицине, образовании, социологии расширяются возможности анализа и прогнозирования количественных и качественных изменений этих характеристик, а также их носителей. Систематическое и комплексное исследование тенденций развития процессов на основе анализа временных рядов (ВР), извлеченных из хранилищ и баз данных, является сферой профессиональной деятельности специалистов различного профиля: менеджеров среднего звена, аудиторов, специалистов в области контроля качества, экономистов, руководителей, маркетологов, аналитиков и др.

Технологии баз данных создают новые возможности извлечения знаний из ВР баз данных для специалистов при решении задач анализа процессов:

1. по качественной оценке текущего и будущего состояния исследуемого процесса,
2. по обнаружению типичных и аномальных типов событий,
3. по выявлению имевших место качественных изменений.

При этом обнаружение тенденций, их качественная оценка и прогноз на основе временных рядов, извлеченных из баз данных предприятий, выступают как отдельная задача анализа, которая приобретает особую актуальность в связи со стремительным ростом и изменением хранимых данных.

Оперативное, адекватное решение указанных задач обуславливает повышение качества принимаемых решений, определяет эффективность деятельности и конкурентоспособность предприятия.

Вместе с тем, выполнение указанных аналитических процедур сопряжено с определенными трудностями, вызванными неготовностью специалистов различного профиля к решению обозначенных задач из-за недостатка времени для обработки множества гетерогенных временных рядов увеличивающейся размерности, хранимых в базах данных, а также из-за отсутствия соответствующей подготовки в области анализа ВР и способов интерпретации его результатов.

Одним из направлений, обеспечивающим «интеллектуальную» поддержку специалистов по решению новых задач анализа ВР баз данных, является интеллектуальный анализ данных или *Data Mining*, в котором анализ поведения и тенденций развития процессов может быть рассмотрен как интеллектуальный анализ временных рядов или *Time Series Data Mining* (TSDM). Основными целями *Time Series Data Mining* являются, во-первых, анализ и моделирование процессов, характеризующихся высокой степенью неопределенности, в том числе «нестохастического» типа, во-вторых, повышение уровня интеллектуальной поддержки современных специалистов, и, в-третьих, выявление скрытых закономерностей и извлечение новых знаний из временных рядов.

В основе новых методов *Time Series Data Mining* лежит нечеткая модель временного ряда, получившая название нечеткого временного ряда (НВР), построенная с привлечением нечетких экспертных оценок и нечетких систем. Такая модель, принципиально являясь более грубой, тем не менее, позволяет использовать дополнительные предметно-зависимые знания и описывать поведение временного ряда в виде качественных оценок изменений и нечетких тенденций. И в этом смысле один и тот же временной ряд в различных предметных областях будет иметь разные нечеткие модели. При моделировании нечетких временных рядов необходимо определить его носитель, объект исследования и решаемые задачи. Носителем нечеткого временного ряда выступает исходный временной ряд, объектом исследования – модель нечеткого временного ряда, совокупность задач *Data Mining* применительно к нечетким временным рядам включает: сегментацию, кластеризацию, классификацию, индексирование, резюмирование, обнаружение аномалий, частотный анализ, прогнозирование, извлечение ассоциативных правил.

Данное новое направление находится в процессе становления, оно обозначено и в основном развивается в научных трудах иностранных ученых: Х. Танаки, К. Сонга, К. Хироты, Я. Капржика, В. Новака, В. Педрича, И. Перфильевой. Среди отечественных ученых данной тематике в области нечетких моделей временных рядов посвящены исследования И. Батыршина, С. Ковалева, К. Дегтярева.

В связи с этим в учебном пособии рассматриваются теоретические основы, приводится систематизация результатов в области нечеткого моделирования временных рядов, а также представлен авторский подход к решению проблемы извлечения новых знаний о качественных изменениях во временных рядах на основе моделей нечетких тенденций. Результаты решения проблемы нечеткого моделирования временных рядов в учебном пособии представлены в многоуровневом виде, включающем как числовые, так и нечеткие (лингвистические) оценки уровней, а также нечеткие оценки динамики временных рядов в терминах нечетких тенденций. Новые знания, извлеченные из временных рядов в виде совокупности нечетких правил «ЕСЛИ-ТО», с одной стороны, востребованы лицами, принимающими решения в условиях неопределенности, а с другой стороны, могут быть использованы в качестве входных данных в системах поддержки проектных и управленческих решений.

Учебное пособие содержит шесть глав.

Первая глава посвящена введению в нечеткие временные ряды: приводится анализ развития применения интеллектуальных технологий, активно использующих экспертные оценки для моделирования процедур поддержки интеллектуальной деятельности лиц, принимающих решение, проектировщиков, экспертов. Рассматриваются отличия экспертного оценивания от измерения, определяются классы экспертных оценок. Формулируется содержательное определение нечеткого временного ряда как временного ряда, построенного на основе четкого временного ряда, уровни которого допускают интервальную качественную оценку.

Во второй главе рассмотрены теоретические основы нечетких временных рядов: нечеткие множества и нечеткие модели, а также перспективы развития теории нечетких множеств и систем, используемых как формальный базис при моделировании нечетких временных рядов.

В третьей главе приведен обзор и анализ направлений в области методов Time Series Data Minig, технологий и методов моделирования временных рядов: статистический, нейросетевой, нечеткий, гибридный. Рассмотрено развивающееся направление интеллектуального анализа и моделирования временных рядов, хранимых в базах данных. Обозначена проблематика нового направления интеллектуального анализа и сформулированы основные положения методологии нечеткого моделирования временных рядов и анализа нечетких тенденций.

В четвертой главе представлены основы теории нечетких тенденций временных рядов. Для оценки параметров нечеткой тенденций временных рядов введены понятие и параметрическая структурно-функциональная модель нечеткой лингвистической шкалы – ACL-шкалы. На основе введенных теоретических положений разработаны методы и алгоритмы построения и применения ACL-шкалы для генерации временного ряда нечетких уровней и нечетких тенденций. Приведены определение и классы нечетких тенденций. Введены модель нечеткой тенденции нечеткого временного ряда, FT-преобразование временного ряда для генерации временных рядов параметров нечетких тенденций и совокупность проекционных моделей временных рядов, предназначенных для решения задач Time Series Data Minig.

В пятой главе обозначен структурно-лингвистический подход в нечетком моделировании временных рядов в терминах нечетких тенденций и рассмотрено применение разработанного формального аппарата моделей нечетких тенденций для решения комплекса задач интеллектуального анализа временных рядов (Time Series Data Minig): сегментации, классификации, кластеризации, прогнозирования, резюмирования, частотного анализа, поиска аномалий. Предложены новые нечеткие модели временных рядов, позволяющие извлекать до-

полнительные знания о закономерностях, выявляемых в поведении нечетких объектов, а также улучшена алгоритмическая реализация нечеткого логического вывода. Введены дополнительные критерии качества нечеткого моделирования, позволяющие оценивать точность моделирования нечетких объектов временных рядов.

Шестая глава посвящена исследованию продуктивности структурно-лингвистического подхода к нечеткому моделированию и анализу нечетких тенденций временных рядов в авторской программной системе. Показана результативность структурно-лингвистического подхода, приведены примеры моделирования и сравнительные характеристики нечетких моделей структурно-лингвистического подхода по сравнению с моделями нечеткого и статистического подхода к анализу временных рядов.

СПИСОК ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ

ACL-шкала (Absolute&Comparative Linguistic) – специальная лингвистическая шкала.

CICO – «четкий вход – четкий выход» – метод CICO (Crisp-Input and CrispOutput).

CiFo – смешанные данные – метод CIFO (Crisp-Inputs and Fuzzy-Outputs).

CWP – гранулярные вычисления (computing with words and perceptions CWP).

DM – новое направление исследований – Data Mining (DM) баз данных временных рядов.

FARIMA - авторегрессионная модель проинтегрированного скользящего среднего (FARIMA)

FAT – Fuzzy Approximation Theorem, согласно которой любая математическая система может быть аппроксимирована системой, основанной на нечеткой логике.

FiFo – «нечеткий вход – нечеткий выход» – метод FIFO (Fuzzy-Inputs and Fuzzy-Outputs).

FLSRA – подход, комбинированный с методом наименьших квадратов FLSRA (Fuzzy least-square regression analysis).

GCL – язык гранулярных вычислений: Generalized Constraint Language.

GTU – обобщенные ограниченные неопределенности (Generalized Theory of Uncertainty).

IFSA – ученые, занимающиеся нечеткой логикой, объединены в International Fuzzy Systems Association.

NL – критерий приемлемости любой теории неопределенности в эпоху построения гуманистических систем, основанных на знаниях.

TPM – теория уточнения значений Theory of Precisiation of Meaning (TPM).

TSDM – интеллектуальный анализ временных рядов Time Series Data Mining (TSDM).

АБС – система на основе процессора с нечеткой логикой.

АКПП – автоматическая коробка переключения передач.

АРПСС (ARIMA) – модель «авторегрессии проинтегрированного скользящего среднего».

БЗ – база знаний.

ВР – временной ряд.

ВРЭТ – временной ряд элементарных тенденций.

ДВС – двигатель внутреннего сгорания.

ЕЯ – естественный язык.

ИНС – искусственные нейронные сети.

КИИ – конференция по искусственному интеллекту.

ЛНТ – локальные нечеткие тенденции.

ЛО – лингвистическая оценка.

ЛПР – лица, принимающие решение.

МНТ – модель нечеткой тенденции ВР.

НВР – нечеткий временной ряд.

НВТ – нечеткая временная тенденция.

НМ – нечеткие множества.

НС – нейронная сеть.

НСМВ – нечеткие системы и мягкие вычисления.

НТ – нечеткая тенденция.

ПК – программный комплекс.

САПР – системы автоматизированного проектирования.

СКО(MSE) – среднее квадратическое отклонение.

ФП – функция принадлежности.

ЭТ – элементарная тенденция.

ГЛАВА 1. ВВЕДЕНИЕ В НЕЧЕТКИЕ ВРЕМЕННЫЕ РЯДЫ

Введение

В современных условиях учет факторов неопределенности и неполноты информации является неотъемлемой принадлежностью сложных организационно-технических систем, в которых значительная доля качества функционирования процессов зависит от человека. Причем традиционный путь учета факторов неопределенности на основе вероятностного и статистического моделирования зачастую оказывается неадекватным решаемым задачам и может привести к неверным результатам, так как функционирование сложных организационно-технических систем на практике характеризуется неопределенностью «нестохастического» типа вследствие:

- неполноты или отсутствия знаний о поведении отдельных входящих в систему элементов и подсистем, а также взаимосвязей между ними;
- невозможности или ограниченной возможности экспериментального исследования процессов, не позволяющей получить достаточную статистическую информацию о наиболее важных характеристиках системы;
- многоцелевого, многокритериального оценивания и гетерогенности структур данных.

Для моделирования, проектирования и анализа таких организационно-технических систем получили распространение экспертные системы, оперирующие экспертными оценками, основанными на опыте и знаниях эксперта (или группы экспертов). Экспертные оценки представляют качественный аспект оцениваемого элемента системы и самой системы, выражаются в лингвистической форме, поэтому в дальнейшем в учебном пособии будет использоваться наряду с термином «экспертная оценка» термин «лингвистическая оценка» как составная часть экспертной оценки.

1.1. Сферы применения нечетких экспертных оценок

Лингвистические оценки (ЛО) являются средством качественного оценивания и сравнения характеристик элементов систем, используемые проектировщиками, менеджерами, лицами, принимающими решения (ЛПР), экспертами. Важным свойством лингвистических оценок является широкое применение на практике для выражения знаний о степени соответствия элемента системы или его характеристики некоторому объективному или субъективному критерию. Указанное свойство определяет класс абсолютных ЛО, отражающий статический аспект оценивания. Примерами таких оценок могут служить оценочные выражения «Удовлетворительно», «Хорошо», «Плохо». Семантика абсолютных лингвистических оценок зависит от контекста среды, в которой они используются. Другое важное свойство лингвистических оценок обусловлено возможностью их ранжирования, что позволяет представить совокупность ЛО в виде некоторой системы с отношениями. Бинарные отношения, образованные на множестве абсолютных ЛО, порождают сравнительные лингвистические оценки по различным критериям, такие как «Больше», «Меньше», «Примерно Равны», «Раньше», «Позже», «Предпочтительнее», «Лучше» и т. д. Сравнительные оценки, построенные на абсолютных ЛО, могут представлять изменения по различным основаниям: в пространстве объектов, во временном пространстве, в пространстве задач и выражают динамический аспект оценивания. Семантика сравнительных оценок также является контекстно-зависимой.

Совокупность лингвистических оценок для конкретной предметной области сформированная экспертами образует множество экспертных оценок.

Можно отметить в последнее десятилетие рост публикаций в области проектирования формальных моделей экспертных лингвистических оценочных выражений, как части естественного языка, в интеллектуальных моделях и системах принятия решений, управления, проектирования.

Ниже приведем публикации, посвященные представлению, использованию и обработке экспертных оценок, сгруппированных по прикладным облас-

тям. Список публикаций будет основываться на материалах тематических российских конференций 2003-2009 гг., таких как КИИ (Конференция по искусственному интеллекту), НСМВ (Нечеткие системы и мягкие вычисления), IAS (Интеллектуальные системы). В список рассматриваемых публикаций также включены родственные материалы из периодических журналов и учебных пособий.

В экономической сфере:

- ✓ Выбор лизинговой компании на основе нечеткой оценки уровня экономического развития предприятия [Новак и др., 2008; Новак, 2008],
- ✓ Использование Н-моделей для управления и прогноза экономики [Напреенко и др., 2008; Аверкин, 1986],
- ✓ Методы искусственного интеллекта в управлении кадрами предприятия [Затылкин, 2008],
- ✓ Оценка эффективности деятельности предприятия [Ярушкина, 2005; Нуруллин, 2008],
- ✓ Недоопределенное календарное планирование [Гофман и др., 2008],
- ✓ Применение экспертных оценок эффективности предпринимательского проекта [Финаев и др., 2008 б],
- ✓ Оценка инвестиций в жилищное строительство с использованием нечетких множеств [Басаев, 2005],
- ✓ Оптимизация фондового портфеля, содержащего опционы (нечеткая постановка задачи) [Недосекин, 2005],
- ✓ Маркетинговые исследования с применением экспертных оценок [Хархаров, 2006],
- ✓ Антикризисное управление [Чернов и др., 2006].

В системах экологической безопасности:

- ✓ Интеллектуальная система управления пожарной безопасностью хранения фрезерного торфа [Палюх, 2008],
- ✓ Прогнозирование распространения лесных пожаров с применением теории нечетких множеств [Абрахин, 2008 а],

- ✓ Моделирование последствий прорыва плотины на реке с использованием теории нечетких множеств [Абрахин, 2008 б],
- ✓ Управление очисткой сточных вод на основе нечеткого моделирования [Ермоленко, 2008],
- ✓ Экспертная система экологической безопасности на предприятии [Яруш-кина и др., 2008].

В транспортных системах:

- ✓ Экспертная система морского мониторинга для оценки и прогнозирования угроз судам, а также динамики изменения состояния их защищенности [Смагин и др., 2008],
- ✓ Идентификация и прогнозирование состояний сложной системы в условиях неопределенности [Кравченко, 2008],
- ✓ Мониторинг динамического воздействия поездов на земляное полотно на основе нечетких систем и распределенных баз знаний [Шабельников и др., 2009],
- ✓ Принятие решений на основе формализма нечеткой логики в транспортных системах [Белый и др., 2002],
- ✓ Оценка аномалий в почвенном покрове на основе темпоральных нечетких моделей и временных рядов [Долгий и др., 2008],
- ✓ Оперативный интеллектуальный прогноз движения поездов [Дулин и др., 2008].

В технологических и производственных процессах:

- ✓ Использование нейронной сети для принятия решения о проведении ремонта оборудования [Линев др., 2008],
- ✓ Обзор интеллектуальных систем управления технологическими объектами [Кулаков и др., 2008],
- ✓ Обоснование актуальности нечетко-регрессионных моделей для моделирования технологических процессов [Павленко, 2008],

- ✓ Анализ применения нечетких регуляторов в составе гибридных для повышения систем автоматического управления [Финаев и др., 2008 а],
- ✓ Прогнозирование дефектов металлопродукции [Кудинов, 2007; Кудинов, 2008],
- ✓ Диагностика функционирования технологического процесса по нечетким критериям [Клячкин, 2003; Макаров и др., 2007],
- ✓ Интеллектуальная система поддержки принятия решений в сфере потребления тепловой энергии в объектах «Теплисис» (версия 1.0) [Головина, 2005],
- ✓ Перспективы использования нечетких временных рядов в управлении сложными процессами [Афанасьева, 2008 а].

В системах принятия решений:

- ✓ Интеллектуальные технологии принятия решений в условиях неопределенности [Штовба, 2007; Чернов, 2007],
- ✓ Нечеткое моделирование [Батыршин и др., 2007 а; Батыршин и др., 2007 б; Борисов и др., 2007],
- ✓ Нечеткая свертка экспертных оценок [Исмагилов, 2008],
- ✓ Виртуальные интеллектуальные предприятия как коллективные рабочие площадки [Тарасов, 2008],
- ✓ Кластеризация экспертных оценок для многокритериального оценивания [Аверкин, 2008; Вятчинин, 2004],
- ✓ Оценка защищенности программного обеспечения на основе нечеткой модели [Цирлов, 2008].

В управлении качеством образования:

- ✓ Анализ учебной деятельности с использованием нечетких оценок и временных рядов [Валеев и др., 2008],
- ✓ Моделирование рейтинговой оценки знаний в нечетко-нейронной системе [Нурматова, 2009; Домрачев, 2001],

- ✓ Мониторинг процесса качества образования [Афанасьев, 2009],
- ✓ Модели прогнозирования в интеллектуальных организациях [Виноградов и др., 2008; Виноградов и др., 2007].

В системах автоматизированного проектирования:

- ✓ Формализация экспертной деятельности в САПР [Ярушкина, 1997; Ярушкина, 2004],
- ✓ Управление процессом проектирования в нечеткой среде [Грачев, 2008],
- ✓ Использование вопросно-ответной среды для принятия коллективного проектного решения [Касапенко и др., 2008; Соснин, 2007],
- ✓ Мониторинг и анализ в области CAD/CAM/CAE-технологий [Аверченков, 2008],
- ✓ Интеллектуальная поддержка проектирования инноваций [Андрейчиков и др., 2008],
- ✓ Автоматизированное проектирование и моделирование вычислительных сетей [Стецко, 2008],
- ✓ Построение временного графика процесса проектирования на основе нечеткого генетического алгоритма [Гладков и др., 2008],
- ✓ Применение нечетких временных рядов в оценке качества процесса проектирования [Афанасьева и др., 2008 б].

Содержательный анализ представленного перечня публикаций позволяет сделать несколько выводов.

Во-первых, основным математическим аппаратом формализации представления и обработки экспертных оценок и высказываний является теория нечетких множеств [Заде, 1974]. Применение аппарата нечеткого множества – это попытка математической формализации нечетких (экспертных) оценок в виде лингвистически поименованных функций для построения моделей обработки этих оценок как композиции указанных функций, имеющих простую лингвистическую интерпретацию. В результате появляется возможность для конечного пользователя оперировать естественными предметно-ориенти-

рованными лингвистическими термами, представляемыми на уровне компьютерных вычислений в виде чисел. Такой подход дает приближенные, но в то же время качественные способы описания поведения сложных и плохо определенных организационно-технических систем. Теоретические же основания данного подхода вполне точны и строги в математическом смысле и не являются сами по себе источником неопределенности. В каждом конкретном случае степень точности решения может быть согласована с требованиями задачи.

Во-вторых, в слабоструктурированных и плохо формализуемых предметных областях при создании систем обработки данных применяют интеллектуальные технологии в виде нечетких экспертных систем: ввод или преобразование данных о предметных областях, в частности, в виде экспертных оценок, хранение экспертных оценок и высказываний в базах знаний и базах данных, обработка экспертных оценок и высказываний на основе нечетких моделей и нейронных сетей [Батыршин и др., 2007 а; Батыршин, 2007 б; Борисов и др., 2007; Ротштейн, 1999] .

В-третьих, экспертные оценки используются в системах управления, проектирования и принятия решений в различных плохо формализуемых предметных областях, которым присущи объективные неопределенности в исходных и/или результирующих данных, а процедуры принятия решений используют экспертные знания. В таких системах с участием человека (гуманистических системах) чаще всего вызывает интерес не только вероятностное описание отношения объекта или события к полной группе объектов или событий, но и уникальность и специфика этого объекта или события.

В-четвертых, основная доля экспертных оценок в рассмотренных выше работах связана с экспертным оцениванием следующих показателей: состояния параметра неизвестного процесса, свойства объекта в условиях неопределенности, критерия качества и поведения сложной системы.

Контрольные вопросы

1. Какие существуют виды неопределенности?

2. Что представляют собой экспертные оценки?
3. Почему в некоторых системах невозможно использовать вероятностное и статистическое моделирование для учета факторов неопределенности?
4. Назовите причины возникновения неопределенности «нестохастического» типа?
5. Моделирование каких систем приводит к применению экспертных систем?
6. Приведите области, в которых используется обработка экспертных оценок.
7. Что такое нечеткое множество?

1.2. Виды экспертных оценок

Вопросам измерения и оценки в широком смысле посвящены исследования в области квалитметрии. Рассмотрим отличие понятий «измерение» и «оценка».

Целью измерения является получение количественной информации о величине исследуемых объектов, под которыми понимаются реально существующие объекты (предметы, явления, процессы), а также взаимодействия между ними. Измерение связывают с определением количественных характеристик объектов на основе нахождения соотношения между измеряемой величиной X и «эталонной единицей» этой величины с помощью физической или теоретической шкалы. При этом значения величины X рассматриваются как детерминированные, точно и четко определенные.

В оценивании обычно используется понятие «критерий оценки», градации которого образуют оценочную шкалу S . Оценка может быть по аналогии с измерением определена как отношение оцениваемой величины X к другой неоднородной величине VK . Тогда формальное представление оценки Mr , как отношения оцениваемой величины X по отношению к выбранному критерию VK , может быть задано в виде

$$Mr = Mark(X, Y, VK, S, MK),$$

где X – оцениваемая величина субъектом Y ;

VK – выбранный формальный критерий (например, эффективность, надежность);

S – известная шкала градаций критерия;

MK – метод операции оценивания.

По виду операции оценивания MK различают оценивание *на основе математических методов* и *экспертное* оценивание. *Оценивание на основе математических методов и моделей* используется для вычисления числовых оценок в условиях стохастической неопределенности и базируется на теории вероятности и математической статистики. *Экспертное оценивание* используется в условиях неопределенности «нестохастического» характера для получения как числовых, так и лингвистических оценок.

Подробное описание способов реализации процедур *экспертного оценивания* приведено в [Яхьяева, 2006].

Оценки, полученные в результате *экспертного оценивания*, в дальнейшем будем называть *экспертными оценками*. Интеллектуальная деятельность, содержанием которой является процесс оценивания параметров, моделей, объектов, состояний, процессов, явлений в условиях «нестохастической» неопределенности относится к *экспертной деятельности*, называемой также *экспертизой*, *экспертным оцениванием*. Результаты экспертной деятельности выражены в виде экспертных оценок и используются в системах поддержки принятия решений. Субъектом экспертной деятельности является эксперт.

Экспертные оценки характеризуются несколькими классификационными признаками [Тоценко, 2006].

Первым является *количество экспертов Y* , участвующих в формировании экспертной оценки. По этому признаку оценки делятся на индивидуальные и групповые. Относительно последних заметим, что групповая оценка – это одна оценка, отражающая мнение всех участвующих в экспертизе экспертов и являющаяся результатом агрегации их индивидуальных оценок. Следующим признаком является *количество критериев VK* , по которым формируется оценка.

По этому признаку оценки делятся на *однокритериальные* и *многокритериальные*. Примером *однокритериальной* может служить оценка ВВП (валового внутреннего продукта), *многокритериальной* – индекс развития человека (Human Development Index), публикуемый в ежегодном докладе ООН. В качестве третьего признака используется *тип оценки*. Тип оценки определяется типом шкалы оценивания S . Так, в случае числовой шкалы, оценки относятся к числовым (количественным) оценкам, при использовании номинальной шкалы – к лингвистическим.

Количественная оценка Mr выражается как числовое значение, например, показатель эффективности процесса, вероятность появления события, оценка параметра модели, интегрированная оценка качества. Такие оценки широко применяются в системах стохастического моделирования для получения новой информации о свойствах оцениваемых объектов, о степени соответствия объекта, явления, процесса некоторому критерию качества, значимому в данный момент времени и для данной ситуации.

Выделим несколько направлений применения количественных оценок:

1. Оценка может быть использована для измерения недетерминированной величины X , к которой применимы понятия: неточно определенная, подверженная случайным или иным искажениям, например, вследствие неточности инструментов и методов измерения.
2. Невозможность практического измерения будущей величины X , поэтому прибегают к оцениванию этой величины, например при автоматизированном проектировании новых систем, при выполнении прогнозов развития процессов.

Так, можно высказать суждение, что оценка, в отличие от «чистого» измерения, носит как более общий, так и более узкий характер. Обобщение связано с тем, что оценка может использоваться для «измерения» не только детерминированных величин, но и не полностью определенных величин и «неизмеряемых» величин. А узкий характер понятия оценка определяется такими ее аспектами, как субъективизм при интерпретации, проблемная ориентация.

Выделим два направления классификации лингвистических экспертных оценок: *содержательное* и *временное*.

Известны несколько аспектов содержательного направления логики оценки: логика абсолютной оценки (например, «хорошо», «плохо», «безразлично»), логика сравнительной оценки (например, «больше, чем», «меньше, чем», «равноценно»). Логика абсолютной оценки формируется на основе номинальной лингвистической шкалы S эксперта, которая может быть по отношению к нему внешней или внутренней.

Логика сравнительной оценки отличается тем, что используется порядковая лингвистическая шкала или сначала лингвистическая шкала S не определена, она формируется экспертом в процессе оценивания, а потом применяется в экспертном оценивании.

При рассмотрении временного направления логики оценки обратим внимание на тот факт, что оцениваемая величина X может изменяться во времени, образуя последовательность значений x_i , которым сопоставлены некоторые моменты времени t_i . Временное направление позволяет выделить в классе экспертных оценок следующие виды оценок:

- 1) оценки величины x_i в конкретный i -тый момент времени (точечные оценки),
- 2) оценки за определенный интервал времени (интервальные оценки),
- 3) оценки, связанные с интервалом, в котором происходит процесс оценивания, и интервалом, на котором проводилось наблюдение за величиной:

а) *pre*-оценка, соответствующая прогнозируемому значению оцениваемой величины, когда интервал, на котором может проводиться наблюдение, еще не наступил;

б) *post*-оценка, соответствующая случаю, когда интервал оценивания следует за интервалом наблюдения, на котором исследуемая величина изменялась;

в) *cur*-оценка, соответствующая ситуации, когда оценивание происходит в процессе наблюдения за изменением исследуемой величины.

Указанные оценки могут формироваться до наблюдения (*pre*-оценка), после наблюдения (*post*-оценка) и в процессе наблюдения (*cur*-оценка) за оцени-

ваемой величиной. *Pre*-оценка может применяться в задачах прогноза, как правило, она формируется на основе и зависит от *post*-оценок и *cur*-оценок.

Каждое из указанных направлений логики оценки основывается на введении нескольких уровней экспертной оценки, проранжированных в порядке возрастания сложности, значимости рассматриваемых свойств. Выражаемые с помощью лингвистических значений границы выделенных уровней экспертной оценки размыты, нечетки.

Подводя итог сравнению понятий измерения и оценки, можно отметить следующее:

- ✓ экспертная оценка – это измерение, основанное на знаниях эксперта;
- ✓ экспертная оценка – это интерпретированное количество;
- ✓ экспертная оценка – это нечеткое отношение по своей природе, выраженное лингвистически.

Получив широкое применение в интеллектуальных системах проектирования, анализа и моделирования сложных организационно-технических систем, лингвистические экспертные оценки развивались в направлении формализации операции оценивания, то есть формализации «интеллектуального» пространства оценок и экспертных шкал, выражающих знания, опыт и интуицию эксперта с целью применения математических моделей и методов для их обработки.

Эволюция экспертных оценок привела к необходимости интеграции количественных и лингвистических оценок для целей уменьшения субъективной составляющей в операциях экспертного оценивания. Такая интеграция была достигнута на основе теории нечеткого множества [Заде, 1974]. Введя понятие лингвистической переменной и допустив, что в качестве ее значений (термов) выступают нечеткие множества, Л. Заде создал математический аппарат для описания процессов интеллектуальной деятельности, включая нечеткость и неопределенность выражений, элементами которых могут выступать нечеткие экспертные оценки.

Определим нечеткую экспертную оценку как

$$F_Mr = F_Mark(X, Y, VK, S, MK),$$

где X – оцениваемая величина субъектом Y ;

VK – выбранный формальный критерий (например, эффективность, надежность);

S – известная нечеткая шкала градаций критерия, выраженная в виде лингвистической переменной. Каждая градация этой шкалы есть нечеткое множество (терм), представленное тройкой $\langle Name, W, MF \rangle$, где $Name$ – лингвистическая оценка (название градации), W – носитель градации (может быть представлен множеством действительных чисел или множеством лингвистических термов), MF – функция принадлежности, определенная на W , сопоставляющая каждому значению $X \in W$ значение из диапазона $[0, 1]$;

MK – метод оценивания на основе операции приведения к нечеткости (фаззификации).

Использование нечетких экспертных оценок полезно по следующим причинам:

1. С их помощью можно поставить на фундамент формальной теории проблематику широкого круга прикладных задач оценивания и выбора в условиях неопределенности, имеющей «нестохастический» характер.

2. Они позволяют формализовать задачи принятия решений, связанные с использованием оценок событий, генерируемых экспертом.

3. Нечеткие экспертные оценки реализуют оценивание явлений единичного характера, для которых отсутствуют вероятностные характеристики.

Контрольные вопросы

1. Какова цель измерения?
2. Чем отличается оценивание на основе математических методов от экспертного оценивания?
3. Сформулируйте определение лингвистической оценки.
4. Что такое абсолютная лингвистическая оценка? Приведите примеры.
5. Что такое сравнительная лингвистическая оценка? Приведите примеры.
6. Дайте определение множеству экспертных оценок.

7. Почему теория нечетких множеств является основным математическим аппаратом формализации представления и обработки экспертных оценок и высказываний?
8. В каких предметных областях и для каких показателей чаще всего применяются экспертные оценки?
9. Определите понятия измерение и оценка. В чем их различие?
10. Запишите формулу для формального представления оценки *Mr*.
11. По какому критерию различают понятия «Оценивание на основе математических методов и моделей» и «Экспертное оценивание»?
12. Сформулируйте определение экспертизы.
13. По какому классификационному признаку экспертные оценки можно разделить на индивидуальные и групповые? Приведите примеры.
14. По какому классификационному признаку экспертные оценки можно разделить на однокритериальные и многокритериальные? Приведите примеры.
15. По какому классификационному признаку экспертные оценки можно разделить на числовые (количественные) оценки и лингвистические? Приведите примеры.
16. Приведите примеры применения количественной оценки *Mr*.
17. Назовите два основных направления классификации лингвистических экспертных оценок.
18. Приведите несколько аспектов содержательного направления логики оценки.
19. В чем различие логики абсолютной оценки и логики сравнительной оценки?
20. Какие виды оценок можно выделить в классе экспертных оценок, использующие временное направление?
21. Определите следующие виды оценок: точечные оценки, интервальные оценки, *pre*-оценка, *post*-оценка, *cur*-оценка.

22. Кому принадлежит идея интеграции количественных и лингвистических оценок?
23. Назовите три основные причины применимости нечетких экспертных оценок.

1.3. Определение нечетких временных рядов

При анализе развития сложных организационно-технических систем во времени необходимо использовать всю полноту знаний об изучаемых данных и прикладной области исследования. Такие знания содержат не только непосредственно временные ряды параметров системы, полученные на основе измерений или наблюдений, но и вербальные описания типичных значений рядов, особых состояний, полученных на основе опыта экспертов.

При анализе временных рядов (ВР) эксперт представляет свои суждения с помощью нечетких оценок, относящихся ко многим объектам [Батыршин и др., 2007 а]:

- ✓ временные области: интервалы времени (несколько дней), абсолютная или относительная позиция на временной шкале (близкое будущее), периодические или сезонные интервалы (неделя до Рождества);
- ✓ ранг значений ВР (высокая цена, очень низкий уровень производства);
- ✓ набор паттернов ВР (быстро растущий, слегка выпуклый);
- ✓ набор ВР, их атрибутов, как элементов системы (фондовый индекс новой компании);
- ✓ набор отношений между ВР, атрибутами или элементами (тесно связанный);
- ✓ множество значений возможности или вероятности (непохоже, очень возможно).

В условиях неопределенности, а также в сложных организационно-технических системах, специалисты сталкиваются с рядом особенностей временных рядов: нестационарность, неоднородность, сложная форма динамики, что затрудняет и ограничивает применимость классических статистических ме-

тодов анализа ВР. Трудности при моделировании временных рядов возникают также в следующих случаях:

- 1) неизвестны вероятностные характеристики стохастического процесса, описывающего ВР;
- 2) неопределенность и неполнота в исходной информации и в информации о функционировании системы;
- 3) нелинейный характер искомой зависимости;
- 4) малое количество элементов выборки.

В этом случае применение находят интеллектуальные методы анализа временных рядов, активно использующие знания экспертов.

Экспертная оценка некоторой величины, описывающей отдельное состояние процесса, системы, объекта, полученная в некоторый момент времени, является по существу качественной интервальной оценкой локального состояния. Упорядоченная во времени последовательность таких оценок состояний системы и объектов представляет собой временной ряд экспертных оценок. Характерной особенностью такого временного ряда является нечеткость его значений, вытекающая из природы экспертных оценок, поэтому такой временной ряд относится к классу *нечетких временных рядов* [Ярушкина, 2004]. В основе интеллектуального анализа нечеткого временного ряда лежит понятие нечеткого множества, моделирующего экспертную оценку, и понятие нечеткой системы, реализующей обработку ВР методами интеллектуального анализа данных.

Подход с точки зрения нечетких моделей позволяет использовать прикладные знания интервального оценивания для нечеткого выражения временного ряда и строить зависимости в виде нечетких функций. Нечеткие описания в виде экспертных оценок можно использовать в разностном уравнении, где зависимые и независимые переменные выражены в нечетком виде. При этом исходные данные могут содержать не только лингвистические описания значений в конкретный момент времени, но и описания изменений параметров во времени. В зависимости от предметной области такие изменения могут обозначаться разными терминами: тренды, динамика, тенденция, траектория, систематиче-

ское движение, поведение и т. д. Классический статистический анализ, выделяющий постоянную систематическую составляющую (тренд), на всем временном ряду. В данном же случае предполагается, что динамика как детерминированная функция наблюдается только на некоторых интервалах времени. Динамика временного ряда на определенном интервале может выражать некоторое состояние изучаемого объекта. Таким образом, предполагая зависимость текущего состояния объекта от прошлых состояний, можно предположить, что имеется зависимость между наблюдаемыми движениями временных рядов в разные периоды времени.

Предположим, что задан процесс, состояния которого описываются n значениями одной переменной.

В результате наблюдения получен временной ряд этой переменной, представляющий последовательность упорядоченных в равноотстоящие моменты времени пар $\{x_i, t_i\}$, таких, что $\forall x_i \in X, X \subset R^1, t_i \in N, i \in [1, n]$. Значение x_i называют *уровнем* временного ряда.

Определение 1.1. *Нечетким временным рядом (НВР) называют упорядоченную в равноотстоящие моменты времени последовательность наблюдений над некоторым процессом, состояния которого изменяются во времени, если значение состояния процесса в момент t_i может быть выражено с помощью нечеткой метки $\tilde{x}_i$.*

Нечеткая метка $\tilde{x}_i$ может быть сформирована непосредственно экспертом или получена на основе некоторого преобразования исходного временного ряда.

Во втором случае она связана с исходным значением временного ряда. Введем для обозначения этой связи функционал *Fuzzy*, такой что

$$\tilde{x}_i = Fuzzy(\mu_{\tilde{x}_i}(w), x_i),$$

где $\tilde{x}_i \in \tilde{X}$, $\tilde{X}$ – множество нечетких меток;

w – носитель (интервал на X) нечеткой метки $\tilde{x}_i$, $x_i \in w$;

$\mu_{\tilde{x}_i}(w) \in [0,1]$ – функция принадлежности нечеткой метки $\tilde{x}_i$ уровню временного ряда x_i , обычно треугольной формы.

Определение 1.2. Носитель нечеткой метки $\tilde{x}_i$ – это четкое множество $w \subseteq B$ таких точек $x_i \in w$, для которых $\mu_{\tilde{x}_i}(w) > 0$, где $B \subset X$ – базовое множество нечетких меток $\tilde{X}$.

Таким образом, нечеткий временной ряд формируется в результате интервального качественного оценивания уровней числового ВР. Интервалы-носители нечетких меток, образованные на множестве X , обязательно пересекаются. Качественный аспект нечеткой метке придает функция $\mu_{\tilde{x}_i}(w) \in [0,1]$.

На рис. 1.1 изображен абстрактный нечеткий временной ряд, каждой нечеткой метке $\tilde{x}_i$ соответствует нечеткое множество, задаваемое функцией принадлежности.

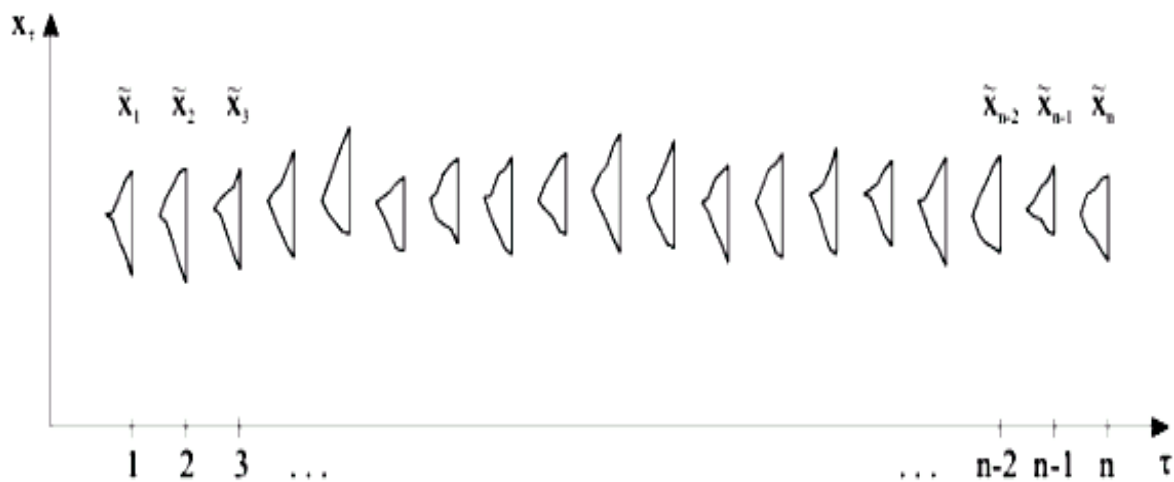


Рис. 1.1. Абстрактный нечеткий временной ряд

Контрольные вопросы

1. Какие объекты при анализе временных рядов можно представить с помощью нечетких оценок?
2. Приведите примеры случаев, когда применение классических статистических методов анализа ВР затруднительно.

3. Сформулируйте определение уровня временного ряда.
4. Сформулируйте определение нечеткого временного ряда.
5. Сформулируйте определение носителя нечеткой метки $\tilde{x}_i$.

1.4. Нечеткие временные ряды в системах поддержки принятия управленческих решений

Существует широкий класс систем управления сложными процессами, функционирование которых базируется на оценках о состоянии входящих в систему объектов. В качестве таких управляемых процессов можно привести следующие [Дубровский, 1984]:

- Процессы проектирования новых объектов, где уже на ранних стадиях необходимо решать вопросы контролепригодности и диагностируемости;
- Процессы испытаний и доработок новых и реинжиниринг функционирующих систем;
- Процессы эксплуатации сложных организационно-технических объектов, где режимы функционирования и обслуживания выбираются в зависимости от их оценки их состояния;
- Процессы оценки качества функционирования систем и др.

Для принятия адекватного решения по управлению в таких процессах применяют локальную сиг-оценку ситуации, включающую данные о состояниях объектов, среды, входных и выходных данных. По результатам локальной оценки выявляются отклонения, проблемы. Локальные оценки совместно с методами прогнозирования используются для последующего принятия решения о том или ином методе управления.

Известны модели и методы [Анфилатов и др., 2003], позволяющие формализовать получение локальной оценки в виде комплексной числовой оценки состояний, например, на основе теории эффективности и функций полезности. При этом рассматривают эффективность как показатель процесса функционирования системы. Под эффективностью процесса понимается комплексное

свойство, отражающее соответствие исхода операции поставленной цели по выбранному критерию эффективности. В зависимости от типа систем и внешних воздействий выделяют три группы показателей и критериев эффективности процессов функционирования:

- В условиях определенности, когда показатель исхода операции отражает один, строго определенный детерминированный исход;
- В условиях риска, если показатели исхода являются случайными величинами с известным законом распределения;
- В условиях неопределенности, если показатели исхода являются случайными величинами с неизвестным законом распределения.

Однако, зачастую, получаемые числовые оценки состояний одного и того же процесса по различным методам являются неоднозначными. Наиболее противоречивыми являются оценки эффективности, получаемые для систем, процессы функционирования которых протекают в условиях риска и неопределенности.

В качестве причины указанной проблемы можно указать на тот факт, что числовые оценки, характеризующие различные аспекты таких процессов, выражены в разных единицах, свертка которых требует дальнейшей экспертной оценки в лингвистических терминах качества, которые относятся к нечетким значениям. В этом случае вывод результата оценивания строится на основе следующей цепочки: «1. Числовая оценка i -того параметра» – «2. Выбор методов свертки» – «3. Свертка оценок по всем параметрам» – «4. Оценка по числовому критерию эффективности результата сверток» – «5. Экспертная оценка в лингвистических терминах». На каждом из указанных этапов происходит количественное наращивание погрешности, которое может привести к противоречивым и неоднозначным результатам оценивания.

С другой стороны, в системах управления сложными организационно-техническими системами процессы принятия решений часто протекают в условиях неопределенности и основываются на внешней и внутренней экспертизе. Экспертиза осуществляется в рамках экспертной деятельности и ее результатом являются экспертные оценки. Формализация экспертных оценок на основе тео-

рии нечетких множеств для принятия решений рассмотрена в [Штовба, 2007; Ротштейн, 1999; Чернов и др., 2006]. При этом вывод результата оценивания может быть короче и понятнее: «1. Экспертная оценка i -того параметра» – «2. Выбор методов свертки» – «3. Свертка экспертных оценок по всем параметрам в интегрированную экспертную оценку». Такие оценки являются нечеткими значениями, применению которых в задачах принятия решений по управлению в условиях неопределенности посвящены работы [Ахрамейко, 2008; Аверкин, 2008; Батыршин, 1989; Беллман и др., 1976; Алиев, 1990].

Для получения *pre*-оценок строят прогнозные модели, учитывающие динамику процессов [Афанасьева, 2008 а; Ярушкина и др., 2007 а]. В таких моделях анализируют не только оценку текущей ситуации (локальную оценку), но и историю ее возникновения, то есть ее поведение во времени, наблюдаемое на основе временных рядов. Статистические методы анализа временных рядов не применимы для построения прогноза поведения процессов, состояния которых заданы нечеткими значениями экспертных оценок.

Таким образом, применительно к системам принятия решений в управлении сложными организационно-техническими системами можно сделать следующие выводы о применимости нечетких временных рядов:

- В процессах управления в условиях «нестохастической» неопределенности применяют экспертные оценки состояний процесса, выраженные нечеткими метками;
- Для моделирования динамики процесса в таких системах целесообразно сформировать из экспертных оценок нечеткий временной ряд;
- Нечеткий временной ряд может служить новым инструментом моделирования динамики и описания процессов, функционирующих в условиях «нестохастической» неопределенности. Для таких процессов показатели исхода определяются экспертными оценками, выраженными нечеткими метками;
- Нечеткий временной ряд может служить новым инструментом моделирования динамики и описания процессов процесса, представленного число-

вым временным рядом, если уровни этого ряда могут быть представлены нечеткими метками.

Контрольные вопросы

1. Приведите примеры управляемых процессов.
2. Какую оценку чаще всего используют для принятия адекватного решения в управляемых процессах?
3. Сформулируйте определение эффективности процесса.
4. Назовите три группы показателей и критериев эффективности процессов функционирования, которые выделяют в зависимости от типа систем и внешних воздействий.
5. Какие оценки получают при построении прогнозные модели?
6. Приведите примеры процессов, при моделировании которых применяют нечеткие временные ряды.

1.5. Нечеткий временной ряд в системах автоматизации проектирования

Двадцать первый век характеризуется усложнением задач-проблем, которые предстоит решать человечеству. К таким задачам можно отнести проектирование новых видов промышленных, пищевых, энергетических, транспортных, технологических, информационных и др. объектов, характеризующихся уровнем организованной сложности. Это усложнение характеризуется большим объемом гетерогенных данных, временных и пространственных связей между ними, ограниченным временем решения и «мобильностью» требований к результатам решения задач. Для эффективного решения новых задач, удовлетворяющих требованиям времени, будут приобретать все большее значение системы автоматизированного проектирования (САПР) сложных организационно-технических объектов, которые должны быть ориентированы на быстроменяющиеся условия внешней среды. Проблемы, с которыми будут сталкиваться

проектировщики таких объектов, обусловлены не только сложностью объектов и недостатком опыта проектирования, но и повышенной степенью риска ошибочных решений и многокритериальностью выбора вариантов. Такие САПР в большей степени будут использовать гибридные (или интегральные) технологии на всех этапах проектирования с привлечением экспертных систем и методов интеллектуального анализа данных, в том числе представленных временными рядами.

Ожидаемыми результатами будут новые классы подсистем САПР, являющиеся, по словам Билла Гейтса, «усилителем интеллекта» экспертной деятельности проектировщика, а также инновационные продукты САПР с элементами искусственного интеллекта в следующих аспектах: в них будут реализованы отдельные операции интеллектуальной деятельности; их функционирование будет активно использовать интеллектуальные технологии ввода, анализа, обработки и вывода данных.

Среди задач автоматизации проектирования объектов с элементами искусственного интеллекта можно выделить задачи, связанные с решением проблем, в основе которых лежат понятия «поведение» объекта, информация о котором представлена временным рядом, и «анализ тенденций»:

- ✓ задача поиска прототипов создаваемых объектов на основе анализа и идентификации перспективных тенденций и достижений,
- ✓ задача формализации описания требований к функционированию (поведению) создаваемого объекта,
- ✓ задача создания новых моделей поведения объектов с элементами искусственного интеллекта и методов анализа их поведения,
- ✓ задача создания новых моделей диагностики поведения объектов с элементами искусственного интеллекта и методов их контроля,
- ✓ задача моделирования процессов управления объектом с учетом тенденции его поведения и др.

Проектирование сложной системы представляет трудоемкий процесс решения совокупности плохо формализуемых задач. Экспертные оценки, аккумулирующие опыт проектировщика, являются основой в принятии проектных решений [Норенков, 2003] и могут относиться к оценке отдельных параметров проекта с учетом динамики, к оценке качества проекта для целей эффективного управления проектированием [Ярушкина, 1997; Ярушкина, 2004; Ярушкина и др., 2007 б; Афанасьева и др., 2008 б; Жирабок, 2001; Борисов, 1985].

1.5.1. Моделирование динамических параметров проектируемой системы

Среди основных параметров проектируемой технической системы следует выделить параметры, изменяемые во времени. Такие параметры определяют динамику, функционирование, жизненный цикл проектируемой технической системы или отдельного ее свойства.

На ранних этапах проектирования в условиях неизвестного закона функционирования, тем не менее, у проектировщика при моделировании динамики заданного параметра, как правило, имеется некоторая информация о его характере изменения. Значения динамики моделируемого параметра могут быть получены по данным, взятым с прототипов проектируемой системы в виде ограниченной числовой последовательности, нечетких значений или качественных интервальных оценок.

Например, в момент времени $t1$ значение параметра должно быть приблизительно равным $z1$, а в интервале $(t2, t3)$ значение параметра может быть $z2$, в момент времени $t4$ значение параметра может быть или равно $z4$ или больше $z5$. Такие оценки относятся к нечетким и для их обработки используют модели на основе нечетких множеств [Борисов, 1985].

Если проектировщик имеет в своем распоряжении числовые характеристики динамики параметра, то они, как правило, представимы короткими временными рядами, поэтому использовать статистические методы для анализа временных рядов динамического параметра на ранних стадиях проектирования

не представляется возможным, так как они «хорошо» работают для больших объемов числовых данных, являющихся выборками из нормально распределенной совокупности значений. Проектировщик в условиях неопределенности, используя экспертное оценивание [Войт, 2008; Малышев и др., 1991], интерпретируя числовые значения короткого временного ряда лингвистическими терминами, переходит от чисел к интервальным качественным оценкам и нечетким временным рядам [Стецко, 2008].

Представляя динамику моделируемого параметра нечетким временным рядом, в дальнейшем можно провести его нечеткое моделирование, получить качественные зависимости, описывающие динамику параметра. Результаты анализа нечетких временных рядов будут способствовать получению новых знаний о динамических свойствах заданного параметра, что позволит на последующих стадиях проектирования сократить пространство проектных решений, пространство действий проектировщика и создаст условия для повышения качества проектирования.

Таким образом, уже на ранних этапах процесса проектирования в условиях неопределенности может быть проведено «мягкое» моделирование динамики параметров на основе нечетких временных рядов.

1.5.2. Нечеткий временной ряд при оценке качества процесса проектирования

Процесс проектирования сложных технических систем происходит преимущественно «сверху-вниз». По мере проработки проекта увеличивается подробность описания, детализации системы. Проектно-конструкторские решения, особенно на ранних стадиях, принимаются на основе данных, которые не всегда обладают точностью, четкостью. Это связано с «грубым», неточным неполным описанием системы на ранних этапах, необходимостью согласования принимаемых решений, использованием упрощенных моделей. Весь ход разработки можно интерпретировать как диалектику двух параллельных процессов:

процесса уточнения, снятия неопределенности, дефаззификации и процесса повышения конструктивной неопределенности за счет формирования качественных обобщений, фаззификации. Отмеченные особенности присущи как практическому, так и автоматизированному проектированию [Вязигин и др., 1989].

При автоматизированном проектировании сложных технических систем решается задача не просто синтеза системы с заданными свойствами, но и синтеза наилучшей системы. Эта задача решается на основе моделирования, анализа, сравнения, экспертной оценки вариантов проектируемой системы, принятия решений на основе некоторых заданных критериев и в условиях недостаточного объема априорной информации. В результате многокритериального оценивания формулируется конкретное проектное решение [Ахрамейко, 2008; Орловский, 1981].

Как правило, найти такие точные и непротиворечивые критерии, которые в полной мере удовлетворяли требованиям затруднительно, поэтому кроме формальных критериев оценки вариантов проектных решений используются опыт и знания проектировщика. В результате, оценка вариантов проектных решений представляет собой некоторую экспертную оценку, формулируемую проектировщиком в нечетких терминах качества на естественном языке, например, «Подходящее», «Удовлетворительное», «Хорошее», «Отличное», «Плохое» и др. Такое «внутреннее» оценивание является средством отображения множества неоднородных критериев в единую интервальную шкалу качества проектных решений. Значимость такого «внутреннего» оценивания проявляется не сразу – или на последующих стадиях проектирования, или, что значительно хуже, при испытании и эксплуатации. Для уменьшения будущих рисков «неоптимальных» решений представляется оправданным хранить нечеткую информацию о процессе проектирования для последующего анализа принятых проектных решений и на этой основе управлять процессом проектирования [Грачев, 2008; Гладков и др., 2008; Соснин, 2007].

В базу данных, хранящую информацию о процессе проектирования сложной технической системы, целесообразно включать не только варианты

проектных решений, выбранные решения и их «внутреннюю» оценку, но и внешнюю экспертную оценку этих решений, задаваемую также в лингвистических терминах качества. Внешние оценки формируются по результатам испытаний [Ярушкина, 2004] и эксплуатации спроектированной системы.

Сопоставление внутренних и внешних оценок проектных решений позволит вывести локальную интегрированную оценку качества проектного решения в некоторый момент времени. К интегрированным оценкам проектных решений, которые являются производными от внутренних и внешних экспертных оценок, применимы также лингвистические категории качества. Использование единой шкалы качества для оценки проектных решений имеет ряд особенностей. Во-первых, появляется возможность сравнения различных решений по одному основанию. Во-вторых, лингвистические оценки выставляются человеком, естественны для него и не требуют дополнительной интерпретации по сравнению с числовыми оценками. В-третьих, лингвистические оценки проектных решений относятся к нечетким значениям по своей природе, следовательно, для их описания и моделирования применим аппарат нечетких множеств.

При развертывании нечетких интегрированных оценок проектных решений во времени по отношению к отдельному процессу проектирования появляется возможность сформировать нечеткий временной ряд оценок. Анализ и идентификация такого нечеткого временного ряда позволит получить интегрированную оценку динамики и качества процесса проектирования в целом, а также обнаружить проблемы и, возможно, спрогнозировать будущие тенденции.

Контрольные вопросы

1. Какие новые задачи будут решаться с помощью систем автоматизированного проектирования (САПР), и какие требования будут предъявляться к их решениям?
2. Приведите примеры задач, связанных с решением проблем, в основе которых лежат понятия «поведение» объекта, информация о котором представлена временным рядом, и «анализ тенденций».

3. Сформулируйте понятие динамических параметров, какие данные системы они могут определять?
4. Почему использование статистических методов для анализа временных рядов динамического параметра на ранних стадиях проектирования не возможно?
5. Опишите переход от числовых характеристик динамики параметра к нечетким временным рядам.
6. Сформулируйте понятия внутреннего и внешнего оценивания вариантов проектных решений, в чем их различие?
7. Сформулируйте определение локальной интегрированной оценки.

1.6. Нечеткие временные ряды в задачах экспертной деятельности

Экспертиза (экспертная деятельность) как неотъемлемая часть функции оценки сложных объектов является основой управленческой и проектной деятельности.

В системах проектирования и управления сложными организационно-техническими системами в условиях неопределенности содержание экспертной деятельности включает решение совокупности экспертных задач [Ярушкина, 2004]: интерпретация, диагностика и мониторинг, прогноз, планирование. Указанные задачи в качестве исходных данных могут использовать нечеткие значения в виде экспертных оценок. Решения задач экспертной деятельности, связанной с анализом поведения объектов, динамики процессов функционирования систем в условиях неопределенности, состояния которых представлены нечеткими значениями, не нашли адекватного решения в настоящее время. Их решение целесообразно искать на основе развития нового направления в области интеллектуального анализа данных – нечетких временных рядов. Использование нечетких временных рядов представляется обоснованным и для моделирования динамики в числовых временных рядах, в том числе в задачах анализа

коротких временных рядов, для которых проблема анализа еще не получила эффективного решения классическими методами.

Рассмотрим постановку некоторых экспертных задач, решение которых целесообразно моделировать на основе НВР.

П 1. Область приложения НВР – интерпретация процессов.

Интерпретация – это процесс обработки данных для описания состояния и динамики процесса с целью определения их смыслового значения. Для различных систем управления и проектирования результатом интерпретации является экспертная оценка, значения которой представлены смысловыми единицами соответствия данных некоторым качественным градациям (интервалам). Эта оценка, задающая отношение принадлежности данных некоторым более общим понятиям, относится к нечетким значениям. В теории временных рядов (ВР) задаче интерпретации соответствует задача идентификации с последующим смысловым описанием.

Постановка задачи интерпретации процессов. Предположим, что задан процесс, наблюдения за которым образуют временной ряд $\{x_i, t_i\}$. Требуется определить тип изменения переменной x в заданном интервале времени $t_0 \leq t_i \leq t_n$, то есть определить тенденцию (систематическую составляющую) развития этой переменной Tr . Значения тенденции Tr могут быть выражены нечеткими лингвистическими термами, такими как «Рост», «Падение», «Стабилизация» [Ярушкина, 2004]. Построение отношений между ВР, задаваемым упорядоченной последовательностью пар $\{x_i, t_i\}$, и тенденцией Tr и является содержанием задачи интерпретации процесса, которую эксперт может решать визуально или на основе экспертной интерпретации значений ВР и преобразования ВР в НВР.

П 2. Область приложений НВР – диагностика процессов.

Диагностика – это процесс поиска неисправностей, проблем, дефектов, аномалий или их отсутствия. При решении задач экспертной деятельности с

целью диагностики процессов, представленных в виде НВР, целесообразно применять методы контроля НВР. Эти методы, как представляется, могут включать сопоставление НВР, отражающего реализованную динамику процесса с НВР с ожидаемой, требуемой динамикой. Интерпретация полученных сопоставлений в виде тенденций нечеткого ряда, учитывающего отклонения между реализованным и требуемым НВР может быть получена при решении задачи интерпретации процессов, рассмотренной выше.

Постановка задачи диагностики процессов. Даны две последовательности нечетких значений одной переменной, наблюдаемых в заданном интервале времени. Каждая последовательность рассматривается как НВР, одна последовательность является требуемой, другая – наблюдаемой. Требуется определить тип соответствия между этими последовательностями в заданном интервале времени, то есть определить тенденцию нового НВР, отражающего отклонения наблюдаемого НВР от требуемого НВР. Значения соответствия может быть представлены нечеткими термами, таким как «Без отклонений», «Значительные отклонения», «Незначительные отклонения» и др. Решение задачи диагностики включает решение задачи интерпретации процессов.

П 3. Область приложения НВР – прогноз развития процессов.

Прогноз – это приближенная оценка будущих изменений, хода событий, поведения на основе модели динамики в прошлом и настоящем.

Постановка задачи прогноза развития процессов. Дана последовательность нечетких значений одной переменной, полученная в результате наблюдения в заданном интервале времени. Требуется идентифицировать тип изменения этой переменной Tr в следующем интервале времени. Тип изменения Tr переменной может быть задан в нечетких термах, используемых в задаче интерпретации «Рост», «Падение», «Стабилизация» и др. Эта задача включает предварительное решение задач интерпретации и диагностики.

П4. Область приложения НВР – планирование.

Завершающим этапом любого экспертного заключения является выработка рекомендаций, которые следует применить для достижения требуемой ди-

динамики исследуемого процесса. Задача планирования является сложной задачей, часто контекст, в котором ведется планирование, известен только приблизительно, так что планирование ведется в условиях неопределенности. При этом возможны различные последствия, оценить которые возможно только на основе моделирования. Использование НВР и представление совокупности данных в виде экспертных оценок, полученных в результате интерпретации, диагностики и прогноза динамики процесса, может служить дополнительным инструментом в принятии более обоснованных решений и выработке рекомендаций.

Постановка задачи планирования. Дана последовательность нечетких значений и тенденция динамики Tr одной переменной, отражающей отклонения динамики процесса от требуемого. Известна тенденция исходного НВР и его прогнозная тенденция. Требуется идентифицировать тип изменения некоторой переменной в следующем интервале времени, так, чтобы отклонения динамики исходного процесса от требуемой были минимальны. Тип изменения переменной может быть задан в нечетких терминах, используемых в задаче идентификации «Рост», «Падение», «Стабилизация» и др.

Контрольные вопросы

1. Какие задачи включает в себя экспертная деятельность в условиях неопределенности?
2. Определите понятие интерпретации процессов и сформулируйте постановку задачи.
3. Определите понятие диагностики процессов и сформулируйте постановку задачи.
4. Определите понятие прогноз развития процессов и сформулируйте постановку задачи.
5. Определите понятие планирование процессов и сформулируйте постановку задачи.

Выводы

Существует класс процессов в системах проектирования и управления сложными организационно-техническими системами, функционирующих в условиях неопределенности, которая может быть описана экспертными (нечеткими) оценками состояний процессов, относящихся к классу интервальных качественных оценок. Применение при моделировании указанных процессов нечеткого временного ряда представляется перспективным направлением, так как анализ нечетких временных рядов позволит экспертам и лицам, принимающим решения, извлекать новые знания об исследуемых процессах, выраженные в нечетких моделях, и, таким образом, повысить эффективность результатов управленческих и проектных решений.

Библиографический список

1. [Абрахин, 2008 а] Абрахин, С. И. Прогнозирование распространения лесных пожаров с применением теории нечетких множеств / С. И Абрахин // Нечеткие системы и мягкие вычисления (НСМВ-2008): сборник научных трудов второй всероссийской научной конференции с международным участием (г. Ульяновск, 27-29 октября, 2008 г.). – Т 2. – Ульяновск : УлГТУ, 2008. – С. 3-9.
2. [Абрахин, 2008 б] Абрахин, С. И. Моделирование последствий прорыва плотины на реке с использованием теории нечетких множеств / С. И Абрахин // Нечеткие системы и мягкие вычисления (НСМВ-2008) : сборник научных трудов второй всероссийской научной конференции с международным участием (г. Ульяновск, 27-29 октября, 2008 г.). – Т. 2. – Ульяновск : УлГТУ, 2008. – С. 9-23.
3. [Аверкин и др., 1986] Нечеткие множества в моделях управления и искусственного интеллекта / А. Н. Аверкин, И. З. Батыршин, А. Ф. Блишун и др. ; Под ред. Д. А. Поспелова. – М. : Наука. Гл. ред. физ.-мат. лит., 1986. – 312 с.

4. [Аверкин и др., 2008] Аверкин, А. Н. Многокритериальный анализ нечетких объектов с кластеризацией экспертных оценок / А. Н. Аверкин, О. В. Костюченко, Н. В. Титова // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия) : Труды конференции. – Т.1. – М. : ЛЕНАНД, 2008. – С. 269-280.
5. [Аверченков и др., 2008] Аверченков, А. В. Построение многоагентной системы мониторинга и анализа информации в области CAD/CAM/CAT-технологий / А. В. Аверченков, В. И. Аверченков // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. – М. : Физматлит, 2008. – Т. 3. – С. 312-314.
6. [Алиев и др., 1990] Алиев, Р. А. Нечеткие модели управления динамическими системами / Р. А. Алиев, Э. Г. Захарова, С. В. Ульянов // Итоги науки и техники. Сер. Техн. кибернетика. – Т. 29. – М. : ВИНТИ АН СССР, 1990. – С. 127-201.
7. [Андрейчиков и др., 2008] Андрейчиков, А. В. Интеллектуальная поддержка проектирования инноваций / А. В. Андрейчиков, О. Н. Андрейчикова // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. – М. : Физматлит, 2008. – Т. 1. – С. 95-100.
8. [Анфилатов и др., 2003] Анфилатов, В. С. Системный анализ в управлении : учеб. пособие / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин ; под ред. А. А. Емельянова. – М. : Финансы и статистика, 2003. – 368 с.
9. [Ахрамейко и др., 2008] Ахрамейко, И. В. Многокритериальные методы обоснования управленческих решений в условиях нестохастической неопределенности данных / И. В. Ахрамейко, И. А. Семенов // Труды V Международной научно-практической конференции «Интегрирован-

- ные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). – Т.2. – М. : Физматлит, 2009. – С. 785-799.
10. [Афанасьев, 2009] Афанасьев, А. Н. Разработка продукционной модели системы мониторинга качества процесса обучения / А. Н. Афанасьев // Труды Седьмой Международной конференции «Математическое моделирование физических, экономических, технических, социальных систем и процессов», 2-5 февраля 2009 года, г. Ульяновск / под ред. д. т. н., проф. Ю. В. Полянского, д. ф.-м. н., проф. В. Л. Леонтьева. – Ульяновск : УлГУ, 2009. – С. 31-33.
 11. [Афанасьева, 2008 а] Афанасьева, Т. В. Нечеткие временные ряды в системах управления сложными процессами / Т. В. Афанасьева // Информационные технологии : межвузовский сборник научных трудов. – Ульяновск : УлГТУ, 2008. – С. 37-39.
 12. [Афанасьева и др., 2008 б] Афанасьева, Т. В. Нечеткие временные ряды в автоматизированном проектировании / Т. В. Афанасьева, Н. Г. Ярушкина // Информационные технологии : межвузовский сборник научных трудов. – Ульяновск : УлГТУ, 2008. – С. 34-37.
 13. [Басаев, 2005] Басаев, М. В. Оценка инвестиций в жилищное строительство с использованием нечетких множеств / М. В. Басаев // Труды III Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 15-17 мая 2005 г.). Т. 2. – М. : Физматлит, 2005.
 14. [Батыршин, 1989] Батыршин, И. З. Принятие решений на базе нечетких отношений предпочтения и функций выбора / И. З. Батыршин // Нечеткие системы поддержки принятия решений. – Калинин : КГУ, 1989. – С. 29-35.
 15. [Батыршин и др., 2007 а] Батыршин, И. З. Модели и методы перцептивного дата майнинга временных рядов для систем поддержки принятия решений / И. З. Батыршин, Л. Б. Шереметов // Нечеткие системы и мягкие вычисления. Т. 2. – 2007. – №1.

16. [Батыршин и др., 2007 б] Батыршин, И. З. Нечеткие гибридные системы. Теория и практика / И. З. Батыршин, А. О. Недосекин, А. А. Стецко и др. – М. : ФИЗМАТЛИТ, 2007. – 208 с.
17. [Беллман и др., 1976] Беллман, Р. Принятие решений в расплывчатых условиях / Р. Беллман, Л. Заде // Вопросы анализа и процедуры принятия решений – М. : Мир, 1976. – С. 172-215.
18. [Белый и др., 2002] Белый, О. В. Архитектура и методология транспортных систем / О. В. Белый, О. Г. Кокаев, С. А. Попов. – СПб. : «Элмор», 2002.
19. [Борисов, 1985] Методы и системы принятия решений. Автоматизированные системы поддержки принятия решений в управлении и проектировании / Под ред. А. Н. Борисова. – Рига : Риж. политехн. ин-т, 1985. – 164 с.
20. [Борисов, 1989] Обработка нечеткой информации в системах принятия решений / А. Н. Борисов, А. В. Алексеев, Г. В. Меркурьева. – М. : Радио и связь, 1989. – 304 с.
21. [Борисов и др., 2007] Борисов, В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия – Телеком, 2007. – 284 с.
22. [Валеев и др, 2008] Валеев, С. Г. Программный комплекс анализа учебной деятельности / С. Г. Валеев, А. В. Мадышев // Нечеткие системы и мягкие вычисления (НСМВ-2008) : сборник научных трудов второй всероссийской научной конференции с международным участием (г. Ульяновск, 27-29 октября, 2008 г.). Т. 2. – Ульяновск : УлГТУ, 2008. – С. 144-151.
23. [Виноградов и др., 2007] Виноградов, Г. П. Модели прогнозирования в интеллектуальных системах / Г. П. Виноградов, Н. А. Семенов. // Программные продукты и системы. – 2007. – №4. – С. 80-82.
24. [Виноградов и др., 2008] Виноградов, Г. П. Модели прогнозирования в интеллектуальных организациях / Г. П. Виноградов, Б. В. Палюх // Один-

- надцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября-3 октября, 2008 г., г. Дубна, Россия) : Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 235-240.
25. [Войт, 2008] Войт, Н. Н. Разработка метода нечеткой оценки проектных характеристик обучаемого инженера для автоматизированных обучающих систем САПР / Н. Н. Войт // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). Т.2. – М. : Физматлит, 2009. – С. 799-808.
26. [Вязигин и др., 1989] Вязигин, В. А. Математические методы автоматизированного проектирования / В. А. Вязигин, В. В. Федоров. – М. : Высш. шк., 1989. – 184 с.
27. [Вятчинин, 2004] Вятчинин, Д. А. Нечеткие методы автоматической классификации / Д. А. Вятчинин. – Минск : Технопринт, 2004. – 219 с.
28. [Гладков и др., 2008] Гладков, Л. А. Нечеткий генетический алгоритм построения временного графика процесса проектирования / Л. А. Гладков, А. Е. Криницкая // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т 1. – М. : Физматлит, 2008. – С. 46-52.
29. [Головина, 2005] Головина, Е. Ю. Интеллектуальная система поддержки принятия решений в сфере потребления тепловой энергии в объектах «Теплисис» / Е. Ю. Головина // Труды III Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 15-17 мая 2005 г.) Т. 2. – М. : Физматлит, 2005.
30. [Гофман и др., 2008] Гофман, И. Д. Актуальные задачи развития технологии недоопределенного календарного планирования / И. Д. Гофман, Д. А. Инишев, А. А. Липатов и др. // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-

- 2008 (28 сентября-3 октября, 2008 г., г. Дубна, Россия) : Труды конференции. Т.1. – М. : ЛЕНАНД, 2008. – С. 68-76.
31. [Грачев, 2008] Грачев, С. А. Модель системы управления процессом проектирования в нечеткой среде / С. А. Грачев // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.) Т. 2. – М. : Физматлит, 2009. – С. 812-822.
32. [Долгий и др., 2008] Долгий, А. Н. Интеллектуальные модели выявления нечетких темпоральных признаков в базах данных геодиагностических систем / А. Н. Долгий, И. Д. Долгий, С. М. Ковалев и др. // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 2. – М. : Физматлит, 2008. – С. 129-135.
33. [Домрачев, 2001] Домрачев, В. Г. Нечеткие модели рейтинговых систем оценки знаний / В. Г. Домрачев, О. М. Полещук, И. В. Ретинская и др. // Телематика'2001. Труды Международной научно-методической конф. – СПб., 2001. – С. 245-246.
34. [Дубровский, 1984] Дубровский, Л. К. Нечеткие измерения при описании состояния объектов / Л. К. Дубровский // Методы и системы принятия решений. Интеллектуальные системы принятия решений. – Рига : Риж. Политехн. Ин-т, 1987. – С. 84-91.
35. [Дулин и др., 2008] Дулин, С. К. Система имитационного моделирования движения железнодорожного транспорта на основе итеративно задаваемых правил организации движения / С. К. Дулин, А. С. Селецкий, В. И. Уманский // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября-3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 1. – М. : ЛЕНАНД, 2008. – С. 77-85.

36. [Жирабок, 2001] Жирабок, А. Н. Нечеткие множества и их использование для принятия решений / А. Н. Жирабок // Соросовский образовательный журнал. – 2001. – №2. – С. 109-115.
37. [Заде, 1974] Заде, Л. А. Основы нового подхода к анализу сложных систем и процессов принятия решений / Л. А. Заде // Математика сегодня. – М. : Знание, 1974. – С. 5-49.
38. [Ермоленко, 2008] Ермоленко, Д. Н. Алгоритм принятия решений по управлению очисткой сточных вод на основе нечеткого моделирования / Д. Н. Ермоленко // Нечеткие системы и мягкие вычисления (НСМВ-2008) : сборник научных трудов второй всероссийской научной конференции с международным участием (г. Ульяновск, 27-29 октября, 2008 г.). Т. 2. – Ульяновск : УлГТУ, 2008. – С. 24-31.
39. [Затылкин, 2008] Затылкин, В. В. Управление кадрами предприятия с применением методов искусственного интеллекта / В. В. Затылкин // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1 – М. : Физматлит, 2008. – С. 343-348.
40. [Исмагилов, 2008] Исмагилов, И. И. Нечеткое прогнозирование временных рядов с использованием процедуры групповой экспертизы / И. И. Исмагилов // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 132-139.
41. [Касапенко и др., 2008] Касапенко, Д. В. принятие решений в экспертных вопросно-ответных средах / Д. В. Касапенко, П. И. Соснин // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т 1. – М. : Физматлит, 2008. – С. 248–255.
42. [Клячкин, 2003] Клячкин, В. Н. Многомерный статистический контроль технологического процесса / В. Н. Клячкин. – М. : Финансы и статистика, 2003.

43. [Кравченко, 2008] Кравченко, Ю. А. Направление разработки адаптивных систем поддержки принятия решений / Ю. А. Кравченко // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1. – М. : Физматлит, 2008. – С. 268-277.
44. [Кудинов, 2008] Кудинов, Ю. И. Принципы построения нечеткой системы прогнозирования дефектов металлопродукции / Ю. И. Кудинов // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 123-131.
45. [Кудинов, 2007] Кудинов, Ю. И. Разработка и идентификация нечетких моделей прогнозирования качества / Ю. И. Кудинов // Мехатроника, автоматизация, управление. – 2007. – №12.
46. [Кулаков и др., 2008] Кулаков, С. М. О функциональной структуре интеллектуальных систем управления технологическими объектами / С. М. Кулаков, В. Б. Трофимов, Н. Ф. Бондарь // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1. – М. : Физматлит, 2008. – С. 353-361.
47. [Линев, 2008] Линев, Н. А. Автоматизация процессов планирования и прогнозирования для задач обслуживания оборудования предприятий / Н. А. Линев // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1. – М. : Физматлит, 2008. – С. 277-280.
48. [Макаров и др., 2007] Макаров, А. Н. Прогнозирование эксплуатационной надежности на основе методов нечеткой логики / А. Н. Макаров, К. Б. Корнеев // Нечеткие системы и мягкие вычисления. Т. 3. – 2007. – №3.

49. [Малышев и др., 2007] Малышев, Н. Г. Нечеткие модели для экспертных систем в САПР / Н. Г. Малышев, Л. С. Бернштейн, А. В. Боженюк – М. : Энергоиздат, 1991. – 136 с.
50. [Напреенко и др., 2008] Напреенко, В. Г. Моделирование экономики Беларуси на основе технологии Н-моделей / В. Г. Напреенко, А. С. Нариньяни, В. Я. Асанович и др. // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 1. – М. : ЛЕНАНД, 2008. – С. 171-178.
51. [Недосекин, 2005] Недосекин, А. Н. Оптимизация фондового портфеля, содержащего опционы (нечеткая постановка задачи) / А. Н. Недосекин // Труды III Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 15-17 мая 2005 г.). Т. 2. – М. : Физматлит, 2005.
52. [Новак и др., 2008] Новак, В. Интегральный метод принятия решений и анализа нечетких временных рядов / В. Новак, И. Перфильева, Н. Ярушкина и др. // Программные продукты и системы. – 2008. – №4(84). – С. 65-68.
53. [Новак, 2008] Новак, В. Применение интегрального метода анализа нечетких временных рядов и функционального моделирования в задаче выбора лизинговой компании / В. Новак // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 149-154.
54. [Норенков, 2002] Норенков, И. П. Основы автоматизированного проектирования : учеб. для вузов / И. П. Норенков. – 2-е изд., перераб. и доп. – М. : Изд-во МГТУ им. Н. Э. Баумана, 2002. – 336 с.
55. [Нурматова, 2009] Нурматова, Е.В. Применение адаптивной нейронечеткой системы для моделирования рейтинговой оценки знаний / Е. В. Нурматова // Труды V Международной научно-практической кон-

- ференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). Т. 2. – М. : Физматлит, 2009. – С. 698-705.
56. [Нуруллин, 2008] Нуруллин, А. Ю. Структура и состав Internet интегрированной среды для экспертизы экономического состояния предприятия на основе системы нечеткого вывода / А. Ю. Нуруллин, И. В. Семушин, А. В. Чекина // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 116-122.
57. [Орловский, 1981] Орловский, С. А. Проблемы принятия решений при нечеткой исходной информации / С. А. Орловский. – М. : Наука, 1981.
58. [Павленко, 2008] Павленко, Е. Н. Моделирование технологических процессов в паровых котлах в условиях неопределенности / Е. Н. Павленко // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т.2. – М. : Физматлит, 2008. – С. 307-312.
59. [Палюх, 2008] Палюх, Б. В. Интеллектуальная система управления пожарной безопасностью хранения фрезерного торфа / Б. В. Палюх, Р. Е. Цветков // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 3. – М. : ЛЕНАНД, 2008. – С. 300-305.
60. [Ротштейн, 1999] Ротштейн, А. П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети / А. П. Ротштейн. – Винница : УНИВЕРСУМ-Винница, 1999. – 320 с.
61. [Смагин и др., 2008] Смагин, А. А. Экспертная система морского мониторинга / А. А. Смагин, С. В. Липатова, А. С. Мельниченко и др. // Нечеткие системы и мягкие вычисления (НСМВ-2008) : сборник научных трудов второй всероссийской научной конференции с международным уча-

- ствием (г. Ульяновск, 27-29 октября, 2008 г.). Т. 2. – Ульяновск : УлГТУ, 2008. – С. 96-104.
62. [Стецко, 2008] Стецко, А. А. Принятие проектных решений на основе анализа нечетких тенденций временных рядов / А. А. Стецко // Программные продукты и системы. – 2008. – №3.
63. [Соснин, 2007] Соснин, П. И. Вопросно-ответное моделирование в разработке автоматизированных систем / П. И. Соснин. – Ульяновск : УлГТУ, 2007. – 333 с.
64. [Тарасов, 2008] Тарасов, В. Б. Построение виртуальных и интеллектуальных предприятий на основе виртуальных рабочих площадок / В. Б. Тарасов, Ю. В. Таратухина // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1.– М. : Физматлит, 2008. – С. 127-132.
65. [Тоценко, 2006] Тоценко, В. Г. Об унификации алгоритмов организации экспертиз / В. Г. Тоценко // Проблемы правовой информатизации. – 2006. – №2(12).
66. [Финаев и др., 2008 а] Финаев, В. И. Структура и задачи гибридных регуляторов / В. И. Финаев, И. С. Коберси // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 1. – М. : Физматлит, 2008. – С. 10-15.
67. [Финаев и др., 2008] Финаев, В. И. Особенности оценки эффективности предпринимательского проекта / В. И. Финаев, С. В. Немченко // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т.1. – М. : Физматлит, 2008. – С. 123-126.
68. [Хархаров, 2006] Хархаров, А. М. Маркетинговые исследования и их роль в организации адаптивного управления строительным предприятием /

- А. М. Хархаров // Транспортное дело России. Специальный научный выпуск. Инновационные проекты на транспорте. – 2006. – №5.
69. [Цирлов, 2008] Цирлов, В. Л. Разработка методики анализа уязвимости программного обеспечения автоматизированных систем / В. Л. Цирлов // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 4-х томах. Т. 3. – М. : Физматлит, 2008. – С. 382-385.
70. [Чернов, 2007] Чернов, В. Г. Модели поддержки принятия решений в инвестиционной деятельности на основе аппарата нечетких множеств / В. Г. Чернов. – М. : Горячая линия – Телеком, 2007. – 312 с.
71. [Чернов и др., 2006] Чернов, В. Г. Нечетко-множественные методы и модели в задачах антикризисного управления / В. Г. Чернов, М. К. Суворов ; под ред. О. И. Кирикова // Научные исследования: информация, анализ, прогноз. – Воронеж : ВГПУ, 2006. – Книга 10. – С. 185-217.
72. [Шабельников и др., 2009] Шабельников, А. Н. Интеллектуальные системы распределенного мониторинга на основе беспроводных сенсорных сетей с использованием системы мобильных объектов / А. Н. Шабельников, В. А. Шабельников, С. М. Ковалев // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.) Т. 1. – М. : Физматлит, 2009. – С. 538-543.
73. [Штовба, 2007] Штовба, С. Д. Проектирование нечетких систем средствами MATLAB/ С. Д. Штовба. – М. : Горячая линия – Телеком, 2007. – 288 с.
74. [Ярушкина, 1997] Ярушкина, Н. Г. Методы нечетких экспертных систем в интеллектуальных САПР / Н. Г. Ярушкина. – Саратов : Изд-во Саратов. ун-та, 1997.
75. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.

76. [Ярушкина и др., 2005] Ярушкина, Н. Г. Компоненто-ориентированная INTERNET интегрированная среда для экспертизы эффективности деятельности крупных предприятий / Н. Г. Ярушкина, И. В. Семушин, А. А. Стецко // Новости искусственного интеллекта. – 2005. – №3.
77. [Ярушкина и др., 2007 а] Ярушкина, Н. Г. Нечеткие временные ряды как инструмент для оценки и измерения динамики процессов / Н. Г. Ярушкина, Т. В. Афанасьева, Т. Р. Юнусов // Датчики и системы. – 2007. – № 12. – С. 46-51.
78. [Ярушкина и др., 2007 б] Ярушкина, Н. Г. Нечеткие временные ряды в задачах экспертной деятельности / Н. Г. Ярушкина, Т. В. Афанасьева // Информационная среда вуза XXI века. Материалы Всероссийской научно-практической конференции. Петрозаводск, (3 - 8 сентября 2007 года). – Петрозаводск, 2007. – С. 88-90.
79. [Ярушкина и др., 2008] Ярушкина, Н. Г. Экспертная система анализа экологической безопасности / Н. Г. Ярушкина, Н. Н. Ястребова, И. С. Ястребов // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября - 3 октября, 2008 г., г. Дубна, Россия): Труды конференции. Т. 2. – М. : ЛЕНАНД, 2008. – С. 278-286.
80. [Яхьяева, 2006] Яхьяева, Г. Э. Нечеткие множества и нейронные сети : учебное пособие / Г. Э. Яхьяева. – М. : Интернет-Университет Информационных технологий: БИНОМ. Лаборатория знаний, 2006. – 316 с.

ГЛАВА 2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ В МОДЕЛИРОВАНИИ НЕЧЕТКИХ ВРЕМЕННЫХ РЯДОВ

Введение

Природа нечетких временных рядов обусловлена использованием экспертных оценок, присущая которым неопределенность относится к классу нечеткости. В отличие от стохастической неопределенности нечеткость затрудняет или даже исключает применение статистических методов и моделей, но может быть использована для принятия предметно-ориентированных решений на основе приближенных рассуждений человека. Формализация интеллектуальных операций, моделирующих нечеткие высказывания человека о состоянии и поведении сложных явлений, образует сегодня самостоятельное направление научно-прикладных исследований, получившее название «нечеткое моделирование». Указанное направление включает комплекс задач, методология решения которых опирается на теорию нечетких множеств, нечеткой логики, нечетких моделей (систем) и гранулярных вычислений. Результаты исследований в этом направлении оформлены в виде методов нечеткого моделирования и представлены в ряде работ [Zadeh, 1965; Заде, 1974; Штовба, 2007; Аверкин, 1986; Борисов, 1989; Ротштейн, 1999; Ярушкина, 2004; Яхьяева, 2006; Павлов, 2006; Батыршин и др., 2007; Борисов и др., 2007].

Основные проблемы, решаемые в нечетком моделировании, связаны с моделированием интеллектуальных операций приближенных рассуждений человека (эксперта), а также объектов, над которыми эти операции выполняются:

1. Объектами интеллектуальных операций, используемых в приближенных рассуждениях человека, являются переменные нового класса – лингвистические переменные, значениями которых являются нечеткие множества. Важным является тот факт, что наименования лингвистической переменной и ее значений должны соответствовать словам, которые использует человек при решении прикладных задач. Таким образом, операндами и результатом интеллектуальных операций являются значения особого вида – *нечеткие множества*.

2. Основными интеллектуальными операциями, используемыми в приближенных рассуждениях человека являются: определение семантики, смысла высказываний, определение «истинности» элементарных и составных высказываний, вывод «истинности» высказывания на основе логических рассуждений. Математические модели выделенных интеллектуальных операций строятся с помощью *операций нечеткой логики*.

3. Алгоритмы вычисления нечетких значений представляют новый класс вычислительных моделей. Такие модели предназначены для манипулирования со значениями, представленными нечеткими множествами на основе операций нечеткой логики, поэтому они классифицируются как нечеткие системы логического вывода. Часто используют сокращенную форму обозначенного класса моделей – *нечеткие модели или нечеткие системы*.

2.1. Нечеткие множества

Теория нечетких множеств, введенная Л. Заде [Zadeh, 1965] для представления нового типа значений, заложила основы моделирования интеллектуальной деятельности человека и явилась начальным толчком к развитию новой математической теории, используемой для описания неопределенностей, относящихся к классу нечеткости высказываний и рассуждений человека. Теория нечетких множеств – это раздел прикладной математики, посвященный методам анализа неопределенных данных, в которых описание неопределенностей реальных явлений и процессов проводится с помощью понятия о множествах, не имеющих четких границ.

Такого вида неопределенности объективно существуют в словах и лингвистических выражениях, к которым относятся экспертные оценки. Лингвистические выражения, в общем, могут быть взяты как имена свойств. Каждое свойство имеет *сущность понятия* – абстрактную конструкцию, характеризующую свойство во всех контекстах. Следовательно, мы можем сделать лингвистические выражения именами сущностей понятий.

В соответствии с этим, в дальнейшем для указания неопределенности экспертных оценок будем использовать эквивалентное выражение – лингвистическая неопределенность.

Вопрос, связанный с математическим моделированием лингвистической неопределенности является ключевым в теории нечетких множеств.

Л. Заде предложил по аналогии с теорией вероятности использовать функцию в качестве математической модели лингвистической неопределенности объекта $x \in X$:

$$Y = \mu(x, B),$$

где Y – результат вычисления функции, выражающий меру неопределенности (нечеткости) для конкретного объекта $x \in X$;

μ – непрерывная функция, такая, что $\mu: X \rightarrow [0, 1]$. Содержательно функция μ определяет распределение неопределенности на X ;

X – область определения функции μ . Область определения задается упорядоченным множеством значений, произвольной природы, называемым *универсальным множеством (или универсумом)*. Носителем функции $\mu(x, B)$ является подмножество $w \subset X$, на котором функция $\mu(x, B)$ принимает значение, отличное от нуля. В качестве универсального множества обычно задается множество действительных чисел;

B – вектор параметров функции, обычно числовых.

Функциональная модель лингвистической неопределенности получила название *нечеткого множества*, так как указанная функция μ рассматривается как характеристическая функция, определенная на множестве объектов X . Таким образом, с математической точки зрения нечеткое множество моделируется параметрической функцией особого класса, называемого классом *функций принадлежности*. В том случае, если значения функции принадлежности нечеткого множества представлены точными числовыми значениями, такие нечеткие множества относят к *нечетким множествам типа 1*. Если значения функции принадлежности нечеткого множества моделируются другими нечет-

кими множествами, то такое нечеткое множество относят к *нечетким множествам типа 2*.

На практике используют несколько способов задания функции принадлежности, среди них выделим следующие:

1. Структурный способ. Данная форма определения нечетких множеств основана на табличном представлении функций. В случае, если известен вектор параметров B , табличное представление функции принадлежности может быть задано явно посредством табулирования функции $Y = \mu(x, B)$ на множестве значений w , являющемся ее носителем. При неизвестном векторе параметров B – путем прямого перечисления множества пар в виде $\mu = \{x_1/y_1, x_2/y_2, \dots, x_n/y_n\}$. Данная форма удобна для графического отображения нечеткого множества и используется часто в тех случаях, когда затруднительно задать математический вид функции $Y = \mu(x, B)$, например, если X не является множеством чисел.
2. Функциональный способ. При этом предполагается, что форма функции принадлежности, моделирующей нечеткое множество, известна и определена на множестве действительных чисел X . Для представления $Y = \mu(x, B)$ используют различные функции (рис. 2.1).

2.1. Гауссова функция принадлежности описывается вектором параметров $B = \{\sigma, c\}$:

$$\mu(x, B) = \exp\left(-\left(\frac{x - c}{\sigma}\right)^2\right),$$

где c – среднее значение;

σ – среднее квадратичное отклонение.

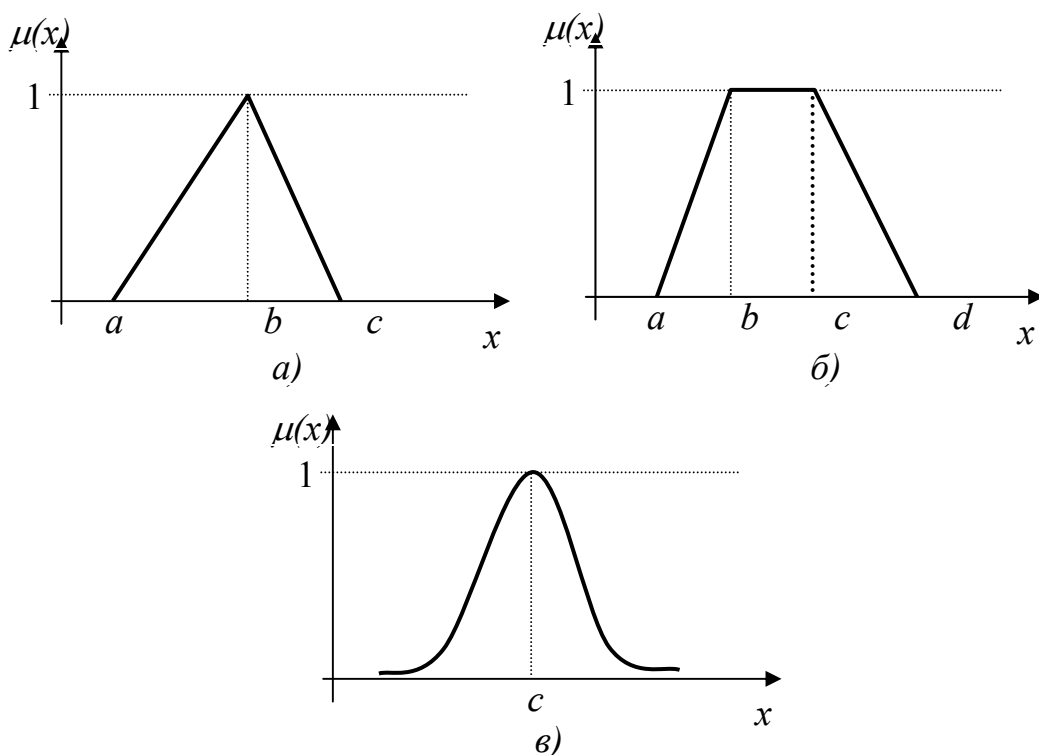


Рис. 2.1. Типовые формы функций принадлежности
 а) треугольная, б) трапецеидальная, в) гауссова

2.2. Треугольная функция принадлежности характеризуется тройкой чисел, $B=\{a, b, c\}$, и вычисляется по формуле:

$$\mu(x, B) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b < x \leq c \\ 0, & x < a, x > c, \end{cases}$$

где b – задает координату вершины треугольника;

a, c – определяют основание треугольника.

2.3. По аналогии задается и трапецеидальная функция принадлежности, которая характеризуется четверкой чисел $B=\{a, b, c, d\}$:

$$\mu(x, B) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x < c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x < a, x > d. \end{cases}$$

На практике часто параметр B явно не указывается для обеспечения более компактной записи функции принадлежности, то есть используется функциональная запись вида $Y=\mu(X)$ вместо $Y=\mu(x, B)$. При дальнейшем изложении в учебном пособии будем использовать компактную запись функции принадлежности.

С каждой функцией принадлежности $\mu(X)$ сопоставляется лингвистическое обозначение нечеткого множества (лингвистический терм) для различения разных нечетких множеств. Тогда функция принадлежности нечеткого множества Z (функциональный способ) будет иметь следующую запись $Z=\mu_Z(X)$. Расширив традиционное понятие множества, Л. Заде построил «математический мостик» при описании свойств понятий в виде нечетких множеств между числом x , свойством Z , выраженным лингвистически, и степенью соответствия числа x свойству Z в виде функции $\mu_Z(x)$. Фактически, обозначение нечеткого множества через Z позволяет именовать функцию принадлежности $\mu_Z(X)$, в общем случае параметрическую, лингвистическими терминами, то есть оперировать с ней как со значениями лингвистической переменной. Часто эти два понятия – Z и $\mu_Z(X)$ – рассматриваются как эквивалентные.

Пусть $X = \{x\}$ – совокупность объектов (универсальное множество), обозначаемых через x . Пусть на X определены три нечетких множества: A =«низкое», B =«удовлетворительное», C =«хорошее», обозначающие имена свойств понятия «качество». Тогда, следуя структурному способу для всех $x \in X$ нечеткое множество A может быть задано совокупностью упорядоченных пар $A = \{x, \mu_A(x)\}$, нечеткое множество B – совокупностью упорядоченных пар $B = \{x, \mu_B(x)\}$, нечеткое множество C – совокупностью упорядоченных пар $C = \{x, \mu_C(x)\}$. Используя функциональный способ записи, приведенный выше,

нечеткие множества будут представлены следующим образом: $A = \mu_A(x)$, $B = \mu_B(x)$, $C = \mu_C(x)$.

Введение лингвистических терминов для обозначения m свойств понятия естественного языка с помощью аппарата нечетких множеств позволяет сформулировать следующие задачи:

- 1) Определение множества лингвистических термов, задающих свойства понятия;
- 2) разбиение множества действительных чисел $X = \{x\}$, на котором определено лингвистическое понятие, на подмножества (лингвистические термы), характеризующие свойства понятия;
- 3) сопоставление лингвистическому терму (слову естественного языка) семантики, выраженной функцией принадлежности;
- 4) определение степени «истинности» нечеткого высказывания.

Контрольные вопросы

1. Почему для нечетких временных рядов затруднительно применять классические статистические методы и модели?
2. Определите понятие нечеткое моделирование. На чем строится решение задач нечеткого моделирования?
3. Назовите ряд особенностей, связанных с нечетким моделированием, включающих следующие понятия: нечеткие множества, операции нечеткой логики, нечеткие модели или нечеткие системы.
4. Сформулируйте понятие лингвистической неопределенности.
5. Напишите функцию в математической модели лингвистической неопределенности объекта.
6. Дайте определение функции принадлежности.
7. К какому типу относятся нечеткие множества, если значения функции принадлежности нечеткого множества представлены точными числовыми значениями?

8. К какому типу относятся нечеткие множества, если значения функции принадлежности нечеткого множества моделируются другими нечеткими множествами?
9. Какие существуют способы задания функции принадлежности?
10. Приведите три примера функции принадлежности, задаваемых функциональным способом.
11. Напишите формулу компактной записи функции принадлежности.

2.2. Лингвистические переменные

Лингвистическая переменная – это переменная, значениями которой являются слова или высказывания естественного или искусственного языка.

«Поскольку слова в общем смысле менее точны, чем числа, понятие лингвистической переменной дает возможность приближенно описывать явления, которые настолько сложны, что не поддаются описанию в общепринятых количественных терминах... высокая точность несовместима с высокой сложностью. Таким образом, быть может, именно по этой причине обычные методы анализа систем и моделирования на ЭВМ, основанные на точной обработке численных данных, по существу не способны охватить огромную сложность процессов человеческого мышления и принятия решений. Отсюда напрашивается вывод о том, что для получения существенных выводов о поведении гуманистических систем придется, по-видимому, отказаться от высоких стандартов точности и строгости, которые мы, как правило, ожидаем при математическом анализе четко определенных механистических систем, и относиться более терпимо к иным подходам, которые являются приближенными по своей природе» [Заде, 1976].

Любая переменная описывается множеством допустимых значений, а лингвистические понятия описываются набором присущих им свойств. Как изменить понятие обычной лингвистической переменной, чтобы ее значения отображали семантический аспект задаваемого ею понятия и были пригодны для

вычисления новых значений? В ответ на этот вопрос Л. Заде расширил понятие обычной лингвистической переменной, допустив, что в качестве ее значений (термов) выступают нечеткие переменные [Заде, 1974].

Формально лингвистическая переменная описывается набором

$$\langle Name, \tilde{X}, X, G, P \rangle,$$

где $Name$ – наименование лингвистической переменной;

X – универсальное множество объектов x ;

$\tilde{X}$ – базовое терм-множество, образующее совокупность термов лингвистической переменной, например, $\tilde{X} = \{\text{«Отличный»}, \text{«Хороший»}, \text{«Плохой»}, \text{«Удовлетворительный» и др.}\}$;

G – синтаксические правила вывода (порождения) новых термов $\tilde{X}^*$, не входящих в базовое терм-множество, задаваемые обычно на основе контекстно-свободной грамматики;

P – семантические правила, контекстно-зависимый способ вычисления смысла на основе функций принадлежности каждого терма из $\tilde{X} \cup \tilde{X}^*$. Пример лингвистической переменной, заимствованный из [Штовба, 2007], представлен на рис. 2.2.

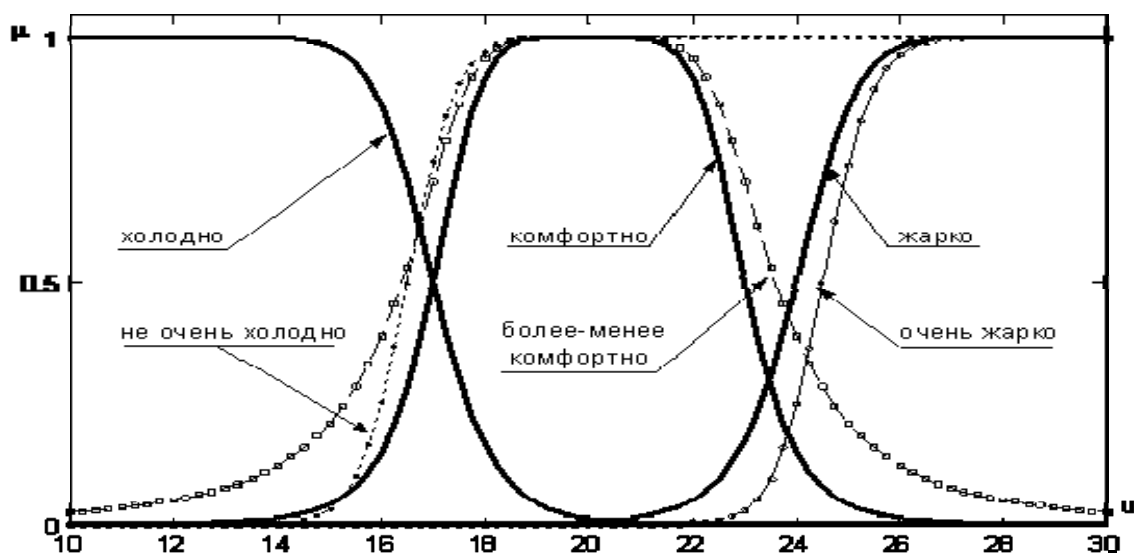


Рис. 2.2. Пример лингвистической переменной «Температура»

Введение лингвистической переменной, обозначающей понятия естественного языка, значения которой описывают свойства этого понятия функциями принадлежности, позволило продвинуться на пути формализации операций вычислений со словами с помощью нечеткой логики.

Контрольные вопросы

1. Сформулируйте определение лингвистической переменной.
2. Опишите набор переменных, с помощью которого описывается лингвистическая переменная?
3. Дайте определения следующим понятиям: X – универсальное множество объектов, $\tilde{X}$ – базовое терм-множество, G – синтаксические правила вывода (порождения) новых термов, P – семантические правила.

2.3. Нечеткая логика

Нечеткая логика – это набор стандартной (Булевой) логики, которая была расширена до обработки понятий частичной истинности для оперирования нечеткими высказываниями и рассуждениями, так как человеческие рассуждения носят приблизительный и нечеткий характер. Л. Заде обосновал следующие необходимые характеристики нечеткой логики:

- В нечеткой логике точное рассуждение рассматривается как частный случай приблизительного рассуждения,
- В нечеткой логике все – вопрос степени,
- Любая логическая система может быть фаззифицирована,
- В нечеткой логике знание интерпретируется как совокупность нечетких ограничений на совокупность переменных.

Важной характеристикой нечеткой логики является то, что любая теория может быть фаззифицирована (fuzzified) и, следовательно, обобщена путем замены понятия четкого множества в теории понятием нечеткого множества. Выигрышем от фаззификации является большая общность и лучшее соответствие модели действительности.

В основе операций нечеткой логики лежит понятие нечеткого множества, выраженного функцией принадлежности. Поэтому операндами и результатами операций нечеткой логики являются также функции, определяющие новые нечеткие множества. В нечеткой логике для моделирования основных логических связок И($\wedge$), ИЛИ($\vee$) над нечеткими множествами используют триангулярные нормы [Батыршин, 2001 а].

Триангулярной нормой (t-нормой) называют отображение $T : [0,1] \times [0,1] \rightarrow [0,1]$, удовлетворяющее следующим условиям:

- 1) $T(0, 0)=0$; $T(x, 1)=x$; $T(1, x)=x$ – ограниченность;
- 2) $T(x, y) \leq T(a, b)$, если $x \leq a$, $y \leq b$ – монотонность;
- 3) $T(x, y) = T(y, x)$ – коммутативность;
- 4) $T(x, T(y, z)) \leq T(T(x, y), z)$ – ассоциативность.

Триангулярной конормой (s-конормой) называют отображение $S : [0,1] \times [0,1] \rightarrow [0,1]$, удовлетворяющее следующим условиям:

- 1) $S(1, 1)=1$; $S(x, 0)=x$; $S(0, x)=x$ – ограниченность;
- 2) $S(x, y) \geq S(a, b)$, если $x \geq a$, $y \geq b$ – монотонность;
- 3) $S(x, y) = S(y, x)$ – коммутативность;
- 4) $S(x, S(y, z)) \leq S(S(x, y), z)$ – ассоциативность.

t-норма и s-конорма в определенном смысле являются двойственными понятиями. Эти функции могут быть получены из друг друга, например, с помощью инволютивного отрицания и законов Де Моргана следующим образом:

$$S(x, y) = n(T(n(x), n(y))) , \quad T(x, y) = n(S(n(x), n(y))) .$$

Простейшими примера t-норм и s-конорм, взаимно связанных этими соотношениями для $n(x)=1-x$, являются следующие:

$T(x, y) = \min\{x, y\}$	(минимум)
$S(x, y) = \max\{x, y\}$	(максимум)
$T(x, y) = xy$	(произведение)
$S(x, y) = x+y-xy$	(вероятностная сумма)
$T(x, y) = \max\{x+y-1, 0\}$	(t-норма Лукасевича)
$S(x, y) = \min\{x+y, 1\}$	(t-конорма Лукасевича ограниченная сумма)

Дальнейшее исследование в области нечеткой логики триангулярных норм связано с введением на них параметрических классов [Ярушкина, 2004; Батыршин и др., 2007].

По существу, все человеческие понятия являются нечеткими, так как они получаются в результате группировки (clumping) точек или объектов, объединяемых по сходству. Тогда нечеткость подобных групп (clumps) есть прямое следствие нечеткости понятия сходства. Простыми примерами таких групп являются понятия «средний возраст», «деловая часть города», «немного облачно», «бестолковый» и др. Данную группу в нечеткой логике называют «гранулой» (*granule*). В естественном языке (ЕЯ) слова играют роль меток гранул и служат для сжатия данных. Сжатие данных с помощью слов является ключевым аспектом человеческих рассуждений и формирования понятий.

В нечеткой логике гранулирование информации лежит в основе понятий лингвистической переменной и нечетких правил типа «ЕСЛИ-ТО», задаваемой операцией импликации.

Операция импликации

В качестве основного математического инструмента при определении импликации $A \rightarrow B$ для нечетких множеств A и B используют композиционное правило Л. Заде [Заде, 1974], являющееся обобщением правила *modus ponens*:

Предпосылка	$A \rightarrow B$
Событие	A^*
Вывод	$A^* \circ (A \rightarrow B)$

Пусть U и V – два универсальных множества с базовыми переменными u и v соответственно. Пусть A и F – нечеткие подмножества множеств U и $U \times V$. Тогда композиционное правило вывода утверждает, что из нечетких множеств A и F следует нечеткое множество $B = A \circ F$. Функция принадлежности результата вычисляется с помощью триангулянтных норм следующим образом

$$\mu_B(v) = S(T(\mu_A(u), \mu_F(u, v))),$$

при моделировании s-конормы и t-нормы операциями $(\vee) \max$ и $(\wedge) \min$ соответственно

$$\mu_B(v) = \bigvee_{u \in U} ((\mu_A(u) \wedge \mu_F(u, v))).$$

Другие формулы, реализующие операцию импликации и математические объекты теории нечетких множеств и нечеткой логики приведены в работах [Ярушкина, 2004; Яхьяева, 2006; Штовба, 2007; Батыршин и др., 2007].

Формализация нечеткой импликации позволила задать правила «ЕСЛИ-ТО» в виде нечетких продукционных правил и заложило основу нечеткого моделирования опыта и знаний экспертов, выраженных в виде приближенных зависимостей.

Контрольные вопросы

1. Сформулируйте определение нечеткой логики.
2. Какие необходимые характеристики нечеткой логики ввел Л. Заде?
3. Что лежит в основе операций нечеткой логики?
4. Какие операции используются для моделирования основных логических связок И($\wedge$), ИЛИ($\vee$) над нечеткими множествами в нечеткой логике?

5. Сформулируйте определения триангулярной нормы и триангулярной конормы (t-норма и s-конорма) и докажите их двойственность. Приведите примеры.
6. Запишите композиционное правило Л. Заде.
7. Сформулируйте содержательное определение операции импликации.

2.4. Нечеткие модели и системы

Модели статических и динамических систем, построение, использование и анализ которых базируется на положениях теории нечетких множеств и нечеткой логики называют *нечеткими моделями* или *нечеткими системами*.

Целью нечеткого моделирования сложных явлений является приближенное описание зависимости (аппроксимация некоторой функции)

$$Y=f(X),$$

где Y – выходная лингвистическая переменная;

X – вектор входных лингвистических переменных размерностью n ;

f – зависимость между X и Y , описываемая совокупностью нечетких продукционных правил.

Нечеткие модели представляют обобщение интервально-оцениваемых моделей, которые, в свою очередь, являются обобщением четких моделей. Классификация и сферы применения нечетких моделей подробно рассмотрены в работе [Борисов и др., 2007].

В основе нечетких продукционных моделей лежат совокупность нечетких правил «ЕСЛИ-ТО», описывающих зависимости между нечеткими переменными предметной области, композиционное правило вывода и способ вычисления значений нечетких переменных (способ нечеткого вывода).

Модель описания поведения систем на естественном (или близком к естественному) языке в виде приближенных рассуждений в теории нечетких множеств и нечеткой логики, основанная на композиционном правиле вывода, называется *системой нечеткого логического вывода*.

В систему нечеткого логического вывода входят следующие объекты (рис. 2.3):

- 1) совокупность нечетких продукционных правил (база правил);
- 2) набор функций принадлежности базы нечетких переменных (база переменных);
- 3) блок фаззификации;
- 4) блок дефаззификации;
- 5) блок вывода.

База правил хранит множество логических правил вывода, а также их порядок (иерархическую структуру) применения. База нечетких переменных содержит названия лингвистических термов и параметры их функций принадлежности. База правил вместе с базой нечетких переменных образуют базу знаний (БЗ) системы нечеткого вывода.

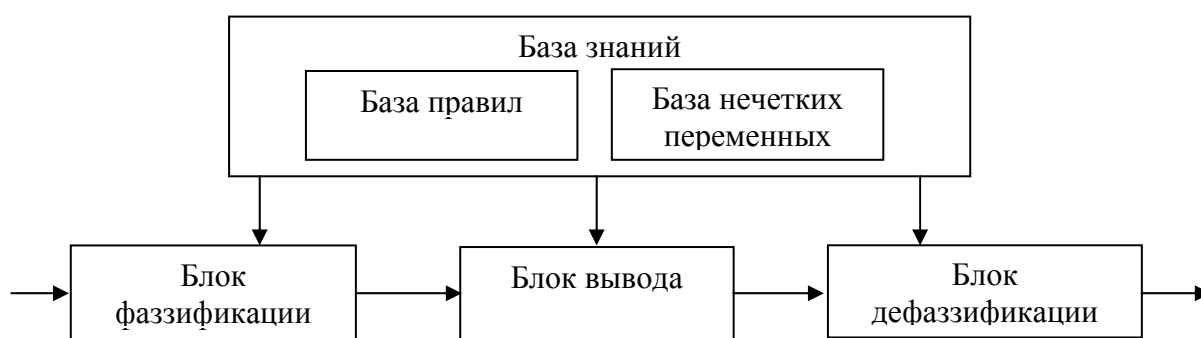


Рис. 2.3. Система нечеткого вывода

Простейшие системы нечеткого логического вывода основаны на правилах вида:

R_i : Если X есть A_i и Y есть B_i , то Z есть C_i ,

R_i : Если X есть A_i и Y есть B_i , то $z=f_i(x,y)$,

где X, Y – входные нечеткие переменные;

Z – выходная нечеткая переменная;

A_i, B_i – входные значения (функции принадлежности);

C_i – выходные нечеткие значения (функции принадлежности);

f_i – некоторые вещественные функции.

При этом должны соблюдаться следующие условия:

- 1) Существует хотя бы одно правило для каждого лингвистического термина выходной переменной.
- 2) Для любого термина входной переменной имеется хотя бы одно правило, в котором этот терм используется в качестве предпосылки (левая часть правила).

В противном случае имеет место неполная база нечетких правил.

Распространены пять способов реализации нечеткого логического вывода [Ярушкина, 2004 а].

Схема 1: Алгоритм Мамдани (Mamdani). Импликация моделируется минимумом, а агрегация – максимумом.

Схема 2: Алгоритм Цукамото (Tsukamoto). Исходные посылки – как у предыдущего алгоритма, но предполагается, что функции принадлежности являются монотонными.

Схема 3. Алгоритм Суджено (Sugeno). Алгоритм предполагает, что правые части правил вывода представлены в виде линейных функций.

Схема 4. Алгоритм Ларсена (Larsen). В алгоритме Ларсена нечеткая импликация моделируется с использованием операции умножения.

Схема 5. Упрощенный алгоритм нечеткого вывода. Исходные правила в данном случае задаются в виде:

$$\text{Если } X \text{ есть } A_i \text{ и } Y \text{ есть } B_i, \text{ то } z = Z_i,$$

где Z_i – четкое значение.

Рассмотрим алгоритм нечеткого вывода по схеме Мамдани для базы правил вида

$$R_i: \text{Если } X \text{ есть } A_i \text{ и } Y \text{ есть } B_i, \text{ то } Z \text{ есть } C_i.$$

1. Фаззификация.

Определяются степени истинности по функциям принадлежности для левых частей каждого правила:

$$a_i = \mu_{A_i}(X)$$

$$b_i = \mu_{B_i}(Y),$$

где a_i – степень принадлежности X к A_i ;

b_i – степень принадлежности Y к B_i ;

$i = [1, r]$;

r – количество правил.

2. Импликация.

Определяется сила каждого правила, t -нормой является логический минимум:

$$\alpha_i = \min(a_i, b_i).$$

Модифицируются функции принадлежности переменной z в каждом правиле:

$$\mu'_i(z) = \min(\alpha_i, \mu_i(z)),$$

где $\mu_i(z)$ – функция принадлежности переменной Z_i .

3. Агрегация.

Объединение выходов каждого правила логическим максимумом (s -конорма):

$$\mu'(z) = \max_i(\mu'_i(z)), i = [1, r].$$

4. Дефазификация.

В теории нечетких множеств процедура дефазификации аналогична нахождению характеристик положения случайных величин в теории вероятности. Простейшим способом выполнения процедуры дефазификации является выбор четкого числа, соответствующего максимуму функции принадлежности. Однако пригодность этого способа ограничивается одноэкстремальными функциями принадлежности. Для многоэкстремальных функций на практике используется часто метод центра тяжести.

Дефаззификация нечеткого множества по методу центра тяжести осуществляется по формуле:

$$x^0 = \frac{\int x \cdot \mu(x) dx}{\int \mu(x) dx}.$$

Физическим аналогом этой формулы является нахождение центра тяжести плоской фигуры, ограниченной осями координат и графиком функции принадлежности нечеткого множества. В случае дискретного универсального множества дефаззификация нечеткого множества по методу центра тяжести осуществляется по формуле:

$$x^0 = \frac{\sum_{i=1}^k x_i \cdot \mu(x_i)}{\sum_{i=1}^k \mu(x_i)}.$$

На рисунке 2.4 графически представлен процесс нечеткого вывода по алгоритму Мамдани.

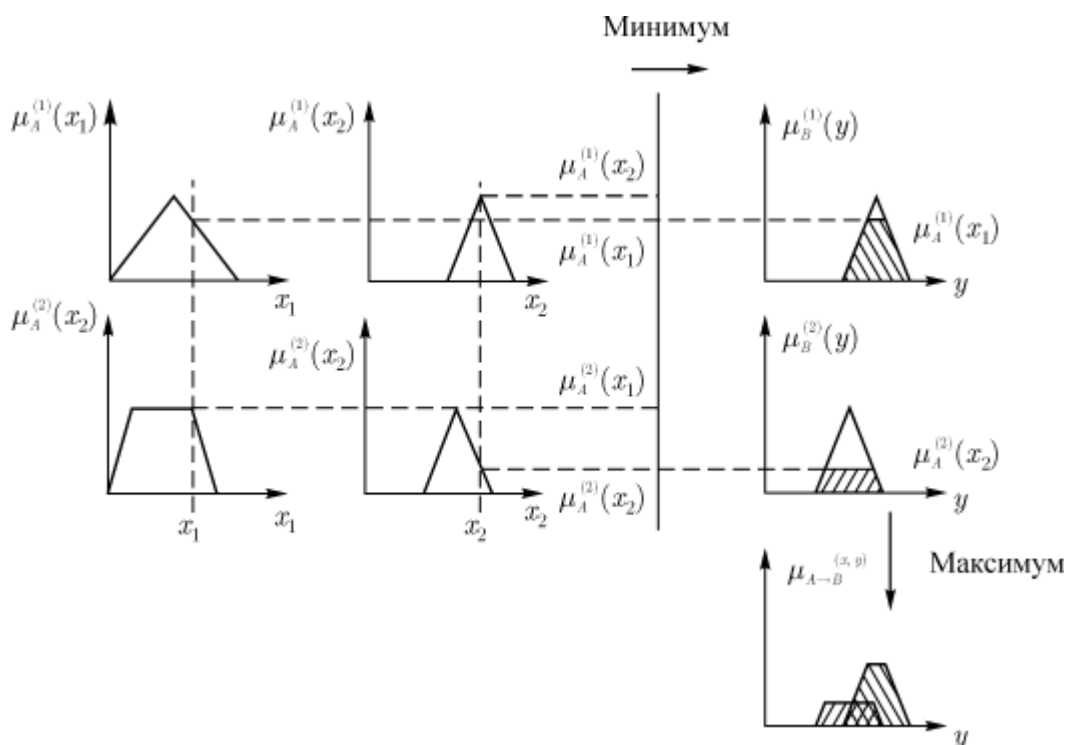


Рис. 2.4. Алгоритм нечеткого вывода Мамдани

Математическую основу для широкого применения нечеткой логики составляет доказательство в конце 80-х Бартоломеем Коско знаменитой теоремы

FAT (Fuzzy Approximation Theorem), согласно которой любая математическая система может быть аппроксимирована системой, основанной на нечеткой логике. Другими словами, с помощью естественно-языковых высказываний «ЕСЛИ-ТО», с последующей их формализацией средствами теории нечетких множеств, можно сколько угодно точно описать произвольную взаимосвязь «входы-выход» без использования сложного аппарата дифференциального и интегрального исчислений, традиционно применяемого в управлении и идентификации.

Свойство универсальности применения систем нечеткого вывода доказано рядом фундаментальных теорем. Так, У. Ванг в 1992 году показал, что справедливо утверждение: если нечеткая импликация основана на использовании операции *min*, функция принадлежности задается гауссовым распределением и используется центроидный метод дефаззификации, то система нечеткого вывода является универсальным аппроксиматором.

В 1995 г. К. Кастро доказал справедливость следующей теоремы: если импликация основана на использовании операции произведения по Ларсену, а функции принадлежности треугольные, то при использовании центроидной дефаззификации нечеткий контроллер является универсальным аппроксиматором.

Более подробное изложение вопросов классификации, проектирования, анализа аппроксимационных свойств нечетких систем приведены в [Ярушкина, 2004; Борисов и др., 2007].

Контрольные вопросы

1. Определите понятия нечетких моделей.
2. Что является целью нечеткого моделирования сложных явлений?
3. Напишите переход от нечетких моделей к четким через интервально-оцениваемые модели.
4. Сформулируйте определение системы нечеткого логического вывода.
5. Какие объекты входят в систему нечеткого логического вывода?
6. Какие условия должны соблюдаться при построении правил на основе системы нечеткого логического вывода?

7. Перечислите пять способов реализации нечеткого логического вывода.
8. Подробно опишите реализацию нечеткого логического вывода с помощью алгоритма Мамдани. Нарисуйте схему процесса нечеткого вывода этого алгоритма.

2.5. Обзор практических применений нечетких моделей

Настоящий обзор составлен с использованием материалов из [Штовба, 2007].

Историческая справка

Началом практического применения теории нечетких множеств можно считать 1975 г., когда Мамдани и Ассилиан (Mamdani and Assilian) построили первый нечеткий контроллер для управления простым паровым двигателем.

Рассмотрим историю создания системы нечеткого логического вывода, реализованного Мамдани [Круглов и др. 2001]. В Англии, в 1973 году, в Лондонском университете профессор и студент пытались стабилизировать скорость небольшого парового двигателя, построенного студентом. У них было современное оборудование, такое как PDP-8 микрокомпьютер и обычные цифровые контроллеры. Но им не удавалось контролировать двигатель так, как им хотелось. Частота вращения двигателя либо превышала целевую скорость и приходила к ней после ряда колебаний или скорость нарастала слишком медленно, медленнее, чем требовалось для быстрого достижения требуемого параметра.

Профессор Мамдани и студент С. Ассилиан решили попробовать применить нечеткую модель для реализации алгоритма управления паровым двигателем на основе правил нечеткого логического вывода. Результат, полученный с использованием контроллера на нечеткой логике, оказался лучше, чем при использовании других способов, скорость приблизилась к желаемой очень быстро, без перепадов, и оставалась стабильной. Это был интересный и важный момент в истории развития науки.

Д-р Мамдани пришел к выводу, что при использовании традиционного подхода требуется большое количество экспериментов (проб и ошибок) для

достижения заданной скорости в требуемом множестве точек. Кроме того, в связи с нелинейностью рабочих характеристик парового двигателя, как только множество точек скорости было изменено, нужно снова повторять пробы и ошибки, чтобы вновь получить эффективное управление. Этого не происходило с нечетким контроллером, который гораздо лучше приспособлен к изменениям, вариациям и нелинейности в системе.

В 1982 Холмблад и Остергад (Holmblad and Osregaad) разработали первый промышленный нечеткий контроллер, который был внедрен в управление процессом обжига цемента на заводе в Дании. Успех первого промышленного контроллера, основанного на нечетких лингвистических правилах «ЕСЛИ-ТО», привел к всплеску интереса к теории нечетких множеств среди математиков и инженеров. Одновременно стало уделяться внимание вопросам построения экспертных систем, построенных на нечеткой логике, разработке нечетких контроллеров. Нечеткие экспертные системы для поддержки принятия решений находят широкое применение в медицине и экономике. Они применяются в автомобильной, аэрокосмической и транспортной промышленности, в области изделий бытовой техники, в сфере принятия управленческих решений и многих других. В бизнесе и финансах нечеткая логика получила признание после того, как в 1988 году экспертная система на основе нечетких правил для прогнозирования финансовых индикаторов единственная предсказала биржевой крах.

Практический опыт разработки систем нечеткого логического вывода свидетельствует, что сроки и стоимость их проектирования значительно меньше, чем при использовании традиционного математического аппарата, при этом обеспечивается требуемый уровень робастности и прозрачности моделей. Помимо этого нечеткие системы позволяют повысить качество продукции при уменьшении ресурсо- и энергозатрат и обеспечивают более высокую устойчивость к воздействию случайных факторов по сравнению с традиционными системами.

Приложения теории нечетких систем в автомобилестроении

(Пример заимствован из работы [Никольский, 2001]).

Первыми на нечеткую логику обратили внимание японские автомобилестроители. В 1991 году компания Nissan впервые применила компоненты нечеткой логики в системе управления пятискоростной автоматической коробкой переключения передач (АКПП), годом позже аналогичная система появилась на автомобилях Honda. Тогда же Mitsubishi Motors представила модель Lancer с системой АБС на основе процессора с нечеткой логикой. Затем концерн General Motors применил подобную систему для управления АКПП, к этому времени на том же Nissan была внедрена нечеткая логика в системах управления впрыском топлива для бензиновых двигателей. Ко второй половине прошлого десятилетия системами с использованием нечеткой логики штатно оснащались машины уже упомянутых Nissan, Mitsubishi и Honda, а BMW, Hyundai, Mazda, Mercedes и Peugeot планировали внедрить такие системы. Типичный пример системы, хорошо поддающейся реализации с помощью нечеткой логики, – АБС – *антиблокировочная тормозная система*. Реализаций АБС существует множество, но в общем случае управление осуществляется по двум входным параметрам: проскальзыванию колеса (отношение скорости автомобиля к мгновенной линейной скорости точки на внешнем радиусе колеса относительно его центра) и радиальному ускорению колеса.

Оба параметра представляются в виде логических переменных с набором из 5-8 термов каждая (например «отсутствует», «слабое», «среднее», «сильное», «очень сильное» и т. п.), на основании которых вычислитель, используя набор правил (их количество равно произведению количества термов входных переменных), получает значение давления в тормозном цилиндре, стремясь к поддержанию оптимального проскальзывания. Подобная задача может решаться и классическими вычислителями с помощью трехмерных таблиц, описывающих плоскость выходного значения в зависимости от двух входных.

Еще один кандидат на «нечеткое управление» – двигатель внутреннего сгорания (ДВС). Сложность систем управления ДВС в последние годы значительно возросла – как в связи с ужесточением экологических норм и требований к снижению расхода топлива, так и вследствие форсирования двигателей.

Несмотря на то, что основных параметров регулирования всего два – подача топлива и момент зажигания – системы управления типа PID-регуляторов в данном случае не годятся, так как алгоритм управления в значительной степени зависит от скорости вращения двигателя и нагрузки. Полная математическая модель ДВС слишком сложна и до сих пор не создана. Из-за этого большинство систем управления ДВС используют табличную модель, полученную экспериментальным путем на испытаниях и с учетом опыта экспертов. Серьезный недостаток такой модели – сложность создания многомерных таблиц и большой объем памяти, требуемый для их записи, если выходной параметр формируется в зависимости от трех и более входных. Нечеткая логика позволяет заменить таблицы правилами (несколько сотен) и реализовать управление по большому числу входных параметров.

В задаче с АКПП интересно то, что одним из «входных сигналов» для системы управления является... человек. Вернее, его действия в процессе движения, которые позволяют с некоторым приближением понять характер дороги или желания водителя. Частота нажатий на педаль акселератора позволяет судить, к примеру, об извилистости дороги, а амплитуда нажатий – о стиле вождения. Рассматриваются и более сложные ситуации, как, например, повторное нажатие педали акселератора в интервале 1-1,5 секунды после первого, что расценивается как недостаточное ускорение и вызывает принудительное включение низшей передачи. Система управления в данном случае выполняет функции эксперта, подстраиваясь под явные или неявные желания водителя и понемногу двигаясь в сторону искусственного интеллекта.

Применение нечетких систем в образовании

(По материалам [Домрачев, 2001]).

Существенной особенностью высшего образования является сложность количественного оценивания процессов обучения и управления. Однозначно понимаемого перечня показателей качества подготовки не существует, так как отсутствуют четкие представления о том, какие количественно измеримые факторы на него влияют, какими достоверно оценивающими показателями оно вы-

ражается, какова достоверность этих показателей и т. д. Нечеткость такого представления не позволяет методам математического моделирования получать адекватные количественные описания исследуемых параметров, а поэтому заставляет искать решения классических задач образовательного процесса неклассическими методами. Использование нечетких множеств для моделирования итоговых рейтинговых балльных оценок, как представляется, позволит более адекватно оценить учебные достижения студентов.

Контрольные вопросы

1. Кто построил первый нечеткий контроллер для управления простым паровым двигателем?
2. Какие полезные свойства были выявлены у первого нечеткого контроллера в отличии от традиционного подхода?
3. В каких отраслях применяются нечеткие экспертные оценки сегодня?
4. Приведите пример использования нечеткой логики в автомобилестроении.
5. Приведите пример использования нечеткой логики в образовании.

2.6. Современные тенденции развития теории нечетких множеств нечеткой логики и нечетких систем

Нечеткое моделирование как научное направление отметило в 2010 г. свое 45-летие. Ученые, занимающиеся нечеткой логикой, объединены в International Fuzzy Systems Association (IFSA), которая проводит один раз в два года мировые конгрессы. Материалы таких конгрессов естественным образом подытоживают работы за два прошедших года. Многие тенденции, разумеется, остаются значимы не в течение двух лет, а в течение больших периодов (5-7 лет). Ранее в статье [Ярушкина, 2003] были проанализированы материалы конгрессов IFSA'97 и IFSA'03 и сделаны некоторые прогнозы развития.

С момента выхода статьи состоялись 3 конгресса IFSA: IFSA'05 (КНР, г. Пекин, 2005) [15] и IFSA'07 (Мексика, г. Канкун, 2007) [16], IFSA-EUSFLAT 2009 (Португалия, г. Лиссабон, 2009).

Особый интерес на конгрессах вызывают доклады пленарных докладчиков. Сам выбор участников обычно позволяет уловить популярную тему в развитии нечеткой логики в настоящий момент. Программы пленарных докладов конгрессов IFSA'05, IFSA'07, IFSA-EUSFLAT 2009, а также программа лекций летних школ приведены в таблицах 2.1, 2.2, 2.3, 2.4.

Таблица 2.1

Пленарные доклады IFSA'05

Автор	Наименования докладов
Л. Заде	Toward a Computational Theory of Precisation of Meaning Based on Fuzzy Logic – The Concept of Cointensive Precisation Перспективы вычислительной теории уточнения смысла, основанной на нечеткой логике – концепция согласованной точности
И. Лиу, Д. Жанг	Order Structure, Topology and Fuzzy Sets Упорядоченные структуры, топология и нечеткие множества
Дж. Клир	From Natural Language to Formalized Language and Back От естественного языка к формализованному языку и обратно
Д. Дюбуа	On the Links between Probability and Possibility Theories О связи между теориями вероятностей и возможностей
В. Педрич	Knowledge-based Clustering for Human-Centric Systems Кластеризация, основанная на знаниях, для гуманистических систем

Таблица 2.2

Пленарные доклады IFSA'07

Автор	Наименования докладов
Л. Заде	Fuzzy Logic as the Logic of Natural Languages Нечеткая логика как логика естественных языков
Р. Ягер	Computing with Words and Granules Вычисления со словами и гранулами
К. Хирота	Control/Robotics Applications based on Soft Computing Technology Роботика и контроллеры, основанные на технологии мягких вычислений
Я. Капржык	Computing with words, usability qualification and linguistic quantifiers: tools for human-centric computing Вычисление со словами, квалификация полезности и лингвистические квантификаторы: инструменты для гуманистических вычислений
В. Педрич	Dynamic and Distributed (D^2) Fuzzy Modeling Динамичное и распределенное нечеткое моделирование

Дж. Мендель	Novel Weighted Averages as a Computing With Words Engine Метод взвешенных средних как машина вычисления со словами
Г. Чен	Uncertainty and Fuzziness in Knowledge Discovery from Large Databases Неопределенность и нечеткость в интеллектуальном анализе данных в больших базах данных
Дж. Келлер (приглашенный лектор)	OWA Operators for Gene Product Similarity, Clustering, and Knowledge Discovery OWA операторы для генетического поиска подобия, кластеризации и интеллектуального анализа данных

Таблица 2.3

Лекции школы «Мягкие вычисления и статистика» IFSA'09

Автор	Наименования докладов
Д. Дюбуа	A Unified View of Uncertainty Theories Обобщенный взгляд на теории неопределенности
В. Педрич	Fuzzy Modeling: Fundamentals, Design and Challenges Нечеткое моделирование: Основы, проектирование и перспективы
Дж. Баздек	Visual Clustering Methods Визуальные методы кластеризации
К. Боргелт	Fuzzy and Probabilistic Clustering Нечеткая и вероятная кластеризация
К. Боргелт	Artificial Neural Networks Искусственные нейронные сети
А. Гегов, Н. Петров	Introduction to Fuzzy Networks Введение в нечеткие сети

Окончание табл. 2.3

Д. Гонсалес-Родригес	Fuzzy Data in Statistics: Formalization and Main Problems Нечеткие данные в статистике: формализация и главные проблемы
Дж. Келлер	Soft Computing for Sensor and Algorithm Fusion Мягкие вычисления для сенсоров и алгоритмов слияния
Ф. Клавонн	Robust Statistics Робастная статистика
Б. Рейш	An Axiomatic Approach to the Notion of Rational Preference Structures Аксиоматический подход к определению структуры рациональных предпочтений
М. Верлейсен	Feature Selection Выбор свойств

Таблица 2.4

Пленарные доклады IFSA- EUSFLAT'09

Автор	Наименования докладов
К. Хирота, Ф. Донг	Casual Communication with Robots using Speech Recognition Module Взаимодействие с роботами, использующее модуль распознавания речи
Е. Халлермайер	Fuzzy Logic in Machine Learning Нечеткая логика в машинном обучении
М. Грабиш	Capacities and the Choquet integral in decision making: a survey of fundamental concepts and recent advances Интегралы Шоке в принятии решений: обзор фундаментальных понятий и текущие достижения
Н. Нурми	Fuzzy Systems, Choice Paradoxes and Optimal Committees Нечеткие системы, парадоксы выбора и оптимальность

Основные положения пленарных докладов. Комментарии

Основные положения докладов Л. Заде

Ключевым моментом конгрессов, конечно, являются доклады Л. Заде.

На IFSA'05 им был прочитан доклад «Перспективы вычислительной теории уточнения смысла основанного на нечеткой логике – концепция согласованной точности», а на IFSA'07 – «Нечеткая логика как логика естественных языков». Доклады Л. Заде, как отца-основателя нечеткой логики, очень хорошо иллюстрируют известную мысль о том, что любое научное направление в начале своего развития богаче идеями, чем в развитом состоянии. Л. Заде на правах живого классика обращается в таких докладах к истокам развития нечеткой логики, вводит новые концепции в парадигму нечетких систем, причем неизменным источником таких идей является семантика естественного языка. В то время как большинство докладчиков ограничиваются узкими задачами теории и практики нечетких систем, не принимая во внимание контекст, связанный с развитием интеллектуальных систем в целом.

Общим свойством обоих докладов Л. Заде является его неизменный интерес к решению основных задач искусственного интеллекта, построению семантики естественного языка. Теория нечетких систем рассматривается сквозь призму моделирования интеллекта с помощью вычислений. Нечеткие множества, предложенные Л. Заде для выражения человеческих понятий, сохраняют для него прежде всего данное назначение. Это свойство нужно особенно подчеркнуть, так как для последних лет развития нечеткой логики характерен сдвиг в область математики. Основная идея доклада Л. Заде на IFSA'05 заключается в развитии гранулярных вычислений. Докладчик начал с обоснованного утверждения о необходимости выбора уровня точности значений, согласованного с требованиями реальной задачи. Развивая это положение, Л. Заде разработал Theory of Precisation of Meaning (TPM), теорию уточнения значений. Основные положения этой теории и были изложены как в этом докладе, так и в статье [Zadeh, 2006]. Интересно сопоставить содержание этой статьи со стартовой

статьей [Zadeh, 1965], чтобы ясно представить себе путь, пройденный нечеткой логикой. Презентация Л. Заде была разделена на следующие разделы:

1. Концепция точности/неточности планов выражения и содержания понятий. Каждое понятие имеет содержание (value), которое может быть задано точно или не точно (*v-precise*, *v-imprecise*). Каждое понятие характеризуется и своей формой значения (meaning), которая также может быть выражена точно или не точно (*m-precise*, *m-imprecise*). Атрибут *m-precise*. Заде использует, как аналог термина «математически определенный». Например, если задана пропозиция

$$p: x \text{ is } X,$$

где X – гауссова случайная переменная с математическим ожиданием m и дисперсией σ ,

m и σ – точные действительные числа;

то говорят, что p имеет атрибуты *v-imprecise* и *m-precise*. Данная концепция лучше всего выражается кратким лозунгом: теория нечетких систем – это точная наука о неточности.

2. Грануляция является необходимым следствием *v-imprecise*.

Для представления неточного значения вместо единичного значения (синглетона) необходимо использовать: интервал, распределение какой-либо функции множества, т. е. гранулу сложной структуры. В общем смысле можно говорить об экстенциональном и интенциональном (*attribute-based*) представлении значений. А возможность выполнять операции над гранулами приводит к гранулярным вычислениям.

3. Для символической записи гранулярных пропозиций предлагается Язык гранулярных вычислений: Generalized Constraint Language (GCL).

Доклад Л. Заде на IFSA'07 служил продолжением доклада IFSA'05, но включал в себя ряд новых положений. В частности, логическая схема доклада может быть представлена, как последовательность следующих рассуждений.

1. Неопределенность – основной атрибут информации. Со времени К. Шеннона изучают прежде всего статистическую природу неопределенности. Но Теория

обобщенных ограничений неопределенности (Generalized Theory of Uncertainty GTU) отличается по существу.

2. Тезис о статистической природе неопределенности заменяется в GTU тезисом о том, что информация – это обобщенные ограничения, а статистическое представление информации – это только частный случай.
3. Бивалентность наличия свойства (в том числе истинности) заменяется степенью проявления свойства.
4. Главная цель GTU – способность описать информацию на естественном языке (NL-capability). В символической форме можно записать: $I(X) = GC(X)$, где X – переменная, определенная на U , $I(X)$ – информация о X , GC – обобщенные ограничения.

Л. Заде NL-capability любой теории неопределенности считает критерием ее приемлемости в эпоху построения гуманистических систем, основанных на знаниях.

Основные положения видео-выступления Л. Заде на IFSA'09 включают два важнейших направления развития нечеткой логики и ее приложений. *Первое направление* – это вычисления с неточными вероятностями и принятие решений с неопределенностью второго порядка, неопределенностью неопределенности. Второе направление – это нечеткая логика и естественный язык. Оба направления тесно связаны, так как неточные вероятности основаны на восприятии и описываются на естественном языке.

По мнению Л. Заде в сообществе ученых, занимающихся нечеткой логикой, данным направлениям не уделено достаточно внимания, хотя и имеются определенные достижения.

Основные положения доклада Д. Дюбуа «О связи между теориями вероятностей и возможностей» на IFSA'05

Доклад Д. Дюбуа «О связи между теориями вероятностей и возможностей» был построен как классификация современных теорий неопределенности

с указанием характерных особенностей, ограничений и областей применения. Рассмотрены и сопоставлены теория возможностей и вероятностей, в том числе теория возможностей и теория субъективных вероятностей; количественная и качественная теории возможностей. Области применения каждой теории рассматривались по их способности описывать гетерогенные данные и знания.

Основные положения докладов В. Педрича на IFSA'05 и IFSA'07

Доклады В. Педрича на IFSA'05 и IFSA'07 представляют собой последовательное развитие идеи формирования гранул информации с помощью кластеризации.

Доклад на IFSA'05 назывался «Кластеризация, основанная на знаниях в гуманистических системах» был основан на хорошо известном FCM-методе кластеризации, адаптированном для формирования гранул информации. Исходными данными являются временные ряды. Доклад IFSA'07 «Динамичное и распределенное нечеткое моделирование» был посвящен динамической кластеризации пространства данных. Динамическая кластеризация пространства рассматривалась как причина динамического развития гранул информации.

Основные положения докладов Дж. Клира на IFSA'05, Р. Ягера на IFSA'07. Я. Капржыка – на IFSA'07

Доклад Дж. Клира «От естественного языка к формализованному языку и обратно», сделанный на IFSA'05, посвящен возможностям нечеткой логики в представлении утверждений естественного языка. Для характеристики такой возможности используются меры информативности нечетких утверждений.

Доклад Р. Ягера на IFSA'07 «Вычисления со словами и гранулами» рассматривал, что именно добавили гранулярные вычисления к Теории Приближенных Рассуждений Заде. Автор доклада обоснованно доказывал, что грануляция задает рамки, как для построения вопросно-ответных интеллектуальных систем, так и для семантического Web.

Пленарный доклад Я. Капржыка на IFSA'07 «Вычисление со словами, квалификация полезности и лингвистические квантификаторы: инструменты для гуманистических вычислений» рассматривал возможности классификации на основе интуиционистских множеств, формирования запросов к базам данных на основе логических связей, лингвистического обобщения временных рядов на основе интегралов Шоке (Choquet). Довольно сложные рассуждения рассматривались на примере нечетких предпочтений избирателей на выборах.

Основные положения доклада К. Хироты «Взаимодействие с роботами, использующими модуль распознавания речи» на IFSA'09

Доклад был посвящен описанию проекта, реализующего совокупность пяти домашних роботов: четырех стационарных (размещенных на телевизоре, информационном терминале, в мини-баре и на машине игры в дарты) и одного мобильного робота. Эти роботы связаны с сервером по RTM. Модуль распознавания речи позволял взаимодействовать с роботами четырем людям (хозяину комнаты, двум гостям и гостю, находящемуся в комнате в движении).

Реакция роботов на команды зависит от ситуации в комнате и описывается нечеткими правилами.

Основные положения доклада Е. Халлермайера «Нечеткая логика в машинном обучении» на IFSA'09

Доклад вызвал большой интерес, так как машинное обучение представляется новой результативной областью приложения теории нечетких систем. Нечеткая логика может служить для выражения неопределенности в индукции и прогностике.

Основные положения докладов М. Грабиша «Интегралы Шоке в принятии решений: обзор фундаментальных понятий и текущие достижения» и Н. Нурми «Нечеткие системы, парадоксы выбора и оптимальность» на IFSA '09

Пленарные доклады М. Грабиша и Н. Нурми хотя и были посвящены разным предметам исследования, но рассматривали общее множество задач принятия решений. Первый из данных докладов описывал множество новых задач из области анализа решений, в которых применение интегралов Шоке может оказаться результативным. Второй доклад, посвященный коллективному принятию решений, рассматривал влияние парадоксальных индивидуальных предпочтений на результаты решения.

Сравнение тематических деревьев (topics) конгрессов

Материалы (Proceedings) конгрессов обширны и, разумеется, необозримы в полном объеме. Чтобы увидеть тенденции развития нечеткой логики, рассмотрим материалы конгрессов с точки зрения группировки статей (наименований тем и секций), которую выполнили программные комитеты. Так как исходными данными научного обзора послужили материалы пяти конгрессов IFSA, то возникает законное опасение, что подбор статей на конгрессы определили не тенденции развития нечеткой логики, а разный состав программных комитетов. Сравнение их составов показывает, что количество постоянных членов программных комитетов достаточно велико: 25 человек из 52 членов программного комитета IFSA'97 работали в комитете IFSA'03. В программных комитетах конгрессах 2005 (52 члена) и 2007 гг. (64 члена) пересечение составляет 16 человек, а 2007 и 2009 (172) – 36.

Причем, в большинстве своем это авторитетные ученые, поэтому можно рассчитывать, что выделенные тенденции и задачи будут определять развитие нечеткой логики в будущем.

Для сравнения тематик пяти конгрессов воспользуемся следующим приемом. Если составить графы тем и секций разных конгрессов, то получим со-

вершенные разные деревья, с одним общим свойством – эти деревья двухуровневые. Для сравнения будем считать, что существует «идеальное» тематическое дерево, с которым будем сравнивать тематику обоих конгрессов. Возьмем в качестве такого дерева структуру журнала «Fuzzy sets and systems».

Обзор тематики статей IFSA'09 дополним данными ранних конгрессов IFSA'97 и IFSA'03, IFSA'05 и IFSA'07 чтобы можно было видеть долгосрочные тенденции, и представим в табл. 2.5.

Таблица 2.5

Тематика статей IFSA'97, 03, 05, 07, 09

Темы Fuzzy sets and systems	Кол-во статей IFSA'97/ Доля рубрики IFSA'97 (%)	Кол-во статей IFSA'03/ Доля рубрики IFSA'03 (%)	Кол-во статей IFSA'05/ Доля рубрики IFSA'05 (%)	Кол-во статей IFSA'07/ Доля рубрики IFSA'07 (%)	Кол-во статей IFSA'09/ Доля рубрики IFSA'09 (%)	Тенденция
1. Основания	33/9	20/8	16/5	20/9	35/11	=↓↑↑
1.1. Основы нечетких множеств и связи	25/7	10/4	11/4	14/7	32/10	↓=↑↑
1.2. Нечеткие отношения и модели предпочтений	8/2	10/4	5/1	6/2	3/1	↑↓==
2. Математика	67/18	44/19	52/17	41/19	67/22	===↑
2.1. Неклассические логики и теория нечетких множеств	7/2	5/2	11/4	4/2	21/7	=↓↑↑

Продолжение табл. 2.5

1	2	3	4	5	6	7
2.2. Нечеткие дифференциальные уравнения и регрессия	5/1	6/3	3/1	2/1	14/5	$\uparrow\downarrow=\uparrow$
2.3. Топология и теория категорий	20/5	1/-	9/3	3/1	4/1	$\downarrow\uparrow\downarrow=$
2.4. Обобщенные меры, теория возможностей и теоретические аспекты нечетких чисел	25/7	26/11	20/6	13/6	25/8	$\uparrow\downarrow=\uparrow$
2.5 Алгебра	10/3	6/3	9/3	19/9	3/1	$=\uparrow\downarrow$
3. Искусственный интеллект	36/10	15/6	25/8	20/9	24/8	$\uparrow\downarrow=$
3.1. Приближенные рассуждения и представление знаний	25/7	9/4	23/7.5	13/6	14/5	$\uparrow\downarrow=$
3.2. Обучение	11/3	6/2	2/0.5	7/3	10/3	$=\downarrow\uparrow=$
4. Обработка информации	50/13	43/18	71/24	48/22	94/30	$\uparrow\uparrow=\uparrow$
4.1. Нечеткие случайные переменные и статистика	8/2	6/3	21/7	0/0	15/5	$=\uparrow\downarrow\uparrow$
4.2. Классификация и обработка образов	23/6	15/6	28/9	16/7	31/10	$=\uparrow\downarrow\uparrow$
4.3. Нечеткие базы данных и информационные системы	19/5	22/9	22/8	32/15	48/15	$\uparrow\uparrow\uparrow=$
5. Нечеткие системы	107/28	44/19	69/23	34/16	43/14	$\downarrow\uparrow\downarrow\downarrow$
5.1. Нечеткое моделирование и нейронечеткие системы	55/15	25/11	26/9	17/8	35/11	$\downarrow\downarrow\downarrow\uparrow$
5.2. Нечеткий контроль	52/14	19/8	43/14	17/8	8/3	$\downarrow\uparrow\downarrow\downarrow$

Окончание табл. 2.5

6. Принятие решений и оптимизация	48/13	26/11	31/10	31/14	34/11	↓=↑↓
6.1. Комбинаторная оптимизация и генетические алгоритмы	15/4	11/5	18/6	19/9	26/8	==↑↓
6.2. Нечеткое математическое программирование	33/9	15/6	13/4	12/5	8/3	↓==↓
7. Приложения	38/10	24/10	32/11	24/11	12/4	===↓
7.1. Гранулярные вычисления и нечеткая инженерия	34/9	18/8	19/7	5/3	8/3	==↓=
7.2. Роботика	2/0.5	5/2	12/4	18/8	4/1	↓↑↑↓

Таблица 2.5 показывает, ряд изменений, а именно – увеличилась доля работ по таким направлениям, как

- нечеткое моделирование и нейро-нечеткие системы;
- основы нечетких множеств и связи;
- неклассические логики и теория нечетких множеств;
- нечеткие дифференциальные уравнения и регрессия;
- обобщенные меры, теория возможностей и теоретические аспекты нечетких чисел;
- нечеткие случайные переменные и статистика;
- классификация и обработка образов.

Сократилась доля работ по направлениям:

- алгебра;
- нечеткий контроль;
- нечеткие системы;
- комбинаторная оптимизация и генетические алгоритмы;
- нечеткое математическое программирование;
- роботика и промышленные приложения.

Новые направления исследований, проявившиеся на IFSA '09

На каждом конгрессе появляются новые работы, которые обещают вырасти в полноценное направление исследований, например, исследования хаотических систем, фракталов (IFSA'97) или исследования в области биоинформатики (IFSA'03), работы по интервальному анализу (BISCSE'05 IFSA'07).

На конгрессе IFSA'09 возросло количество работ по нечетким системам и статистике, в том числе по обработке нечетких временных рядов. Среди работ, посвященных искусственному интеллекту, выросла доля работ по машинному обучению.

Перечень специальных сессий IFSA '09

На IFSA'05 и BISCSE'05 специальные сессии не проводились, зато на IFSA'07 и IFSA'09 такие события были широко представлены. На IFSA'07 на 14 специальных сессиях было сделано 128 докладов, а на IFSA'09 было организовано уже 32 специальных сессии (240 докладов). Заявленные темы специальных сессий можно укрупнить, выделяя большое количество исследований в области баз данных (нечеткие базы данных, текстовые базы данных, Data Mining, семантический Web), в одну группу (табл. 2.6).

Но с IFSA'05 можно выделить новое направление исследований – *Data Mining* (DM) баз данных временных рядов. Такие темы специальных сессий, как история или философские и гуманистические аспекты мягких вычислений, свидетельствуют о зрелости нечеткой логики, как научной теории и поисках ее места в общем научном мировоззрении.

Контрольные вопросы

1. Что такое International Fuzzy Systems Association (IFSA)?
2. Какие три основные идеи выделяет Л. Заде в своих докладах?
3. Выделите два направления в докладе Л. Заде на IFSA'09.
4. Выделите основные положения доклада Д. Дюбуа «О связи между теориями вероятностей и возможностей» на IFSA'05.

5. Выделите основные положения докладов В. Педрича на IFSA'05 и IFSA'07.
6. Выделите основные положения докладов Дж. Клира на IFSA'05, Р. Ягера на IFSA'07. Я. Капржыка IFSA'07.
7. Выделите основные положения доклада К. Хироты «Взаимодействие с роботами, использующими модуль распознавания речи» на IFSA'09.
8. Выделите основные положения доклада Е. Халлермайера «Нечеткая логика в машинном обучении» на IFSA'09.
9. Выделите основные положения докладов М. Грабиша «Интегралы Шоке в принятии решений: обзор фундаментальных понятий и текущие достижения» и Н. Нурми «Нечеткие системы, парадоксы выбора и оптимальность» на IFSA'09.
10. Опишите новые активно исследуемые прикладные направления.

Выводы

Несомненное достоинство нечетких моделей заключается в возможности параллельного оперирования гетерогенной информацией, представленной в виде сложных качественных лингвистических описаний и количественных данных. Анализ состояния и тенденций в области нечетких моделей и систем позволяет определить перспективы их развития и выделить новые активно исследуемые прикладные направления:

1. Нечеткая логика и статистика;
2. Нечеткая логика и нейронные сети;
3. Моделирование нечетких временных рядов;
4. Исследование нечетких баз данных, в том числе баз данных нечетких временных рядов;
5. Построение интеллектуальных информационных систем.

Специальные сессии IFSA '09

Тема	Организаторы
Recent advances in Evolving Fuzzy Systems Текущие достижения в Оценивающих нечетких системах	E. Lughofer, D. Filev, P. Angelov
Advances in Soft Computing Applied to Databases and Information Systems Достижения в мягких вычислениях, применяемых к базам данных и информационным системам	P. Bosc, A. Hadjali, O. Pivert
Transforms, Time Series and Other Applications Преобразования, временные ряды и другие приложения	I. Perfilieva, V. Novak, N. Yarushkina
Fuzzy and Possibilistic Optimization Нечеткая и возможностная оптимизация	M. Inuiguchi, W. A. Lodwick, M. Luhandjula
Fuzzy Differential Equations Нечеткие дифференциальные уравнения	Y. Chalco-Cano, W. A. Lodwick
Soft-computing for Web 2.0 and Semantic Web Мягкие вычисления для Web 2.0 и семантический Web	R. Yager, M. Reformat
Solvability of Fuzzy Relation Equations and Fuzzy Interpolation Разрешимость нечетких разностных уравнений и нечеткая интерполяция	I. Perfilieva, M. Stepnicka
Aggregation Operators Операторы агрегации	H. Bustince, T. Calvo, R. Mesiar
Fuzzy Sets in Computational Biology Нечеткие множества в вычислительной биологии	U. Bodenhofer, E. Huellermeier, F. Klawonn
Mathematical Fuzzy Logic Математическая нечеткая логика	P. Cintula, C. Noguera
Machine Learning and Data Mining Машинное обучение и интеллектуальный анализ данных	P. Angelov, E. Huellermeier, F. Klawonn, D. Sanchez
Type-2 Fuzzy Logic, Advances and Applications Нечеткая логика типа 2, достижения и применения	A. Celikyilmaz, I. B. Turksen
Inter-relation Between Interval and Fuzzy Techniques Взаимоотношения между интервальной и нечеткой техниками	V. Kreinovich
Advances in Soft Computing for Spatiotemporal Information Systems Достижения в мягких вычислениях для пространственно-временных информационных систем	G. De Tré, R. Ribeiro, J. Dujmovic
Interpretability of Fuzzy systems: Theory and Applications Интерпретируемость нечетких систем: теория и приложения	J. M. Alonso, L. Magdalena
Computing With Words, Actions and Perceptions Вычисления со словами, действиями и восприятиями	S. Guadarrama
New trends in Fuzzy Reasoning of Robotic Systems Новые тенденции в нечетких рассуждениях робототехнических систем	P. J. Sequeira Gonçalves, L. F. Mendonça

Fuzzy Numbers and Fuzzy Arithmetic Нечеткие числа и нечеткая арифметика	P. Grzegorzewski, L. Stefanini
Intuitionistic Fuzzy Sets Интуиционистские нечеткие множества	E. Szmidt, J. Kacprzyk
New Advances on Genetic Fuzzy Systems Новые достижения в генетических нечетких системах	Y. Nojima, R. Alcalá
Measures and Integrals Меры и интегралы	M. Grabisch
Fuzzy Geographical Information Нечеткая географическая информация	C. C. Fonte, J. Santos, M. Caetano, L. Gonçalves
Soft Computing in Image Processing and Computer Vision Мягкие вычисления в обработке образов и компьютерном зрении	Soft Computing in Image Processing and Computer Vision
Models and Fuzzy Arithmetic in Economics and Business Модели и нечеткая арифметика в экономике и бизнесе	M. L. Guerra and L. Stefanini
Soft Computing in Finance Мягкие вычисления в финансах	R. J. Almeida, M. Lovric, V. Milea
Topics in Decision-Making Using Fuzzy Sets Задачи в принятии решений, использующие нечеткие множества	D. Ralescu
Decision Making in Fuzzy Environments Принятие решений в нечетких условиях	M. T. Lamata, D. Pelta
Soft Computing in Medical Imaging Мягкие вычисления в медицинской обработке образов	I. K. Vlachos, G. Schaefer
Medical Concepts in Soft Computing Медицинские понятия и мягкие вычисления	C. Schuh, R. Seising
Philosophical, Sociological and Economical Thinking Философское, социологическое и экономическое мышление	E. Trilla, R. Seising, H. Nurmi

Библиографический список

1. [Zadeh, 1965] Zadeh, A. Lotfi. Fuzzy Sets / Lotfi A. Zadeh // Information and Control. – 1965.
2. [Аверкин и др., 1986] Нечеткие множества в моделях управления и искусственного интеллекта / А. Н. Аверкин, И. З. Батыршин, А. Ф. Блишун и др. ; Под ред. Д. А. Поспелова. – М. : Наука. Гл. ред. физ.-мат. лит., 1986. – 312 с.
3. [Батыршин и др., 2007 б] Батыршин, И. З. Нечеткие гибридные системы. Теория и практика / И. З. Батыршин, А. О. Недосекин, А. А. Стецко и др. – М. : ФИЗМАТЛИТ, 2007. – 208 с.

4. [Блюмин и др., 2005] Блюмин, С. Л. Окрестные системы / С. Л. Блюмин, А. М. Шмырев. – Липецк : Липецкий эколого-гуманитарный институт. – 2005. – 132 с.
5. [Борисов и др., 2007] Борисов, В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия – Телеком, 2007. – 284 с.
6. [Домрачев, 2001] Домрачев, В. Г. Нечеткие модели рейтинговых систем оценки знаний / В. Г. Домрачев, О. М. Полещук, И. В. Ретинская и др. // Телематика'2001. Труды Международной научно-методической конф. – СПб., 2001. – С. 245-246.
7. [Заде, 1974] Заде, Л. А. Основы нового подхода к анализу сложных систем и процессов принятия решений / Л. А. Заде // Математика сегодня. – М. : Знание, 1974. – С. 5-49.
8. [Круглов, 2001] Круглов, В. В. Нечеткая логика и искусственные нейронные сети / В. В. Круглов, М. И. Дли, Р. Ю. Голунов. – М. : Физматлит, 2001. – 224 с.
9. [Никольский, 2001] Никольский, С. Нечетко едешь – дальше будешь / С. Никольский // Компьютера. – 2001. – №38.
10. [Ротштейн, 1999] Ротштейн, А. П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети / А. П. Ротштейн. – Винница : УНИВЕРСУМ-Винница, 1999. – 320 с.
11. [Павлов и др., 2006] Павлов, А. Н. Принятие решений в условиях нечеткой информации : учеб. пособие / А. Н. Павлов, Б. В. Соколов ; ГУАП – СПб., 2006. – 72 с.
12. [Штовба, 2007] Штовба, С. Д. Проектирование нечетких систем средствами MATLAB / С. Д. Штовба. – М. : Горячая линия – Телеком, 2007. – 288 с.
13. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
14. [Яхьяева, 2006] Яхьяева, Г. Э. Нечеткие множества и нейронные сети : учебное пособие / Г. Э. Яхьяева. – М. : Интернет-Университет Информационных технологий: БИНОМ. Лаборатория знаний, 2006. – 316 с.

ГЛАВА 3. ОБЗОР НАПРАВЛЕНИЙ И ПОДХОДОВ В МОДЕЛИРОВАНИИ ВРЕМЕННЫХ РЯДОВ

Введение

Анализ временных рядов представляет собой самостоятельную, обширную и одну из наиболее интенсивно развивающихся областей исследования прикладной математики.

Временной ряд (ВР) – это последовательность дискретных упорядоченных в неслучайные равноотстоящие моменты времени измерений (показателей, наблюдений) $y(t_1), y(t_2), \dots, y(t_N)$, характеризующих уровни состояний изучаемого процесса, протекающего в условиях неопределенности.

Целью анализа временного ряда является достижение понимания причинных механизмов, обусловивших поведение изучаемого процесса, построение моделей временных рядов, которые не только объясняют поведение процесса, но и могут быть использованы для оценки прогноза развития изучаемого процесса.

Детерминированные процессы характеризуются достаточной информацией для определения функциональной зависимости $y = f(t)$.

В том случае, если процесс протекает в условиях неопределенности, традиционно используют стохастические модели временных рядов и изучаются числовые временные ряды, содержащие как систематическую, так и случайную компоненту. Данный подход базируется на теоретических предпосылках теории вероятности и прикладной статистики, основан на принципе многомодельности и использует накопленный опыт моделирования детерминированных процессов в виде функциональных зависимостей.

Процессы, функционирующие в условиях «нестохастической» неопределенности, как было показано в главе 1, могут моделироваться нечеткими временными рядами, теоретический базис которых сформирован теорией нечетких множеств и нечетких моделей, рассмотренных в главе 2.

Для моделирования процессов в реальных сложно-организованных системах, функционирование которых подвержено неопределенностям разного вида и характера, могут быть использованы гибридные модели временных рядов, описывающие стохастические и нечеткие (лингвистические) типы неопределенности.

Задача моделирования временных рядов в общем виде может быть сформулирована следующим образом.

Пусть заданы значения временного ряда $Y = \{y(1), y(2), \dots, y(N)\}$, где $y(t)$ – значение показателя исследуемого процесса, зарегистрированного в t -м такте времени ($t = 1, 2, \dots, N$). Требуется построить оценки будущих значений ряда $\hat{Y} = \{\hat{y}(N+1), \hat{y}(N+2), \dots, \hat{y}(N+\tau)\}$, $1 \leq \tau \leq N$, где τ – горизонт прогнозирования.

Основная идея, объединяющая подходы к моделированию временных рядов, базируется на выделении систематических (регулярных) зависимостей и анализе по фиксированным критериям полученных остатков. Независимо от применяемого метода предполагается, что закономерность изменений, выявленная для определенного периода временного ряда в прошлом, сохранится на ограниченном отрезке времени в будущем.

3.1. Статистический подход к моделированию временных рядов

При практическом изучении временных рядов исследователь (эксперт) на основании наблюдений – временного ряда конечной длины – должен сделать выводы о свойствах этого ряда и о вероятностном механизме, порождающем этот ряд. Статистический подход к моделированию ВР основывается на восстановлении по конкретному числовому временному ряду y_t приближенной модели, отражающей статистическую зависимость, для описания и численного прогноза поведения исследуемого процесса [Кендэл, 1981; Айвазян и др., 1998].

Общей статистической моделью числового временного ряда служит модель вида

$$y_t = f(x_t, a) + \varepsilon_t.$$

В этой модели наблюдаемый ряд y_t рассматривается как сумма некоторой систематической компоненты $f(x_t, a)$, где a – параметр, и случайной компоненты ε_t , рассматриваемой как независимые реализации случайного процесса типа «белый шум» с постоянным математическим ожиданием, постоянной и малой дисперсией.

Систематическая составляющая $f(x_t, a)$ временного ряда может быть представлена в виде линейной комбинации компонент и декомпозирована на трендовую, периодическую, сезонную компоненты, явно зависящие от времени (при $x_t = t$), и на авторегрессионную компоненту (при $x_t = y_{t-p}$), которая описывает зависимость между текущим значением уровня временного ряда и прошлым значением, сдвинутыми на лаг p .

Временные ряды, модель которых явно выражает зависимость от времени t и представляется в виде $y_t = f(t, a)$, относят к классу детерминированных временных рядов. Временные ряды, модель которых описывает поведение стационарных и нестационарных процессов и представляется в виде $y_t = f(y_{t-p}, a) + \varepsilon_t$, относят к классу стохастических.

Стационарный временной ряд отличается от нестационарного следующими свойствами: его математическое ожидание, дисперсия и ковариация не зависят от момента времени, в котором они вычисляются. Фундаментальным утверждением, служащим одновременно и ограничением статистического подхода, является теорема Вальда (1938) о разложении, согласно которой любой стационарный случайный процесс представляется в виде суперпозиции некоторого регулярного процесса $f(y_{t-p}, a)$ и белого шума ε_t .

В целом, статистический подход к анализу временных рядов заключается в выявлении и моделировании его детерминированных компонент на основе аддитивной (или мультипликативной) параметрической функциональной модели, приведении остатков к стационарному виду, при моделировании которых полученные ошибки ε_t удовлетворяли ограничениям модели (по теореме Вальда).

Методология моделирования временных рядов в рамках статистического подхода сводится к итеративному решению следующих задач (см. рис. 3.1):

1. Постулирование общего класса модели временного ряда.

Модель временного ряда рассматривается в общем виде как система, элементами которой являются линейные параметрические функции.

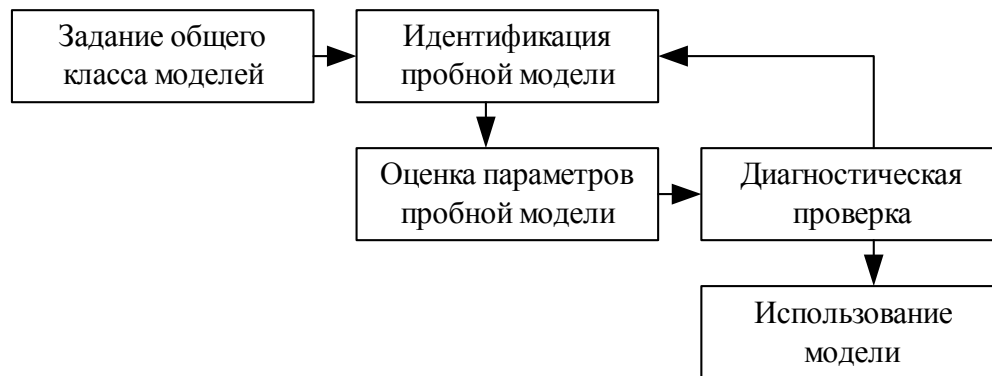


Рис. 3.1. Этапы итеративного подхода к построению моделей

Каждая функция моделирует поведение отдельных компонент (в зарубежной литературе используется термин паттерн (шаблон) вместо термина компонента) временного ряда. Выбор функций и разложение временного ряда по функциям решается специалистом-аналитиком (экспертом) на основе методов вычисления дополнительных функций и/или визуального распознавания типов возможных паттернов (визуальный Data mining).

2. Структурно-параметрическая идентификация модели.

Эта задача включает идентификацию каждой модели линейной функциональной зависимостью, описывающей поведение паттернов временного ряда. При рассмотрении функциональной зависимости как структуры, состоящей из неизвестного количества элементов, неизвестных значений параметров и типов связей, структурная идентификация заключается в последовательном постулировании типа связей, определяющего класс модели паттерна (например, аддитивная модель) и оценивании количества элементов, задающих размерность модели. Структурная идентификация моделей паттернов обычно решается экспертно путем определения количества элементов модели или на основе переборных методов.

Оценивание параметров составляет содержание этапа параметрической идентификации выбранной структуры модели и основывается на решении задачи минимизации остатков (ошибок модели) методом наименьших квадратов. Полученные оценки должны обладать свойствами состоятельности, несмещенности, эффективности и достаточности [Валеев, 2001].

3. Анализ адекватности модели моделируемому временному ряду.

Для обеспечения точности и достоверности результатов прогнозирования необходима оценка адекватности полученной при решении предыдущих задач модели ВР. Обычно для этих целей исследуют остатки на независимость и нормальность распределения, а также анализируются их свойства стационарности и отсутствия автокоррелированности на основе проверки статистических гипотез и статистических критериев, таких как, например, критерий Фишера, критерий Дарбина-Уотсона и др. На практике для проверки стационарности ряда остатков и оценки его дисперсии чаще всего используют автокорреляционную и частную автокорреляционную функции. Результаты решения задачи проверки уровня адекватности модели, основанные на анализе соответствия построенной модели предположениям и ограничениям общей модели, представлены в различных шкалах и требуют экспертной оценки с целью принятия окончательного решения.

4. Применение и исследование модели.

Верификация и анализ качества построенной модели ВР в соответствии с системной методологией в статистическом подходе основаны на проведении имитационных экспериментов, в процессе которых реализуется многокритериальное оценивание качества модели. При этом вычисляются внутренние и внешние меры качества модели на основе стандартизованных критериев. К таким критериям относят, в первую очередь, критерии точности моделирования, позволяющие сравнивать конкурирующие модели ВР.

Наиболее распространенные критерии точности моделирования временных рядов представлены в таблице 3.1, где y_i – реальные значения ВР, $\hat{y}_i$ – смоделированные значения ВР, n – количество членов ряда.

Критерии точности моделей временных рядов

Критерий	Формула расчета
Средняя квадратичная ошибка (СКО)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Квадрат из средней квадратичной ошибки	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Средняя относительная ошибка	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \cdot 100 \%$
Симметричная средняя относительная ошибка	$SMAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{(y_i + \hat{y}_i) / 2} \right \cdot 100 \%$

Для повышения точности при моделировании ВР используют методы фильтрации и подгонки функции. Большинство методов исследования временных рядов включает различные способы фильтрации шума. Многие монотонные временные ряды можно хорошо приблизить линейной функцией. Если же имеется явная нелинейная компонента, то данные вначале следует преобразовать, чтобы устранить нелинейность. Обычно для этого используют логарифмическое, экспоненциальное или (менее часто) полиномиальное преобразование данных.

Точность является показателем качества модели временного ряда, которая представляет результат процесса его моделирования. Отметим ряд других не менее важных показателей качества моделей временных рядов, которые характеризуют процесс моделирования, влияют на его результат и определяют в конечном счете выбор модели:

- ✓ Показатель трудоемкости (математической и алгоритмической сложности).
- ✓ Показатель временных затрат на построение модели.
- ✓ Показатель уровня автоматизации процесса построения модели.
- ✓ Уровень квалификации разработчика модели.
- ✓ Уровень квалификации пользователя модели.

- ✓ Показатель информативности и интерпретируемости области полученной модели в терминах решаемой задачи предметной.

По оценкам зарубежных и отечественных систематиков прогностики, насчитывается свыше ста методов прогнозирования. Число базовых методов, которые в тех или иных вариациях повторяются в других методах, гораздо меньше. Часть из этих методов относятся скорее к отдельным приемам или процедурам прогнозирования, другие представляют набор отдельных эвристических алгоритмов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения.

Так для моделирования тренда (сглаживания) используют регрессионные модели или методы сглаживания временных рядов. Для анализа сезонного эффекта применяют специальные модели сезонного сглаживания и сезонной авторегрессии. Колебания относительно тренда выявляются применением гармонического и спектрального анализа, а для описания и прогнозирования таких процессов используют гармонические модели или модели авторегрессии, скользящего среднего.

В последнее время возникло и оформилось целое научное направление, связанное с вейвлет-анализом [Витязев, 2001]. Вейвлеты широко применяются для фильтрации и предварительной обработки данных, анализа и прогнозирования временных рядов.

При моделировании временных рядов широкое распространение получили метод регрессионного моделирования и комплексные модели. Среди комплексных моделей выделим модель «авторегрессии проинтегрированного скользящего среднего» (АРПСС (p,d,q) или ARIMA (p,d,q)) [Бокс-Дженкинс, 1974; Канторович, 2002]. Популярность модели класса ARIMA обусловлена несколькими факторами: она позволяет упростить процесс разработки модели временного ряда, получить модели широкого класса временных рядов (стационарных и нестационарных) с приемлемыми показателями точности, снизить требования к уровню квалификации пользователя и реализована в распространенных статистических пакетах.

В качестве модели стационарных временных рядов используются модели ARIMA(p,d,q), в которых параметры структуры p, d, q , определяющие порядок модели, могут принимать нулевые значения:

1. Модель авторегрессии AR(p) связывает текущие значения временного ряда с прошлыми значениями и соответствует модели ARIMA(p,0,0). Формально модель авторегрессии AR(p) записывается в виде взвешенной суммы:

$$X(t) = f_0 + f_1 * X(t-1) + f_2 * X(t-2) + \dots + f_p * X(t-p) + E(t),$$

где $X(t)$ – текущее значение уровня ряда в момент времени t ;

$f_0, f_1, f_2, \dots, f_p$ – оцениваемые параметры;

p – порядок авторегрессии;

E_t – ошибка от влияния переменных, которые не учитываются в данной модели.

Задача заключается в том, чтобы оценить параметры $f_0, f_1, f_2, \dots, f_p$. Их можно оценить различными способами, например, через систему уравнений Юла-Уолкера, для составления этой системы потребуется расчет значений автокорреляционной функции, или методом наименьших квадратов.

2. Модель скользящего среднего MA(q) связывает текущие значения уровня ряда со значениями предыдущих ошибок и соответствует модели ARIMA(0,0,q). Формально модель MA(q) представима в виде взвешенной суммы:

$$Z(t) = m + \varepsilon(t) - w_1 * \varepsilon(t-1) - w_2 * \varepsilon(t-2) - \dots - w_q * \varepsilon(t-q),$$

где $Z(t)$ – текущее значение уровня ряда в момент времени t ;

m – константа, определяющая математическое ожидание временного ряда;

$w_0, w_1, w_2, \dots, w_q$ – оцениваемые параметры.

3. Комбинированные модели стационарных временных рядов ARMA(p,q) соответствуют модели ARIMA(p,0,q) и представляют собой объединение моделей AR(p) и MA(q).

Нестационарные временные ряды, приводящиеся к стационарным удалением тренда (или «взятием разности»), описываются моделью ARIMA(p,d,q), где параметр d указывает количество вычислений разности соседних уровней ВР

На практике не всегда удастся построить адекватные модели нестационарных временных рядов. Этому препятствует недостаточный объем наблюдений и изменяющаяся со временем статистическая структура временного ряда.

Отметим следующие проблемы статистического подхода в моделировании временных рядов:

1. Не существуют однозначных и эффективных критериев и методов определения факта наличия детерминированного тренда. Существуют статистические критерии проверки гипотезы о наличии тренда [Кендэл, 1981]. Но эти критерии используют двухальтернативный базис: тренд или случайная компонента (метод восходящих-нисходящих серий), тренд или периодическая компонента, регулярная или случайная компонента.

Традиционной проблемой является выбор наилучшего вида параметрической модели тренда. В качестве такого критерия отбора может быть использована доля объясненной дисперсии, называемая коэффициентом детерминации, который для некоторых временных рядов может быть незначим.

Поэтому, как правило, привлекается экспертный критерий (визуализация экспертом) или используются несколько критериев, дающих результат с различной эффективностью. Следует также отметить существующую нечеткость при моделировании тренда, обусловленную тем, что выбор метода выделения тренда определяется экспертом, а различные методы выделения тренда в принципе генерируют различные временные ряды остатков, что приводит, в конечном счете, к построению различных моделей временных рядов.

2. Другой проблемой является моделирование нестационарных ВР, для которых характерны нелинейность поведения при отсутствии детерминированного тренда, сезонности и цикличности.

Если в стационарном случае есть доказательная уверенность в асимптотической состоятельности оценок той или иной статистики, то в нестационарном случае отсутствует само понятие генеральной совокупности, что делает неприменимым весь развитый аппарат современной математической статистики, кроме тех случаев, когда априори задана функциональная принадлежность модели процесса. Однако на практике часто бывает не известно, к какому клас-

су принадлежит распределение и является ли оно стационарным, причем оба этих фактора могут быть определены лишь с некоторой доверительной вероятностью – корректно определенной, однако, только для стационарных процессов.

Нестационарные процессы, изучаемые в относительно малом числе публикаций, чаще всего относятся к определенным функциональным классам, проверка принадлежности к которым реальных процессов является гораздо более трудной задачей, чем проверка их на стационарность.

В адаптивных методах исследования рядов, про которые априори не известно, являются ли они (ряды) стационарными или нет, не решен вопрос, по выборке какого объема следует проводить скользящее усреднение, чтобы получить наименьшую ошибку прогноза. Решение этой проблемы в существующих критериях оставляется на усмотрение эксперта в соответствии с его квалификацией и опытом.

2. В задачах статистического анализа и прогноза стохастической компоненты временного ряда стремятся получить оптимальную модель из класса заданных, среднеквадратическое отклонение остатков которой минимально. Существуют две схемы решения.

✓ В первой схеме модель постулируется, то есть выбирается экспертно, например, ARIMA (АРПСС). Затем ВР преобразовывается таким образом, чтобы удовлетворять ограничениям этой модели. В дальнейшем проводятся параметрическая оптимизация этой модели (по числу и значению параметров), прогноз и обратное преобразование. На практике оказывается, что качество (в смысле точность прогноза) моделей в первом подходе будет разным для различных ВР одного класса, то есть зависит от квалификации пользователя, его опыта, а также от класса ВР. Кроме того, прогнозные значения таких моделей не всегда корректно отображают ожидаемую тенденцию изменения, так как используют меру качества, основанную на усреднении квадрата разностей отклонений MSE (СКО).

✓ Вторая схема использует модификацию первой схемы на основе принципа передачи функции эксперта по постулированию модели системе анализа и прогноза ВР. В отличие от первой схемы в этой схеме идет дополнительная оптимизация по набору статистических моделей [Валеев, 2001]. При этом сначала выбираются оптимальные модели определенного класса, а затем строится наилучшая комплексная модель.

Несмотря на обилие математических моделей, методов и критериев, которые во многих случаях позволяют получать высокоточные модели, статистический подход к моделированию временных рядов не лишен недостатков и ограничений, к которым относят:

– Ограничение на класс моделируемых процессов.

Последовательность наблюдений в статистическом подходе рассматривается как реализация последовательности статистически независимых случайных величин, имеющих нормальное совместное распределение. Проблемы, связанные с прогнозированием нестационарных временных рядов и с нарушением предположения о независимости и нормальности распределения наблюдений приведены в работах [Канторович, 2002; Осьминин, 2008; Комиссарова, 2006];

– Ограничение на длину временного ряда.

Статистические модели характеризуются невысоким качеством при моделировании коротких временных рядов (количество наблюдений меньше 40) [Бокс и др., 1974; Khashei, 2008];

– Ограничения ресурсов.

Высокая трудоемкость процесса построения адекватных моделей временных рядов и адаптации моделей к новым наблюдениям, высокая квалификация разработчика модели, нередко требующая диссертационного исследования [Канторович, 2002; Сергейчик, 2007; Беляков, 2005; Осьминин, 2008];

– Ограничения информативности.

Модель представлена в виде набора числовых параметров структуры, числовых оценок коэффициентов, числовых оценок адекватности модели. Результаты моделирования представлены в виде числовых оценок уровней временного ряда и показателей точности модели. Все эти числовые оценки требуют дополнительного анализа и вербальной оценки квалифицированного эксперта;

– Ограничения неопределенности.

Рассматривается и моделируется неопределенность одного типа – стохастическая.

Контрольные вопросы

1. Что такое временной ряд?
2. Какова цель анализа временных рядов?
3. В каком случае используются стохастические модели временных рядов?
4. В каком случае используются гибридные системы?
5. Сформулируйте задачу моделирования временных рядов.
6. На чем основывается статистический подход к моделированию ВР?
7. Какова общая статистическая модель числового ВР?
8. Чем отличается стационарный временной ряд от нестационарного?
9. Опишите сущность задачи структурно-параметрической идентификации модели.
10. Приведите критерии точности моделей ВР.
11. Какие модели получили широкое распространение при моделировании ВР?
12. Приведите проблемы статистического подхода в моделировании ВР.

Библиографический список

1. [Айвазян и др., 1998] Айвазян, С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998. – 1024 с.

2. [Афанасьев и др., 2001] Афанасьев, В. Н. Анализ временных рядов и прогнозирование : учебник / В. Н. Афанасьев, М. М. Юзбашев. – М. : Финансы и статистика, 2001. – 228 с.
3. [Бокс и др., 1974] Бокс, Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс ; Под ред. В. Ф. Писаренко ; Пер. с англ. – М. : Мир, 1974. – 406 с.
4. [Беляков, 2005] Беляков, С. С. Использование агрегирования в методах нелинейной динамики для анализа и прогнозирования временных рядов котировок акций / С. С. Беляков // Автореферат диссертации на соискание ученой степени к. э. н. – 2005.
5. [Валеев, 2001] Валеев, С. Г. Регрессионное моделирование при обработке данных / С. Г. Валеев. – Казань : ФЭН, 2001.
6. [Витязев, 2001] Витязев В. В. Вейвлет-анализ временных рядов / В. В. Витязев. – СПб. : Изд-во С.-Петербур. Ун-та, 2001.
7. [Канторович, 2002] Канторович, Г. Г. Анализ временных рядов / Г. Г. Канторович // Экономический журнал ВШЭ . – 2002. – №№1-2.
8. [Комиссарова, 2006] Комиссарова, К. А. Экономико-математическое моделирование деятельности страховых компаний методами нелинейной динамики / К. А. Комиссарова // Автореферат диссертации на соискание ученой степени к. э. н. – 2006.
9. [Кендэл, 1981] Кендэл, М. Временные ряды / М. Кендэл ; Пер. с англ. и предисл. Ю. П. Лукашина. – М. : Финансы и статистика, 1981. – 199 с.
10. [Носко, 2002] Носко, В. П. Эконометрика. Введение в регрессионный анализ временных рядов / В. П. Носко. – М. : НФПК, 2002. – 273 с.
11. [Осьминин, 2008] Осьминин, К. П. Алгоритмы прогнозирования нестационарных временных рядов / К. П. Осьминин // Автореферат диссертации на соискание ученой степени к.ф.-м. н. – 2008.
12. [Сергейчик, 2007] Сергейчик, О. И. Модели и алгоритмы спектрального анализа обработки кардиологических временных рядов / О. И. Сергейчик // Автореферат диссертации на соискание ученой степени канд.техн. наук. – 2007.

3.2. Нейросетевой подход к моделированию временных рядов

В нейросетевом подходе задача прогнозирования временных рядов формулируется как задача распознавания образов, для решения которой формируется обучающая последовательность данных временного ряда, и нейронная сеть обучается распознавать соответствующие образы.

Искусственные нейронные сети (ИНС) – это вычислительные модели, построенные по принципу организации и функционирования биологических нейронных сетей. ИНС представляют собой систему соединенных и взаимодействующих между собой простых процессоров (искусственных нейронов). Важная особенность ИНС состоит в возможности параллельной обработки информации и способности к обучению и обобщению накопленных знаний. Широкое распространение нейронных сетей в разных областях объясняется тем, что во многих случаях формализация процедур решения сложных задач в экономике, медицине, технике, военном деле зачастую оказывается очень трудоемкой, либо невозможной.

3.2.1. Основы моделирования временных рядов нейронными сетями

Основной элемент нейронной сети – это формальный нейрон, осуществляющий операцию нелинейного преобразования суммы произведений входных сигналов на весовые коэффициенты

$$y = F\left(\sum_{i=1}^n w_i x_i\right) = F(WX),$$

где $X = (x_1, x_2, \dots, x_n)^T$ – вектор входного сигнала;

$W = (w_1, w_2, \dots, w_n)$ – вектор весовых коэффициентов (оцениваемые параметры);

F – функция нелинейного преобразования.

Способности нейронной сети к прогнозированию ВР напрямую следуют из ее способности к обобщению и выделению скрытых зависимостей между входными и выходными данными. После обучения сеть способна предсказать

будущее значение уровня временного ряда на основе нескольких предыдущих значений.

Моделирование ВР в рамках нейросетевого подхода сводится к задаче наилучшей аппроксимации нелинейной функции от многих переменных по набору примеров, заданных историей временного ряда:

$$\hat{y}_{k+1} = \varphi(y_k, \dots, y_{k-n+1}) + \varepsilon_{k+1},$$

где $\hat{y}_{k+1}$ – прогнозируемое значение уровня временного ряда;

$y_k, \dots, y_{k-n+1}$ – наблюдаемые значения уровней временного ряда;

$\varphi(y_k, \dots, y_{k-n+1})$ – некоторая нелинейная функция, параметрической моделью которой служит нейронная сеть;

ε_{k+1} – ошибка прогноза;

n – порядок модели.

Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения – одно из главных преимуществ нейронных сетей перед традиционными численными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами при минимизации среднеквадратичного отклонения ошибки ε_{k+1} . В процессе обучения нейронная сеть способна выявлять сложные нелинейные зависимости между входными данными и выходными, а также выполнять обобщение. Это значит, что, в случае успешного обучения, сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке.

Математическую основу нейросетевого подхода при моделировании и прогнозировании временного ряда образуют ряд теорем [Ширяев, 2007]. Доказана обобщенная аппроксимационная теорема [Колмогоров, 1956]: с помощью операций сложения, умножения и суперпозиции можно из произвольного нелинейного элемента получить устройство, вычисляющее любую непрерывную функцию с любой наперед заданной точностью. По теореме Такенса [Takens, 1981], если временной ряд порождается динамической системой, то есть значения y_k есть произвольная функция состояния этой системы, то существует

«глубина погружения» n , которая обеспечивает однозначное предсказание следующих значений уровней временного ряда с помощью некоторого функционального преобразования, явно не зависящего от k . Согласно теореме о полноте [Горбань, 1998] любая непрерывная функция на замкнутом ограниченном множестве может быть равномерно приближена функциями, вычисляемыми нейронными сетями, если функция активации нейронов дважды непрерывно дифференцируема и нелинейна. Японским ученым Фунахаши была доказана теорема о нейронной сети как функциональном универсальном аппроксиматоре [Bothe, 1997]. Это означает, что нелинейная функция нейрона может быть произвольной: от сигмоидальной до произвольного волнового пакета или вейвлета, синуса или полинома. От выбора нелинейной функции может зависеть сложность конкретной сети, но с любой нелинейностью сеть остается универсальным аппроксиматором и при правильном выборе структуры может сколь угодно точно аппроксимировать функционирование любого непрерывного автомата.

Таким образом, задача прогнозирования временных рядов с помощью ИНС сводится к задаче восстановления оценки нелинейной функции $\varphi(y_k, \dots, y_{k-n+1})$ по набору примеров, заданных историей ВР и реализуется в виде последовательности этапов:

- сбор данных для обучения;
- подготовка и нормализация данных;
- выбор топологии нейронной сети;
- экспериментальный подбор характеристик нейронной сети;
- экспериментальный подбор параметров обучения;
- обучение нейронной сети;
- проверка адекватности обучения;
- корректировка параметров, окончательное обучение;
- вербализация сети с целью дальнейшего использования.

Известны разнообразные типы нейронных сетей, отличающиеся способом реализации отдельных этапов моделирования.

Так, в сетях прямого распространения (Feedforward) все связи направлены строго от входных нейронов к выходным. Примерами таких сетей являются персептрон Розенблатта, многослойный персептрон, сети Ворда.

В рекуррентных нейронных сетях сигнал с выходных нейронов или нейронов скрытого слоя частично передается обратно на входы нейронов входного слоя (обратная связь). Рекуррентная сеть, сеть Хопфилда, «фильтрует» входные данные, возвращаясь к устойчивому состоянию и, таким образом, позволяет решать задачи компрессии данных. Частным случаем рекуррентных сетей является двунаправленные сети. В таких сетях между слоями существуют связи как в направлении от входного слоя к выходному, так и в обратном. Классическим примером двунаправленных сетей является нейронная сеть Коско.

Известны и другие типы сетей: сеть Джордана, сеть Элмана, сеть Хэмминга, сеть Кохонена, когнитрон, неокогнитрон, хаотическая нейронная сеть, осцилляторная нейронная сеть, сеть встречного распространения, сеть радиальных базисных функций (RBF-сеть), сеть обобщенной регрессии, вероятностная сеть, сямская нейронная сеть, сети адаптивного резонанса [Ярушкина, 2004; Ширяев, 2007].

3.2.2. Обучение методом обратного распространения ошибок

С математической точки зрения обучение нейронных сетей (НС) – это многопараметрическая задача нелинейной оптимизации.

Для обучения сети используются различные алгоритмы обучения и их модификации [Барский, 2004; Бутенко, 2002; Гусак, 2002]. Наиболее распространенным является алгоритм обратного распространения ошибки. Алгоритм минимизирует среднеквадратичную ошибку нейронной сети. Для этого с целью настройки синаптических связей используется метод градиентного спуска в пространстве весовых коэффициентов и порогов нейронной сети. Следует отметить, что в настройке синаптических связей сети используются не только метод градиентного спуска, но и методы сопряженных градиентов, Ньютона, квазиньютоновский метод [Новиков, 2002].

Рассмотрим алгоритм обратного распространения ошибки ИНС. Для упрощения обозначений ограничимся ситуацией, когда сеть имеет только один скрытый слой. Матрицу весовых коэффициентов от входов к скрытому слою обозначим W , а матрицу весов, соединяющих скрытый и выходной слой – как V . Для индексов примем следующие обозначения: входы будем нумеровать только индексом i , элементы скрытого слоя – индексом j , а выходы, соответственно, индексом k .

Пусть сеть обучается на выборке (X_a, Y_a) , $a = 1 \dots p$. Активности нейронов будем обозначать малыми буквами y с соответствующим индексом, а суммарные взвешенные входы нейронов – малыми буквами x .

Алгоритм обратного распространения ошибки ИНС:

Шаг 0. Начальные значения весов всех нейронов всех слоев $V(t=0)$ и $W(t=0)$ полагаются случайными числами.

Шаг 1. Сети предъявляется входной образ X_a , в результате формируется выходной образ $y^1 \dots Y_a$. При этом нейроны последовательно от слоя к слою функционируют по следующим формулам: скрытый слой $x_j = \sum_i W_{ij} X_i^a$; $y_j = f(x_j)$, выходной слой $x_k = \sum_j V_{jk} y_j$; $y_k = f(x_k)$.

Здесь $f(x)$ – сигмоидальная функция.

Шаг 2. Функционал квадратичной ошибки ИНС для данного входного образа имеет вид: $E = \frac{1}{2} \sum_k (y_k - Y_k^a)^2$.

Данный функционал подлежит минимизации. Классический градиентный метод оптимизации состоит в итерационном уточнении аргумента согласно формуле: $V_{jk}(t+1) = V_{jk}(t) - h \frac{\partial E}{\partial V_{jk}}$.

Функция ошибки в явном виде не содержит зависимости от веса V_{jk} , поэтому можно использовать формулы неявного дифференцирова-

ния сложной функции:

$$\begin{aligned}\frac{\partial E}{\partial y_k} &= \delta_k = (y_k - Y_k^a); \\ \frac{\partial E}{\partial x_k} &= \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_k} = \delta_k \cdot y_k(1 - y_k); \\ \frac{\partial E}{\partial V_{jk}} &= \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_k} \cdot \frac{\partial x_k}{\partial V_{jk}} = \delta_k \cdot y_k(1 - y_k) \cdot y_j.\end{aligned}$$

Здесь учтено полезное свойство сигмоидальной функции $f(x)$: ее производная выражается только через само значение функции, $f'(x) = f(1 - f)$. Таким образом, все необходимые величины для подстройки весов выходного слоя V получены.

Шаг 3. На этом шаге выполняется подстройка весов скрытого слоя. Используем градиентный метод: $W_{ij}(t+1) = W_{ij}(t) - h \frac{\partial E}{\partial W_{ij}}$.

Вычисления производных выполняются аналогично, за исключением некоторого усложнения формулы для ошибки d_j :

$$\begin{aligned}\frac{\partial E}{\partial x_k} &= \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_k} = \delta_k \cdot y_k(1 - y_k); \\ \frac{\partial E}{\partial y_j} &= \delta_j = \sum_k \frac{\partial E}{\partial x_k} \cdot \frac{\partial x_k}{\partial y_j} = \sum_k \delta_k \cdot y_k(1 - y_k) \cdot V_{jk}; \\ \frac{\partial E}{\partial W_{ij}} &= \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_j} \cdot \frac{\partial x_j}{\partial W_{ij}} = \delta_j \cdot y_j(1 - y_j) \cdot X_i^a = \left[\sum_k \delta_k \cdot y_k(1 - y_k) \cdot V_{jk} \right] \cdot [y_j(1 - y_j) \cdot X_i^a]\end{aligned}$$

При вычислении d_j здесь и был применен принцип обратного распространения ошибки: частные производные берутся только по переменным последующего слоя. По полученным формулам модифицируются веса нейронов скрытого слоя. Если в нейронной сети имеется несколько скрытых слоев, процедура обратного распространения применяется последовательно для каждого из них, начиная со слоя, предшествующего выходному, и далее до слоя, следующего за входным. При этом формулы сохраняют свой вид с заменой элементов выходного слоя на элементы соответствующего скрытого слоя.

Шаг 4. Шаги 1-3 повторяются для всех обучающих векторов. Обучение завершается по достижении заданного уровня ошибки или максимально допустимого числа итераций.

Как видно из описания шагов 2-3, обучение сводится к решению задачи оптимизации функционала ошибки градиентным методом.

Параметр h имеет смысл темпа обучения и выбирается достаточно малым для обеспечения сходимости метода.

Невысокий темп сходимости является «генетической болезнью» всех градиентных методов, так как локальное направление градиента не всегда совпадает с направлением к минимуму. Подстройка весов выполняется независимо для каждой пары образов обучающей выборки. При этом улучшение функционирования на некоторой заданной паре может иногда приводить к ухудшению работы на предыдущих образах. В этом смысле нет достоверных (кроме весьма обширной практики применения метода) гарантий сходимости.

Исследования показывают, что для представления произвольного функционального отображения, задаваемого обучающей выборкой, достаточно всего два слоя нейронов. Однако на практике, в случае сложных функций, использование более чем одного скрытого слоя может давать экономию полного числа нейронов [Ярушкина, 2004].

Большую роль для повышения эффективности обучения сети играет архитектура НС [Горбань, 1990]. Точность определяется числом нейронов в скрытом слое, но при слишком большой размерности скрытого слоя может наступить явление, называемое перетренировкой (переобучение) сети. Для устранения этого недостатка необходимо, чтобы число нейронов в промежуточном слое было значительно меньше, чем число тренировочных образов. С другой стороны, при слишком маленькой размерности скрытого слоя можно попасть в нежелательный локальный минимум.

В [Горбань, 1990] предложены способы настройки числа нейронов в процессе обучения, которые обеспечивают построение нейронной сети для решения задачи и дают возможность избежать избыточности. Эти способы адапта-

ции структуры НС можно разделить на две группы: конструктивные алгоритмы и алгоритмы сокращения.

В основе алгоритмов сокращения лежит принцип постепенного удаления из нейронной сети синапсов и нейронов. В начале работы алгоритма обучения с сокращением число нейронов в скрытых слоях сети заведомо избыточно.

Существуют два подхода к реализации алгоритмов сокращения: метод штрафных функций и метод проекций.

Для реализации первого в целевую функцию алгоритма обучения вводятся штрафы за то, что значения синаптических весов отличны от нуля, например в виде

$$C = \sum_{i=1}^N \sum_{j=1}^M w_{ij}^2,$$

где w_{ij} – синаптический вес;

i – номер нейрона;

j – номер входа;

N – число нейронов;

M – размерность входного сигнала нейронов.

Метод проекций реализуется следующим образом. Синаптический вес обнуляется, если его значение попало в заданный диапазон

$$w_{ij} = \begin{cases} 0, & -\varepsilon \leq w_{ij} \leq \varepsilon \\ w_{ij}, & w_{ij} < -\varepsilon, w_{ij} > \varepsilon \end{cases},$$

где ε – некоторая константа.

В [Головко, 2001] показано, что для систем прогноза ВР на базе ИНС наилучшие качества показывает гетерогенная сеть, состоящая из скрытых слоев с нелинейной функцией активации нейронных элементов и выходного линейного нейрона. Недостатком большинства рассмотренных нелинейных функций активации является то, что их область выходных значений ограничена отрезком $[0,1]$ или $[-1,1]$. Это приводит к необходимости масштабирования данных.

Примеры использования ИНС для решения задач моделирования и прогнозирования ВР рассмотрены в работах [Ширяев, 2007; Козадаев, 2008]

Таким образом, искусственные нейронные сети являются результативным инструментом моделирования и прогнозирования временных рядов, позволяющим снизить требования к квалификации пользователя и обеспечивающим значительное снижение трудоемкости процесса создания модели. Однако создаваемые модели временных рядов с помощью ИНС невозможно интерпретировать в терминах предметной области.

Контрольные вопросы

1. Что такое искусственные нейронные сети?
2. Основным элементом нейронной сети – это ...
3. К чему сводится задача прогнозирования временных рядов с помощью ИНС?
4. Приведите примеры типов нейронных сетей.
5. Что включает в себя алгоритм обратного распространения ошибки ИНС?
6. Какие подходы существуют к реализации алгоритмов сокращения?

Библиографический список

1. [Bothe, 1997] Bothe, H.-H. Fuzzy Neural Network / H.-H. Bothe. – Prague : IFSA, 1997.
2. [Takens, 1981] Takens, T. Detecting strange attractors in turbulence / T. Takens // Lec. Notes in Math., 1981.
3. [Барский, 2004] Барский, А. Б. Нейронные сети: распознавание, управление, принятие решений / А. Б. Барский. – М. : Финансы и статистика, 2004. – 176 с.
4. [Бутенко, 2002] Бутенко, А. А. Обучение нейронной сети при помощи алгоритма фильтра Калмана / А. А. Бутенко // Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: сб. докл. – М. : Ин-

- ститут проблем управления им. В. А. Трапезникова РАН, 2002. – С. 1120-1125.
5. [Головко, 2001] Головко, В.А. Нейронные сети: обучение, организация и применение. Кн.4 : учебн. пособие для вузов / В. А. Головко ; Общая ред. В. А. Галушкина. – М. : ИПРЖР, 2001. – 256 с.
 6. [Горбань и др., 1998] Горбань, А. Н. Нейроинформатика / А. Н. Горбань, В. Л. Дунин-Барковский, А. Н. Курдин и др. – Новосибирск : Наука, 1998. – 296 с.
 7. [Горбань, 1990] Горбань, А. Н. Обучение нейронных сетей / А. Н. Горбань. – М. : СП «ПараГраф», 1990. – 159 с.
 8. [Гусак, 2002] Гусак, А. Н. Подход к послойному обучению нейронной сети прямого распространения / А. Н. Гусак // Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» сб. докл. – М. : Институт проблем управления им. В. А. Трапезникова РАН, 2002. Автореферат диссертации на соискание ученой степени к. э. н. – С. 931-933.
 9. [Козадаев, 2008] Козадаев, А. С. Математические модели временных рядов на основе аппарата искусственных нейронных сетей и программный комплекс для их реализации / А. С. Козадаев // Автореферат диссертации на соискание ученой степени к. т. н. – 2008.
 10. [Колмогоров, 1956] Колмогоров, А. Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных / А. Н. Колмогоров // Докл. АН СССР. – 1956. – Т. 108. – №2. – С. 179-182.
 11. [Новиков, 2002] Новиков, А. В. Метод поиска экстремума функционала оптимизации для нейронной сети с полными последовательными связями / А. В. Новиков // Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» : сб. докл. – М. : Институт проблем управления им. В. А. Трапезникова РАН, 2002. – С. 1000-1006.
 12. [Ширяев, 2007] Ширяев, В. И. Финансовые рынки и нейронные сети / В. И. Ширяев. – М. : Издательство ЛКИ, 2007. – 224 с.

13. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.

3.3. Нечеткий подход к моделированию временных рядов

Нечеткое моделирование временных рядов представляет новую научную область, специфика которой по отношению к статистическому и нейросетевому моделированию ВР определяется нечеткими уровнями нечеткого временного ряда (НВР), а по отношению к нечетким моделям – более сложной структурной организацией обрабатываемых нечетких значений.

Подход с точки зрения нечетких моделей позволяет использовать прикладные знания для нечеткого выражения уровней временного ряда, строить нечеткие временные ряды и выявлять зависимости в виде нечетких продукционных правил.

Представление временных рядов в классе нечетких временных рядов основывается на предположении, что возможна лингвистическая интерпретация значений временного ряда, основанная на понятии нечетких множеств. Эта семантически значимая интерпретация значений ВР, относящаяся как к его уровням, так и к временным моментам, выраженная в нечетких лингвистических оценках, зависит от сущности и контекста свойств наблюдаемого объекта, а также от восприятия эксперта, выполняющего интерпретацию. Восприятие интегрирует компетентностную, временную и пространственную позицию эксперта. Заметим, что деятельность эксперта по лингвистической оценке значений ВР, позволяющая построить инструмент для преобразования исходного ВР в нечеткий временной ряд, является расширением его деятельности на этапе разведывательного анализа данных предметной области и определения ограничений переменных. В то же время, нечеткий временной ряд может быть получен и на основе абстрактных лингвистических оценок, так как работа эксперта дорогостояща и трудоемка.

Прикладной аспект проблематики анализа нечетких временных рядов определяется возможностью расширения множества прикладных задач обработки

ВР, множества технологий их решения и области результатов за счет оперирования не только числовой, но и качественной информацией, выраженной лингвистическими терминами.

Примером прикладных задач, которые образуют расширение множества прикладных задач обработки НВР, могут выступать задачи, связанные с анализом тенденций, решаемые в экспертной деятельности, в процессе проектирования, управления и принятия решений.

Расширение множества технологий решения прикладных задач обработки ВР связано с необходимостью обработки новых типов данных с нечеткими значениями.

В целом, расширение области результатов решения задачи анализа временных рядов за счет нечеткого моделирования ВР позволит принимать более обоснованные решения на основе обработки качественной информации. Ценность полученного результата обработки НВР заключается в том, что в нем выражена семантически значимая интерпретация сущности и контекста объектов предметной области и их развития в виде естественных и понятных человеку лингвистических оценок.

3.3.1. Основы нечеткого моделирования временных рядов

В отличие от традиционного временного ряда значениями нечеткого ВР являются нечеткие множества, а не действительные значения уровней ВР. В 1993 году Song и Chissom [Song, 1993] предложили нечеткие модели детерминированных (time-variant) и авторегрессионных (time-invariant) временных рядов первого порядка (first-order) и применили разработанные модели для прогнозирования количества регистрирующихся студентов университета штата Алабама (США), фаззифицировав предварительно четкий временной ряд. Это было первое применение нечетких моделей при моделировании ВР и первое определение моделей нечетких временных рядов.

Пусть $X_t, (t = 1, \dots) \subset R^1$ – универсальное множество, на котором определены нечеткие множества $y_t^i, (i = 1, 2, \dots)$ и Y_t – коллекция $y_t^i, (i = 1, 2, \dots)$. Тогда Y_t называется нечетким ВР.

На практике в большинстве временных рядов последовательные наблюдения зависимы, так что:

$$R = \{(y_t, y_{t-1}), (y_{t-1}, y_{t-2}) \dots\} \subseteq Y_t \times Y_{t-1},$$

где Y_t, Y_{t-1} обозначают переменные;

y_t, y_{t-1} – наблюдаемые значения этих переменных.

Наиболее частой моделью зависимости является явная функция:

$$f: Y_{t-1} \rightarrow Y_t,$$

представленная линейной функцией (марковским процессом, модель AR(1)):

$$y_t = f(y_{t-1}, \phi, \varepsilon) = \phi y_{t-1} + \varepsilon_t,$$

где ε_t – случайная ошибка, шум.

В случае нечеткого временного ряда в качестве модели авторегрессии используется нечеткое разностное уравнение:

$$\begin{aligned} y_t^j &= y_{t-1}^i \circ R_{ij}(t, t-1), \\ y_t^i &\in Y_t, \quad y_{t-1}^i \in Y_{t-1}, \quad i \in I, \quad j \in J, \end{aligned}$$

где $\circ$ – обозначает операцию композиции из теории нечетких множеств;

$R(t, t-1) = \bigcup_{i,j} R_{ij}(t, t-1)$ – система нечетких отношений, которая символически может быть записана в виде $Y_t \rightarrow Y_{t-1}$.

Систему отношений R в выражении $Y_t = Y_{t-1} \circ R(t, t-1)$ называют *моделью нечеткого временного ряда* первого порядка, данная модель – важный частный случай общей модели порядка p :

$$\begin{aligned} Y_t &= (Y_{t-1} \times Y_{t-2} \times \dots \times Y_{t-p}) \circ R(t, t-p), \\ R(t, t-p) &= \max_p \left\{ \min_{j, i_1, i_2, \dots, i_p} \{y_t^j, y_{t-1}^{i_1}, \dots, y_{t-p}^{i_p}\} \right\}. \end{aligned}$$

Математическую основу нечеткого моделирования временных рядов образуют нечеткие модели и теоретические выводы, рассмотренные в разделе 2.4, в частности, теорема FAT (Fuzzy Approximation Theorem), согласно которой любая математическая система может быть аппроксимирована системой, основанной на нечеткой логике. Другими словами, с помощью естественно-языковых высказываний «ЕСЛИ-ТО», с последующей их формализацией средствами теории нечетких множеств, можно сколько угодно точно описать произвольную взаимосвязь «входы-выход».

3.3.2. Нечеткое сглаживание временного ряда

Нечеткое сглаживание временных рядов – методика, разработанная И. Перфильевой [Перфильева, 2003], которая может быть отнесена к методикам нечеткого приближения на основе нечеткого преобразования.

Нечеткое преобразование (F-преобразование) представлено для непрерывных функций и функций на дискретном наборе точек. В этом случае F-преобразование называется дискретным нечетким преобразованием, которое и используется для анализа временных рядов.

F-преобразование предполагает задание нечеткого разбиения универсального множества. В качестве последнего выбирается конечный интервал $[a, b]$ действительной прямой. Зафиксируем n ($n \geq 2$) узлов $x_1, \dots, x_n$ на $[a, b]$ и предположим, что $x_1 \prec \dots \prec x_n$, причем $a = x_1, b = x_n$.

Определение 3.1 [Perfilieva, 2006]. Под нечетким разбиением $[a, b]$ будем понимать совокупность n функций $A_1, \dots, A_n : [a, b] \rightarrow [0, 1]$, удовлетворяющих следующим свойствам:

- 1) $A_k : [a, b] \rightarrow [0, 1], A_k(x_k) = 1$;
- 2) $A_k(x) = 0$, если $x \notin (x_{k-1}, x_{k+1})$, где для единообразия обозначения мы положим $x_0 = a, x_{n+1} = b$;
- 3) $A_k(x)$ непрерывна;
- 4) $A_k(x), k = 2, \dots, n$ строго возрастает на $[x_{k-1}, x_k]$ и $A_k(x), k = 1, \dots, n-1$, строго убывает на $[x_k, x_{k+1}]$;

5) для всех $x \in [a, b]$ $\sum_{k=1}^n A_k(x) = 1$.

Функции $A_1, \dots, A_n$ называются *базисными функциями*. Базисные функции $A_1, \dots, A_n$ могут служить также функциями принадлежности нечетких подмножеств $A_1, \dots, A_n$ (обозначения функций и множеств унифицированы). Отметим, что форма базисных функций может быть уточнена дополнительно и согласована с такими требованиями к модели, как, например, гладкость.

Следующие формулы представляют нечеткое разбиение отрезка $[x_1, x_n]$, полученное совокупностью функций:

$$A_1(x) = \begin{cases} 1 - \frac{(x - x_1)}{h_1}, & x \in [x_1, x_2], \\ 0, & \text{иначе} \end{cases}$$

$$A_k(x) = \begin{cases} \frac{(x - x_{k-1})}{h_{k-1}}, & x \in [x_{k-1}, x_k], \\ 1 - \frac{(x - x_k)}{h_k}, & x \in [x_k, x_{k+1}], \\ 0, & \text{иначе} \end{cases}$$

$$A_n(x) = \begin{cases} 1 - \frac{(x - x_{n-1})}{h_{n-1}}, & x \in [x_{n-1}, x_n], \\ 0, & \text{иначе} \end{cases}$$

где $k = 1, \dots, n-1$, и $h_k = x_{k+1} - x_k$.

Предположим, что функция f имеет своей областью определения множество $P = \{p_1, \dots, p_l\} \subset [a, b]$, где $l > n$. Множество P считается плотным относительно нечеткого разбиения $A_1, \dots, A_n$, если выполнено условие:

$$(\forall k) (\exists j) A_k(p_j) > 0.$$

Пусть $A_k(p_j) = a_{kj}$, $k = 1, \dots, n$; $j = 1, \dots, l$, тогда матрица $A_{n \times l} = (a_{kj})$ называется матрицей нечеткого разбиения для P , для которой справедливы свойства:

$$1) (\forall k) (\forall j) a_{kj} \in [0, 1];$$

$$2) (\forall j) \sum_{k=1}^n a_{kj} = 1.$$

Отождествляя функцию $f : P \rightarrow R$ с множеством ее значений на P , т.е. $f = (f_1, \dots, f_l)$, где $f_j = f(p_j)$, $j = 1, \dots, l$, отметим, что $f \in R^l$, где R^l – множество l – мерных векторов с действительными координатами.

Определение 3.2 [Perfilieva, 2006]. F -преобразованием вектора $f \in R$, определяемым матрицей нечеткого разбиения A , назовем вектор $F_n[f] \in R^n$, где

$$F_n[f] = (F_1, \dots, F_n) \text{ и } F_i = \frac{\sum_{j=1}^l a_{ij} f_j}{\sum_{i=1}^l a_{ij}}.$$

Координаты вектора $F_n[f]$ назовем компонентами F -преобразования.

Обозначим $a_i = \sum_{j=1}^l a_{ij}$, $i = 1, \dots, n$; тогда $(a_1 F_1, \dots, a_n F_n)^T = A \cdot f$. Компоненты F -преобразования являются точками минимума функции, задающей критерий взвешенного среднеквадратичного отклонения.

Теорема 3.1 [Perfilieva, 2006]. Пусть имеют место все вышеизложенные предположения относительно $A_1, \dots, A_n, P, f$.

Тогда компоненты $F_n[f] = (F_1, \dots, F_n)$ F -преобразования функции f минимизируют следующую функцию действительных переменных $y_1, \dots, y_n$

$$\Phi(y_1, \dots, y_n) = \sum_{i=1}^n \sum_{j=1}^l (f_j - y_i)^2 \cdot a_{ij}.$$

Определение 3.3 [Perfilieva, 2006]. Пусть $f \in R^l$, и $F_n[f] \in R^n$ есть F -преобразование f , определяемое матрицей $A_{n \times l} = (a_{kj})$. Обратным F -преобразованием $F_n[f]$ назовем вектор $f_{F,n} \in R^l$, вычисляемый по формуле $f_{F,n}^T = F_n[f] \cdot A$.

Можно доказать, что если n возрастает, тогда $f_{F,n}(p_j)$ сходится к $f(p_j)$, $j = 1, \dots, N$.

F -преобразование имеет (кроме прочих) следующие свойства, важные для использования в качестве сглаживания временных рядов: (а) у него прекрасные

фильтрующие свойства; (b) его легко вычислять; (c) F-преобразование стабильно относительно выбора точек $p_1, \dots, p_N$. Это означает, что при выборе других точек p_k (и, возможно, изменяя их число N), результирующая функция $f_{F,n}$ значительно не меняется. Подробное формальное описание F-преобразования, включая необходимые теоремы, можно найти в [Perfilieva, 2006].

Использование F-преобразования в моделировании ВР рассмотрено в работе [Афанасьева и др., 2009].

3.3.3. Обзор методов моделирования нечетких временных рядов

Моделирование нечетких временных рядов в соответствии с нечеткой моделью, предложенной в работе [Song, 1993 a], состоит в реализации следующих шагов:

1. Определение нечетких переменных – разбиение данных на множество интервалов (носителей нечетких множеств), определение лингвистических значений нечетких множеств и их функций принадлежности.
2. Формирование логических отношений $Y_t \rightarrow Y_{t-1}$.
3. Фаззификация входных данных – определение степени принадлежности входных данных входным нечетким переменным.
4. Вычисление результата применения нечеткого правила $R_{ij}(t, t-1)$ для каждой импликации.
5. Вычисление результирующего отношения R как объединение $\bigcup_{i,j} R_{ij}(t, t-1)$.
6. Применение полученной модели к входным данным и получение выходных нечетких результатов.
7. Дефаззификация нечетких результатов.

Предложенная Сонгом [Song, 1993 a] модель НВР имеет следующие недостатки:

1. Эвристическое задание количества входных, выходных переменных и параметров нечетких множеств.
2. При реализации нечеткой максиминой композиции модели, требуется большое количество вычислений, особенно, когда нечеткое отношение очень велико.
3. Отсутствуют возможности проверки на полноту базы знаний и поиска наилучшей модели, что приводит к недостаточной точности модели.
4. Не реализован анализ паттернов (компонент) временного ряда.

Чен [Chen, 1996], полагая, что метод Сонга [Song, 1993] слишком сложен для применения, предложил использовать арифметические операции вместо логической максиминной композиции.

После этих работ началось всестороннее исследование предложенных нечетких моделей ВР, было разработано множество расширений и выявлены проблемы НВР.

Одной из проблем в нечетком моделировании ВР является отсутствие четких рекомендаций на первом этапе построения модели по выбору количества и параметров нечетких множеств, моделирующих входные и выходные переменные, в частности по определению их носителей (длины интервалов). Данные задачи выполняются экспертом, и, как показывают исследования, от выбора интервалов сильно зависит результат исследования.

Проблема длин интервалов ставилась, но должным образом не обсуждалась, пока эффективные длины интервалов не были исследованы в работе [Huarng, 2001 a]. Исследование показало, что различные длины интервалов могут привести к различным нечетким отношениям и, в свою очередь, породить различные модели и результаты прогноза.

Chen и Hsu [Chen, 2004] предложили метод прогнозирования для нестационарных (переменных во времени) нечетких временных рядов применительно к данным регистрации студентов университета штата Алабама. В методе они использовали повторное деление изначально выбранных интервалов на 4, 3, 2 части в зависимости от того, в каком интервале содержится большее коли-

чество наблюдаемых данных. Они использовали эвристические правила, позволяющие определить тенденцию на интервале.

В работе [Chen, 2006] Chen и Chung предложили метод определения длины интервалов, основой которого является генетический алгоритм.

В работе [Jilani, 2007] был исследован метод выбора эффективной длины интервалов на основе среднего значения разницы между соседними уровнями временного ряда и знаний эксперта, формализованных в виде таблицы, по которой определяется интервал на основе полученного среднего.

Для повышения точности модели ВР авторы работы [Khashei, 2008] использовали способ построения функций принадлежности, в частности интервалов, на которых они строятся, в зависимости от частотности соответствующих этим интервалам наблюдений. Также приводятся приложения методов анализа НВР к прогнозу абитуриентов университета штата Алабама (эти данные используют практически все исследователи для сравнения результатов), для прогнозирования аварий на дорогах в Бельгии, для прогноза погоды. Затем показано, что предлагаемый метод превосходит по точности другие методы моделирования НВР.

В работе [Chen, 2002] Чен предложил метод прогноза регистрации студентов, основанный на нечетких временных рядах старших порядков: второго, третьего, четвертого и пятого. Сонг [Song, 2003] продолжил исследование Чена, используя функцию автокорреляции как меру зависимости между нечеткими данными для выбора подходящего порядка в модели нечетких временных рядов. Он пришел к выводу, что применение моделей второго и третьего порядка более эффективно, чем первого.

Оун и Ю (Own и Yu) [Own, 2005] предложили эвристическую модель старших порядков, введя эвристическую функцию. Tsaur и др. [Tsaur, 2005] использовал понятие энтропии, чтобы измерить степень нечеткости системы и определять время T , для которого данные приближаются к устойчивому состоянию.

В работе [Şah, 2004] предложена модификация метода Сонга, в которой вместо значений временного ряда для составления универсума, на котором построены нечеткие множества, используются приращения (изменения соседних уровней временного ряда).

Huarng [Huarng, 2001б] усовершенствовал модель Чена, используя эвристики – неформальные, интуитивные стратегии, – которые выражают ожидания экспертов/аналитиков относительно тренда регистрации студентов в будущем году (так называемые тренды роста/снижения/стабильности).

Таким образом, уже в этих исследованиях намечается новое направление, связанное с повышением точности моделей НВР за счет применения алгоритмов поиска оптимальных носителей нечетких множеств, выбора порядка моделей и выделением правил, описывающих изменения тенденций в структуре нечеткого временного ряда.

3.3.4. Перспективы в моделировании нечетких временных рядов

Приведенный выше обзор подходов к нечеткому моделированию позволяет сделать некоторые выводы и обозначить проблемы.

Нечеткие временные ряды появились как эволюционное развитие формализма нечетких множеств в пространство математических моделей анализа поведения временных рядов.

Для временных рядов различной природы моделирование и анализ их поведения с привлечением дополнительных знаний, описывающих неопределенность на основе нечетких множеств, как представляется, позволит не только решать традиционные задачи анализа числовых ВР, но и существенно расширить их круг за счет обработки данных нового типа.

Анонсируемым достоинством программных систем моделирования ВР в виде нечеткого ВР, отмечаемым практически всеми исследователями этого направления, является их продуктивность в качестве альтернативного инструмента моделирования числовых временных рядов (сравнение нечетких моделей со статистическими моделями приведено в работе [Дегтярев, 2007]). Лингвистиче-

ские термы НВР, моделируемые нечеткими множествами, могут явно отражать семантику объектов прикладной области, формализация которой в модели с одной стороны повышает степень ее адекватности, а с другой стороны, улучшает ее понимание прикладными пользователями.

Указанная возможность модели НВР позволяет надеяться, что решение задач анализа временных рядов различной природы станет более доступным для широкой категории заинтересованных пользователей, для которых извлечение нечетких правил из ВР, выраженных и в лингвистических термах, наиболее предпочтительна. К таким пользователям относятся пользователи, основная профессиональная компетенция которых связана с принятием решений: лица, принимающие решения, эксперты прикладных областей, менеджеры, проектировщики.

Таким образом, отметим следующий ожидаемый эффект от использования нечетких моделей ВР:

1. Расширение пользователей программных систем анализа ВР.
2. Расширение видов обрабатываемых данных.
3. Решение новых задач выявления нелинейных зависимостей, выраженных в лингвистических терминах.

В то же время нечеткие модели временных рядов, представленные в виде нечетких временных рядов, требуют дальнейшего исследования и развития, так как в приведенных выше работах не рассматривались прогностические возможности нечетких моделей ВР по внешним показателям качества, вычисляемым на тестовых примерах.

3.3.5. Пример моделирования временного ряда в нечетком подходе

Приведем пример нечеткого моделирования временного ряда для прогнозирования прямых валютных котировок ЦБ USD/RUB за июнь 2005 года, описанный в работе [Дегтярев, 2008] и представляющий модификацию метода Сонга [Song, 1994], отличающуюся (а) использованием изменений (прираще-

ний) данных прошлого вместо реальных числовых значений (регистрации или валютного курса), и (б) вычислением отношений R_j для предсказания будущих состояний.

Рассматриваемый в работе метод был изначально успешно применен к временному ряду, характеризующему количество поступающих в Алабамский университет, которые являются бенчмаркингом при сравнении методов моделирования нечетких временных рядов [Şah, 2004]. Анализ результативности предлагаемого метода по показателю точности средней относительной ошибки аппроксимации для 6 нечетких множеств (MAPE=2,42) показал, что предложенный метод превышает аналогичный показатель для этого ВР, полученный методом Сонга и Чена (см. рис. 3.2).

Применительно к проблеме прогнозирования валютного курса USD/RUB пошаговое описание предлагаемого метода нечеткого моделирования ВР можно свести к следующему:

Шаг 1: Задание области определения (универсального множества U) проблемы, исходя из вычисленных приращений валютного курса в течение рассматриваемого интервала времени.

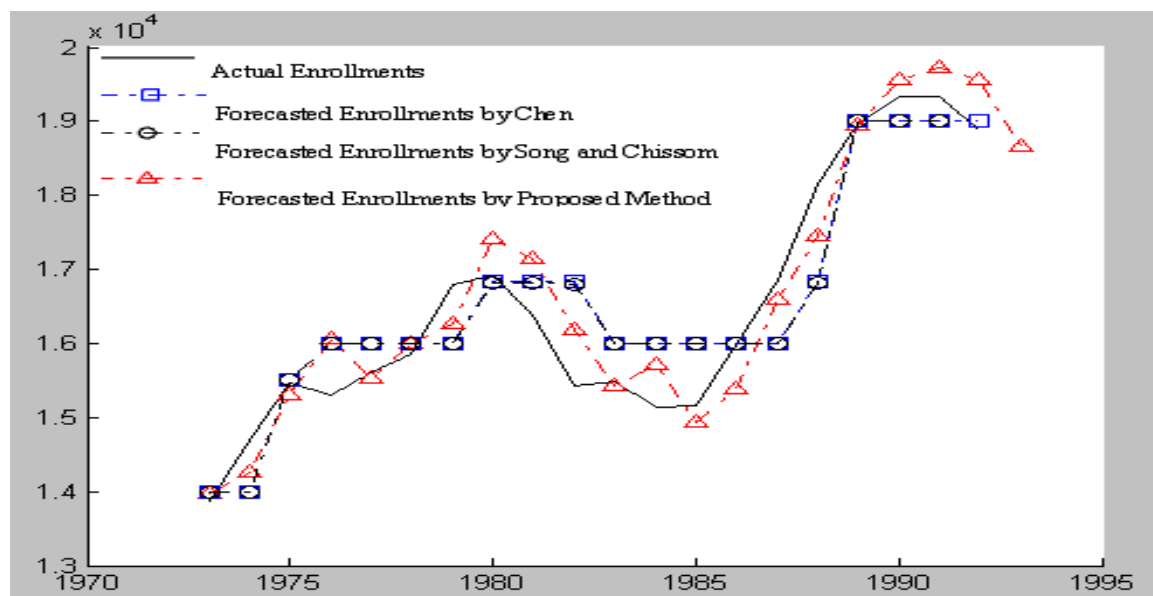


Рис. 3.2. Сравнение результатов нечеткого моделирования

Наибольшее положительное приращение курса доллара по отношению к российскому рублю наблюдается в феврале 2002 года, т. е. по сравнению с январским значением рост составляет 0.3679 (более 36 копеек/месяц). В ноябре-декабре 2004 года происходит самое значительное за трехлетний период наблюдений падение котировки доллара почти на 70 копеек (-0.6949). В результате, с целью упрощения последующего разбиения на равновеликие интервалы полученные граничные значения (-0.6949 и +0.3679) слегка корректируются. Например, в случае использования 6-ти подинтервалов U может быть представлена отрезком $[-0.7, -0.5]$.

Шаг 2: Разбиение множества U на интервалы одинаковой длины.

Если мы оперируем с шестью нечеткими множествами, то область определения делится на 6 интервалов $u_i, i = \overline{1,6}$, $u_1 = [-0.7, -0.5]$, $u_2 = [-0.5, -0.3]$, ..., $u_6 = [0.3, 0.5]$ (в действительности, количество нечетких множеств не обязательно должно совпадать с числом интервалов разбиения).

Шаг 3: Определение нечетких множеств A_i .

Предположим, что лингвистическая переменная «изменение валютного курса» характеризуется терм-множеством, образуемым следующими значениями: A_1 (значительное уменьшение), A_2 (уменьшение), A_3 (без изменений/флэт), A_4 (увеличение), A_5 (значительное увеличение), A_6 (очень большое увеличение).

Для шести построенных выше интервалов $u_i, i = \overline{1,6}$, факт принадлежности каждого конкретного u_i определенному множеству $A_j, j = \overline{1,6}$ выражается действительным числом из единичного интервала $[0,1]$ (предполагается, что элементы, отсутствующие в представлении множеств A_j , характеризуются нулевой степенью принадлежности):

$$\begin{aligned}
A_1 &= \{1/u_1 + 0.5/u_2\} \\
A_2 &= \{0.5/u_1 + 1/u_2 + 0.5/u_3\} \\
A_3 &= \{0.5/u_2 + 1/u_3 + 0.5/u_4\} \\
A_4 &= \{0.5/u_3 + 1/u_4 + 0.5/u_5\} \\
A_5 &= \{0.5/u_4 + 1/u_5 + 0.5/u_6\} \\
A_6 &= \{0.5/u_5 + 1/u_6\},
\end{aligned}$$

где $u_i \subset U$ – элементы универсума U , а число, стоящее в числителе каждого элемента нечеткого множества, представляет собой степень принадлежности $\mu(u_i)$ этого элемента нечеткому множеству A_j , $j = \overline{1,6}$.

Шаг 4: Фаззификация приращений, полученных на шаге 1.

Считаем, что если приращение года t есть $p \in u_i$, и существует лингвистическое значение (нечеткое множество A_j) с максимальной степенью принадлежности, приходящейся на элемент u_i , тогда p фаззифицируется как A_j . Например, приращение за март 2002 по сравнению с предыдущим месяцем составляет +0.2476 – это значение попадает в интервал u_5 , и фаззифицированное приращение становится равным A_5 . Аналогичным образом производятся попарные сравнения каждого последующего и предыдущего месяцев, приводящие к формированию последовательности A_6 (февраль 2002), A_5 (март 2002), A_5 (апрель 2002), A_4 (май 2002), A_5 (июнь 2002), A_5 (июль 2002), A_4 (август 2002) и т. д.

Шаг 5: Формирование логических отношений $A_i \rightarrow A_j$.

Для построения последовательности логических отношений, мы рассматриваем попарно последовательные фаззифицированные приращения (февраль – март, март – апрель, и т. д.), определенные на шаге 4. Исключая повторяющиеся комбинации, окончательный список отношений принимает вид:

$$\begin{aligned}
&A_1 \rightarrow A_2, A_1 \rightarrow A_4 \\
&A_2 \rightarrow A_1, A_2 \rightarrow A_2, A_2 \rightarrow A_3, A_2 \rightarrow A_4, A_2 \rightarrow A_5 \\
&A_3 \rightarrow A_2, A_3 \rightarrow A_3, A_3 \rightarrow A_4 \\
&A_4 \rightarrow A_2, A_4 \rightarrow A_3, A_4 \rightarrow A_4, A_4 \rightarrow A_5 \\
&A_5 \rightarrow A_3, A_5 \rightarrow A_4, A_5 \rightarrow A_5, A_5 \rightarrow A_6 \\
&A_6 \rightarrow A_5, A_6 \rightarrow A_4.
\end{aligned}$$

Мы предполагаем, что нечеткое импликативное отношение $D = B \rightarrow C$ для произвольных векторов B и C интерпретируется как нечеткая импликация Мамдани, следовательно, элементы матрицы D вычисляются по формуле $d_{ij} = b_i^T \times c_j = \min(b_i, c_j)$, где b_i и c_j – элементы векторов B и C , соответственно.

Шаг 6: Объединение логических отношений (шаг 5), имеющих одинаковые левые части, в группы, и вычисление отношений $R_i, i = \overline{1,6}$ для каждой сформированной группы. Можно обратить внимание на то, что группы отношений уже практически построены (см. шаг 5), и выглядят они следующим образом:

$$\begin{aligned} A_1 &\rightarrow A_2, A_4 \\ A_2 &\rightarrow A_1, A_2, A_3, A_4, A_5 \\ A_3 &\rightarrow A_2, A_3, A_4 \\ A_4 &\rightarrow A_2, A_3, A_4, A_5 \\ A_5 &\rightarrow A_3, A_4, A_5, A_6 \\ A_6 &\rightarrow A_5, A_4. \end{aligned}$$

Результирующие отношения $R_i, i = \overline{1,6}$, представляют собой объединения логических отношений, попавших в i -тую группу:

$$\begin{aligned} R_1 &= A_1^T \times A_2 \cup A_1^T \times A_4 \\ R_2 &= A_2^T \times A_1 \cup A_2^T \times A_2 \cup A_2^T \times A_3 \cup A_2^T \times A_4 \cup A_2^T \times A_5 \\ R_3 &= A_3^T \times A_2 \cup A_3^T \times A_3 \cup A_3^T \times A_4 \\ R_4 &= A_4^T \times A_2 \cup A_4^T \times A_3 \cup A_4^T \times A_4 \cup A_4^T \times A_5 \\ R_5 &= A_5^T \times A_3 \cup A_5^T \times A_4 \cup A_5^T \times A_5 \cup A_5^T \times A_6 \\ R_6 &= A_6^T \times A_4 \cup A_6^T \times A_5. \end{aligned}$$

Шаг 7: Прогнозирование и дефаззификация получаемых результатов.

Вычисленные отношения R_i используются в модели прогнозирования

$$A_i = A_{i-1} \circ R_i,$$

где A_i – нечеткое множество, выражающее прогнозное приращение месяца i ;

A_{i-1} – известное приращение предшествующего $(i-1)$ -го месяца (если

$A_{i-1} = A_j$, то $R_i = R_j, j = \overline{1,6}$);

$\circ$ – обозначает композиционный «max-min» оператор.

Например, приращение валютного курса за февраль 2004 года при известном приращении (-0.5714) за январь месяц того же года вычисляется по формуле $F(02.2004) = A_1 \circ R_1$, где R_1 имеет вид, показанный в первой строке, а A_i – фаззифицированное приращение января 2004 года.

Шаг 8: Вычисление прогнозных валютных котировок USD/RUB.

Этот этап предусматривает преобразование полученных на шаге 7 нечетких прогнозных приращений в целые числа. В значительной степени такой процесс зависит от особенностей рассматриваемой задачи, и одним из критериев выбора процедуры дефаззификации является ее вычислительная простота.

После того как получено числовое приращение для рассматриваемого месяца, оно суммируется с уже имеющимся значением обменного курса предыдущего месяца. Рассмотренный метод нечеткого моделирования может быть отнесен к числу полуавтоматических процедур, поскольку большинство выполняемых шагов, включая построение универсума на основании множества исходных данных задачи, могут быть эффективно воплощены в программной форме, однако участие аналитика (эксперта) при формировании интервалов разбиения и соответствующих нечетких множеств играет также огромную роль.

3.3.6. Пример построения нечеткой системы прогнозирования дефектов металлопродукции

В работе [Кудинов и др., 2007] рассматриваются основные принципы построения нечеткой системы прогнозирования дефектов металлопродукции, включающие анализ технологических процессов, постановку задачи прогнозирования, разработку нечетких моделей и алгоритмов обучения, а также опытную проверку системы на одном из металлургических предприятий.

Отличительными особенностями таких производств являются нестационарность процессов, исключительная сложность и нелинейность связей между переменных, огромный объем информации, подверженной влиянию помех и погрешностей измерения. В этих условиях надежное прогнозирование дефектов возможно лишь с помощью нечеткой модели.

Постановка задачи прогнозирования. Обозначим вектор входных переменных через $\mathbf{x}_r^g = (x_{1r}^g, x_{2r}^g, \dots, x_{jr}^g, \dots, x_{mr}^g) = (\mathbf{z}^g, \mathbf{u}_r, \mathbf{z}_r)$ и исключим из рассмотрения вектор ненаблюдаемых входных переменных $\mathbf{e}_r$, учитываемых при идентификации нечеткой модели. Входные воздействия $x_j(t_{jk})$, $j = \overline{1, m}$, действующие в моменты времени t_{jk} и формирующие дефект $y_i(k)$ в момент времени t_k , можно записать в виде вектора

$$\mathbf{x}(k) = (x_1(k), x_2(k), \dots, x_m(k)) = (x_1(t_{1k}), x_2(t_{2k}), \dots, x_m(t_{mk})),$$

показывая связь входов $x_j(k)$, $j = \overline{1, m}$ и выхода $y_i(k)$ в момент времени t_k .

Таким образом, для прогнозирования качества металлопродукции требуется разработать нечеткие многосвязные модели, рассчитывающие значения дефектов $y_i(k)$ слывов, полученных в r -ом ручье из g – ой марки стали в моменты времени t_k , $k = \overline{1, N}$:

$$\hat{y}_i(k) = f(\mathbf{x}(k), \xi_i), \quad \forall i = \overline{1, q},$$

где $\xi_{r,i}^g$ – вектор параметров и структурных элементов модели.

Точность прогноза, т. е. близость расчетного $\hat{y}_i(k)$ к измеренному $y_i(k)$ значению i -го дефекта оценивается величиной критерия

$$J_i = \frac{1}{N} \sum_{k=1}^N |y_i(k) - \hat{y}_i(k)|, \quad \forall i = \overline{1, q}.$$

Разработка нечеткой модели прогнозирования. Для прогнозирования дефектов целесообразно использовать нечеткие модели, состоящие из совокупности продукционных правил, в правой части которых находятся линейные уравнения:

$$R_i^\theta : \text{если } x_1(k) \text{ есть } X_{i1}^\theta, x_2(k) \text{ есть } X_{i2}^\theta, \dots, x_j(k) \text{ есть } X_{ij}^\theta, \dots, x_m(k) \text{ есть } X_{im}^\theta, \\ \text{то } y_i^\theta(k) = c_{i0}^\theta + c_{i1}^\theta x_1(k) + \dots + c_{ij}^\theta x_j(k) + \dots + c_{im}^\theta x_m(k), \theta = \overline{1, n}, i = \overline{1, q},$$

где X_{ij}^θ – нечеткие множества, описываемые функциями принадлежности (ФП) $X(x_j, \mathbf{d}_{ij}^\theta)$, зависящими от входных переменных x_j и векторов параметров $\mathbf{d}_{ij}^\theta = (d_{ij1}^\theta, d_{ij2}^\theta, \dots, d_{ij\sigma}^\theta)$, $\theta = \overline{1, n}$, $j = \overline{1, m}$, $i = \overline{1, q}$;

θ, n_i – номер и количество продукционных правил нечеткой модели, прогнозирующей i -й показатель качества y_i ;

c_{ij}^θ – коэффициенты линейных уравнений.

Продукционные правила совместно с операциями фазификации, нечеткого вывода и фазификации образуют нечеткую модель, имеющую следующую аналитическую форму:

$$y_i(k) = \sum_{\theta=1}^{n_i} (c_i^\theta)^T x_i^\theta(k, d_i^\theta),$$

где $(c_i^\theta)^T = (c_{i0}^\theta, c_{i1}^\theta, \dots, c_{im}^\theta)$ – вектор коэффициентов линейного уравнения θ -го правила;

$\theta = \overline{1, n_i}$; $\tilde{x}_i^\theta(k) = (\beta_i^\theta, \beta_i^\theta x_1(k), \dots, \beta_i^\theta x_m(k))$ – расширенный входной вектор θ -го правила, содержащий в качестве множителя нелинейную нечеткую функцию параметров d_i^θ

$$\beta_i^\theta(d_i^\theta) = \frac{w_i^\theta}{\sum_{\theta=1}^{n_i} w_i^\theta},$$

где w_i^θ – значение истинности θ -го правила, равное произведению функций принадлежности

$$w_i^\theta = X_i^\theta(x_1) \cdot X_i^\theta(x_2) \cdot \dots \cdot X_i^\theta(x_m).$$

Теперь в векторе ξ , уточняемом алгоритмом идентификации, можно указать параметрические (коэффициенты линейных уравнений $c = \{c_{ij}^\theta\}$, параметры ФП $d = \{d_{ij}^\theta\}$, $\theta = \overline{1, n}$, $i = \overline{1, q}$, $j = \overline{1, m}$) и структурные элементы (количество правил n и число m и состав переменных).

Опытная проверка системы. Программный комплекс (ПК) осуществляет построение, обучение более 900 нечетких моделей прогнозирования 13 дефектов слябов, изготовленных из 7 групп марок сталей в 10 ручьях.

В течение двух недель в АСУ «Качество» конвертерного производства ОА «Северсталь» проводилась опытная проверка ПК на основании разработанной программы испытаний. ПК признается пригодным для прогнозирования дефекта i , если величина точности его прогноза $J_{\Pi}^r(i)$ не ниже допустимого значения $J_{\Pi}^{\partial} = 0.8$, т. е. удовлетворяет условию

$$J_{\Pi}^r(i) \geq J_{\Pi}^{\partial}.$$

Фрагменты результатов испытаний ПК представлены в таблице 3.2.

Таблица 3.2

Оценка качества нечеткой модели прогнозирования
дефектов металлопродукции

Сталь	Ручей	Дефект	Балл	Количество данных (N_{Π})	Совпадение		Точность прогноза балла	Точность прогноза дефекта
					полное ($N_{\Pi c}$)	по дефекту ($N_{\Pi d}$)		
1	1		0	0	0	0	0	0,95
			1	0	0	0	0	
			2	5	4	1	0,95	
			3	0	0	0	0	
3	2		0	0	0	0	0	0,89
			1	0	0	0	0	
			2	11	6	5	0,89	
			3	0	0	0	0	
6	5		0	0	0	0	0	0,86
			1	20	17	2	0,92	
			2	160	51	100	0,79	
			3	0	0	0	0	
6	7		0	3	3	0	1	0,87
			1	11	5	6	0,86	
			2	1	0	1	0,75	
			3	0	0	0	0	

Из таблицы 3.2 следует, что при достаточном количестве данных можно считать нечеткую модель пригодной для прогнозирования дефектов изготовленных из трех марок сталей 1, 3, 6 в ручьях 1, 2, 5, 7.

Контрольные вопросы

1. Сформулируйте особенности нечеткого моделирования ВР.
2. Приведите примеры прикладных задач обработки нечетких ВР.
3. Что означает нечеткое сглаживание ВР?
4. Дайте определение нечеткому разбиению.
5. Опишите основные этапы алгоритма моделирования нечетких ВР в соответствии с нечеткой моделью Сонга.
6. В чем преимущество использования моделей нечетких ВР?

Библиографический список

1. [Song, 1993 a] Song, Q. Fuzzy time series and its models / Q. Song, B. Chissom // Fuzzy Sets and Systems. – №54 (1993) – P. 269-277.
2. [Song, 1993 б] Song, Q. Forecasting enrollments with fuzzy time series – Part I / Q. Song, B. Chissom // Fuzzy Sets and Systems. – №54 (1993) – P. 1-9.
3. [Song, 1994] Song, Q. Forecasting enrollments with fuzzy time series – Part II / Q. Song, B. Chissom // Fuzzy Sets and Systems. – №64 (1994) – P. 1-8.
4. [Chen, 1996] Chen, S. M. Forecasting enrollments based on fuzzy time series / S.M. Chen // Fuzzy Sets and Systems. – № 81 (1996) – P. 311–319.
5. [Huarng, 2001 a] Huarng, K. Effective lengths of intervals to improve forecasting in fuzzy time series / K. Huarng // Fuzzy Sets and Systems, 2001. – №123). – P. 387-394.
6. [Huarng, 2001 б] Huarng, K. Heuristic models of fuzzy time series for forecasting / K. Huarng // Fuzzy Sets and Systems. – №123 (2001). – P. 369-386.
7. [Chen, 2002] Chen, S. M. Forecasting enrollments based on high-order fuzzy time series / S.M. Chen // Cybernetics and Systems: An International Journal. – №33 (2002). – P. 1-16.
8. [Song, 2003] Song, Q. A note on fuzzy time series model relation with sample autocorrelation functions / Q. Song // Cybernetics and Systems: An International Journal. – №34 (2003). – P. 93-107.

9. [Chen, 2004] Chen, S. M. A new method to forecast enrollments using fuzzy time series / S. M. Chen // International Journal of Applied Sciences and Engineering. – №2 (3) (2004) . – P. 234-244.
10. [Khashei, 2008] Khashei, M. Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks / M. Khashei, M. Bijari , G. Rassi Ardali // Neurocomputing, 2008.
11. [Own, 2005] Own, C. M. Forecasting fuzzy time series on a heuristic high-order model / C. M. Own, P. T. Yu // Cybernetics and Systems: An International Journal. – №36 (2005). – P. 705-717.
12. [Перфильева, 2003] Перфильева, И. Нечеткое преобразование. / И. Перфильева // Нечеткая логика. – Амстердам, 2003. – С. 275-300.
13. [Perfilieva, 2006] Perfilieva, I. Fuzzy transforms: Theory and applications / I. Perfilieva // Fuzzy Sets and Systems, 2006. – №157.
14. [Şah, 2004] Şah, M. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series / M. Şah, K. Y. Degtiarev // Proc. Int. Conf. Computational Intelligence (ICCI) (2004). – P. 375-378
15. [Tsaur, 2005] Tsaur, R. C. Fuzzy relation analysis in fuzzy time series model / R. C. Tsaur, J. C. O. Yang, H. F. Wang // Computer and Mathematics with Applications.– №49 (2005) . – P. 539-548.
16. [Chen, 2006] Chen, S. M. Forecasting enrollments of students by using fuzzy time series and genetic algorithms / S. M. Chen, N. Y. Chung // Information and Management Sciences, Vol. 17.– №3, September, 2006. – P. 1-17.
17. [Jilani, 2007] Jilani, Tahseen Ahmed. Fuzzy Metric Approach for Fuzzy Time Series Forecasting based on Frequency Density Based Partitioning / Tahseen Ahmed Jilani, Syed Muhammad Aqil Burney, Cemal Ardil. // Proceedings of World Academy of Science, Engineering and Technology, Vol. 23, 2007.
18. [Xihao, 2008] Xihao, Sun. Average-based fuzzy time series models for forecasting Shanghai compound index / Sun Xihao, Li Yimin // World Journal of Modelling and Simulation, Vol. 4 (2008) .– №2. – P. 104-111.

19. [Дегтярев, 2007] Дегтярев, К. Ю. Применение специализированных компьютерных программ и методов, основанных на нечетких временных рядах для краткосрочного прогнозирования USB/RUB котировок / К. Ю. Дегтярев. [Доступно по адресу: <http://www.exponenta.ru/educat/news/degtyarev/paper.pdf>; дата обращения 30.12.2009].
20. [Дегтярев, 2008] Дегтярев, К. Ю. Прогнозирование валютных котировок с использованием модифицированного стационарного метода, основанного на нечетких временных рядах / К. Ю. Дегтярев [Доступно по адресу: <http://www.exponenta.ru/educat/news/degtyarev/paper2.pdf>; дата обращения 30.12.2009].
21. [Кудинов и др., 2007] Кудинов, Ю. И. Разработка и идентификация нечетких моделей прогнозирования качества / Ю. И. Кудинов, К. С. Иванченко, И. Ю. Кудинов // Мехатроника, автоматизация, управление, 2007.– №12. – С. 12-15.
22. [Афанасьева и др., 2009] Афанасьева, Т. В. F-преобразование в прогнозировании временных рядов / Т. В. Афанасьева, А. А. Ивахина, И. Г. Перфильева // ИННОВАТИКА 2009 : Труды Международной конференции. – Ульяновск : УлГУ, 2009. – С. 459-461.
23. [Новак, 2008] Новак, В. Интегральный метод принятия решений и анализа нечетких временных рядов / В. Новак, И. Перфильева, Н. Ярушкина и др. // Программные продукты и системы. – 2008.– №4 . – С. 65-68.

3.4. Гибридные модели временных рядов

Опыт последних лет показал, что применение однородных методов, то есть методов, соответствующих одной научной парадигме, для решения сложных задач, к которым несомненно относится задача моделирования ВР, далеко не всегда приводит к успеху. В гибридной архитектуре систем анализа ВР, объединяющей несколько парадигм, эффективность одного подхода может компенсировать слабость другого. Поэтому одной из активно развивающихся тен-

денций в настоящее время является создание интегрированных, гибридных и синергетических систем, объединяющих различные методы и технологии в интересах достижения более глубокого понимания причинных механизмов в поведении временных рядов

3.4.1. Нечетко-статистический подход в моделировании временных рядов

Нечеткий регрессионный анализ

В области прикладной статистики, анализа временных рядов и принятия решений в условиях неопределенности накоплен богатый опыт исследований и существует множество моделей, начиная от простейших линейных регрессионных моделей поиска тренда временного ряда и заканчивая сложными многоуровневыми авторегрессионными и адаптационными моделями. Регрессионный анализ, основанный на методе наименьших квадратов (Least-square), является очень удобным методом построения моделей, позволяющих численно оценивать зависимость интересующего исследователя параметра от воздействующих на него факторов. При анализе зависимости нечетких оценок от воздействующих факторов зачастую исследователям приходится иметь дело с важной информацией, которая не может быть задана точно. Некоторые наблюдения могут быть описаны только лингвистическими выражениями (типа «удовлетворительный», «хороший» и «превосходный»). Для таких данных аппаратом формализации может служить теория нечетких множеств.

Возможность аппроксимации нечетких данных, авторегрессия нечетких данных исследовались с 1982 по настоящее время во многих работах. Были разработаны различные нечеткие регрессионные модели, основой которых является модель нечеткой линейной регрессии. Работы разных лет опираются на эту модель, развивая, уточняя и дополняя ее.

В нечеткой регрессионной модели параметры представляются триангулярными нечеткими числами и являются коэффициентами в нечеткой линейной

функции. Неопределенность (vagueness) системы представляется суммарным разбросом («шириной») параметров (нечетких коэффициентов).

Построение модели состоит в нахождении оптимальных в некотором смысле коэффициентов с учетом нечеткой информации об объекте и субъективных представлений исследователя.

Базовые предположения нечеткой регрессии заключаются в том, что остатки, полученные как разность между наблюдениями и их оценками, продуцируются не случайными ошибками измерения, а неопределенностями (типа нечеткость) при вычислении параметров модели.

Можно выделить два основных подхода к построению моделей нечеткой линейной регрессии (рис. 3.3).

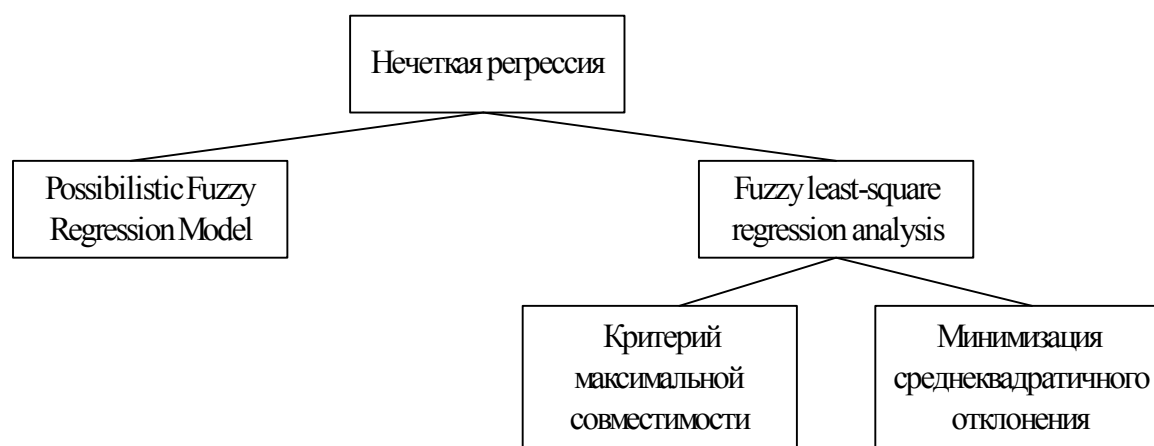


Рис. 3.3. Методы нечеткой регрессии

Первым подходом является нечеткая регрессия, основанная на критерии минимизации нечеткости (Possibilistic Fuzzy Regression Model) [Tanaka, 1982]. Вторым является подход, комбинированный с методом наименьших квадратов и получивший название FLSRA (Fuzzy least-square regression analysis) [Diamond, 1988; Celmiņš, 1987].

Этот метод, в свою очередь, имеет две разновидности, в одной из которых используется критерий максимальной совместимости, а в другой – критерий минимизации квадратичного отклонения.

Следует особенно отметить, что все три метода могут в качестве исходной информации об исследуемом параметре использовать как нечеткую ин-

формацию, выраженную в виде функций принадлежности, так и полностью детерминированную информацию, что существенно расширяет область их использования.

Большинство работ, посвященных нечеткой регрессии были основаны на следующих базовых определениях.

Пусть дано множество наблюдений: $(y_j, x_{j1}, \dots, x_{jn}), j = 1, \dots, m$, необходимо найти нечеткую модель по следующей форме:

$$\tilde{Y} = A_0 + A_1 x_1 + \dots + A_n x_n,$$

где $A_i(a_i^c, s_i^L, s_i^R), i = 1, \dots, n$ – триангулярные нечеткие числа;

a_i^c – среднее значение A_i ;

s_i^L, s_i^R – показывают левый и правый разброс соответственно.

Используются два критерия определения нечетких коэффициентов модели:

1. Для всех наблюдений принадлежность значения y_j к его нечеткой оценке Y_j должна быть как минимум $Y_j(y_j) \geq h, j = 1, \dots, m$, где h – уровень доверия, выбранный лицом, принимающим решения.

2. Общая нечеткость предсказываемого значения зависимой переменной должна быть минимизирована. Это может быть достигнуто минимизацией суммы разбросов нечетких чисел для всех наборов данных. Итак, проблему настройки нечеткой модели с заданными данными $(y_j, x_{j1}, \dots, x_{jn}), j = 1, \dots, m$ можно решить как эквивалентную задачу линейного программирования:

Найти $a_-^c = (a_0^c, \dots, a_n^c), s_-^L = (s_0^L, \dots, s_n^L), s_-^R = (s_0^R, \dots, s_n^R)$, которые минимизируют

$$Z = m(s_0^L + s_0^R)s_0^L + \sum_{i=1}^n \left[(s_i^L + s_i^R) \sum_{j=1}^m |x_{ij}| \right].$$

Чтобы оценить качество настройки нечеткой регрессии, используют метод наименьших квадратов. Для нечеткой регрессии среднеквадратичное отклонение (MSE) определяется следующим образом:

$$MSE = \frac{1}{m} \sum_{j=1}^n [y_j - def(Y_j)]^2,$$

где $def(Y_j)$ – дефаззифицированное значение зависимой переменной.

Развивая направление нечеткой регрессии, исследователями были выделены различные варианты методов на основе классификации «вход – выход»: «четкий вход – четкий выход» метод CICO (Crisp-Input and CrispOutput), «нечеткий вход – нечеткий выход» метод FIFO (Fuzzy-Inputs and Fuzzy-Outputs) и смешанные данные – метод CIFO (Crisp-Inputs and Fuzzy-Outputs) [D'Urso, 2003; Hojati, 2005; Bissierier, 2009].

Интерес представляют работы [Таранцев, 1997; Bardossy, 1990; Sabic, 1991], посвященные разработке, исследованию и применению нечеткого регрессионного анализа. В работе [Полещук, 2000] приведена нечеткая линейная регрессионная модель и показано, что ее применение к моделированию зависимостей в реальных данных образовательного процесса на 40% улучшает прогноз по сравнению с классической линейной регрессионной моделью.

Нечеткая ARIMA модель

Нечеткая регрессионная модель временного ряда генерировала интервальные численные оценки прогнозных значений, при этом в условиях наличия выбросов ширина интервала прогнозной оценки являлась достаточно большой. Для устранения указанного недостатка была предложена нечеткая авторегрессионная модель проинтегрированного скользящего среднего (FARIMA) [Tsen-ga, 2001]. Авторы представили модель временного ряда, комбинирующую модель нечеткой регрессии и модель Бокса-Дженкинса ARIMA с целью построения моделей коротких ВР. Применение модели FARIMA для прогнозирования изменения валютных котировок показало ее продуктивность и лучшие показатели мер качества по сравнению с моделью ARIMA и моделью нечеткой регрессии.

В работе [Khashei, 2008] представлен метод прогнозирования ВР, основная идея которого заключается в расширении возможности метода ARIMA для

моделирования нелинейных зависимостей за счет использования нечеткой логики и искусственной нейронной сети. Показано, что гибридный метод ARIMA + нечеткие временные ряды + нейронные сети превосходит по точности отдельно взятый метод при прогнозировании финансовых рынков.

3.4.2. Нечетко-нейронный подход к моделированию временных рядов

Учитывая, что нейронные сети обладают большим потенциалом в моделировании нелинейных зависимостей, многие авторы комбинируют их для получения более адекватных результатов моделирования временных рядов [Ярушкина, 2007; Стецко, 2008].

Так, в работах [Huarng, 2006; Yu, 2008] рассмотрены гибридные модели моделирования временных рядов на основе применения нейронных сетей для прогнозирования нечетких временных рядов.

В работе [Alizadeh, 2009] заявляется нейро-нечеткий подход для прогнозирования курса доллара по отношению к японской йене. Авторы предлагают адаптивную нейро-нечеткую систему. Результаты применения предлагаемого нейронно-нечеткого подхода к моделированию временных рядов сравнивались со следующими моделями: модель множественной регрессии, искусственной нейронной сетью и нечеткой моделью, реализованной на основе алгоритма Суджено. Анализ результатов применения адаптивной нейро-нечеткой системы показал результативность и лучшие показатели точности по сравнению со сравниваемыми моделями.

Рассмотренный метод прогнозирования уровня продаж в работе [Кюо, 2001] основан на использовании нечеткой нейронной сети с генетической настройкой начальных значений весов и показывает более точные результаты при сравнении с искусственной нейронной сетью. В работе [Глебов, 2006] рассматривается применение нечеткой нейронной сети для краткосрочного прогнозирования временных рядов, значения которых характеризуют уровень электропотребления, а в работах [Ярушкина и др., 2007; Стецко, 2008] нечеткая

нейронная сеть применяется для анализа временного ряда уровней трафика вычислительных сетей.

Нечеткие нейронные сети

Нечеткой нейронной сетью (НС) обычно называют четкую нейросеть, которая построена на основе многослойной архитектуры с использованием специальных «И»-, «ИЛИ»-нейронов [Ярушкина, 2004].

Нечеткая нейросеть функционирует стандартным образом на основе четких действительных чисел, нечеткой является только интерпретация результатов.

Нечеткие нейронные сети осуществляют выводы на основе аппарата нечеткой логики, а параметры функций принадлежности настраиваются с использованием алгоритмов обучения НС. Поэтому для подбора параметров таких сетей применим метод обратного распространения ошибки, изначально предложенный для обучения многослойного персептрона. Нечеткая нейронная сеть, как правило, состоит из четырех слоев: слоя фазификации входных переменных, слоя агрегирования значений активации условия, слоя агрегирования нечетких правил и выходного слоя.

Наибольшее распространение в настоящее время получили архитектуры нечеткой НС вида ANFIS и TSK. Доказано, что такие сети являются универсальными аппроксиматорами. Быстрые алгоритмы обучения и интерпретируемость накопленных знаний – эти факторы сделали сегодня нечеткие нейронные сети одним из самых перспективных и эффективных инструментов мягких вычислений. Структуры и методы обучения нечетких нейронных сетей приведены в работах [Ярушкина, 2004; Борисов и др., 2007; Ковалев, 2007; Батыршин и др., 2007].

Контрольные вопросы

1. Какие существуют подходы к построению моделей нечеткой линейной регрессии?

2. Какие существуют критерии для определения нечетких коэффициентов модели?
3. Какие вы знаете варианты методов на основе классификации «вход – выход»?
4. В чем отличие модели ARIMA от ее нечеткого представления?
5. Что такое нечеткие нейронные сети?

Библиографический список

1. [Alizadeh, 2009] Alizadeh, M. Forecasting Exchange Rates. A Neuro-Fuzzy Approach / M. Alizadeh // *IFSA-EUSFLAT 2009*.
2. [Bardossy, 1990] Bardossy, A. Note on fuzzy regression / A Bardossy // *Fuzzy Sets and Systems*. – 1990. – №37 – P. 65-75.
3. [Bisserier, 2009] Bisserier, Amory. An Interval Approach for Fuzzy Linear Regression with Imprecise Data / Amory Bisserier, Reda Boukezzoula, Sylvie Galichet // *IFSA-EUSFLAT 2009*.
4. [Celmiņš, 1987] Celmiņš, A. Least squares model fitting to fuzzy vector data / A. Celmiņš, // *Fuzzy Sets and Systems*. – 1987. – № 22(3). – P. 245-269.
5. [Diamond, 1988] Diamond, P. Fuzzy least squares / P. Diamond // *Information Sciences*. – 1988. – №46(3). – P. 141-157.
6. [D'Urso, 2003] D'Urso, P. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data / P. D'Urso // *Computational Statistics & Data Analysis*. – 2003. – №42 (1-2). – P. 47-72.
7. [Hojati, 2005] Hojati M. A simple method for computation of fuzzy linear regression / M. Hojati, C. R. Bector, K. Smimou // *European Journal of Operational Research*. – 2005. – №166. – P. 172-184.
8. [Huarng, 2006] Huarng, K. The application of neural networks to forecast fuzzy time series / K. Huarng // *Physica*. – 2006. – A 336. – P. 481-491.
9. [Krashei, 2008] Krashei, A. Improvement of Auto-Regressive Integrated Moving Average Models using Fuzzy logic and Artificial Neural Networks / A. Krashei // *Neurocomputing*, 2008.

10. [Kuo, 2001] Kuo, R. J. A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm / R. J. Kuo // European Journal of Operational Research Volume 129, Issue 3, 16 March 2001, P. 496-517.
11. [Sabic, 1991] Sabic, D.A. Evaluation on fuzzy linear regression models / D. A. Sabic, W. Pedrycr // Fuzzy Sets and Systems. – 1991 – №23. – P. 51-63.
12. [Tanaka, 1982] Tanaka, H. Linear regression analysis with fuzzy model / H. Tanaka, S. Uejima, K. Asai // IEEE Transactions on Systems, Man and Cybernetics. – 1982. – №12(6). – P. 903-907.
13. [Tsenga, 2001] Tsenga, F. M. Fuzzy ARIMA model for forecasting the foreign exchange market / F. M. Tsenga, G. H. Tzengb, H. C. Hsiao-Cheng Yua // Fuzzy Sets and Systems. – 2001. – №118.
14. [Yu, 2008] Yu, T. A bivariate fuzzy time series model to forecast the TAIEX / T. Yu, K. Huarng // Expert systems with Applications. – 2008. – Vol. 34, Issue 4.
15. [Батыршин и др., 2007] Батыршин, И. З. Нечеткие гибридные системы. Теория и практика / И. З. Батыршин, А. О. Недосекин, А. А. Стецко и др. – М. : ФИЗМАТЛИТ, 2007. – 208 с.
16. [Борисов и др., 2007] Борисов, В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия – Телеком, 2007. – 284 с.
17. [Глебов, 2006] Глебов, А. А. «Модель краткосрочного прогнозирования электропотребления с помощью нейро-нечетких систем» / А. А. Глебов // Южно-Российский вестник геологии, географии и глобальной энергии. – 2006. – №7(20). – С. 142-146.
18. [Полещук, 2000] Полещук, О. М. Выявление существенных показателей при работе с нечеткой информацией / О. М. Полещук // Автоматизация и компьютеризация информационной техники и технологии. Научные труды. Вып. 308. – М. : МГУЛ, 2000. – 220 с.

19. [Стецко, 2008] Стецко, А. А. Принятие проектных решений на основе анализа нечетких тенденций временных рядов / А. А. Стецко // Программные продукты и системы. – 2008. – №3.
20. [Таранцев, 1997] Таранцев, А. А. Принципы построения регрессионных моделей при исходных данных с нечетким описанием / А. А. Таранцев // Автоматика и телемеханика. – 1997. – №11. – С. 27-32.
21. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
22. [Ярушкина, 2007] Ярушкина, Н. Г. Моделирование трафика терминал-сервера на основе анализа нечетких тенденций временных рядов / Н. Г. Ярушкина, Т. Р. Юнусов, Т. В. Афанасьева // Программные продукты и системы. – 2007. – №4. – С. 15-19.

3.5. Интеллектуальный анализ баз данных временных рядов

3.5.1. Понятие гранулированного временного ряда

Гранулированное представление является дальнейшим развитием нечетких моделей ВР.

В отличие от нечеткого подхода к моделированию временных рядов, в котором преобразование исходного числового ВР в нечеткий временной ряд имеет целью получение числового прогноза, в гранулированном временном ряду результат нечеткого моделирования выражается в виде предложения естественного языка, моделирующего высказывания эксперта.

Гранулирование информации лежит в центре человеческих рассуждений, взаимодействий и формирования понятий. В естественном языке (ЕЯ) слова играют роль меток гранул. В этой роли они служат для сжатия данных.

Развитие гранулярных вычислений привело к формированию понятия гранулированного временного ряда [Ярушкина, 2009]. Традиционная сегмента-

ция (дискретизация) ВР выполняется методом скользящего окна заданной ширины k на X .

Пусть $W(x, k) = \{w_i | i = 1, 2, \dots, n - k + 1\}$ обозначает множество всех k -широких окон на X . Зададим меру, расстояние между двумя подпоследовательностями w_i . Если выполнить любым из известных способов кластеризацию таких подпоследовательностей, получим s кластеров: $C_i (i = 1, 2, \dots, s)$. Алфавит $\Omega = \{a_i | i = 1, 2, \dots, s\}$ представляет символы образцов ВР. Дискретную версию временного ряда $D(x) = \{a_{j_1}, a_{j_2}, \dots, a_{j_m}\}$ называют символьным ВР.

Информационные гранулы, построенные на исходном временном ряду, могут быть одномерными и многомерными. Одномерные гранулы представляют высказывания, нечеткие по уровням и четкие по времени, двумерные – образуются высказываниями на основе нечетких меток уровней и нечетких временных меток.

Переход к гранулярному ВР позволяет ввести нечеткость по времени дополнительно к уровням временного ряда для лучшего соответствия описания временного ряда с лингвистической оценкой поведения, обычно формируемой человеком. В результате формат нечетких высказываний «ЕСЛИ-ТО» включает анализ нечетких временных меток. Формирующееся в теории моделирования нечетких временных рядов направление, моделирующее временной аспект, получило название *темпоральный анализ* [Ковалев, 2009].

Двумерные гранулы образуют различные типы кластеров, используемые для извлечения знаний из нечетких временных рядов на основе гранулярных вычислений (*computing with words and perceptions CWP*) [Zadeh, 2001; Batyrshin, 2004].

Для представления неточного значения вместо единичного значения (синглтона) необходимо использовать интервал или распределение какой-либо функции множества, т. е. гранулу сложной структуры. В общем смысле можно говорить об экстенциональном и интенциональном (attribute-based) представлении значений.

Для определения информационной гранулы используется *принцип обобщенных ограничений* (generalized constraint) [Zadeh, 2006]. Обобщенное ограничение задается в виде

$$X \text{ is } r R,$$

где X – ограниченная переменная;

r – тип модальности;

R – ограничивающее (нечеткое) отношение.

Типы ограниченных переменных:

- X – n -арная переменная, $X = (X_1, \dots, X_n)$;
- X – это пропозиция;
- X – функция другой переменной: $X = f(Y)$;
- X – обусловлена другой переменной X/Y ;
- X имеет структуру, например, $X = \text{Location}(\text{Residence}(\text{Carol}))$;
- X – обобщенное ограничение $X: Y \text{ is } r R$;
- X – групповая переменная $G[A]: (\text{Name}_1, \dots, \text{Name}_n)$, с каждым элементом группы Name_i , $i = 1, \dots, n$, ассоциируется атрибут A_i .

Типы обобщенных ограничений $X \text{ is } r R$:

$r: =$ ограничение эквивалентности: $X = R$, аббревиатура $X \text{ is } = R$;

$r: \leq$ ограничения неэквивалентности: $X \leq R$;

$r: \subset$ ограничения вложенности: $X \subset R$;

$r: \text{blank}$ возможностное ограничение: $X \text{ is } R$, R – распределение возможности на X ;

$r: v$ истинностное ограничение: $X \text{ is } v R$, R – распределение истины на X ;

$r: p$ вероятностное ограничение: $X \text{ is } p R$, R – распределение вероятностей на X ;

$r: \text{bm}$ бимодальное ограничение: X случайная переменная $R \text{ is } \text{bm}$;

$r: rs$ ограничение случайных множеств: $X isrs R$, R – распределение вероятностей на X ;

$r: fg$ ограничения нечеткого графа: $X isfg R$, X функция и R – ее нечеткий граф;

$r: u$ ограничения «традиции» (привычной практики usually): $X isu R$ «обычно означает» ($X is R$);

$r: g$ групповое ограничение: $X isg R$ означает, что R ограничивает все значения атрибутов.

Для символической записи гранулярных пропозиций предлагается язык гранулярных вычислений: *Generalized Constraint Language (GCL)*.

Композитные ограничения могут служить основой для извлечения ассоциативных правил из ВР. Такие правила извлекают на основе четырех методов:

- 1) ассоциативные правила на основе мер доверия;
- 2) ассоциативные правила на основе частотности;
- 3) ассоциативные правила на основе корреляции ВР;
- 4) ассоциации локальных трендов.

Исследования в области гранулярных вычислений применительно к нечетким временным рядам обозначили формирование нового научного направления: извлечения знаний из нечетких временных рядов на основе гранулярных вычислений. Методология вычисления со словами и восприятиями (*computing with words and perceptions CWP*) определяет основную задачу анализа гранулированных ВР: распознавание паттернов ВР (восприятий perception) и извлечения ассоциативных правил в лингвистической форме. Форма правил определяется принципом обобщенных ограничений (*generalized constraints*). В состав правил входят переменные, принимающие гранулированные значения.

3.5.2. Извлечение знаний из временных рядов (*Time Series Data Mining*)

Исследования данных и методов их анализа в последние десятилетия оформились в виде отдельного направления, называемого *интеллектуальным анализом данных* или *Data Mining*, в котором анализ временных рядов получил

название *интеллектуального анализа временных рядов* или *Time Series Data Mining (TSDM)*.

Извлечение знаний обычно определяется как методология для получения знаний из баз данных, но какие формальные методы используются для интеграции этого знания с базой знаний – остается открытой проблемой.

Для решения этой проблемы необходимо представление результатов *Data Mining* в форме, используемой в человеческих знаниях. Человеческое знание основано на образах и формулируется лингвистически. Вычисления со словами и образами (CWP) дают методологию для управления информацией и для разработки систем, основанных на знаниях. *Data Mining* интегрирует методы, основанные на образах, дает возможность для извлечения информации из баз данных в лингвистической форме, подходящей для их использования в методах принятия решений. Методы и технологии извлечения знаний с использованием временных рядов должны оперировать паттернами временных рядов, отыскивать ассоциации между ними и извлекать знания.

На основе новой методологии *Data Mining* решается расширенная совокупность задач анализа временных рядов, определенных в работе [Batyrrshin, 2007]:

- 1) *сегментация* – разбиение ВР на значимые сегменты [Graves, 2009];
- 2) *кластеризация* – поиск группировок ВР или их паттернов [Giove, 2009];
- 3) *классификация* – назначение ВР или их паттернам одного из заранее определенных классов [Herbst, 2009];
- 4) *индексирование* – построение индексов для эффективного выполнения запросов к базам данных ВР;
- 5) *резюмирование (summarization)* – формирование краткого описания ВР, содержащего существенные черты с точки зрения решаемой задачи [Kasprzyk, 2009];
- 6) *обнаружение аномалий* – поиск новых, не типичных паттернов ВР;
- 7) *частотный анализ* – поиск часто проявляющихся паттернов ВР;
- 8) *прогнозирование* – прогноз очередного значения ВР на основе истории ВР;

9) *извлечение ассоциативных правил* – поиск правил, относящихся к паттернам ВР.

Традиционное выделение паттернов ВР было связано с выделением участков с постоянным знаком первой и второй производной: *возрастающий и выпуклый, убывающий и гладкий* и др. Различные шкалы и методы гранулярных вычислений Л.Заде использовались для описания паттернов линейных трендов: *рост, падение, резкий рост, медленное падение* и т. д. Параметрические методы выпукло-гладкой модификации линейных функций и нечеткая грануляция выпукло-гладких паттернов позволили получить лингвистическое описание для ВР, подобное следующему: *медленно убывающий и строго гладкий*.

В рамках описанного выше направления TSDM акцент делается на поиск и извлечение правил из ВР, при этом полагаются на следующие основные принципы [Ковалев, 2007]:

- 1) поиск правил нацелен на получение понимаемых результатов и не обязательно самых точных прогнозов;
- 2) важнейшим шагом на пути извлечения интерпретируемых знаний является порождение описаний фрагментов ВР в форме темпоральных образов, допускающих естественно-языковое толкование.

Требование к модели представления – способность отражать основные темпорально-логические концепты знаний (наиболее общие представления экспертов о логических и временных особенностях в данных).

Виды темпорально-логических концептов приведены в работе [Ковалев, 2007]:

- 1) концепт временной продолжительности – присутствие определенного паттерна или признака ВР на определенном интервале времени;
- 2) концепт очередности – порядок следования паттернов ВР во времени;
- 3) концепт одновременности – совпадение во времени темпоральных событий (паттернов различных ВР);
- 4) концепт нечеткости – нечеткость выраженности темпоральных событий и отношений.

В работе [Ковалев, 2007] рассмотрен подход к сегментации ВР на известные заранее нечеткие тенденции, называемые нечеткими темпоральными образами. Данный подход базируется на концепции информационно-теоретического подхода: поиск границ темпоральных образов на основе анализа энтропии распределения вероятностей появления следующих друг за отсчетов ВР. Внутри образа энтропия для начальных позиций максимальна и уменьшается по мере продвижения к концу образа. Энтропия скачкообразно возрастает с выходом из образа.

В основе построения модели положено понятие нечеткого темпорального события, характеризующего поведение ВР на временном интервале:

$$\Phi(t_a, t_b) = (y(t_a, t_b) = \tau) \& ((t_a, t_b) = q).$$

Такое событие имеет смысловое описание «на временном интервале $[t_a, t_b]$ с нечеткой продолжительностью q наблюдается тренд временного ряда с нечетким значением τ . Автор рассматривает частные случаи нечетко-темпоральной модели, где каждое событие описывается своим методом расчета. Например, нечеткая переменная «тренд» определяется тангенсом угла наклона прямой линии, аппроксимирующей монотонные участки ВР. В качестве инструментов анализа предлагается использовать нечеткие гиперграфы, нечеткие нейронные сети.

В соответствии с методологией СWP основные направления работ сгруппированы в следующие классы [Ярушкина, 2009]:

- ✓ Уточнение (precisiation) паттернов ВР, основанных на восприятии;
- ✓ Обработка ВР на основе **принципа обобщенных ограничений** (generalized constraints);
- ✓ Извлечение ассоциативных правил;
- ✓ Преобразование ассоциаций на основе **принципа обобщенных ограничений** (generalized constraints);
- ✓ Использование экспертных знаний в системах поддержки принятия решений.

Открытыми проблемами нового научного направления *Time Series Data Mining* являются следующие:

- 1) извлечение знаний о поведении временных рядов в форме нечетких тенденций [Ярушкина, 2004];
- 2) прогнозирование на основе распознанной тенденции нечеткого ВР: *If trend is F then next point is Y* [Ярушкина, 2004];
- 3) использование извлеченных правил в экспертных системах и системах поддержки принятия решений;
- 4) лингвистическая трансляция результатов кластеризации паттернов ВР в правила [Yu, 2005];
- 5) резюмирование совокупности нечетких ВР [Batyrrshin, 2004].

Контрольные вопросы

1. Дайте определение гранулированного ВР.
2. Дайте определение темпорального анализа ВР.
3. Что понимается под интеллектуальным анализом данных или Data Mining?
4. Какие задачи решаются на основе Data Mining?
5. Приведите виды темпорально-логических концептов ВР.
6. Назовите проблемы нового научного направления Time Series Data Mining.

Выводы

Анализ методологических подходов к решению моделирования временных рядов, приведенный в главе 3, позволяет обобщить основные принципы, лежащие в их основе:

1. Принцип «разделяй и властвуй».

Это теоретический принцип декомпозиции общей модели временного ряда на модели, выражающие predetermined классы поведения временного ряда. Другими словами, это экспертная декомпозиция исходного временного ряда на совокупность однородных по типу и одновременных временных рядов,

поведение каждого может быть описано отдельной моделью, композиция которых образует общую модель поведения ВР.

2. Принцип многомодельности.

Принцип является системным по своему содержанию, он основан на многомодельности и отношениях между моделями, описывающими функционирование отдельных компонент ВР. Каждая компонента может быть реализована отдельно взятыми независимыми модулями, реализующими тот или иной метод ее моделирования. Данный принцип не распространяется на многомодельное представление данных.

3. Принцип неточности.

Этот принцип является следствием нескольких причин: использование приближенных моделей и численных алгоритмов, искажения в исходных данных в результате измерений, изменения внешней среды, нечеткость интерпретации (оценки) исходных, промежуточных и выходных данных, нечеткость оценивания моделей, порожденные экспертной деятельностью при их проектировании.

4. Принцип адаптации (обучения).

Обучение применительно к системам моделирования временных рядов может рассматриваться как важный процесс, значительно влияющий на качество создаваемых моделей. Существуют два аспекта в проблеме обучения: обучение эксперта построению моделей временных рядов и обучение модели. Системы моделирования создаются как закрытые архитектурные решения в виде лицензионных комплексов дорогостоящих программ, имеющие специфичные интерфейсы, закрытые форматы данных и критерии, нацеленные на моделирование либо отдельных компонент, либо всего ВР. Использование и настройка таких архитектур требует высокой квалификации эксперта. Обучение моделей временных рядов рассматривается как их подгонка на новых данных путем перестроения общей модели.

В таблице 3.3 приведено сравнение рассмотренных в настоящей главе подходов к моделированию временных рядов с позиции «вход/выход» и типов решаемых задач.

Таким образом, итоги и перспективы основных направлений исследований в области моделирования временных рядов связаны с развитием методов интеллектуального анализа, использующих гибридные технологии, нечеткое моделирование ВР и гранулярные вычисления.

Несмотря на достигнутые результаты, многие задачи нечеткого моделирования и анализа нечетких временных рядов остаются нерешенными, в частности задачи анализа такого объекта временного ряда, как нечеткая тенденция, и генерации правил распознавания нечетких тенденций.

В 2004 году Н. Г. Ярушкиной [Ярушкина, 2004] было введено понятие нечеткой тенденции нечетких временных рядов, определены новые задачи и методы их решения. Данное направление связано с извлечением новых знаний о закономерностях изменения нечетких тенденций во временных рядах, решением новых задач анализа НВР, построением новых математических моделей и методов анализа временных рядов.

Таблица 3.3

Сравнение подходов к моделированию временных рядов

Подходы	Тип исходных/ выходных данных	Решаемые задачи
Статистический (стохастическая неопределенность)	Числовой/ числовая оценка	1. Построение математической модели статистической зависимости 2. Оценивание параметров 3. Прогноз
Нейросетевой (стохастическая неопределенность)	Числовой/ числовая оценка	1. Построение нейронной сети 2. Обучение нейронной сети 3. Прогноз
Нечеткий (неопределенность нечеткости)	Числовой/ числовая оценка. Лингвистический/ лингвистическая оценка Числовой/ лингвистическая оценка Лингвистический/ числовая оценка	1. Построение нечетких множеств 2. Построение нечеткой модели 3. Прогноз

Нечетко-статистический (стохастическая и нечеткая неопределенность)	Числовой/ числовая оценка	1. Построение математической модели статистической зависимости 2. Оценивание нечетких параметров 3. Прогноз
Нечетко-нейронный (стохастическая и нечеткая неопределенность)	Числовой/ числовая оценка. Лингвистический/ лингвистическая оценка Числовой/ лингвистическая оценка Лингвистический/ числовая оценка	1. Построение нечетких множеств 2. Построение нечеткой модели 3. Построение нейронной сети 4. Обучение нейронной сети 5. Прогноз
Интеллектуальный анализ баз данных ВР (стохастическая, нечеткая и ассоциативная неопределенность)	Комбинированный, содержащий числовой и лингвистический тип данных	1. Сегментирование 2. Классификация 3. Частотный анализ 4. Прогноз 5. Поиск аномалий 6. Резюмирование 7. Индексирование 8. Извлечение ассоциативных связей 9. Кластеризация

Отличительной чертой данного направления является тот факт, что результаты решения задач анализа ВР могут быть выражены не только в числовой форме, но и в лингвистической, выражающей тенденции развития в прошлом и будущем. Указанное свойство особенно важно, так как создает возможность представлять результаты в терминах онтологии предметной области и актуально для задач поддержки проектных и управляющих решений в различных предметных областях, в которых человеческий фактор имеет определяющее значение.

Уже достигнутые результаты в области нечеткого моделирования и гибридных моделей ВР позволяют рассматривать нечеткую логику как перспективный инструмент направления интеллектуального анализа временных рядов. Современные исследования в этом направлении при нечетком моделировании ВР обозначают ряд проблем, главной из которых является отсутствие методологии нечеткого моделирования и анализа нечетких временных рядов.

В основу такой методологии в нечетком моделировании временных рядов, как представляется, целесообразно положить формализм нечетких шкал, обеспечивающих построение и преобразования нечетких объектов временного ряда (нечетких уровней и нечетких тенденций), методы и задачи *Time Series Data Mining*, позволяющие извлекать знания о тенденциях ВР, выраженные в форме нечетких продукционных правил, и лингвистически интерпретируемые информационные гранулы для описания паттернов ВР.

В следующих главах будут раскрыты обозначенные положения методологии нечеткого моделирования.

Библиографический список

1. [Batyrrshin, 2004] Batyrrshin, I. Construction of granular derivatives and solution of granular initial value problem / I. Batyrrshin // Fuzzy Partial Differential Equations and Relational Equations. Studies in Fuzziness and Soft Computing, Vol. 142, Springer-Verlag, 2004. – P. 285-307.
2. [Batyrrshin, 2007] Batyrrshin, I. Perception Based Time Series Data Mining for Decision Making / I. Batyrrshin // IFSA'07 Fuzzy Logic, Soft Computing and Computational Intelligence.
3. [Giove, 2009] Giove, S. Fuzzy logic and Clustering methods for time series analysis / S. Giove // 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009).
4. [Graves, 2009] Graves, D. Multivariate Segmentation of Time Series with Differential Evolution / D. Graves, W. Pedrycz // 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009).
5. [Herbst, 2009] Herbst, G. Online Recognition of fuzzy time series patterns / G. Herbst, S. F. Bocklisch // 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009).

6. [Kacprzyk, 2009] Kacprzyk, J. Using Fuzzy Linguistic summaries for the comparison of time series / J. Kacprzyk, A. Wilbik // 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009).
7. [Zadeh, 2001] Zadeh, Lotfi A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. / Lotfi A. Zadeh // Fuzzy Sets and Systems, Vol. 90. – 1997. – P. 111-127.
8. [Zadeh, 2006] Zadeh, Lotfi A. Generalized theory of uncertainty (GTU) – principal concepts and ideas / Lotfi A. Zadeh // Computational statistic & Data analysis. – 2006. – №51. – P. 15-46.
9. [Yu, 2005] Yu, F. Finding Fuzzy Rules from Granular Time series / F. Yu, W. Pedrycz, J. Yuan // IFSA'05 Fuzzy Logic, Soft Computing and Computational Intelligence.
10. [Ковалев, 2007] Ковалев, С. М. Гибридные нечетко-темпоральные модели временных рядов в задачах анализа и идентификации слабо формализованных процессов / С. М. Ковалев // Интегрированные модели и мягкие вычисления в искусственном интеллекте: сборник трудов IV-й международной научно-практической конференции. В 2-х томах. Т.1 – М. : Физматлит, 2007. – 354 с.
11. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем: учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
12. [Ярушкина, 2009] Ярушкина, Н. Г. Современный интеллектуальный анализ нечетких временных рядов / Н. Г. Ярушкина // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). В 2-х т. Т.1. – М. : Физматлит, 2009. – С. 19-30.

ГЛАВА 4. ОСНОВЫ ТЕОРИИ НЕЧЕТКИХ ТЕНДЕНЦИЙ НЕЧЕТКИХ ВРЕМЕННЫХ РЯДОВ

Введение

В главе 1 была определена содержательная модель нечеткого временного ряда, как временного ряда, уровни которого выражены нечеткой меткой, соответствующей нечеткой экспертной оценке:

$$F_Mr = F_Mark(X, Y, VK, S, MK),$$

где X – оцениваемая величина субъектом Y ;

VK – выбранный формальный критерий (например, эффективность, надежность);

S – известная нечеткая шкала градаций критерия, выраженная в виде лингвистической переменной. Каждая градация этой шкалы есть нечеткое множество (терм), представленное тройкой $\langle Name, W, MF \rangle$, где $Name$ – лингвистическая оценка (название градации), W – носитель градации (может быть представлен множеством действительных чисел или множеством лингвистических термов), MF – функция принадлежности, определенная на W , сопоставляющая каждому значению $X \in W$ значение из диапазона $[0, 1]$;

MK – метод оценивания на основе операции приведения к нечеткости (фаззификации).

Степень субъективности оценки F_Mr определяется ролью эксперта при идентификации всех ее компонент.

Анализ современных подходов к моделированию нечетких временных рядов (НВР), приведенный в третьей главе, позволяет сделать вывод, что основное направление исследований в этой области связано с сокращением степени субъективности на всех этапах, в том числе и при определении множества нечетких меток уровней временного ряда и их параметров на основе применения формальных методов. Такой подход является ключевым при проектировании систем моделирования НВР с целью получения числового прогноза, а так-

же в задачах интеллектуального анализа одновременных НВР, извлекаемых из баз данных, так как позволяет не только сократить время решения задач, но и повысить степень адекватности их результатов. Формализация всех компонент введенной нечеткой экспертной оценки F_{Mr} для получения абсолютной оценки нечеткого уровня ВР, позволит не только достичь эффекта сокращения субъективности при моделировании НВР, но и создаст возможность извлечения новых знаний о поведении НВР, отражающих динамику изменения нечетких уровней НВР в виде нечетких тенденций.

Содержательно, нечеткие тенденции были введены в работе [Ярушкина, 2004].

Нечеткой тенденцией (НТ) нечеткого временного ряда будем называть нечеткую метку, выражающую характер изменения (систематическое движение) последовательности нечетких уровней НВР в заданном интервале времени. Нечеткая тенденция выражает поведение НВР в лингвистическом виде, например: «Рост», «Падение», «Стабилизация», «Колебания», «Хаос». Для нечетких термов, обозначающих тенденцию, возможно применение модификаторов «очень», «более-менее» и т. д. Отметим важное свойство нечетких экспертных оценок, обусловленное возможностью их ранжирования, что позволяет представить их совокупность в виде некоторой системы (шкалы) с отношениями. Бинарные отношения, образованные на множестве нечетких экспертных оценок, порождают сравнительные оценки по различным критериям, такие, как «Больше», «Меньше», «Примерно Равны», «Рост», «Падение», «Предпочтительнее», «Лучше». Такие сравнительные оценки представляют изменения (различие) нечетких меток в различных пространствах: в пространстве объектов, во временном пространстве, в пространстве задач и характеризуют тенденции.

Изменения нечетких меток во временном пространстве порождают нечеткий временной ряд с нечеткой тенденцией.

Анализ тенденций развития процессов является важной составляющей экспертной деятельности. В задачах принятия управленческих и проектных решений эксперт оперирует не только значениями оценки состояния, но и в значительной степени учитывает тенденции их изменения. Обозначенный факт

применительно к нечеткому временному ряду актуализирует задачу формализации понятий нечеткой экспертной оценки как нечеткого уровня ВР, и нечеткой оценки ее изменений, то есть нечеткой тенденции. Для определения модели нечеткой тенденции важным является то, что она строится на множестве оценок нечетких уровней ВР.

В соответствии с логикой оценивания будем считать оценку нечеткого уровня ВР – *абсолютной нечеткой оценкой*, а оценку изменения нечетких уровней (нечеткую тенденцию) – *сравнительной нечеткой оценкой*.

В настоящее время исследование, формальное описание, моделирование и обработка сравнительных нечетких оценок на множестве абсолютных нечетких оценок не получило достаточного развития.

Оценки уровней ВР, являясь качественными оценками, характеризуют состояние моделируемого процесса в некоторый момент времени и обычно задаются экспертом, лицом, принимающим решение (ЛПР) или процедурой, моделирующей оценочную деятельность эксперта. Экспертная деятельность подобного рода базируется на использовании внутренней, нечеткой, лингвистически выраженной шкалы, сформированной на основе накопленного опыта эксперта выполнения процедур кластеризации, классификации и ранжирования.

Исходя из естественного предположения о связи абсолютных и сравнительных оценок, выражаемых в шкалах, авторы предлагают для решения задачи формализации нечетких тенденций специальную лингвистическую шкалу.

Формализация нечеткой тенденции НВР позволит переформулировать все задачи интеллектуального анализа временных рядов (Time Series Data Mining) в терминах нечетких тенденций и обозначить новое направление в нечетком моделировании временных рядов.

4.1. Концептуальная модель ACL-шкалы для генерации нечетких оценок

В настоящем разделе предлагается специальная лингвистическая шкала в качестве инструмента как абсолютного, так и сравнительного нечеткого оцени-

вания – *ACL-шкала* (*Absolute & Comparative Linguistic*) [Афанасьева, 2008а]. Абсолютные оценки, полученные по *ACL-шкале*, соответствуют нечетким оценкам (меткам) уровней НБР, а сравнительные оценки – нечетким тенденциям НБР.

Формально шкалой называется кортеж из трех элементов:

$$\langle X, \varphi, Y \rangle,$$

где $X = \{x_i, R_x\}$ – реальный объект со свойствами $x_i, i \in [1, m]$, на которых задано отношение R_x ;

$Y = \{\varphi(x_i), R_y\}$ определяет шкалу как знаковую систему с отношением R_y ; $\varphi \in \Phi$ гомоморфное отображение X на Y так, что $\{\varphi(x_i)\} \in R_y$ только тогда, когда $\{x_i\} \in R_x$ для всех $i \in [1, m]$ [Анфилатов, 2003].

Тип шкалы определяется по $\Phi = \{\varphi(x_i)\}$ и множеству допустимых операций.

4.1.1. Структурная модель *ACL-шкалы*

Введем следующие предположения при определении модели *ACL-шкалы*:

1. Множество оцениваемых объектов $x \in X$ образует носитель *ACL-шкалы* и может быть любой природы. Объекты множества X обладают свойством упорядоченности, то есть на X определено бинарное отношение $x \leq y$, обладающее следующими свойствами:

- (I) рефлексивность: $x \leq x, \quad \forall x \in X$
- (II) транзитивность: если $x \leq y$ и $y \leq z$, то $x \leq z, \quad \forall x, y, z \in X$
- (III) антисимметричность: если $x \leq y$ и $y \leq x$, то $x = y, \quad \forall x, y \in X, x \neq y$.

2. Градации *ACL-шкалы* задают лингвистические наименования нечетких экспертных оценок $\tilde{x}$, образующих конечное множество $\tilde{X}$, элементы которого также частично упорядочены в силу природы нечетких экспертных оценок $\tilde{x}$. На множестве $\tilde{X}$ определено бинарное отношение $\tilde{x} \leq \tilde{y}$, обладающее следующими свойствами:

- (I) рефлексивность: $\tilde{x} \leq \tilde{x}, \quad \forall \tilde{x} \in \tilde{X}$
- (II) транзитивность: если $\tilde{x} \leq \tilde{y}$ и $\tilde{y} \leq \tilde{z}$, то $\tilde{x} \leq \tilde{z}, \quad \forall \tilde{x}, \tilde{y}, \tilde{z} \in \tilde{X}$

(III) антисимметричность: если $\tilde{x} \leq \tilde{y}$ и $\tilde{y} \leq \tilde{x}$, то $\tilde{x} = \tilde{y}$, $\forall \tilde{x}, \tilde{y} \in \tilde{X}$.

3. Считаем, что каждый элемент $\tilde{x}_i \in \tilde{X}$ моделируется функцией $\mu_{\tilde{x}_i}(x)$ так, что $\tilde{X}$ покрывает множество X . Функция $\mu_{\tilde{x}_i}(x)$ задает семантику нечеткой оценки $\tilde{x}_i \in \tilde{X}$ объекта $x \in X$ и называется функцией принадлежности (соответствия) объекта x нечеткому множеству $\tilde{x}_i$.

Таким образом математическим объектом нечеткой оценки является нечеткое множество $\tilde{x}_i = \{x, \mu_{\tilde{x}_i}(x)\}$.

Определение 4.1. Структурная модель *ACL-шкалы* S_x для определения абсолютных нечетких оценок представима в виде лингвистической переменной

$$S_x = \langle \text{Name_} S_x, \tilde{X}, X, G, P \rangle,$$

где $\text{Name_} S_x$ – имя ACL-шкалы (или название критерия VK , по которому производится оценивание объектов $x \in X$);

X – универсальное множество объектов x , образующее область определения шкалы. В дальнейшем будем рассматривать конечное множество X , имеющее точную нижнюю и верхнюю грани: $\text{min} = \inf(X)$, $\text{max} = \sup(X)$;

$\tilde{X}$ – базовое конечное терм-множество абсолютных нечетких оценок (лингвистических названий градаций шкалы, порождающих оценки F_Mr), например, $\tilde{X} = \{\text{Плохой, Удовлетворительный, Хороший, Отличный и др.}\}$, $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$;

G – синтаксические правила вывода (порождения) цепочек оценочных высказываний (множеств производных термов $\tilde{X}^*$, не входящих в базовое терм-множество);

P – семантические правила, определяющие функции принадлежности для каждого терма $\tilde{x}_i \in \tilde{X} \cup \tilde{X}^*$, $i \in [1, m^*]$.

Пример множества $\tilde{X} = \{A0-1, A0, A1, A2, A2+1\}$, его носителя $X = [-10, 50]$ и соответствующих треугольных функций принадлежности $p_i \in P$, образующих основу шкалы S_x , приведен на рис. 4.1.

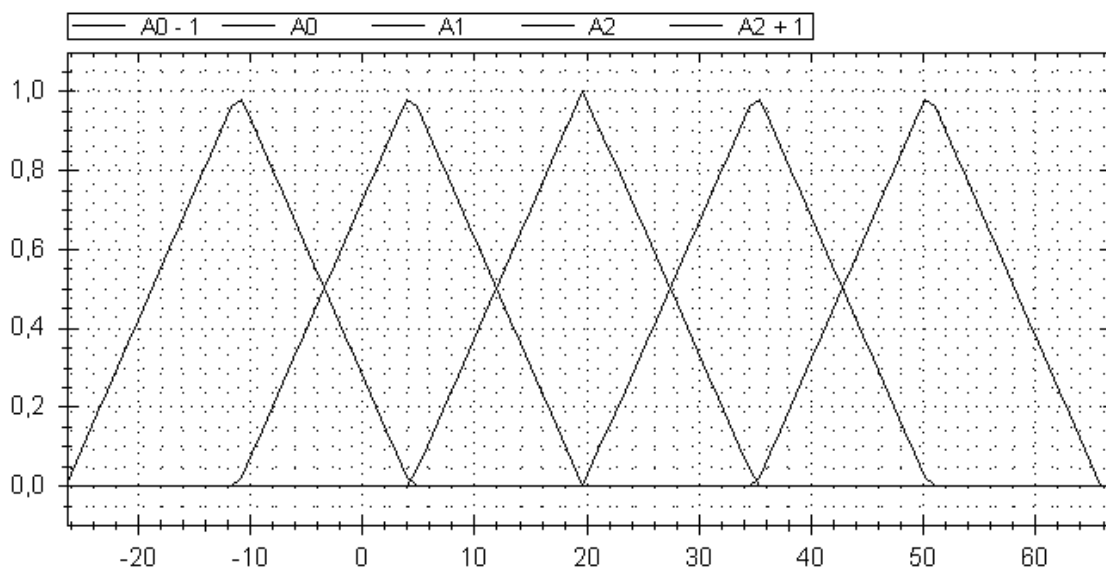


Рис. 4.1. Пример нечеткой шкалы

Так как на множестве $\tilde{X}$ определено бинарное отношение $\tilde{x} \leq \tilde{y}$, обозначим это отношение как $TTend(\tilde{x}_i, \tilde{x}_j)$, по содержанию – это лингвистическое отношение, фиксирующее тип изменения между двумя нечеткими оценками $\tilde{x}_i$, $\tilde{x}_j$ шкалы.

Отношение $TTend(\tilde{x}_i, \tilde{x}_j)$ является нечетким лингвистическим отношением, применяемым для определения сравнительной нечеткой оценки $\tilde{v}_{ij} = TTend(\tilde{x}_i, \tilde{x}_j)$, характеризующей направление изменения (увеличение или уменьшение) значения абсолютной нечеткой оценки $\tilde{x}_i$ по отношению к оценке $\tilde{x}_j$, которое может быть представлено лингвистическими выражениями, например, значениями из множества {РОСТ, ПАДЕНИЕ, СТАБИЛЬНОСТЬ}. Пример отношения $TTend$ для трех термов из $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3\}$ представлен в таблице 4.1.

Отношение $TTend(\tilde{x}_i, \tilde{x}_j)$

	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$
$\tilde{x}_1$	Стабильность	Рост	Рост
$\tilde{x}_2$	Падение	Стабильность	Рост
$\tilde{x}_3$	Падение	Падение	Стабильность

Отметим, что каждая нечеткая оценка $\tilde{v}_{ij} = TTend(\tilde{x}_i, \tilde{x}_j)$ представима своим нечетким множеством, семантика которого задается экспертно или на основании некоторой формализованной процедуры.

Отношение $TTend$ антирефлексивно, антисимметрично и транзитивно:

$$\begin{aligned} \forall \tilde{x} \in \tilde{X} \quad TTend(\tilde{x}, \tilde{x}) &= 0, \\ \forall \tilde{x}, \tilde{y} \in \tilde{X} \quad (\tilde{x} \neq \tilde{y}) \quad TTend(\tilde{x}, \tilde{y}) \wedge TTend(\tilde{y}, \tilde{x}) &= 0 \\ \forall \tilde{x}, \tilde{y}, \tilde{z} \in \tilde{X} \quad TTend(\tilde{x}, \tilde{z}) &> TTend(\tilde{x}, \tilde{y}) \wedge TTend(\tilde{y}, \tilde{z}) \end{aligned}$$

Указанные свойства отношения $TTend$ позволяют классифицировать его как отношение порядка.

Тогда совокупность всех возможных нечетких оценок $\tilde{V} = \{\tilde{v}_{ij}\}$ образует нечеткую порядковую шкалу $S_v = \langle Name_TTend, \tilde{V}, \tilde{X}, X, G_v, P_v \rangle$.

Предположим, что существует бинарное отношение $RTend(\tilde{x}_i, \tilde{x}_j)$ – лингвистическое отношение, фиксирующее интенсивность различия между двумя нечеткими оценками $\tilde{x}_i, \tilde{x}_j$ шкалы. Отношение $RTend(\tilde{x}_i, \tilde{x}_j)$ является также нечетким лингвистическим отношением, применяемым для определения сравнительной нечеткой оценки $\tilde{\alpha}_{ij} = RTend(\tilde{x}_i, \tilde{x}_j)$, характеризующей степень различия, «неметрическое расстояние» между нечеткими оценками $\tilde{x}_i, \tilde{x}_j$, которое может быть выражено лингвистически, например, значениями из множества {БОЛЬШОЕ, СРЕДНЕЕ, МАЛОЕ, ОТСУТСТВУЕТ}. Пример отношения $RTend$ для четырех термов из $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4\}$ представлен в таблице 4.2.

Таблица 4.2

Отношение $RTend(\tilde{x}_i, \tilde{x}_j)$				
	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$
$\tilde{x}_1$	Отсутствует	Малое	Среднее	Большое
$\tilde{x}_2$	Малое	Отсутствует	Малое	Среднее
$\tilde{x}_3$	Падение	Малое	Отсутствует	Малое
$\tilde{x}_4$	Большое	Среднее	Малое	Отсутствует

Семантика оценки $\tilde{\alpha}_{ij}$ также представима своим нечетким множеством с функцией принадлежности, задаваемой экспертно или на основании некоторой формализованной процедуры.

Отношение $RTend$ антирефлексивно и симметрично:

$$\begin{aligned}\forall \tilde{x} \in \tilde{X} \quad RTend(\tilde{x}, \tilde{x}) &= 0, \\ \forall \tilde{x}, \tilde{y} \in \tilde{X} \quad RTend(\tilde{x}, \tilde{y}) &= RTend(\tilde{y}, \tilde{x}).\end{aligned}$$

Указанные свойства отношения $RTend$ позволяют классифицировать его как отношение различия, при этом совокупность всех возможных оценок $\tilde{A} = \{\tilde{\alpha}_{ij}\}$ образует нечеткую шкалу $S_a = \langle Name_RTend, \tilde{A}, \tilde{X}, X, G_a, P_a \rangle$.

Введем в состав ACL -шкалы S_x нечеткие термы сравнительных оценок $\tilde{V}$, $\tilde{A}$, семантические правила которых $P_v, P_a \in P$ определяются отношениями $TTend, RTend$, тогда получим ее расширение для «оценивания» изменений, то есть для определения не только абсолютных, но и сравнительных нечетких оценок. Структурную модель расширенной ACL -шкалы S_x представим в виде лингвистической переменной

$$S_x = \langle Name_S_x, \tilde{X}, X, G, P, \tilde{V}, \tilde{A} \rangle.$$

Таким образом, *ACL-шкала* S_x нечеткого оценивания является двухуровневой. На первом уровне иерархии *ACL-шкала* S_x позволяет определять нечеткие оценки $\tilde{x}_i$ для значений $x \in X$. Такие нечеткие оценки относятся к классу абсолютных нечетких оценок. А на втором уровне иерархии для значений $\tilde{x}_i$ и $\tilde{x}_j$ – нечеткие оценки их изменений $(\tilde{v}_{ij}, \tilde{\alpha}_{ij})$, характеризующие качественные аспекты различий или «разности первого порядка» по шкалам S_v, S_a . Такие нечеткие оценки относятся к сравнительным нечетким оценкам, которые могут быть рассмотрены как параметры нечетких тенденций.

Рассмотрим особенности *ACL-шкалы*. Предлагаемая лингвистическая *ACL-шкала* S_x относится к классу нечетких оценочных шкал, входящих в класс порядковых шкал, в ней дополнительно можно оценивать тип различия и степень различия. Это свойство позволяет рассматривать лингвистическую оценочную *ACL-шкалу* S_x как «квазиинтервальную» и определить для нее «оценочные» и «вычислительные» операции.

4.1.2. Функциональная модель *ACL-шкалы*

Введем операционный базис $F = \{ F_T, F_C, F_Er \}$ *ACL-шкалы*, в его состав включим функции для реализации операций: оценивания – F_T , вычисления оценок – F_C и определения погрешностей – F_Er .

Множество F_T «оценочных» операций *ACL-шкалы* S_x , порождающих нечеткие оценки, включает:

1. Операцию определения абсолютной нечеткой оценки $\tilde{x}_i$ по значению оцениваемого объекта x_j

$$\tilde{x}_i = Fuzzy(x_j), x_j \in X, \tilde{x}_i \in \tilde{X}.$$

2. Операцию определения значения оцениваемого объекта x_j по абсолютной нечеткой оценке $\tilde{x}_i$

$$x_j = DeFuzzy(\tilde{x}_i), x_j \in X, \tilde{x}_i \in \tilde{X}.$$

3. Операцию определения типа различия (сравнительной нечеткой оценки)

$$\tilde{v}_{ij} = TTend(\tilde{x}_i, \tilde{x}_j), \tilde{x}_i \in \tilde{X}, \tilde{x}_j \in \tilde{X}.$$

Операция *TTend* некоммутативна.

4. Операцию определения интенсивности различия (сравнительной нечеткой оценки)

$$\tilde{\alpha}_{ij} = RTend(\tilde{x}_i, \tilde{x}_j), \tilde{x}_i \in \tilde{X}, \tilde{x}_j \in \tilde{X}.$$

Операция *RTend* коммутативна.

Для нечетких значений $\tilde{v}_{ij}$ и $\tilde{\alpha}_{ij}$ допустима операция *DeFuzzy* для получения приближенных значений типа и интенсивности различия.

5. Операцию вычисления новой абсолютной нечеткой оценки

$$\tilde{x}_j = Comp(\tilde{x}_i, \tilde{v}_{ij}, \tilde{\alpha}_{ij}).$$

Таким образом, операции оценивания образуют множество

$$F_T = \{ Fuzzy, DeFuzzy, TTend, RTend, Comp \}.$$

Обозначенные операции оценивания определяют отображения, представленные в таблице 4.3.

Таблица 4.3

Операции оценивания *ACL-шкалы*

Операции оценивания	Вид отображения
ACL-шкалы	
$Fuzzy: X \rightarrow \tilde{X}$	«количество» $\rightarrow$ «качество»
$DeFuzzy: \tilde{X} \rightarrow X$	«качество» $\rightarrow$ «количество»
$TTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{V}$	«качество» $\rightarrow$ «тип изменений качества»
$RTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{A}$	«качество» $\rightarrow$ «тип качества изменений»
$Comp: \tilde{X} \times \tilde{V} \times \tilde{A} \rightarrow \tilde{X}$	«качество» $\rightarrow$ «новое качество»

Определим множество F_C как совокупность допустимых «вычислительных» операций *ACL-шкалы*, выраженных функциями для вычисления новых нечетких сравнительных оценок на основе имеющихся:

1. Операция разности интенсивностей различий

$$\tilde{\alpha}_{ij} = \text{Diff}(\tilde{\alpha}_i, \tilde{\alpha}_j).$$

2. Операция объединения интенсивностей различия

$$\tilde{\alpha}_{ij} = \text{Union}(\tilde{\alpha}_i, \tilde{\alpha}_j).$$

3. Операция пересечения интенсивностей различий

$$\tilde{\alpha}_{ij} = \text{Inter}(\tilde{\alpha}_i, \tilde{\alpha}_j).$$

Операции *Diff*, *Union*, *Inter* коммутативны, ассоциативны, ограничены.

Реализация введенных операций над нечеткими множествами может быть выполнена с помощью систем нечеткого логического вывода, рассмотренных в главе 2.

Совокупность «вычислительных» операций образует множество допустимых операций *ACL-шкалы*: $F_C = \{\text{Diff}, \text{Union}, \text{Inter}\}$.

Рассмотрим операции вычисления погрешностей (ошибок) $F_Er = \{Er_v, Er_a, Er_x, Er_x'\}$, возникающих при оценивании по *ACL-шкале*, и формирующие следующие показатели:

1. Показатель «точности» определения типа различия: сравнивается тип различия v_{ij} , полученный по «оценочной» операции *ACL-шкалы* и тип v'_{ij} , полученный по «вычислительной» операции *ACL-шкалы*

$$Er_v_j = TTend(\tilde{x}_j, \tilde{x}'_j),$$

где $\tilde{x}_j$ – значение «оценочной» операции *ACL-шкалы Fuzzy*;

$\tilde{x}'_j$ – значение абсолютной оценки, полученное на основе «вычислительной» операции *Comp*.

Результат оценивает характеристику несовпадения типов изменений.

Количество несовпадений определяется при оценивании множества значений. Это количество кроме числовой формы может быть выражено и в виде нечеткой метки («Большое», «Незначительное», «Малое» и т. д.).

2. Показатель «точности» определения интенсивности различия

$$Er_a_j = RTend(\tilde{x}_j, \tilde{x}'_j),$$

где $\tilde{x}_j$ – значение, оцененное по «оценочной» операции *ACL-шкалы Fuzzy*;

$\tilde{x}'_j$ – значение абсолютной оценки, полученное на основе «вычислительной» операции *Comp*.

Результат может быть представлен или в числовой форме или в лингвистической форме.

3. Показатель «точности» абсолютного оценивания $\tilde{x}_i$

$$Er_x = Er_x(\tilde{x}_j, \tilde{x}'_j).$$

4. Показатель «точности» оценивания исходного объекта x

$$Er_x = Er(x, x'),$$

где $x' = DeFuzzy(\tilde{x})$;

x – исходное (наблюдаемое) значение оцениваемого объекта $x \in X$;

$\tilde{x}$ – абсолютная оценка значения x по ACL-шкале.

Для конечных пользователей *ACL-шкала* S_x обеспечивает внешнее представление оценок в естественной лингвистической форме, внутреннее представление которых – семантика – формализуется на основе аппарата приближенных вычислений теории нечетких множеств и нечетких систем. Очевидно, что совокупность введенных функций *ACL-шкалы* можно рассматривать как функциональную модель *ACL-шкалы*, а ее применение для оценивания – как модель экспертной оценочной деятельности.

4.1.3. Параметрическая модель ACL-шкалы

Введенная структурная модель совместно с определенной функциональной моделью *ACL-шкалы* образуют структурно-функциональную модель *ACL-шкалы*.

В том случае, если все компоненты структурно-функциональной модели *ACL-шкалы* задаются экспертно, такая шкала будет соответствовать экспертной шкале. Оценки, полученные на ее основе, будут относиться к классу контекстно-зависимых. В то же время введенные формальные определения операций при алгоритмическом уточнении обеспечат генерацию абсолютных нечетких

оценок и введенных бинарных отношений, а значит и сравнительных нечетких оценок с помощью автоматизированных процедур. Таким образом, комплекс алгоритмически реализованных определенных в функциональной модели операций является условием и базой для автоматизированной процедуры построения модели *ACL-шкалы* на заданном универсальном множестве X . Такая шкала будет частично зависима от контекста среды.

Вследствие этого представляется обоснованным ввести параметрическую модель *ACL-шкалы*, параметры которой будут обеспечивать генерацию подмножеств возможных структурно-функциональных моделей *ACL-шкал*, построенных на одном и том же универсальном множестве X , каждая из которых будет отражать контекст предметной области. Обозначенное актуализирует задачу определения параметрической модели *ACL-шкалы* при фиксированной структурно-функциональной модели.

Параметризация *ACL-шкалы* полезна, с одной стороны, как инструмент настройки шкалы на специфику предметной области, а с другой – для реализации оптимизационных процедур с целью минимизации погрешности оценивания. В условиях невозможности решения задачи параметрической оптимизации, параметры модели *ACL-шкалы* могут устанавливаться и изменяться экспертно.

Определим параметрическую модель *ACL-шкалы* S_x в виде

$$\Psi = \{E, d, MF, nmin, nmax\},$$

где E – тип нечеткой шкалы (номинальная, порядковая или «квазиинтервальная (равномерная/неравномерная)»);

d – параметр, определяющий носитель типа изменения нечетких градаций «Стабильность», то есть длину интервала на X , все значения x в котором могут рассматриваться с позиции данной шкалы, как одинаковые, неразличимые;

MF – тип функций принадлежности, моделирующих нечеткие оценки $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$, например, треугольного вида;

$$nmin = \inf(X), nmax = \sup(X).$$

Тогда количество нечетких градаций шкалы (мощность множества $\tilde{X}$) вычисляется: $m = \frac{2 \cdot (n_{\max} - n_{\min})}{d} + 1$, с последующим округлением до ближайшего целого.

4.1.4. Параметрическая структурно-функциональная модель ACL-шкалы

Введенные структурная, параметрическая и функциональная модели образуют параметрическую структурно-функциональную модель ACL-шкалы, которую представим в виде алгебраической системы:

$$C = \{H, \Omega, \Psi\},$$

где H – множество объектов: $H = \{X, \tilde{X}, G, P, \tilde{V}, \tilde{A}\}$;

Ω – множество операций на множестве H , заданных функциями:

$$\Omega = \{F_T, F_C, F_Er\},$$

$$F_T = \{Fuzzy, DeFuzzy, Comp, TTend, RTend\},$$

$$F_C = \{Diff, Union, Inter\},$$

$$F_Er = \{Er_v, Er_a, Er_x, Er_x\};$$

Ψ – множество параметров $\Psi = \{E, d, MF, nmin, nmax\}$.

Использование единого базиса в виде ACL-шкалы для порождения абсолютных и соответствующих им сравнительных нечетких оценок позволит оперировать совместимыми значениями таких нечетких оценок объектов и проектировать нечеткие модели, обладающие дополнительными возможностями. К таким возможностям следует отнести контекстную адаптацию ACL-шкалы путем ее модификации (расширение, преобразование, сжатие), анализ как статических, так и динамических семантических свойств объектов различной природы и их последовательностей в рамках однородных знаковых структур, критериев и целей оценки. На рис. 4.2 с учетом вышеизложенного приведена концептуальная модель ACL-шкалы, позволяющей выполнять нечеткое оценивание и генерировать абсолютные и сравнительные нечеткие оценки.

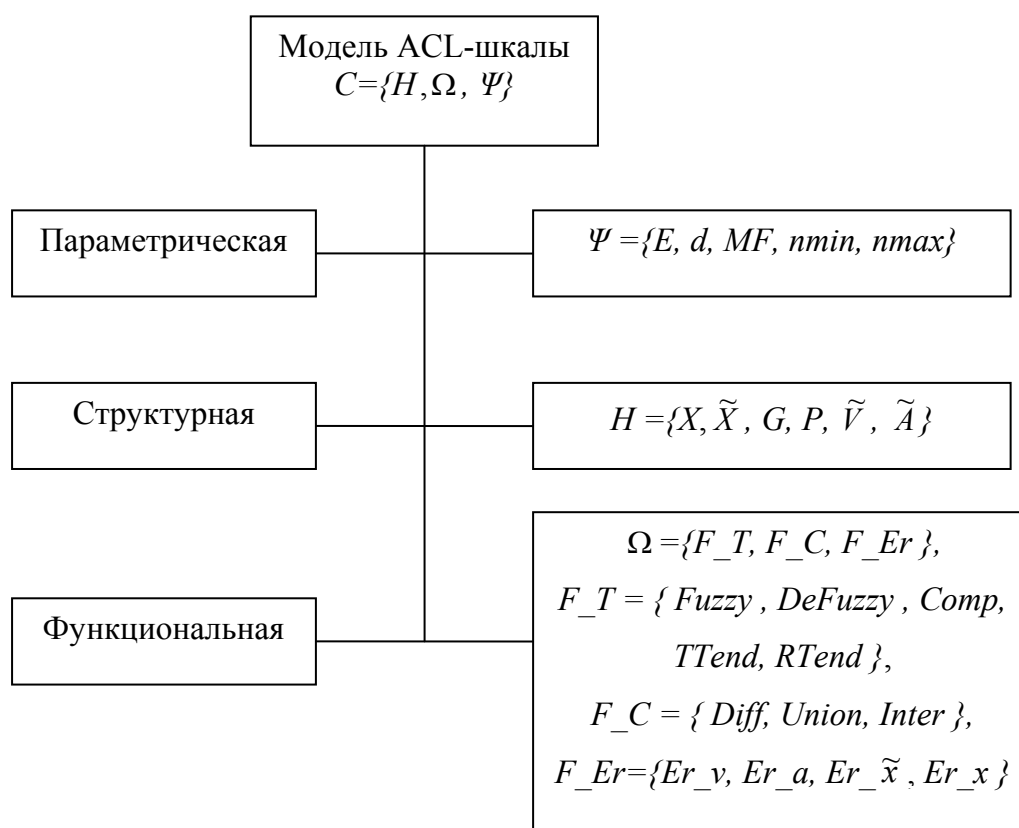


Рис. 4.2. Концептуальная модель ACL-шкалы для генерации нечетких оценок

Введенная ACL-шкала по своей природе является нечеткой, гибридной, многомерной и нелинейной. Нечеткий аспект шкалы связан с определением ее градаций и процедур оценивания в терминах нечетких множеств. Гибридный характер шкалы выражается в объединении нескольких видов традиционных шкал: так, по одному измерению она соответствует *номинальной шкале* (название нечеткого множества) с операциями сравнения на равенство, по второму измерению – *интервальной шкале*, задающей изменение базовых значений переменной (с числовыми операциями), по третьему измерению – *интервальной шкале*, задающей изменение функции принадлежности базового значения соответствующему нечеткому множеству (с операциями для нечетких множеств), по четвертому измерению – *порядковой шкале*, определяющей отношение порядка между элементами номинальной шкалы.

4.1.5. Метод построения ACL-шкалы для нечеткого оценивания уровней временного ряда

При построении ACL-шкалы важным с точки зрения «допустимого уровня погрешности» является определение количества ее градаций, выраженных в виде нечетких множеств (НМ) и их носителей.

Определение 4.2. Мощностью ACL-шкалы будем называть количество нечетких множеств (функций принадлежности), задающих количество ее градаций m .

Пусть известно универсальное множество X . Обозначим задачу автоматического определения такого количества одинаковых функций принадлежности (ФП), покрывающих множество X , при котором ошибка оценивания по ACL-шкале находится в допустимом интервале, длина которого не превышает определенного уровня ε . Фактически, это задача о минимальном разбиении универсума X на покрывающие диапазон интервалы, при котором каждая абсолютная ошибка оценивания по ACL-шкале по модулю не превышает заданный уровень ε .

Обратим внимание на источники, генерирующие погрешность $\varepsilon = |x - x'|$ при переходе от четкого значения x к нечеткому множеству A и обратно к четкому значению x' . Источником является усреднение множества четких значений, принадлежащих носителю нечеткого множества (НМ) длины d , и замена их другим четким значением x' , как правило, определяемым через центр тяжести нечеткого множества A .

Следовательно, длина интервала d , равная $\varepsilon/2$, на котором определен носитель нечеткого множества, может служить оценкой значения генерируемой погрешности $\varepsilon' \leq 2 \cdot d$ для каждого значения носителя нечеткого множества A . При бесконечном уменьшении длины интервала d носителя НМ погрешность ε будет стремиться к нулю. Отметим, что в таком представлении d является еще и показателем уровня размытости, нечеткости объекта, моделируемого нечетким множеством.

Теорема 4.1. *Теорема о мощности ACL-шкалы.* Мощность ACL-шкалы, построенной для временного ряда, обратно пропорциональна относительной погрешности оценивания и количеству членов временного ряда.

Доказательство:

Применительно к ВР будем считать, что количество m равномерно заданных ФП одинаковой формы ACL-шкалы можно определить на основе задаваемой погрешности ε оценки уровня ВР по функции:

$$m = \frac{2 \cdot (n_{\max} - n_{\min})}{d} + 1,$$

где $d = \varepsilon/2$ – длина интервала носителя нечеткого множества;

$n_{\max}, n_{\min}$ – задают минимальное и максимальное значение уровней ВР.

Значение m с последующим округлением до ближайшего целого и определяет количество равномерно заданных ФП одинаковой формы ACL-шкалы.

Значение погрешности ε может быть использовано для автоматического построения ФП и при кластерном разбиении диапазона ACL-шкалы. Это значение погрешности ε будет определять максимальное отклонение от центра кластера.

Рассмотрим вопрос оценивания погрешности аппроксимации ВР. При оценивании используются различные меры, используем меру, называемую средней абсолютной процентной ошибкой

$$MAPE = \frac{1}{k} \left(\sum_{i=1}^k (abs(X_i - X'_i) / X_i) \right),$$

где k – количество членов временного ряда;

$(X_1, \dots, X_k)$ – ненулевые уровни ВР, полученные с помощью наблюдений;

$(X'_1, \dots, X'_k)$ – уровни ВР, полученные в результате применения операций

Fuzzy/DeFuzzy ACL-шкалы.

Как видно, под знаком суммы числитель представляет собой абсолютную ошибку. Предполагая, что все они не превышают некоторой максимальной оценки ε , получим

$$MAPE = \frac{\varepsilon}{k} \sum_{i=1}^k \frac{1}{X_i}.$$

Обозначим $S = \sum_{i=1}^k \frac{1}{X_i}$, тогда $MAPE = S \cdot \frac{\varepsilon}{k}$.

Задавая экспертно желаемый уровень относительной погрешности MAPE, например, $MAPE = \delta$, получим $\delta = S \cdot \frac{\varepsilon}{k}$.

Для исходного ВР на этой основе можно определить оценку $\varepsilon = \frac{k \cdot \delta}{S}$ и оценку длины интервала носителя каждого НМ.

Таким образом, по исходным значениям ВР $(X_1, \dots, X_k)$ и задаваемой погрешности δ можно определить длину носителя нечетких множеств ACL-шкалы: $d = \frac{k \cdot \delta}{2 \cdot S}$.

Подставляя эти значения в формулу вычисления количества нечетких множеств ACL-шкалы

$$m = \frac{2 \cdot (n_{\max} - n_{\min})}{d} + 1,$$

определяющих ее мощность, получим:

$$m = \text{int} \left(\frac{4 \cdot (X_{\max} - X_{\min})}{k \cdot \delta} \cdot S \right) + 1,$$

где $X_{\max} - X_{\min}$ – размах уровней ряда;

$(X_1, \dots, X_k)$ – ненулевые уровни ВР;

$$S = \sum_{i=1}^k \frac{1}{X_i};$$

k – количество членов временного ряда;

δ – задаваемый допустимый уровень погрешности оценивания (критерий MAPE) по ACL-шкале.

Теорема доказана.

Заметим, что для одного и того же ВР при уменьшении погрешности δ мощность ACL-шкалы (количество НМ) должно возрастать, так как между ними зависимость обратно пропорциональная.

Сформулируем постановку задачи построения ACL-шкалы следующим образом. Пусть имеется универсальное множество X . Требуется построить нечеткую лингвистическую шкалу, содержащую в качестве градаций упорядоченные нечеткие метки качественных оценок $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$, такую, что каждому элементу $x \in X$ на шкале можно сопоставить одну или несколько нечетких меток с разной степенью соответствия.

Для построения на множестве X ACL-шкалы необходимо задать множество параметров, образующих ее параметрическую модель, а затем генерировать структурную модель шкалы. Моделирование экспертной деятельности решения различных задач включает процедуру *нечеткой кластеризации* накопленных знаний о предметной области универсального множества X . Кластеры на X будут определять пересекающиеся подмножества объектов $x \in X$, таких, что каждому кластеру эксперт сопоставляет некоторую экспертную оценку $\tilde{x}_i$. Множество сформированных экспертных оценок образует множество градаций шкалы $\tilde{X} = \{\tilde{x}_i\}$, $i \in [1, m]$. Очевидно, что количество кластеров m соответствует количеству экспертных оценок.

В простейшем случае, когда универсальное множество есть ограниченное множество действительных чисел, кластеры представляются интервалами, в общем случае различной длины. В условиях отсутствия знаний о предметной области универсального множества X , эксперт может задать кластеры в виде интервалов одинаковой длины и задать количество градаций шкалы m .

Для построения отношений типа и интенсивностей различий или изменений в нечетких экспертных оценках применяют процедуры ранжирования.

Рассмотрим обобщенный алгоритм решения задачи построения ACL-шкалы.

1. Анализ типа универсального множества X : числовое, символическое. Допустим, для определенности универсальное множество — числовое.

2. Задание множества параметров $\Psi = \{E, d, MF, nmin, nmax\}$. Пусть шкала относится к классу $E = \text{«квазиинтервальная и равномерная»}$. Тип функций принадлежности $MF = \text{«треугольные»}$.

3. Разбиение универсального множества на интервалы. Если длина интервала d , на котором изменения считаются несущественными, задана, то определение мощности ACL-шкалы, то есть количества нечетких множеств $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$, образующих градации шкалы, по формуле $m = \frac{2 \cdot (nmax - nmin)}{d} + 1$, с последующим округлением до ближайшего целого.

Если длина интервала d не задана, то должно быть задано явно количество градаций m . Затем необходимо вычислить длину интервала d . В предположении, что разбиение универсума X равномерное, то $d = (nmax - nmin) / m$.

При неравномерном разбиении универсума X для определения соответствующих длин интервалов целесообразно использовать кластерный метод.

4. Построение m функций принадлежности $\tilde{x}_i = \{x, \mu_{\tilde{x}_i}(x)\}$, $x \in X$, $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$ класса MF (например, треугольных) на интервалах d универсального множества X . Треугольная функция принадлежности определяется тройкой чисел (a, b, c) , и ее значение в точке x вычисляется согласно выражению:

$$\mu(x) = \begin{cases} \frac{x - a}{b - a}, & a \leq x \leq b \\ \frac{c - x}{c - b}, & b < x \leq c \\ 0, & x < a, x > c. \end{cases}$$

5. Генерация отношений $TTend, RTend$, задающих изменения на множестве упорядоченных нечетких градаций $\tilde{X}$ и соответствующих нечетких шкал S_v, S_a , построение функций принадлежности.

Отметим, что генерация отношений $TTend, RTend$ сводится к построению квадратных матриц по таблицам, вид которых представлен таб-

лицами 4.1, 4.2. Размерность матриц, задающих отношения $TTend$, $RTend$, одинакова и равна m – количеству нечетких оценок $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$. Элементы матрицы для отношения $TTend$ задаются из фиксированного множества {Рост, Падение, Стабильность}, а при задании лингвистических элементов матрицы $RTend$ имеет смысл использовать лингвистические термы эксперта или абстрактные лингвистические термы, выражающие «неметрическое расстояние» между двумя нечеткими метками. При числовом универсальном множестве значения элементов отношения $RTend$ могут быть определены как центры тяжести нечетких множеств (нечетких кластеров) $\tilde{x}_i$. В условиях выполнения предположения об упорядоченности нечетких меток построение таких матриц не представляется затруднительным.

В результате, сформирована структурная модель ACL-шкалы со встроенным функциональным базисом, которую можно использовать для оценивания и генерации абсолютных и сравнительных нечетких оценок.

4.1.6. Алгоритм генерации нечетких оценок на основе ACL-шкалы

Сформулируем постановку задачи генерации нечетких оценок на основе ACL-шкалы как задачу моделирования нечетких экспертных оценок [Афанасьева и др., 2009].

Пусть имеется универсальное множество X и элемент $x \in X$.

Требуется решить следующие задачи:

A1) Вычисление абсолютной нечеткой оценки. Данная задача в экспертном варианте решения основывается на процедуре *нечеткой классификации* объекта $x \in X$. Постановку задачи нечеткой классификации сформулируем следующим образом. Для элемента $x \in X$ определить соответствующую ему нечеткую метку $\tilde{x}_i \in \tilde{X}$ и ее степень соответствия. Если таких меток будет несколько, определить ту, которая имеет максимальную степень соответствия.

A2) *Вычисление сравнительных оценок.* Задаче вычисления сравнительных оценок, выражающих различия двух нечетких объектов $\tilde{x}_i \in \tilde{X}, \tilde{x}_j \in \tilde{X}$, естественно сопоставить также задачу нечеткой классификации. Сформулируем постановку задачи в следующем виде: для нечетких меток $\tilde{x}_i, \tilde{x}_j \in \tilde{X}$ определить нечеткую метку типа $\tilde{v}_{ij} \in \tilde{V}$ и нечеткую метку интенсивности $\tilde{\alpha}_{ij} \in \tilde{A}$ изменений (различий). Если таких меток будет несколько, определить ту, которая имеет максимальную степень соответствия.

Обобщенный алгоритм решения задачи вычисления абсолютной нечеткой оценки, в основе которого лежит использование операции $Fuzzy(x)$ по ACL-шкале, представлен ниже.

1. С помощью операции $Fuzzy: X \rightarrow \tilde{X}$ определяем для оцениваемого объекта $x \in X$ множество нечетких оценок $\tilde{x}_i = Fuzzy(x), x \in X, \tilde{x}_i \in \tilde{X}, i = [1, k]$.
2. Среди полученного множества нечетких оценок $\tilde{x}_i$ определяем номер i оценки с максимальной степенью соответствия по значениям функций принадлежности:

$$i = \arg \max (\mu_{\tilde{x}_i}(x)), \text{ при изменении } i = [1, k].$$

Аналогично поступаем для вычисления любой другой нечеткой оценки $\tilde{x}_j$.

Задача вычисления сравнительной оценки типа и интенсивности изменений с помощью операций $TTend, RTend$ по ACL-шкале алгоритмически реализуется нечеткими системами логического вывода, например, на основе нечеткой модели Мамдани следующим образом:

Формируется база правил R , реализующих операции $TTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{V}$ для определения типа изменения $\tilde{v}_{ij} = TTend(\tilde{x}_i, \tilde{x}_j), \tilde{x}_i \in \tilde{X}, \tilde{x}_j \in \tilde{X}$ и $RTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{A}$ для определения интенсивности изменения $\tilde{\alpha}_{ij} = RTend(\tilde{x}_i, \tilde{x}_j), \tilde{x}_i \in \tilde{X}, \tilde{x}_j \in \tilde{X}$ следующей структуры:

$$R_1 := \text{IF } X_t \text{ is } A_{11} \text{ AND } X_{t+1} \text{ is } A_{12} \text{ THEN } v_t \text{ is } B_1 \text{ AND } a_t \text{ is } N_1$$

.....

$$R_m := \text{IF } X_t \text{ is } A_{m1} \text{ AND } X_{t+1} \text{ is } A_{m2} \text{ THEN } v_t \text{ is } B_m \text{ AND } a_t \text{ is } N_m$$

Семантика правил R вычисления сравнительных оценок отображена в таблице 4.4.

Пример вычисления сравнительных оценок

$\tilde{x}_i$	$\tilde{x}_j$	$\tilde{v}_{ij} = TTend(\tilde{x}_i, \tilde{x}_j)$	$\tilde{\alpha}_{ij} = RTend(\tilde{x}_i, \tilde{x}_j)$
at	S	at	Ze
	G	G	Ze
o	E	E	Ze
x	B	B	Ze
ad	Z	Z	Ze
e	Z	B	Sm
e	Z	S	Me
e	Z	G	Bi
e	Z	E	VeBi
ad	B	Z	Sm
ad	B	S	Sm
ad	B	G	Me
ad	B	E	Bi
at	S	Z	Me
at	S	B	Sm
at	S	G	Sm
at	S	E	Me
o	G	Z	Bi
o	G	B	Me

	G		S	Dec	Sm
o		at			
	G		E	Inc	Sm
o		x			
	E		Z	Dec	VeBi
x		e			
	E		B	Dec	Bi
x		ad			
	E		S	Dec	Me
x		at			
	E		G	Dec	Sm
x		o			

В таблице 4.4 были использованы следующие сокращения для нечетких значений НБР: *Ze(Отсутствует)*, *Bad(Плохой)*, *Sat(Удовлетворительный)*, *Go(Хороший)*, *Ex(Отличный)*, для значений типов изменений *Inc(РОСТ)*, *Dec(ПАДЕНИЕ)*, *St(СТАБИЛЬНОСТЬ)*, для значений интенсивности изменений – *Bi(Большой)*, *Me(Средний)*, *Sm(Малый)* и модификаторы *Ve(Очень)*, *Si(Значительно)*, *No(Не)*.

В результате, определен алгоритмический базис параметрической структурно-функциональной модели ACL-шкалы, позволяющей формализовать для любых двух нечетких оценок, заданных на одном универсальном множестве, сравнительную оценку, характеризующую тип и интенсивность их различия (изменения). Такая сравнительная оценка и модель ACL-шкалы будут служить фундаментом при формализации модели нечеткой тенденции нечетких временных рядов.

4.1.7. Алгоритм построения нечеткого временного ряда с использованием ACL-шкалы

Напомним содержательное определение нечеткого временного ряда [Ярушкина, 2004]. *Нечетким временным рядом* (НБР) называют упорядоченную в равноотстоящие моменты времени последовательность наблюдений над

некоторым процессом, состояния которого изменяются во времени, если значение состояния процесса в момент t_i выражено с помощью нечеткой метки $\tilde{x}_i$.

Для преобразования четкого временного ряда, уровни которого могут быть любой природы, в нечеткий временной ряд предлагается использовать алгоритмический базис параметрической структурно-функциональной ACL-шкалы.

Сформулируем постановку задачи.

Дан временной ряд в виде последовательности упорядоченных в моменты времени пар $Y = \{t_i, x_i\}$, таких что $\forall x_i \in R^1, t_i \in N, i \in [1, n]$.

Требуется построить нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}$, то есть для каждого значения x_i пары $\{x_i, t_i\}$ определить соответствующее значение нечеткой оценки $\tilde{x}_i \in \tilde{X}, i \in [1, n]$ и соответствующую функцию принадлежности $\mu_{x_i}(x_i)$.

Построение НВР будем осуществлять в два этапа. На первом этапе по исходному ВР построим ACL-шкалу, на втором – оценим по этой шкале уровни исходного ВР.

Процедура построения ACL-шкалы по временному ряду $Y = \{t_i, x_i\}$ аналогична обобщенному алгоритму построения ACL-шкалы на универсальном множестве $X = [x_{\max}, x_{\min}]$, где $x_{\max}, x_{\min}$ определяют размах уровней ВР (см. раздел 4.1.5).

Процедура нечеткого оценивания уровней временного ряда – обобщенному алгоритму вычисления абсолютной нечеткой оценки, примененному к каждому уровню ВР, рассмотренному в разделе 4.1.6.

Таким образом, построенный на основе исходного ВР, нечеткий временной ряд с помощью ACL-шкалы представляется в виде упорядоченных по времени совокупностей значений

$$\tilde{Y} = \{t_i, \tilde{x}_i, \mu_{x_i}(x_i)\},$$

где $\tilde{x}_i$ – лингвистическое наименование нечеткой оценки;

$\mu_{x_i}(x_i)$ – функция принадлежности, соответствующая этой нечеткой оценке.

Замечание 4.1. Так как в соответствии с определением нечетких множеств $\tilde{x}_i = \{x_i, \mu_{x_i}(x_i)\}$, то следующие определения нечеткого временного ряда

$$(1) \quad \tilde{Y} = \{t_i, \tilde{x}_i, \mu_{x_i}(x_i)\};$$

$$(2) \quad \tilde{Y} = \{t_i, \tilde{x}_i\};$$

$$(3) \quad \tilde{Y} = \{t_i, \mu_{x_i}(x_i)\};$$

$$(4) \quad \tilde{Y} = \{\tilde{y}_t\}$$

в дальнейшем будем рассматривать как эквивалентные.

Утверждение 4.1. Для каждого исходного временного ряда $Y = \{t_i, x_i\}$ можно построить множество нечетких временных рядов $\{\tilde{Y}_k\}$ в зависимости от выбора параметров *ACL-шкалы*, алгоритмов реализации ее операций и способов построения.

Утверждение 4.2. Для фиксированных параметров, способов построения и реализации операций *ACL-шкалы* по исходному ВР $Y = \{t_i, x_i\}$ можно построить единственный нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}$.

Утверждение 4.3. Для нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}$, построенного с помощью *ACL-шкалы* на исходном временном ряду $Y = \{t_i, x_i\}$, существует множество четких временных рядов $\{Y_s\}$, полученных путем применения различных способов реализации операции $\text{DeFuzzy}(\tilde{Y})$ *ACL-шкалы*, являющихся результатом нечеткой аппроксимации. Качество нечеткой аппроксимации исходного временного ряда определяется на основе операции Er_x *ACL-шкалы*.

Утверждение 4.4. Построение нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}$ по исходному временному ряду $Y = \{t_i, x_i\}$ с помощью *ACL-шкалы*, позволяющей генерировать оценки изменения (различия) между любыми нечеткими метками

$\tilde{x}_i$, обеспечивает необходимый формальный аппарат для оценки поведения НВР в терминах нечетких тенденций.

Контрольные вопросы

1. Дайте определение шкалы.
2. Приведите формальное определение ACL-шкалы.
3. Что такое $TTend(\tilde{x}_i, \tilde{x}_j)$?
4. Дайте определение отношения $RTend(\tilde{x}_i, \tilde{x}_j)$.
5. Что включает в себя множество «оценочных» операций ACL-шкалы, порождающих нечеткие оценки?
6. Для чего нужна параметризация ACL-шкалы?
7. Дайте алгебраическое представление параметрической структурно-функциональной модели ACL-шкалы.
8. Приведите теорему о мощности ACL-шкалы.
9. Перечислите этапы обобщенного алгоритма решения задачи построения ACL-шкалы по ВР.
10. Сформулируйте постановку задачи генерации нечетких оценок на основе ACL-шкалы как задачу моделирования нечетких экспертных оценок.
11. Можно ли преобразовать четкий временной ряд, уровни которого могут быть любой природы, в нечеткий временной ряд?

4.2. Классификация нечетких тенденций

Согласно определению нечеткой тенденции [Ярушкина, 2004], каждая нечеткая тенденция τ_k нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}, i \in [1, n], \tilde{x}_i \in \tilde{X}$ может быть представлена нечеткой меткой, выражающей характер поведения НВР.

В рамках теории нечетких множеств нечеткая тенденция представима нечетким множеством $\tau_k = \{\tilde{X}, \mu_{\tau_k}(\tilde{X})\}, k \in [1, kn]$, где $\mu_{\tau_k}(\tilde{X})$ – функция принадлеж-

ности нечеткой тенденции τ_k , определенная на множестве нечетких уровней $\tilde{X}$ нечеткого временного ряда.

Содержательно термин «тенденция» ассоциируется с изменением и используется в предложениях совместно с общеупотребительными лингвистическими оценками изменений, связанными с протяженностью, типом и интенсивностью, например, «длительная тенденция роста», «ярко выраженная тенденция падения», «тенденция стабильного высокого качества» и т. д.

Выделим следующие свойства нечеткой тенденции (НТ):

1. *Нечеткость*. Это свойство обозначает факт, что НТ построена на основе нечетких значений НВР и наследует нечеткость этих значений.
2. *Протяженность*. Это свойство связано с различной длительностью НТ.
3. *Типичность*. Свойство типичности НТ позволяет различать классы, типы НТ, внутри которых нечеткие тенденции будут рассматриваться как однородные.
4. *Значимость*. Для различия нечетких тенденций одного типа и одинаковой длительности целесообразно использовать характеристику степени значимости или интенсивности НТ.
5. *Ориентированность во времени*. Это свойство обозначает, что нечеткие тенденции определяются в направлении увеличения временных отсчетов.
6. *Лингвистическая интерпретируемость*. Данное свойство нечеткой тенденции следует из определения нечеткой тенденции. По определению, НТ есть нечеткая метка, которой сопоставляется лингвистически терм.

Формальным аппаратом для извлечения знаний из нечетких временных рядов, позволяющим представлять поведение ВР в терминах лингвистических оценок тенденций, являются *ACL-шкала* и ее функциональное наполнение, связанное с генерацией сравнительных нечетких оценок изменений.

Рассмотрим виды *ACL-шкал* и нечетких оценок изменений, получаемых на этих шкалах с целью формализации свойств нечеткой тенденции нечеткого временного ряда.

Так, в случае номинальной *ACL-шкалы* нечеткая тенденция задается бинарной оценкой $E0=(\text{Есть изменения}, \text{Нет изменений})$.

Для ранговой *ACL-шкалы* определены понятия *БОЛЬШЕ*, *МЕНЬШЕ*, *РАВНО*, которые задают тип тенденции. Поэтому, если *ACL-шкала* ранговая, то тенденция может быть выражена в нечетких оценках, отражающих тип тенденции $E1=(\text{Нет Изменений}, \text{Есть Изменения (в сторону увеличения, в сторону уменьшения)})$ или $E1=(\text{Стабильность}, \text{Рост}, \text{Падение})$.

В том случае, если имеется возможность определить степень выраженности типа изменения, логика оценок уточняется за счет включения этого дополнительного параметра, для оценки которого должна быть определена своя шкала, например, бальная, ранговая или числовая. Тогда результат «измерения» на *ACL-шкале* изменений может быть представлен в виде $E2=(\text{Стабильность}, \text{Рост}(\tilde{A}), \text{Падение}(\tilde{A}))$, где $\tilde{A}$ – оценки для «измерения» степени интенсивности тенденции.

В более общем случае каждая оценка $E2$ представима следующим образом: $(\nu_k (p_{ks}))$, где ν_k – тип изменения, k – количество типов изменения, p_{ks} – параметры типов изменения, s – количество параметров. Следует отметить, что при оценивании двух нечетких уровней нечеткого временного ряда каждая оценка $E0$, $E1$, $E2$ наследует свойства нечеткости этих значений, то есть вычисленную при фаззификации степень принадлежности.

Отметим особенности оценивания тенденций в базисе оценок $E2$ для последовательности нечетких значений, представленных в виде нечеткого временного ряда (НВР).

Первая особенность заключается в том, что исходный НВР порождает временной ряд степеней принадлежности нечетких тенденций.

Рассмотрим вторую особенность. При использовании *ACL-шкал* для оценивания тенденций в последовательности нечетких значений НВР, результаты формируются в виде множеств, соответствующих типам ν_k оценок $E2$. Каждое из этих множеств характеризуется функцией распределения выраженности $\alpha_{\nu k}$ и мощностью, которая в простейшем случае сопоставима с длительностью

k -того типа нечеткой тенденции Δt_{vk} . Указанные функции α_{vk} , Δt_{vk} образуют параметры типов тенденций, а последовательность типов тенденций определяет структуру тенденции НВР.

Следовательно, при определении тенденции НВР на основе оценок $E0$, $E1$, $E2$ модель нечеткой тенденции τ может быть представлена с разной степенью детализации:

$$\begin{aligned}\tau_k(\tilde{X}, E0) &= (v_k, \mu_k), \\ \tau_k(\tilde{X}, E1) &= (v_k, \mu_k), \\ \tau_k(\tilde{X}, E2) &= (v_k, \mu_k, \alpha_{vk}, \Delta t_{vk}),\end{aligned}$$

где τ_k – наименование нечеткой тенденции;

v_k – тип нечеткой тенденции, например, для оценки $E0$ из базового множества типов {«Есть изменения», «Нет изменения»}, для оценок $E1$ и $E2$ из базового множества {«Рост», «Падение», «Стабильность»}. Базовое множество может быть расширено за счет производных типов тенденций;

μ_k – функция принадлежности нечеткой тенденции, которая формируется на основе образующих ее нечетких значений ВР;

α_{vk} – интенсивность каждого типа нечеткой тенденции, контекстное расширение типа тенденции;

Δt_{vk} – длительность каждого типа нечеткой тенденции.

В дальнейшем будем рассматривать модель нечеткой тенденции НВР в виде оценок $E2$ как наиболее информативную.

Назовем нечеткие метки «Рост», «Падение», «Стабильность» базовыми типами тенденций, а «Колебания», «Хаос» – производными типами, так как вывод относительно их типа формируется на основе базовых.

Тогда нечеткое множество ACL -шкалы $\tilde{V}$ может быть использовано для обозначения типов тенденций НВР. Представим $\tilde{V} = \tilde{V}1 \cup \tilde{V}2$, где $\tilde{V}1$ определяет множество базовых типов тенденций НВР, которое конечно и ограничено тремя типами $\tilde{V}1 = \{\tilde{v}1, \tilde{v}2, \tilde{v}3\}$, например, $\tilde{v}1 = \text{«Стабилизация»}$, $\tilde{v}2 = \text{«Падение»}$, $\tilde{v}3 = \text{«Рост»}$, $\tilde{V}2$ – множество производных типов тенденций НВР, которое

также конечно и ограничено, например, $\tilde{V}2 = \{\tilde{v}4, \tilde{v}5\}$, где $\tilde{v}4 = \text{«Колебания периодические»}$, $\tilde{v}5 = \text{«Колебания хаотические»}$.

Пусть нечеткое множество $\tilde{A}$ для заданного нечеткого временного ряда генерируется в процессе построения *ACL-шкалы* автоматизированной процедурой.

Для построения лингвистических оценок протяженности тенденции построим универсальное множество ΔT , такое, что $\forall \Delta t \in \Delta T$. На этом универсуме построим *ACL-шкалу* $S_{\Delta T}$ для генерации лингвистической нечеткой оценки длительности нечеткой тенденции $\Delta \tilde{t} \in \Delta \tilde{T}$.

Утверждение 4.5. Каждая нечеткая тенденция $\tau \in \mathfrak{T}$ нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}$ может быть представлена структурной моделью в виде кортежа отношения, построенного на декартовом произведении свойств нечеткой тенденции $\tilde{V} \times \tilde{A} \times \Delta \tilde{T} \rightarrow \mathfrak{T}$:

$$\tau = \langle \tilde{v}, \tilde{a}, \Delta \tilde{t}, \mu \rangle,$$

где τ — наименование нечеткой тенденции (идентификатор) из множества $\mathfrak{T}$, $\tau \in \mathfrak{T}$;

$\tilde{v}$ — тип нечеткой тенденции (тип изменений), $\tilde{v} \in \tilde{V}$;

$\tilde{a}$ — степень изменения, то есть интенсивность нечеткой тенденции, $\tilde{a} \in \tilde{A}$;

$\Delta \tilde{t}$ — продолжительность нечеткой тенденции, $\Delta \tilde{t} \in \Delta \tilde{T}$;

μ — функция принадлежности участка нечеткого временного ограниченного $\Delta \tilde{t}$ нечеткой тенденции τ .

При значении функции принадлежности, равном единице, нечеткая тенденция относится к четким тенденциям.

Определение 4.3. К однородным нечетким тенденциям относятся нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, для которых верно $\tilde{v}_i = \tilde{v}_s$.

Определение 4.4. Объединением однородных нечетких тенденций $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ является тенденция $\tau_j = \tau_i \cup \tau_s$, такая что $\tau_j = \langle \tilde{v}_j, \tilde{a}_j, \Delta \tilde{t}_j, \mu_j \rangle$, для которой $\tilde{v}_j = \tilde{v}_s$, $\tilde{a}_j = \text{Union}(\tilde{a}_i, \tilde{a}_s)$, $\mu_j = \mu_i \cup \mu_s$, длительность тенденции $\Delta \tilde{t}_j = \text{Union}(\Delta \tilde{t}_i, \Delta \tilde{t}_s)$.

Определение 4.5. К эквивалентным нечетким тенденциям относятся однородные нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{a}_i = \tilde{a}_s$, $\Delta \tilde{t}_i = \Delta \tilde{t}_s$, $\mu_i = \mu_s$.

Определение 4.6. К неоднородным нечетким тенденциям относятся нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{v}_i \neq \tilde{v}_s$.

Определение 4.7. К подобным нечетким тенденциям относятся однородные нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ такие, что $\tilde{a}_i = \tilde{a}_s$, $\Delta \tilde{t}_i = \Delta \tilde{t}_s$.

Семантика введенных компонент нечеткой тенденции и операций их объединения, сравнения определяется их нечетко-множественной природой и моделируется на основе отдельных *ACL-шкал*.

Рассмотрим нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}, i \in [1, n]$, n – количество членов нечеткого ВР. В том случае, если модель нечеткой тенденции определяется для двух соседних уровней НВР, такую нечеткую тенденцию τ отнесем к классу *элементарных нечетких тенденций* $\tau \in \mathfrak{T}$, если модель нечеткой тенденции определяется между начальным и конечным уровнем НВР, такую нечеткую тенденцию будем классифицировать как *общую (основную) тенденцию* $G\tau \in G\mathfrak{T}$ НВР, в противном случае, нечеткие тенденции, определяемые для последовательностей уровней НВР, будем считать *локальными тенденциями* $N\tau \in N\mathfrak{T}$.

Содержательно *элементарная нечеткая тенденция* моделирует изменения между двумя соседними значениями нечеткого временного ряда $\tilde{x}_i, \tilde{x}_{i+1}$ и может быть сопоставлена с моментом времени t_i .

Локальная нечеткая тенденция определяется между двумя выбранными значениями нечеткого временного ряда $\tilde{x}_i$ и $\tilde{x}_j$, при $i < j$ может быть сопоставлена с моментом времени t_i . Любая локальная НТ может быть выражена последовательностью элементарных нечетких тенденций $N\tau = \{\tau_1, \tau_2, \dots, \tau_m\}$ при $m = j - i, m < n$.

Общая нечеткая тенденция характеризует поведение всего НВР и представима в виде последовательности локальных, а значит и элементарных нечетких тенденций $G\tau = \{\tau_1, \tau_2, \dots, \tau_m\}$ при $m=n$.

Последовательность элементарных тенденций, образующая локальную или общую нечеткую тенденций НВР, может быть представлена моделью нечеткой тенденции в результате применения «вычислительной» операции *Union* ACL-шкалы и операции объединения элементарных тенденций.

Таким образом, для любого НВР может быть построен временной ряд элементарных НТ и временной ряд локальных НТ.

Утверждение 4.6. Исходный нечеткий временной ряд может быть представлен в виде нечеткого временного ряда элементарной тенденций $\tau = \{t_i, \tau_i\}, i \in [1, n]$, нечеткого временного ряда локальных тенденций $N\tau = \{t_j, N\tau_j\}, j \in [1, n1], n1 < n$.

Резюмируя, отметим следующее:

1. Модели элементарной, локальной, общей НТ нечеткого временного ряда обладают общей структурой.
2. Локальная и общая нечеткие тенденции НВР могут быть выражены через временной ряд элементарных НТ.
3. Временной ряд элементарных НТ является инвариантным средством лингвистического представления поведения любого НВР.
4. Представление временных рядов в виде нечетких временных рядов и временных рядов элементарных нечетких тенденций позволяет учитывать при их анализе дополнительные знания в виде семантики прикладной области за счет использования контекстно-зависимых нечетких меток.

Гранулярная трактовка нечеткой тенденции

В рамках направления Data Mining временных рядов, нечеткие уровни нечеткого временного ряда интерпретируются с позиции теории нечеткой логики *point-wise* нечеткими ограничениями (*pw-ограничения*).

Пример на рис 4.3, взятый из работы [Batyrrshin, 2007], иллюстрирует пример распространения *rw-ограничения* на временной домен с помощью расширения по принципу Л. Заде.

В работе [Batyrrshin, 2007] введен класс нечетких ограничений, называемый *window-wise* ограничением (*ww-ограничение*). Данный класс ограничений определен на интервале между временными точками или может быть распространен на интервал между временными точками и соответствует в нашем представлении понятию нечеткой тенденции.

С помощью различных классов нечетких ограничений, определенных на НВР, таким образом, подчеркивается их различная семантика.

Так как любое нечеткое ограничение имеет лингвистическую интерпретацию дополнительно к функциональному представлению, то в качестве модели высказывания, соответствующего композиции нечетких ограничений, могут использоваться информационные гранулы [Zadeh, 2006].

В гранулярном представлении НВР элементарная тенденция есть информационная гранула первого порядка, отражающая поведение НВР на единичном интервале изменения некоторой величины. Интенсиал информационной гранулы элементарная тенденция есть ее структурная модель, а экстенсиал образован тенденциями с различными значениями элементов структурной модели, обнаруженных на НВР.

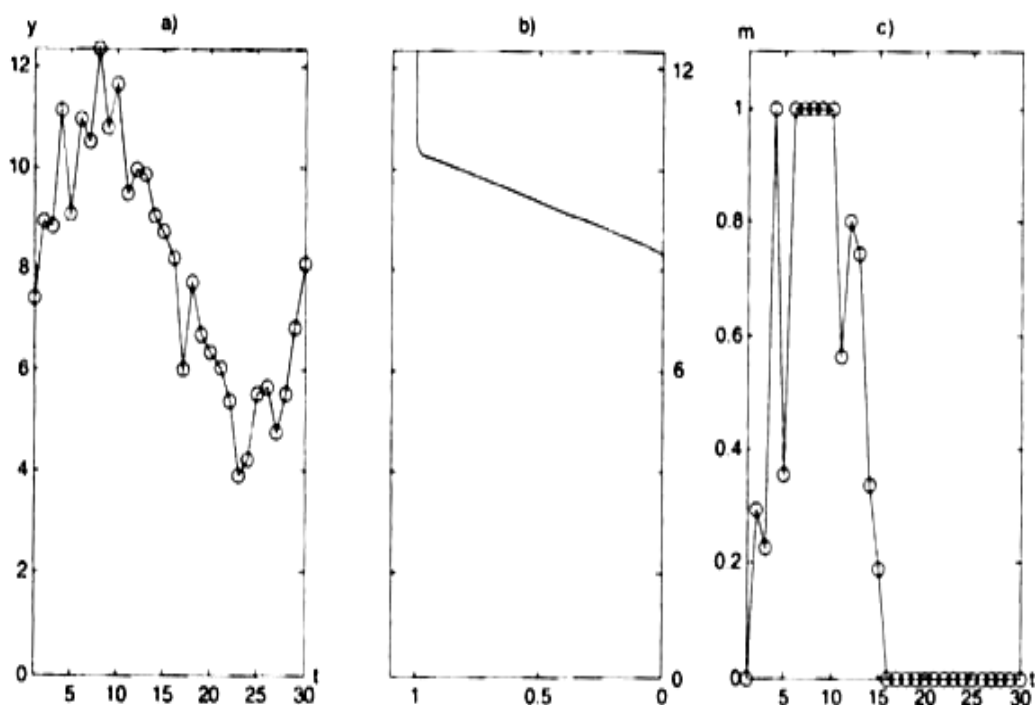


Рис. 4.3. а) Временной ряд значений добычи нефти; б) нечеткое ограничение, определенное на линии значений добычи нефти с помощью высказывания (перцептива) *высокий уровень добычи нефти*; с) нечеткое ограничение дни с высоким уровнем добычи нефти

Локальная тенденция – есть информационная гранула второго порядка, полученная на основе соединения ЭТ по определенному правилу, например, на основе конкатенации соседних элементарных тенденций $N\tau = \{\tau_1, \tau_2, \dots, \tau_m\}$.

На рис. 4.4 приведена классификация нечетких тенденций.

Контрольные вопросы

1. Перечислите свойства нечеткой тенденции.
2. Дайте определение однородным нечетким тенденциям.
3. Определите общее и отличное между элементарной, локальной и общей тенденциями.
4. Приведите гранулярную трактовку нечетких тенденций.

4.3. Модель элементарной тенденции нечеткого временного ряда

Модель элементарной тенденции $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, компоненты которой получены в результате оценивания по ACL-шкале, отнесем к классу *нечетких структурно-лингвистических моделей*.

В случае, если в модели элементарной тенденции $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ компоненты представлены числами, полученными применением операции *DeFuzzy ACL-шкалы* или иным способом, то такую модель отнесем к классу *структурно-лингвистических моделей* [Афанасьева, 2009].



Рис. 4.4. Классификация нечетких тенденций

Определение 4.8. Элементарная тенденция (ЭТ) – это такая нечеткая тенденция НВР $\tau_k \in \mathfrak{T}$, выражающая характер изменения на участке НВР между двумя соседними нечеткими метками НВР $\tilde{x}_i, \tilde{x}_{i+1}$.

Отметим, что типы элементарных тенденций являются базовыми типами нечетких тенденций НВР из множества $\tilde{V}1 = \{\tilde{v}1, \tilde{v}2, \tilde{v}3\}$, $\tilde{v}1 = \text{«Стабилизация»}$, $\tilde{v}2 = \text{«Падение»}$, $\tilde{v}3 = \text{«Рост»}$.

Определения, введенные для нечетких тенденций, справедливы и для элементарной тенденции, за исключением некоторых уточнений, рассмотренных ниже.

Определение 4.9. К противоположным элементарным тенденциям относятся неоднородные элементарные тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{v}_i \neq C$ и $\tilde{v}_s \neq C$.

Тенденция типа $\tilde{v} = C$ противоположной тенденции не имеет.

Определение 4.10. К непротивоположным элементарным тенденциям относятся неоднородные элементарные тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что $(\tilde{v}_i = C \text{ и } \tilde{v}_s \neq C)$ или $(\tilde{v}_i \neq C \text{ и } \tilde{v}_s = C)$.

Определение 4.11. Объединением непротивоположных элементарных тенденций $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, $\tilde{v}_i \neq C$ и $\tilde{v}_s = C$ является тенденция $\tau_j = \tau_i \cup \tau_s$, такая, что $\tau_j = \langle \tilde{v}_j, \tilde{a}_j, \Delta \tilde{t}_j, \mu_j \rangle$, для которой $\tilde{v}_j = \tilde{v}_s$, $\tilde{a}_j = \text{Union}(\tilde{a}_i, \tilde{a}_s)$, $\mu_j = \mu_i \cup \mu_s$, длительность тенденции $\Delta \tilde{t}_j = \text{Union}(\Delta \tilde{t}_i, \Delta \tilde{t}_s)$.

Определение 4.12. Финальной элементарной тенденцией $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ является такая элементарная тенденция, которая образована на последней паре соседних значений $\tilde{x}_{i+1}, x_i$ НВР.

Определение 4.13. Начальной элементарной тенденцией $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ является такая элементарная тенденция, которая образована на первой паре соседних значений $\tilde{x}_{i+1}, x_i$ НВР.

Математическим аппаратом преобразования любого нечеткого временно-го ряда во временной ряд нечетких элементарных тенденций является параметрическая структурно-функциональная модель ACL-шкалы. Приведенные определения позволяет сделать вывод о представимости любого НВР инвариантами в виде последовательности особого класса нечетких тенденций – элементарных тенденций.

Теорема 4.2. Для каждой временной точки нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}, i \in [1, n]$, кроме первой и последней, могут быть определены две элементарные тенденции.

Доказательство. Так как все соседние интервалы, на которых идентифицируются элементарные тенденции, кроме первого и последнего, имеют общие точки, которые для одного интервала являются начальными, а для другого конечными, то эти общие точки одновременно принадлежат двум соседним интервалам НВР. А так как для каждого такого интервала может быть определена отдельная элементарная тенденция, то для таких общих точек могут быть определены две элементарные тенденции. Для начальной и конечной точки НВР отсутствуют соседние точки, а поэтому для них могут быть определены только по одной элементарной тенденции.

При рассмотрении нового объекта – нечеткой элементарной тенденции – на нечетком временном ряду интерес представляет реализация операции сравнения, результат которой будет представлен не только качественно, но и количественно.

Ниже введем для этих целей количественные меры сходства и различия нечетких элементарных тенденций.

Пусть заданы две нечеткие элементарные тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$.

Введем для любых двух ЭТ τ_i и τ_s меру $q: \tau_i \times \tau_s \rightarrow [0,1]$, отражающую степень их сходства, и меру $\rho: \tau_i \times \tau_s \rightarrow [0,1]$, отражающую степень их различия (расстояние между ЭТ). Предположим, что существует следующее соотношение между введенными мерами в виде $q(\tau_i, \tau_s) = 1 - \rho(\tau_i, \tau_s)$, следовательно, будем считать, что при $q(\tau_i, \tau_s) = 0, \rho(\tau_i, \tau_s) = 1$ сходство между τ_i и τ_s отсутствует, при $q(\tau_i, \tau_s) = 1, \rho(\tau_i, \tau_s) = 0$ – полное сходство, и тенденции τ_i и τ_s будут относиться к классу эквивалентных.

Теорема 4.3. Теорема о степени различия нечетких элементарных тенденций. Степень различия двух ЭТ τ_i и τ_s определяется формулой:

$$\rho(\tau_i, \tau_s) = \rho(\tilde{v}_i, \tilde{v}_s) \cdot \sqrt{\rho(\tilde{a}_i, \tilde{a}_s)^2 + \rho(\mu_i, \mu_s)^2}.$$

Доказательство: С целью определения вида функции $\rho(\tau_i, \tau_s)$, рассмотрим типы нечетких элементарных тенденций: противоположные $\tilde{v}_i \neq \tilde{v}_s$, эквивалентные $\tilde{v}_j = \tilde{v}_s$, $\tilde{a}_i = \tilde{a}_s$, $\mu_i = \mu_s$ и однородные $\tilde{v}_j = \tilde{v}_s$ элементарные тенденции.

Будем считать, что противоположные ЭТ являются объективно различными, для них положим $\rho(\tau_i, \tau_s) = \rho(\tilde{v}_i, \tilde{v}_s) = 1$. Для не противоположных элементарных тенденций положим $\rho(\tau_i, \tau_s) = \rho(\tilde{v}_i, \tilde{v}_s) = 0,5$. Для эквивалентных элементарных нечетких тенденций определим максимальную степень сходства $\rho(\tau_i, \tau_s) = \rho(\tilde{v}_i, \tilde{v}_s) = 0$, тогда для однородных элементарных нечетких тенденций справедливо $0 \leq q(\tau_i, \tau_s) \leq 1$, $0 \leq \rho(\tau_i, \tau_s) \leq 1$.

Так как две однородные тенденции τ_i, τ_s могут различаться и по интенсивности $\tilde{a}$, и по степени принадлежности μ , предположив, что $\mu_s = \mu_i$, определим функцию различия $\rho(\tau_i, \tau_s)$ как функцию расстояния интенсивностей ЭТ $\rho(\tilde{a}_i, \tilde{a}_s)$, то есть $\rho(\tau_i, \tau_s) = \rho(\tilde{a}_i, \tilde{a}_s)$, $\mu_s = \mu_i$, $\tilde{v}_j = \tilde{v}_s$.

Функция расстояния интенсивностей ЭТ $\rho(\tilde{a}_i, \tilde{a}_s)$ может быть выражена в виде нечеткой оценки ACL-шкалы, порождаемой операцией разности интенсивностей $\tilde{a}_{is} = \text{Diff}(\tilde{a}_i, \tilde{a}_s)$.

Для формирования приближенного количественного представления функции $\rho(\tilde{a}_i, \tilde{a}_s)$ имеет смысл использовать операцию дефаззификации ACL-шкалы применительно к нечеткой оценке различия в интенсивностях анализируемых элементарных нечетких тенденций τ_i, τ_s , или к самим нечетким значениям интенсивностей, с последующим отображением на отрезок $[0,1]$. Возможны различные варианты приближенного количественного выражения функции различия $\rho(\tilde{a}_i, \tilde{a}_s)$, например:

1. $\rho(\tilde{a}_i, \tilde{a}_s) = \text{Defuzzy}(\text{Diff}(\tilde{a}_i, \tilde{a}_s)) / \text{Defuzzy}(\max(\tilde{a}_i, \tilde{a}_s))$.
2. $\alpha_i = \text{Defuzzy}(\tilde{a}_i)$, $\alpha_i = \alpha_i / \max(\alpha_i, \alpha_s)$, $\alpha_s = \text{Defuzzy}(\tilde{a}_s)$, $\alpha_s = \alpha_s / \max(\alpha_i, \alpha_s)$,

тогда $\rho(\tilde{\alpha}_i, \tilde{\alpha}_s)$ может быть вычислено как расстояние в метрическом пространстве, например, $\rho(\tilde{\alpha}_i, \tilde{\alpha}_s) = |a_i - a_s|$, или на основе других числовых метрик.

Очевидно, что для функции $\rho(\tilde{\alpha}_i, \tilde{\alpha}_s)$ выполнимы все аксиомы числовой метрики, а ее значения удовлетворяют неравенству $0 \leq \rho(\tau_i, \tau_s) \leq 1$.

Таким образом, мера различия однородных элементарных нечетких тенденций задается мерой различия их интенсивностей $\rho(\tilde{\alpha}_i, \tilde{\alpha}_s)$, то есть $\rho(\tau_i, \tau_s) = \rho(\tilde{\alpha}_i, \tilde{\alpha}_s)$, $\mu_s = \mu_i$, $\tilde{v}_j = \tilde{v}_s$, а мера сходства — $q(\tau_i, \tau_s) = 1 - \rho(\tau_i, \tau_s)$.

Далее рассмотрим среди однородных элементарных тенденций такие, у которых интенсивности равны, то есть $\tilde{\alpha}_i = \tilde{\alpha}_s$, а степени принадлежности различны $\mu_s \neq \mu_i$.

Тогда очевидна количественная интерпретация степени различия тенденций $\rho(\tau_i, \tau_s)$ в виде функции над степенями принадлежности тенденций, имеющей смысл функции расстояния: $\rho(\tau_i, \tau_s) = \rho(\mu_i, \mu_s)$.

В качестве функции расстояния для двух элементарных тенденций $\rho(\mu_i, \mu_s)$, степень принадлежности которых выражена числом μ , можно использовать числовые метрики: Эвклида, Хемминга и другие [Юрков и др., 2007; Пивкин и др., 2006].

Другой подход учитывает тот факт, что степень принадлежности μ порождена функцией принадлежности нечеткого множества и может рассматриваться не только как число, но и как одноточечная функция принадлежности. В этом случае функцию расстояния целесообразно выразить посредством операции разности нечетких множеств [Ярушкина, 2004; Яхьяева, 2006]: $\rho(\mu_i, \mu_s) = \mu_i \& (1 - \mu_s) = \min(\mu_i; (1 - \mu_s))$.

Можно использовать для функционального выражения $\rho(\mu_i, \mu_s)$ и тот факт, что минимум из двух функций принадлежности μ_i, μ_s содержит их общую часть, тогда различие в функциях принадлежности μ_i, μ_s может быть выражено в виде $\rho(\mu_i, \mu_s) = \max(\mu_i, \mu_s) - \min(\mu_i, \mu_s)$. Данная мера различия может быть по-

лучена и при интерпретации различия как разницы расстояний нечетких элементарных тенденций τ_i и τ_s до четкой ЭТ, для которой $\mu = 1$.

При одноточечных функциях принадлежности, характерных для нечетких элементарных тенденций, нетрудно заметить, что мера $\rho(\mu_i, \mu_s)$ сводится к формуле вычисления расстояния в метрическом пространстве E^1 : $\rho(\mu_i, \mu_s) = |\mu_i - \mu_s|$. Нетрудно показать, что в связи с тем, что $\mu \in [0, 1]$, то и значение $\rho(\mu_i, \mu_s)$ также принадлежит отрезку $[0, 1]$.

В результате, мера различия двух однородных нечетких элементарных тенденций при условии $\tilde{a}_i = \tilde{a}_s$ может быть выражена следующим образом $\rho(\tau_i, \tau_s) = \rho(\mu_i, \mu_s)$.

Мера различия однородных элементарных нечетких тенденций при $\tilde{a}_i \neq \tilde{a}_s$, $\mu_i \neq \mu_s$ может быть обобщена на двумерный случай и выражена согласно формуле евклидова расстояния в следующем виде:

$$\rho(\tau_i, \tau_s) = F(\rho(\tilde{a}_i, \tilde{a}_s), \rho(\mu_i, \mu_s)) = \sqrt{\rho(\tilde{a}_i, \tilde{a}_s)^2 + \rho(\mu_i, \mu_s)^2}.$$

Модифицируем данную формулу с целью учета в оценке различия двух ЭТ влияния типа тенденции в виде пороговой функции $\rho(\tau_i, \tau_s)$, тогда

$$\rho(\tau_i, \tau_s) = \rho(\tilde{v}_i, \tilde{v}_s) \cdot \sqrt{\rho(\tilde{a}_i, \tilde{a}_s)^2 + \rho(\mu_i, \mu_s)^2}.$$

Понятие различия (сходства) для локальных НТ

Локальные нечеткие тенденции (ЛНТ) есть результат соединения элементарных НТ.

Для удобства будем рассматривать две ЛНТ, содержащие одинаковое количество ЭТ, имеющие одинаковые временные отсчеты. Временным отсчетам анализируемых нечетких тенденций сопоставим целые числа, соответствующие номерам входящих в них элементарных тенденций: $t \in [1, m]$. Такие ЛНТ могут быть получены на основе одного НВР или на основе двух одновременных НВР.

Будем считать, что даны две ЛНТ $N\tau_1 = \{\tau_{11}, \tau_{21}, \dots, \tau_{m1}\}$ и $N\tau_2 = \{\tau_{12}, \tau_{22}, \dots, \tau_{m2}\}$, содержащие m упорядоченных по времени появлений элементарных нечетких тенденций τ_{ij} .

Требуется определить, являются ли $N\tau 1$ и $N\tau 2$ различными и какова степень их различия $\rho(N\tau 1, N\tau 2)$ и степень их сходства $q(N\tau 1, N\tau 2)$.

Естественным подходом к решению задачи определения различия и сходства двух ЛНТ является подход на основе попарного сравнения и вычисления степени различия элементарных тенденций, входящих в состав локальных нечетких тенденций. Можно предложить аддитивную свертку степеней различия элементарных НТ для ЛНТ $N\tau 1$ и $N\tau 2$ на временном отрезке $t \in [1, m]$:

$$\rho(N\tau 1, N\tau 2) = \sum_{t=1}^m \frac{\rho(\tau_{it}, \tau_{st})}{\max(\rho(\tau_{it}, \tau_{st}))}$$

и, соответственно, степень сходства

$$q(N\tau 1, N\tau 2) = 1 - \rho(N\tau 1, N\tau 2),$$

где $\rho(\tau_{it}, \tau_{st}) = \rho(\tilde{v}_{it}, \tilde{v}_{st}) \cdot \sqrt{\rho(\tilde{a}_{it}, \tilde{a}_{st})^2 + \rho(\mu_{it}, \mu_{st})^2}$ – согласно теореме о степени различия элементарных нечетких тенденций.

Контрольные вопросы

1. Изобразите графическое представление классификации нечетких тенденций.
2. Сформулируйте определение эквивалентных и противоположных нечетких тенденций.
3. Приведите определение однородных нечетких тенденций.
4. Сформулируйте теорему о степени различия нечетких элементарных тенденций.
5. Дайте понятие различия (сходства) нечетких тенденций.

4.4. FT-декомпозиция нечеткого временного ряда в базисе элементарных тенденций

В дальнейшем для однозначности будем позиционировать оценку тенденции с началом временного интервала, на котором она определена. Это позволит декомпозировать НВР на несколько ВР, соответствующих компонентам

модели нечеткой тенденции. Так, для структурно-лингвистической модели элементарной тенденции в результате декомпозиции НВР получим временной ряд типов элементарных тенденций $X_v = \{t_i, v_i\}$, временной ряд степени принадлежности элементарных тенденций $X_\mu = \{t_i, \mu_i\}$ и для каждого типа элементарной тенденции v_k временные ряды интенсивностей $X_a = \{t_i, a_i\}$ и временные ряды длительностей $X_{\Delta t} = \{t_i, \Delta t_i\}$, где $i \in [1, n-1]$, n – количество членов нечеткого ряда.

Назовем декомпозицию исходного ВР в совокупность временных рядов компонент элементарной тенденции *FuzzyTend-декомпозицией* или *FT-декомпозицией* [Афанасьева, 2008 б].

Процедура *FT-декомпозиции* исходного временного ряда X может быть позиционирована в последовательности преобразований временных рядов следующим образом:

$$X \rightarrow Fuzzy(X) \rightarrow \tilde{X} \rightarrow FT(\tilde{X}, E, T) \rightarrow \{X\}_m \rightarrow Func(\{X\}_m) \rightarrow RFunc,$$

где $\{X\}_m$ обозначает множество временных рядов, полученных в результате декомпозиции, *Func* определяет процедуру решения конкретной задачи анализа НВР на основе НТ, например, идентификацию нечетких тенденций в виде общей тенденции. К результатам *RFunc* такого анализа можно также отнести кодифицированные знания о природе изменений исходного ВР, выраженные в структуре, интенсивности и длительности тенденций. С другой стороны, знания о природе изменений и общей тенденции могут быть использованы не только для идентификации НТ временного ряда, но и для прогноза параметров его будущих тенденций.

Контрольные вопросы

1. Что такое FuzzyTend-декомпозиция?
2. Какова цель процедуры FT-декомпозиции?

4.5. Структурно-лингвистическая модель нечеткого временного ряда в базисе элементарной тенденции

Data Mining временных рядов – быстро развивающаяся область исследований, основанная на использовании интеллектуального анализа хранимых данных. Для хранения временных рядов используются хранилища и базы данных, в основе которых лежит реляционная модель информационных объектов. В соответствии с этим рассмотрим представления временного ряда и его нечетких моделей на основе реляционной модели информационного объекта баз данных.

Для этих целей введем в рассмотрение структурную модель временного ряда (ВР), являющегося информационным носителем НВР.

Определение 4.14. Структурную модель носителя нечеткого временного ряда определим как реляционную модель в виде отношения

$$Y = \langle T, X \rangle,$$

где T – домен времени, задаваемый упорядоченными по возрастанию моментами времени;

X – домен, хранящий уровни ВР.

Нечеткой метке $\tilde{x}_i$ сопоставим абсолютную лингвистическую оценку, полученную по лингвистической ACL -шкале S_x , построенной на универсальном множестве X .

Определение 4.15. Нечеткий временной ряд $\tilde{Y}_x$ по атрибуту X есть результат лингвистического оценивания по ACL -шкале уровней X носителя – временного ряда Y . Структурно-лингвистическая модель уровней нечеткого временного ряда $\tilde{Y}_x$ есть расширение отношения Y следующего вида

$$Y_x = \langle X, \tilde{X}, \mu_{\tilde{X}}(X) \rangle,$$

где $\tilde{X}$ – задает абсолютную лингвистическую оценку по ACL -шкале уровней;

$\mu_{\tilde{X}}(X)$ – степень принадлежности X лингвистической метке $\tilde{X}$.

Определение 4.16. Нечеткий временной ряд $\tilde{Y}_T$ по атрибуту T есть результат лингвистического оценивания по *ACL-шкале* моментов времени S_T носителя – временного ряда Y . Структурно-лингвистическая модель моментов времени нечеткого временного ряда $\tilde{Y}_T$ есть расширение отношения Y следующего вида

$$Y_T = \langle T, \tilde{T}, \mu_{\tilde{T}}(T) \rangle,$$

где $\tilde{X}$ – задает абсолютную лингвистическую оценку по *ACL-шкале* моментов времени;

$\mu_{\tilde{T}}(T)$ – степень принадлежности T лингвистической метке $\tilde{T}$.

Определение 4.17. Нечеткий временной ряд $\tilde{Y}$ есть результат реляционной операции соединения

$$\tilde{Y} = Join(Y, \tilde{Y}_x, \tilde{Y}_T)$$

определенных выше отношений Y , $\tilde{Y}_x$ и $\tilde{Y}_T$, порождающей структурно-лингвистическую модель *НВР*

$$\tilde{Y} = \langle T, X, \tilde{T}, \mu_{\tilde{T}}(T), \tilde{X}, \mu_{\tilde{X}}(X) \rangle.$$

Отметим, что в этом отношении домен моментов времени T является ключевым.

Каждому моменту времени в структурно-лингвистической модели *НВР* можно сопоставить нечеткую тенденцию τ , определяемую на основе «оценочных» операций *ACL-шкалы* уровней X .

Определение 4.18. Модель нечеткой тенденции есть отношение

$$\tau = \langle \nu, \mu, \alpha, \Delta t \rangle,$$

где τ – наименование нечеткой тенденции;

ν – тип нечеткой тенденции, определяемый на основе операции *TTend* *ACL-шкалы* уровней X . Последовательность типов нечетких тенденций моделирует структуру изменений *НВР*;

μ – степень принадлежности нечеткой тенденции *НВР*, которая формируется на основе образующих ее нечетких значений уровней;

α – интенсивность нечеткой тенденции, контекстное расширение тенденции, определяемое операцией $RTend\ ACL$ -шкалы уровней X ;

Δt – длительность данного типа нечеткой тенденции.

Используя введенную модель, определим временной ряд нечетких тенденций.

Определение 4.19. *Временной ряд нечетких тенденций*, построенный на нечетком временном ряду $\tilde{Y}$, есть отношение

$$\tilde{Y}_\tau = \langle T, \tau \rangle,$$

где $\tau = \langle \nu, \mu, \alpha, \Delta t \rangle$.

Определение 4.20. *FT-расширение* структурно-лингвистической модели НВР в базисе нечетких тенденций есть *расширенная структурно-лингвистическая модель НВР* $\tilde{Y}'$, полученная на основе реляционной операции соединения

$$\tilde{Y}' = Join(\tilde{Y}, \tilde{Y}_\tau).$$

Расширенная структурно-лингвистическая модель НВР относится к классу реляционных моделей представления данных, для которого применимы операции реляционной алгебры: объединение, пересечение, выборка, проекция, соединение, вычитание, декартово произведение. Рассмотрим проекции модели НВР и соответствующие им проекционные модели НВР [Афанасьева, 2009]:

1. Проекция расширенной модели $\tilde{Y}'$ по моментам времени и типам НТ есть отношение

$$Pr1 = Proj_{T, \nu}(\tilde{Y}'),$$

представляющее временной ряд типов нечетких тенденций. Этот ВР назовем *структурной моделью нечеткой тенденции НВР*, а процедуру получения структурной модели – *структурной идентификацией модели нечеткой тенденции НВР*. Эта структурная модель может быть описана на основе формальных методов как последовательность слов некоторого языка тенденций для сокращения набора нечетких правил, для генерации этих правил, используемых при решении задач анализа и резюмирования НВР, для оценки прогноза в тер-

минах типов тенденций, для определения подобных, различных и эквивалентных НВР.

2. Проекция расширенной модели $\tilde{Y}'$ по моментам времени и степени нечеткости НТ есть отношение

$$Pr2 = Proj_{T,\mu}(\tilde{Y}'),$$

моделирующее в виде временного ряда функцию принадлежности нечеткой тенденций НВР. Эта функция, являясь *параметром структурной модели тенденции НВР*, в дальнейшем может использоваться как нечеткая мера тенденции НВР.

3. Проекция расширенной модели $\tilde{Y}'$ по моментам времени и степени интенсивности НТ при фиксированном типе нечеткой тенденции ν (ν -фильтр) есть отношение

$$Pr3 = Proj_{T,\alpha,\nu}(\tilde{Y}'),$$

моделирующее в виде временного ряда функцию выраженности соответствующего типа тенденции ν . Эта функция, являясь *параметром структурной модели тенденции НВР*, в дальнейшем может использоваться для определения интегральной характеристики соответствующего типа тенденции.

4. Проекция расширенной модели $\tilde{Y}'$ по нечетким меткам моментов времени и нечетким меткам НВР есть отношение

$$Pr4 = Proj_{\tilde{T},\tilde{X}}(\tilde{Y}'),$$

моделирующее НВР в гранулярном представлении лингвистических меток НВР.

5. Проекция расширенной модели $\tilde{Y}'$ по нечетким меткам моментов времени и нечетким тенденциям НВР есть отношение

$$Pr5 = Proj_{T,\nu,\mu,\alpha,\Delta t}(\tilde{Y}'),$$

представляющее *структурно-параметрическую модель нечеткой тенденции нечеткого временного ряда по моментам времени $\tilde{T}$. Структурно-параметрической идентификацией модели нечеткой тенденции НВР* назовем процедуру получения проекции $Pr5$. Структурно-параметрическая модель нечеткой тенденции НВР по моментам времени $\tilde{T}$ образует гранулярное представление НВР в лингвистических терминах нечетких тенденций.

6. Все проекционные модели $Pr1, Pr2, Pr3, Pr4, Pr5$, образующие новые временные ряды, могут быть использованы для решения задач Time Series Data Mining, таких как сегментации НВР, прогноза по отдельным компонентам модели тенденций, и в целом, для резюмирования НВР, для определения сходства и различий между несколькими НВР в задачах контроля и диагностики. В таблице 4.5 приведены варианты применения обозначенных проекционных моделей, полученных на основе структурно-лингвистической модели НВР для решения задач направления Time Series Data Mining.

7. Результаты решения указанных задач на основе проекционных моделей, выраженные в терминах тенденций, могут быть преобразованы в лингвистические метки НВР и числовые уровни ВР на основе обратной операции FT^1 -расширения.

Таблица 4.5

Связь проекционных моделей ВР и задач Time Series Data Mining

Задачи	$Pr1$	$Pr2$	$Pr3$	$Pr4$	$Pr5$
Сегментация ВР	+	+		+	
Резюмирование		+			+
Прогноз	+	+	+		
Кластеризация		+		+	
Классификация	+			+	
Частотный анализ	+		+	+	
Поиск аномалий	+	+	+	+	
Извлечение ассоциативных правил	+	+	+	+	

Основные положения структурно-лингвистического подхода

1. Терминальным символам грамматики языка $LANG$ сопоставим нечеткие уровни и элементарные нечеткие тенденции $v_T = \tilde{X} \cup \tilde{Z}$ нечеткого ВР $\tilde{Y}$.

2. Лексика грамматики языка *LANG* определяется локальными нечеткими тенденциями $V_N = N\tau$ НБР. Аксиома грамматики есть общая тенденция $S = G\tau$.

3. Синтаксис грамматики языка *LANG* зададим в виде продукционных правил следования нечетких тенденций $P = Rule_et$, определяемых на НБР как результат зависимости нечетких тенденций от значений тенденций в предыдущие моменты времени: «ЕСЛИ тенденция(i), ТО затем тенденция (j)».

4. Семантике лингвистических выражений языка *LANG* сопоставим нечеткие множества $\tilde{X}, \tilde{V}, \tilde{A}$, задаваемые функциями принадлежности в ACL-шкале.

5. Решение задач интеллектуального анализа временных рядов (TSDM), определенных в главе 4 как задачи извлечения знаний из временных рядов, сведем к решению задач анализа и синтеза грамматики языка *LANG*:

- Сегментация – представление НБР в виде последовательности терминальных символов (нечетких тенденций) грамматики языка *LANG*.
- Кластеризация – построение лексики языка *LANG*: образование локальных нечетких тенденций.
- Поиск ассоциативных правил – определение синтаксиса языка *LANG* на основе извлечения синтаксических правил следования нечетких тенденций.
- Классификация – распознавание общей нечеткой тенденции НБР как вывода нечеткой аксиомы грамматики языка *LANG*.

Используем сгенерированную грамматику в режиме порождения для построения гранулярного представления НБР и решения следующих задач извлечения знаний из временных рядов:

- Частотный анализ – формирование компонент гранулярного описания НБР в виде часто встречающихся нечетких тенденций и их паттернов.
- Прогнозирование – получение многоуровневого прогноза: в терминах общей, локальной, элементарной тенденций, в терминах нечетких и числовых уровней ВР.

- Резюмирование – генерация гранулярного описания поведения НВР, определение семантики гранул и трансляция гранулярного описания НВР в виде предложения на естественном языке.
- Поиск аномалий – выявление лексических, синтаксических и семантических ошибок и нетипичных нечетких объектов и правил на основе грамматики языка *LANG*.

Таким образом, структурно-лингвистический подход нацелен на решение комплекса задач интеллектуального анализа временных рядов на основе нечеткого моделирования ВР и анализа нечетких тенденций.

Контрольные вопросы

1. Дайте определение структурной модели носителя нечеткого временного ряда.
2. Дайте определение расширенной структурно-лингвистической модели временного ряда.
3. Дайте формальное определение модели нечеткой тенденции.
4. Приведите соотношение проекционным моделям задачам Data Mining.
5. Приведите основные положения структурно-лингвистического подхода.

Библиографический список

1. [Batyrshin, 2007] Batyrshin, I. Perception Based Time Series Data Mining for Decision Making / I. Batyrshin // IFSA'07 Fuzzy Logic, Soft Computing and Computational Intelligence.
2. [Zadeh, 2006] Zadeh, Lotfi A. Generalized theory of uncertainty (GTU) – principal concepts and ideas / Lotfi A. Zadeh // Computational statistic & Data analysis. – 2006. – №51. – P. 15-46.
3. [Анфилов, 2003] Анфилов, В. С. Системный анализ в управлении : учеб. пособие / В. С. Анфилов, А. А. Емельянов, А. А. Кукушкин ; Под ред. А. А. Емельянова. – М. : Финансы и статистика, 2003. – 368 с.

4. [Афанасьева, 2008 а] Афанасьева, Т. В. Модель ACL-шкалы для генерации лингвистических оценок в принятии решений / Т. В. Афанасьева // Вопросы современной науки и практики. Университет им. В. И. Вернадского. Т.2. Серия «Технические науки» – 2008. – №4(14). – С. 91-97.
5. [Афанасьева, 2008 б] Афанасьева, Т. В. FT-преобразование нечетких временных рядов / Т. В. Афанасьева // Нечеткие системы и мягкие вычисления (НСМВ-2008) : сборник научных трудов второй всероссийской научной конференции с международным участием (г. Ульяновск, 27-29 октября, 2008 г.). В 2 т. Т.1. – Ульяновск : УлГТУ, 2008. – С. 122-126.
6. [Афанасьева, 2009] Афанасьева, Т. В. Модель элементарной тенденции нечеткого временного ряда / Т. В. Афанасьева // Труды Седьмой Международной конференции «Математическое моделирование физических, экономических, технических, социальных систем и процессов», 2-5 февраля 2009 года, г. Ульяновск / под ред. д. т. н., проф. Ю. В. Полянского, д. ф.-м. н., проф. В. Л. Леонтьева. – Ульяновск : УлГУ, 2009. – С. 35-37.
7. [Афанасьева и др., 2009] Афанасьева, Т. В. Моделирование лингвистических оценок на основе ACL-шкалы / Т. В. Афанасьева, А. О. Козлов, А. А. Ивахина // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). – Т.2. – М. : Физматлит, 2009. – С. 298-304.
8. [Пивкин и др., 1997] Пивкин, В. Я. Нечеткие множества в системах управления / В. Я. Пивкин, Е. П. Бакулин, Д. И. Кореньков [доступно по адресу <http://works.tarefer.ru/46/100085/index.html>; дата обращения 30.12.2009].
9. [Юрков и др., 2007] Юрков, В. Ю. Нечеткая геометрия как модель и средство развития визуального мышления / В. Ю. Юрков, О. В. Лукина // Наука в образовании. Электронное научное издание, 2007. – №3.
10. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.

11. [Яхьяева, 2006] Яхьяева Г. Э. Нечеткие множества и нейронные сети : учебное пособие / Г. Э. Яхьяева. – М. : Интернет-Университет Информационных технологий: БИНОМ. Лаборатория знаний, 2006. – 316 с.

ГЛАВА 5. СТРУКТУРНО-ЛИНГВИСТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ НЕЧЕТКИХ ТЕНДЕНЦИЙ ВРЕМЕННЫХ РЯДОВ

Введение

В настоящей главе рассматриваются в рамках направления Time Series Data Mining (TSDM) основы нового структурно-лингвистического подхода для извлечения знаний о поведении временных рядов, выраженных в форме зависимостей нечетких тенденций.

Знания о нечетких тенденциях, извлекаемые из временных рядов, будут представлены в виде нечетких моделей и базы нечетких продукционных правил, допускающих как лингвистическую интерпретацию, так и числовую оценку уровней ВР. Они могут дополнять знания, получаемые в различных подходах к построению моделей временных рядов, а в случаях трудностей при их построении служить средством «интеллектуальной поддержки» специалистов различной квалификации.

При разработке структурно-лингвистического подхода были использованы структурно-функциональная модель ACL-шкалы, FT-преобразование временного ряда в нечеткий временной ряд и модель нечетких тенденций, рассмотренные в главе 4.

Основная идея структурно-лингвистического подхода заключается в идентификации нечеткой грамматики языка *LANG*, задающей форму представления знаний о поведении ВР в виде нечетких продукционных правил, при этом исследуемый временной ряд рассматривается как предложение на этом языке.

5.1. Структурно-лингвистический подход при анализе временных рядов

Язык, обычно используемый для представления модели временного ряда, подчиняется правилам формальной грамматики языка математических формул и правилам семантической интерпретации, позволяющим на основе исходных числовых значений получать результирующие числовые значения. Модель синтаксиса такого языка представима формальной грамматикой

$$G = \langle V_N, V_T, P, S \rangle,$$

где V_N – множество нетерминальных символов;

V_T – множество терминальных символов;

S – аксиома, начальный символ $S \in V_N$;

P – конечное множество правил подстановки (продукций) вида $u \xrightarrow{p} v$; $u, v \in (V_N \cup V_T)$, $p \in P$, позволяющих определять синтаксически правильные цепочки терминальных символов.

Множество терминальных символов языка формул образовано символами алфавита, описывающими переменные, числа, знаки операций, разделители. Множество нетерминальных символов применительно к аналитической форме модели временных рядов включает понятия «идентификатор», «выражение», «процедура».

Универсальное представление модели временного ряда в форме аналитической зависимости – удобное средство, позволяющее на основе математически определенных и программно реализованных элементов языка формул (его синтаксиса и семантики) и математических методов обработки ВР определять числовые значения временного ряда в любой заданный момент времени.

Преобразование временного ряда в нечеткий временной ряд с помощью нечеткой *ACL-шкалы* позволит уже на начальной стадии построения модели ВР использовать качественные оценки и предметно-ориентированные знания. Результаты анализа нечеткого временного ряда, полученные на основе нечетких

моделей, формируются в терминах качественных характеристик, выраженных лингвистически, допускающих числовую интерпретацию, обладают большей информативностью, понятны специалистам и удобны при принятии управленческих решений, в проектировании, контроле, диагностике, медицине и т. д.

В связи с этим сформулируем задачу анализа временного ряда как задачу анализа нечеткого временного ряда (НВР) для идентификации модели поведения НВР, порождающую проблему определения языка представления структуры и параметров этой модели.

Так как уровни нечеткого временного ряда представимы значениями лингвистической переменной, то естественно представить структуру НВР как цепочку семантически определенных лингвистических элементов, терминальных символов V_T грамматики некоторого языка $LANG$. Отношения следования (предшествования) между этими термами, обнаруживаемые на исследуемом нечетком временном ряду, выражают синтаксические правила, правила подстановки P . Промежуточные структуры, составленные из последовательности термов, образуют нетерминальные символы V_N искомой грамматики, которые на основе свертки порождают аксиому грамматики S .

Параметрами рассматриваемой структурной модели могут выступать нечеткие ограничения, представимые функциями принадлежности, а также и числовые значения.

Таким образом, при рассмотрении проблемы определения языка $LANG$ для представления модели исследуемого НВР интерес представляет решение следующих задач:

1. Возможно ли свести задачу структурной идентификации НВР к задаче определения синтаксиса грамматики некоторого языка $LANG$?
2. Как определить семантические правила «вычисления» новых значений в языке $LANG$?

3. Какие информационные технологии и математические методы наиболее эффективны при реализации языка представления модели НБР?
4. Какие задачи анализа НБР могут быть определены на основе языка *LANG*?

Сформулированная выше совокупность задач позволяет обозначить структурно-лингвистический подход в решении задачи анализа ВР как задачи структурно-параметрической идентификации модели поведения НБР, задаваемой грамматикой языка *LANG*, выраженной в форме лингвистических зависимостей понятных прикладному пользователю [Афанасьева, 2008].

Одной из первых задач при определении языка *LANG* и его грамматики является определение терминальных символов. Так как эта задача касается языка, предназначенного для описания модели поведения нечеткого временного ряда, то целесообразно в качестве терминальных символов языка *LANG* выбрать понятия, обозначающие нечеткие объекты НБР: нечеткие уровни, нечеткие тенденции, нечеткие временные интервалы.

Основные положения структурно-лингвистического подхода в анализе НБР

Рассмотрим предложенную в главе 4 расширенную структурно-лингвистическую модель НБР при условии нечетких уровней и четких моментов времени:

$$\tilde{Y} = \langle T, X, \tilde{X}, \mu_{\tilde{X}}(X), \tau \rangle,$$

где $\tau = \langle \tilde{v}, \mu, \tilde{\alpha}, \Delta t \rangle$,

T – атрибут времени, задаваемый упорядоченными по возрастанию моментами времени;

X – атрибут, хранящий уровни ВР;

$\tilde{X}$ – задает множество лингвистических термов, обозначающих нечеткие уровни ВР;

$\mu_{\tilde{X}}(X)$ – степень принадлежности X лингвистическому терму $\tilde{X}$;

τ – наименование нечеткой тенденции;

$\tilde{v}$ – тип нечеткой тенденции, определяемый на основе операции *TTend ACL-шкалы*. Последовательность типов нечетких тенденций моделирует структуру изменений НБР;

$\tilde{a}$ – интенсивность нечеткой тенденции, контекстное расширение тенденции, определяемое операцией *RTend ACL-шкалы*;

μ – степень принадлежности нечеткой тенденции нечеткому временному ряду, при $\mu=1$ нечеткая тенденция рассматривается как четкая тенденция;

Δt – длительность данного типа нечеткой тенденции.

Сформулируем основные положения структурно-лингвистического подхода:

1. Терминальным символам грамматики языка *LANG* сопоставим нечеткие уровни и элементарные нечеткие тенденции $v_T = \tilde{X} \cup \tilde{\tau}$ нечеткого ВР $\tilde{Y}$.

2. Лексика грамматики языка *LANG* определяется локальными нечеткими тенденциями $V_N = N\tau$ НБР. Аксиома грамматики есть общая тенденция $S = G\tau$.

3. Синтаксис грамматики языка *LANG* зададим в виде продукционных правил следования нечетких тенденций $P = Rule_et$, определяемых на НБР как результат зависимости нечетких тенденций от значений тенденций в предыдущие моменты времени: «ЕСЛИ тенденция(i), ТО затем тенденция (j)».

4. Семантике лингвистических выражений языка *LANG* сопоставим нечеткие множества $\tilde{X}, \tilde{V}, \tilde{A}$, задаваемые функциями принадлежности в ACL-шкале.

5. Решение задач интеллектуального анализа временных рядов (TSDM), определенных в главе 4 как задачи извлечения знаний из временных рядов, сведем к решению задач анализа и синтеза грамматики языка *LANG*:

- Сегментация – представление НБР в виде последовательности терминальных символов (нечетких тенденций) грамматики языка *LANG*.
- Кластеризация – построение лексики языка *LANG*: образование локальных нечетких тенденций.

- Поиск ассоциативных правил – определение синтаксиса языка *LANG* на основе извлечения синтаксических правил следования нечетких тенденций.

- Классификация – распознавание общей нечеткой тенденции НВР как вывода нечеткой аксиомы грамматики языка *LANG*.

Используем сгенерированную грамматику в режиме порождения для построения гранулярного представления НВР и решения следующих задач извлечения знаний из временных рядов:

- Частотный анализ – формирование компонент гранулярного описания НВР в виде часто встречающихся нечетких тенденций и их паттернов.

- Прогнозирование – получение многоуровневого прогноза: в терминах общей, локальной, элементарной тенденций, в терминах нечетких и числовых уровней ВР.

- Резюмирование – генерация гранулярного описания поведения НВР, определение семантики гранул и трансляция гранулярного описания НВР в виде предложения на естественном языке.

- Поиск аномалий – выявление лексических, синтаксических и семантических ошибок и нетипичных нечетких объектов и правил на основе грамматики языка *LANG*.

Таким образом, структурно-лингвистический подход нацелен на решение комплекса задач интеллектуального анализа временных рядов на основе нечеткого моделирования ВР и анализа нечетких тенденций.

Далее рассмотрим подходы к решению вышеперечисленных задач.

Контрольные вопросы

1. Перечислите задачи Data Mining, которые возникают при исследовании модели НВР на основе структурно-лингвистического подхода.
2. Сформулируйте основные положения структурно-лингвистического подхода.

3. Как моделируется структура нечеткого ВР?

5.2. Сегментация НВР в терминах нечетких тенденций

Традиционно сегментация (дискретизация) ВР выполняется методом скользящего окна заданной ширины k на Y .

На основе введенной классификации нечетких тенденций в главе 4 рассмотрим особенности сегментации при извлечении знаний из нечетких временных рядов для нечетких элементарных тенденций (ЭТ) и тенденций, образованных как последовательности элементарных нечетких тенденций: локальных однородных и неоднородных нечетких тенденций. Последние образуют класс производных нечетких тенденций (НТ).

5.2.1. Алгоритм сегментации НВР в терминах элементарных тенденций

Рассмотрим применение структурно-лингвистического подхода к решению задачи сегментации НВР в терминах элементарных тенденций.

Временной ряд элементарных тенденций (ВРЭТ) позволяет каждому временному полуинтервалу единичной длины сопоставить нечеткую элементарную тенденцию.

Для построения ВРЭТ введем функционал $ETend: \tilde{X} \times \tilde{X} \rightarrow \mathfrak{T}$, такой, что $\tau_t = ETend(\tilde{X}_t, \tilde{X}_{t+1})$ при изменении t на отрезке $[1..n-1]$ с шагом 1, $\mathfrak{T} = \{\tau_t\}$, где $\tau_t = (\tilde{v}_t, \mu_t, \tilde{a}_t)$.

В соответствии с моделью элементарной тенденции временной ряд элементарной тенденции представляется в виде совокупности одновременных временных рядов, определяемых проекциями расширенной структурно-лингвистической модели ВР, введенной в главе 4.

Функционал $ETend$ реализуется на основе операций $TTend$ и $RTend$ ACL-шкалы, построенной для исходного ВР (алгоритм построения шкалы см. в главе 4).

Ниже представлено описание алгоритма сегментации НВР в терминах нечетких ЭТ.

Исходные данные. Нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}, i \in [1, n]$, где $\tilde{x}_i$ – нечеткое множество, определенное для значения x_i некоторого временного ряда $Y = \{t_i, x_i\}$, t_i – значение момента времени;

n – количество членов НВР.

Выходные данные. Временной ряд элементарных тенденций как совокупность одновременных временных рядов, образованных нечеткими уровнями и параметрами нечетких элементарных тенденций:

$$\{t_i, x_i, \tilde{x}_i, \tilde{v}_i, \tilde{a}_i, \mu_i\},$$

где $\{t_i, x_i\}$ – временной ряд, носитель нечеткого временного ряда;

$\tilde{x}_i$ – название нечеткого уровня НВР;

$\tilde{v}_i, \mu_i$ – тип тенденции между $\tilde{x}_i, \tilde{x}_{i+1}$ и его степень принадлежности;

$\tilde{a}_i$ – интенсивность типа тенденции $\tilde{v}_i$.

Алгоритм будем строить на основе предварительного сглаживания НВР, применив для этого метод F-преобразования [Perfilieva, 2006].

1. $x_i \rightarrow x'_i$, где x'_i получается с помощью F-преобразования.
2. Определение погрешности $\Delta x_i = ErI(x_i, x'_i)$. Эта функция может быть вычислена одним из стандартных способов, например, с использованием критерия среднеквадратичного отклонения (СКО).
3. Определение остатков $\Delta x_i = x_i - x'_i$.
4. Таким образом, получаем фрагмент выходной модели $\{t_i, x_i, x'_i, \Delta x_i\}$.

5. Вычисление абсолютной нечеткой оценки для x'_i на основе операции $Fuzzy: X \rightarrow \tilde{X}$ ACL-шкалы и определение нечеткого множества $\tilde{x}_i$ с максимальной степенью принадлежности: $\mu_i = \max(\mu_{\tilde{x}_j}(x'_i))$, $\tilde{x}_j \in \tilde{X}$, $j = [1, m]$.
Таким образом, получаем $\{t_i, x_i, x'_i, \Delta x_i, \tilde{x}_i, \mu_i\}$.
6. Далее анализируем пары соседних значений $\tilde{x}_i, \tilde{x}_{i+1}$. Этот анализ осуществляем на базе отношений различия ACL-шкалы: $\tilde{v}_i = TTend(\tilde{x}_i, \tilde{x}_{i+1})$ для определения типа тенденции и $\tilde{a}_i = RTend(\tilde{x}_i, \tilde{x}_{i+1})$ для определения интенсивности тенденции.
7. Используем для этого нечеткие правила типа:

ЕСЛИ x_i есть $\tilde{x}_{11}$ и x_{i+1} есть $\tilde{x}_{12}$ то $\tilde{v}_i$ есть $TTend(\tilde{x}_{11}, \tilde{x}_{12})$,

ЕСЛИ x_i есть $\tilde{x}_{11}$ и x_{i+1} есть $\tilde{x}_{12}$ то $\tilde{a}_i$ есть $RTend(\tilde{x}_{11}, \tilde{x}_{12})$.

Так получаем выходную структурно-лингвистическую модель НВР $\{t_i, x_i, \tilde{x}_i, \tilde{v}_i, \tilde{a}_i, \mu_i\}$, а при дефаззифицированных значениях интенсивностей $\tilde{a}_i$ элементарных нечетких тенденций ($\alpha_i = DeFuzzy(\tilde{a}_i)$) в виде:

$$\{t_i, x_i, \tilde{x}_i, \tilde{v}_i, a_i, \mu_i\}.$$

Результат сегментации НВР при дефаззифицированных значениях лингвистических меток параметров ЭТ представим в виде одновременных ВР, полученных на основе проекций $Pr2, Pr3$ структурно-лингвистической модели ВР (ν -фильтр) (рис. 5.1).

Структурная модель сегментированного НВР в базисе элементарных тенденций, полученная с помощью ACL-шкалы, приведена в таблице 5.1.

Таблица 5.1

Нечеткие характеристики ряда

Время t_i	Ряд x_i	ФП μ_i	НМ $\tilde{x}_i$	ТТend $\tilde{v}_i$	RTend $\tilde{a}_i$
1971	13055	1,00	A0	Рост	Небольшой
1972	13563	0,76	A1	Стабильность	Отсутствует
1973	13847	0,88	A1	Рост	Небольшой
1974	14696	0,68	A2	Рост	Небольшой



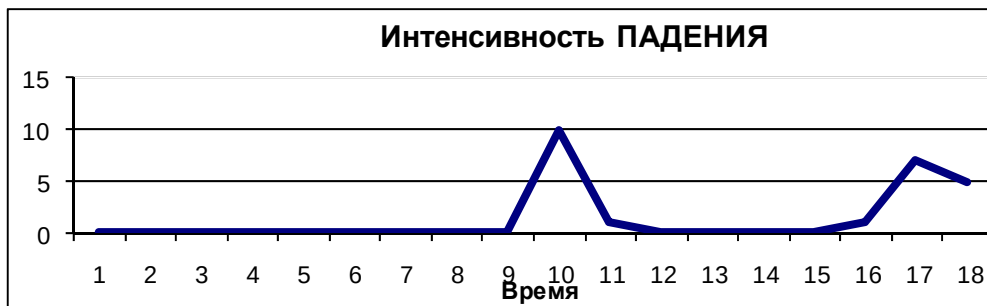


Рис. 5.1. Результат сегментирования НВР в терминах элементарных тенденций

5.2.2. Алгоритм сегментации НВР в терминах локальных однородных нечетких тенденций

Процедура сегментации нечеткого временного ряда в терминах локальных однородных нечетких тенденций зависит от типа локальных тенденций. В ситуации, если тип локальной тенденции относится к базовым, то есть кодируется нечеткими метками «Рост», «Падение», «Стабильность», то сегментация в терминах локальных однородных тенденций основывается на результатах сегментации НВР в терминах элементарных НТ, представленной в предыдущем разделе. Другая ситуация связана с сегментацией НВР в терминах локальных НТ, тип которых относится к классу производных НТ.

Сегментация НВР в терминах локальных однородных НТ с базовым типом

Отметим, что на основе утверждения о представимости временного ряда в виде временного ряда элементарных тенденций существует объективная возможность представления временного ряда в виде временного ряда локальных НТ с базовыми типами «Рост», «Падение», «Стабильность». Поскольку понятию локальной однородной НТ соответствует последовательность элементарных однородных тенденций $N\tau = \{\tau_1, \tau_2, \dots, \tau_m\}$, выполним сегментацию НВР в терминах локальных однородных НТ с базовым типом на основе операции конкатенации соседних однородных элементарных тенденции. Введем функционал

$STend: \mathfrak{T} \times \mathfrak{T} \rightarrow N\mathfrak{T}$, вычисляющий результат объединения однородных элементарных тенденций, такой что $N\tau_i = STend(\tau_i, \tau_{i+1})$.

Введенный функционал $STend$ определяет семантику результата объединения однотипных элементарных тенденций на основе «вычислительной» операции объединения интенсивностей $Union$ ACL-шкалы и свертки последовательности элементарных тенденций в структурную модель нечеткой тенденции $\tau = \langle \tilde{v}, \tilde{a}, \Delta \tilde{t}, \mu \rangle$. При этом объединяются только нечеткие тенденции с одинаковыми типами.

Обобщенная структура системы правил вычисления интенсивностей $\tilde{a}$ объединенной локальной тенденции имеет вид:

$$\begin{array}{l}
R_{11} := \text{IF } v_t \text{ is Inc THEN} \\
\quad \text{IF } a_t \text{ is } A_{11} \text{ AND } a_{t+1} \text{ is } A_{12} \text{ THEN } TInc \text{ is } B_1 \\
\text{.....} \\
R_{m1} := \text{IF } v_t \text{ is Inc THEN} \\
\quad \text{IF } a_t \text{ is } A_{m1} \text{ AND } a_{t+1} \text{ is } A_{m2} \text{ THEN } TInc \text{ is } B_m, \\
R_{12} := \text{IF } v_t \text{ is Dec THEN} \\
\quad \text{IF } a_t \text{ is } A_{11} \text{ AND } a_{t+1} \text{ is } A_{12} \text{ THEN } TDec \text{ is } B_1 \\
\text{.....} \\
R_{m2} := \text{IF } v_t \text{ is Dec THEN} \\
\quad \text{IF } a_t \text{ is } A_{m1} \text{ AND } a_{t+1} \text{ is } A_{m2} \text{ THEN } TDec \text{ is } B_m
\end{array}$$

Семантика системы правил функционала *STend* приведена в таблице 5.2.

Таблица 5.2

Правила вычисления функционала $STend$

	$\tilde{a}_t$	$\tilde{a}_{t+1}$	TInc (TDec)
1	Высокий	Высокий	Самый Высокий
2	Очень Малый	Очень малый	Малый
3	Средний	Средний	Не Большой
4	Средний	Высокий	Очень Высокий
5	Высокий	Малый	Более чем Высокий
6	Средний	Малый	Выше Среднего
7	Малый	Малый	Средний
	...	...	...

Сегментация ВР в терминах производных нечетких тенденций

Производные тенденции – это такие нечеткие тенденции, которые представлены паттернами неоднородных локальных тенденций $N\tau$ и для грамматики языка $LANG$ представляются нетерминальными символами. Будем считать, что производные тенденции являются предметно-ориентированными. К таким производным тенденциям относят «бычий подъем» в финансовых временных рядах, «загрузка/простой» во временных рядах телекоммуникационного трафика.

Задачу сегментации ВР в терминах производных нечетких тенденций можно сформулировать и как задачу распознавания паттернов производных локальных тенденций [Ярушкина, 2004], задаваемых экспертно.

Производную локальную НТ можно представить как график функций или набор упорядоченных данных, определенных на универсуме всевозможных графиков или наборов данных. С этой точки зрения задача распознавания НТ решается как задача распознавания образов, например, нейронной сетью.

Нечеткая логика в задаче распознавания НТ позволяет использовать системы логического вывода. Эксперту при этом необходимо описать НТ в виде последовательности нечетких меток, которые формируют правила распознавания.

Задача сегментации на производные локальные нечеткие тенденции является основополагающей в построении системы анализа данных, решение которой позволит приступить к выявлению связей между параметрами исследуемого объекта.

Для простоты обозначим $\tau = N\tau$.

Данная задача состоит из следующих подзадач:

- 1) определение лингвистических переменных нечетких тенденций;
- 2) построение временного ряда нечеткой тенденции.

Решение задачи обозначим функционалом $Tend$:

$$\tau = Tend[\tilde{Y}_{\Delta}].$$

Для определения лингвистической переменной, выражающей производную нечеткую тенденцию, требуется выделить типичное поведение нечеткого временного ряда, то есть составить терминальное множество нечетких тенденций. Универсальное множество является множеством нечетких временных рядов, поэтому универсум можно представить как множество всевозможных графиков (функций), отображающих динамику. Задача эксперта – выделить типичные (систематические) тенденции нечеткого временного ряда. При этом необходимо определить длину ряда m – интервал временной шкалы, на котором будет описываться тенденция. Для построения временного ряда нечеткой тенденции необходимо не только выделить их (обозначить лингвистические термы), но также дать описание и найти тенденцию на нечетком временном ряду. Таким образом, функционал $Tend$ включает в себя инструмент описания нечеткой тенденции, алгоритм нахождения соответствия нечетких временных рядов выбранному описанию, то есть алгоритм нечеткого оценивания.

Для анализа и построения четкого временного ряда по нечеткой модели необходима операция получения нечеткого временного ряда из оценки тенденции. Решение данной задачи обозначим функционалом $deTend$, являющимся обратным функционалом $Tend$:

$$\tilde{Y}'_{\Delta} = deTend[\tau], \tilde{Y}'_{\Delta} \in \tilde{Y}_{\Delta},$$

где $\tilde{Y}'_{\Delta} = \{\tilde{y}'_{\Delta}\}$ – множество типичных (характерных) нечетких временных рядов для НТ, имеющих наибольшую степень принадлежности $\mu_{\tau}(\tilde{y}'_{\Delta}) = \max(\mu_{\tau}[\tilde{Y}_{\Delta}])$. Таким образом, каждому виду нечеткой тенденции может сопоставляться несколько интервальных нечетких временных рядов.

Применяя операцию $Y = DeFuzzy(\tilde{Y}'_{\Delta})$, получим четкие числовые оценки уровней ВР.

В качестве инструмента построения функционала $Tend$ в виде нечетких отношений и системы нечетких продукционных правил воспользуемся аппаратом нечетких нейронных сетей (ННС).

Для этого используем классические нечеткие нейроны, в которых операции сложения и умножения заменяются триангулярными нормами:

- И-нейрон $\beta = T(S(x_1, w_1), S(x_2, w_2))$: импликация предпосылок правила;
- ИЛИ-нейрон $\varphi = S(T(\beta_1, z_1), T(\beta_2, z_2))$: агрегация правил.

Интерпретация этого в лингвистической форме приводит к высказываниям:

Если $(x_1$ или w_1) и $(x_2$ или w_2), то β ,

Если $(\beta_1$ и $z_1)$ или $(\beta_2$ и $z_2)$, то φ .

Веса w и z можно рассматривать как степени влияния соответствующего входа на выход.

На основе таких нейронов можно построить сеть логического вывода по Мамдани для нахождения нечеткой тенденции. Каждому виду производной НТ соответствует своя сеть. ННС состоит из двух слоев (если считать за слой нейроны): слой конъюнкции нечетких входов и слой дизъюнкции правил. Формализованно такую сеть можно выразить следующим образом:

$$\tau' = \bigvee_{j=1}^r [T(\beta_j, z_j)], \quad \beta_j = \bigwedge_{i=1}^p [T[S(\tau_{t-k}^i, w_{j,i,k})]],$$

где τ' – значение, рассчитанное для некоторого вида тенденции τ ;

T – оператор конъюнкции;

S – оператор дизъюнкции;

r – количество правил;

l, p – параметры модели нечеткой тенденции.

Основная идея обучения ННС состоит в итерационной процедуре оптимизации весов $(z_j, w_{j,i,k})$ и устранении незначимых связей (сокращении сети), в результате которого формируется необходимый состав и количество правил.

Контрольные вопросы

1. Дайте определение временному ряду элементарных тенденций.
2. Опишите алгоритм сегментации НВР в терминах нечетких элементарных и локальных тенденций.
3. Что такое функционал STend?
4. Какие тенденции называются производными?
5. Сформулируйте задачу сегментации ВР в терминах производных нечетких тенденций?
6. Дайте определение функционалу deTend?

5.3. Кластеризация НВР в терминах нечетких тенденций

5.3.1. Кластеризация НВР в терминах элементарных нечетких тенденций

Задача кластеризация НВР в терминах ЭТ основывается на результатах решения задачи сегментации. То есть исходными данными для кластерного анализа является множество элементарных тенденций, распределенных на временном отрезке.

Сформулируем постановку задачи кластеризации в терминах элементарных тенденций. Пусть задано множество элементарных тенденций $\mathcal{T} = \{\tau_i\}$, где $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$.

Требуется разбить множество $\mathcal{T}$ на подмножества – кластеры $s_1, s_2, \dots, s_k$ так, чтобы каждый кластер состоял из элементарных тенденций, близких по метрике ρ , а элементарные тенденции разных кластеров существенно отличались.

На множестве $\mathcal{T}$, согласно теореме о различии элементарных нечетких тенденций, можно построить для любых двух ЭТ τ_i и τ_s функцию $\rho \rightarrow [0, 1]$, выражающую степень их различия (расстояние между ЭТ), которая может служить метрикой на множестве $\mathcal{T}$.

Если кластеры $s_1, s_2, \dots, s_k$ понимать как точки в параметрическом пространстве ЭТ, то задача кластеризации ВР в терминах ЭТ формулируется как выделение «сгущений элементарных тенденций» и разбиение множества $\mathcal{T}$ на однородные подмножества элементарных тенденций.

Алгоритм кластеризации будет заключаться в построении функции $f_Clus: \mathcal{T} \rightarrow S$, которая любому объекту $\tau \in \mathcal{T}$ ставит в соответствие кластер $s \in S$.

Так как элементарные тенденции могут относиться к трем инвариантным по отношению к предметным областям и временным рядам базовым типам: «Рост», «Падение», «Стабильность», то целесообразно выбрать иерархический метод кластерного анализа, в процессе которого сначала строятся три кластера, а затем каждый кластер разбивается дополнительно на несколько кластеров, в зависимости от значения интенсивности тенденции $\tilde{a}_t$. При использовании нечеткого представления $\tilde{a}_t$ необходим нечеткий метод кластерного анализа, в основу которого целесообразно положить нечеткое бинарное отношение между элементарными тенденциями или лингвистическое нечеткое отношение. Тогда задав необходимый порог, характеризующий значимый уровень различия, можно сформировать различные кластеры. Отметим, что в случае использования лингвистического нечеткого отношения, метки кластеров будут наследовать лингвистическое значение элемента этого отношения, например, $s_1 =$ «Рост, Большой». Для лингвистических нечетких отношений удобным средством кластеризации может являться реализация алгоритма на основе деревьев решений.

Вышеизложенное может быть расширено и на класс локальных нечетких тенденций, для которых их длительность Δt выступает дополнительным параметром при кластеризации. С другой стороны, для целей извлечения знаний из временных рядов иногда требуется исследовать отдельные компоненты нечетких тенденций, тогда логика кластерного анализа ВР в терминах НТ может быть сужена до группировки нечетких тенденций по типам или по интенсивностям, или по длительностям.

Поскольку количество нечетких лингвистических меток, обозначающих типы и интенсивности ЭТ конечно и ограничено градациями ACL-шкалы, то принципиально можно построить ограниченное количество кластеров. Пусть мощность множества S всевозможных кластеров, построенных на лингвистических метках элементарных тенденций, равна k . После выполнения процедуры кластеризации множества $\mathfrak{T}$ сформировано k кластеров: $S = \{s_1, s_2, \dots, s_k\}$. Тогда определим на множестве S функцию $g_Clus: S \rightarrow R$, где $R \in E^I$, вычисляющую для каждого кластера его мощность, то есть количество элементарных тенденций, входящих в каждый кластер. Функция g_Clus будет полезна при рассмотрении задачи частотного анализа нечетких тенденций.

Контрольные вопросы

1. Сформулируйте постановку задачи кластеризации в терминах элементарных тенденций.
2. Какова суть алгоритма кластеризации?
3. Дайте определение функции g_Clus .

5.4. Частотный анализ

Рассматривая свойства поведения НВР через призму нечетких тенденций, интерпретируемых в рамках структурно-лингвистического в виде символов абстрактного языка $LANG$, естественно использовать технику частотного анализа, используемого для анализа текстов. На основе частотного анализа текстов обычно решаются задача определения авторства (распознавание типа поведения автора при генерации текстов), строятся словари и таблицы распределения частот слов.

Проводя аналогию между нечетким временным рядом, описанным лингвистическими терминами нечетких тенденций, каждая из которых кодируется словом и текстом, сформулируем задачу частотного анализа нечетких тенденций на НВР.

Пусть дан временной ряд $\{t_i, z_i\}$, $i \in [1, n]$, представленный временным рядом нечетких меток $z_i \in Z$. В качестве нечетких меток $z_i \in Z$ могут выступать лингвистические обозначения уровней ВР, элементарных, локальных нечетких тенденций, кластеры нечетких тенденций, отдельные компоненты нечетких тенденций: типы, интенсивности, длительности.

Требуется определить кластеры $S = \{s_1, s_2, \dots, s_k\}$ подобных нечетких меток z_i и для каждого кластера $s \in S$ вычислить его мощность. Мощность кластера будет определять количество входящих в него элементов, тогда частотная характеристика нечеткой метки $z_i \in s$ будет определяться как отношение мощности кластера s к общему количеству кластеров.

Подход к решению задачи может основываться на реализации рассмотренной выше функции $f_Clus: Z \rightarrow S$, которая любому объекту $z_i \in Z$ ставит в соответствие некоторый кластер $s \in S$.

Пусть нечеткая метка z_i задает нечеткий объект – элементарную тенденцию на ВР.

Определим на множестве S функцию $g_Clus: S \rightarrow R$, где $R \subset E^1$, вычисляющую для каждого кластера его мощность, то есть количество элементарных тенденций, входящих в каждый кластер.

Применение данного подхода при интерпретации z_i в виде типов элементарных нечетких тенденций и точек смены тенденций рассмотрим ниже. Исходный временной ряд приведен на рис. 5.2.



Рис. 5.2. Пример временного ряда изменения финансового индекса

После оценивания по ACL-шкале, построенной по значениям уровней данного временного ряда с количеством нечетких множеств 5 для выявления элементарных тенденций, получим кусочно-линейное представление (см. рис. 5.3). Частотная характеристика базовых типов нечетких тенденций приведена на рис. 5.4.

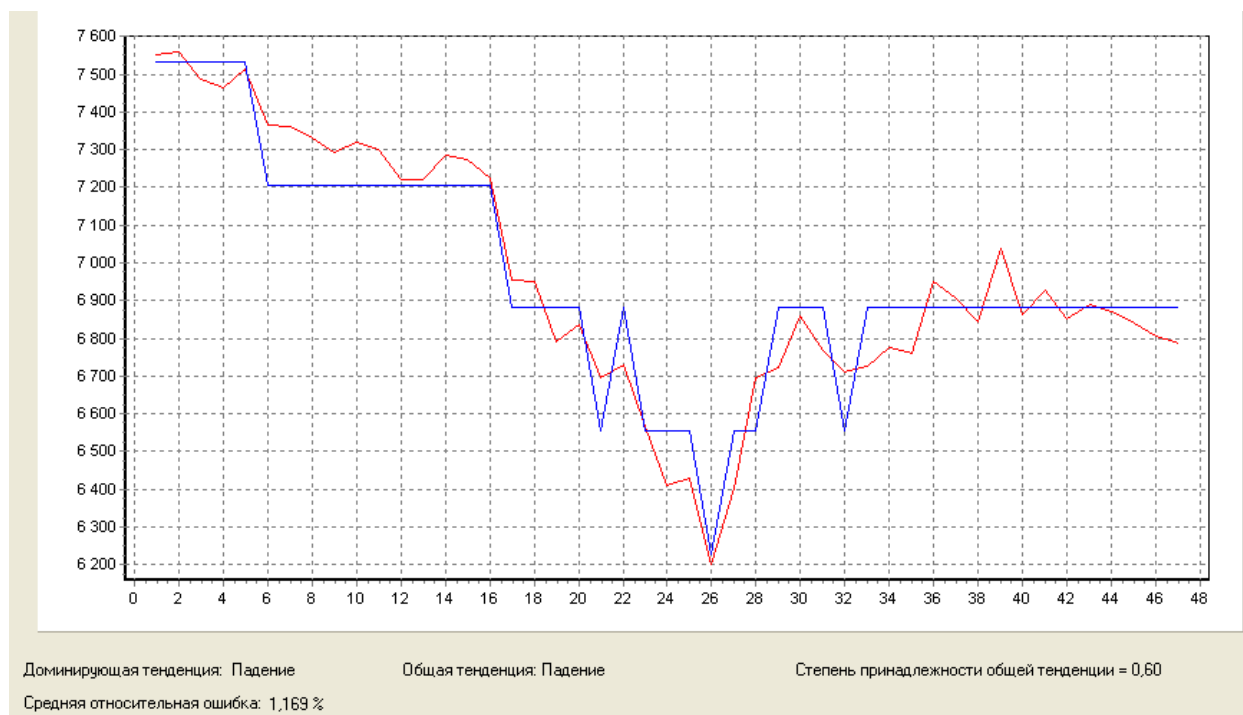


Рис. 5.3. Временной ряд изменения финансового индекса с нечетким профилем на основе элементарных тенденций

начальное значение ряда	время: 1, значение: 7552, кластер: Отличный
конечное значение ряда	время: 47, значение: 6787, кластер: Удовлетворительный
минимальное	время: 26, значение: 6200, кластер: Очень плохой
максимальное	время: 2, значение: 7560, кластер: Отличный
среднее	время: 24, значение: 6964,14893617021
количество особых точек	4
количество особых точек отрицательного перегиба	3
количество особых точек положительного перегиба	1
количество особых точек отрицательного скачка	0
количество особых точек положительного скачка	0
количество участков падения	6
количество участков роста	4

Рис. 5.4. Частотная характеристика базовых типов нечетких тенденций для временного ряда изменения финансового индекса

Контрольные вопросы

1. Для чего нужен частотный анализ?
2. Сформулируйте задачу частотного анализа нечетких тенденций.

5.5. Классификация НВР в терминах нечетких тенденций

Содержательно задача классификации НВР сводится к отнесению временного ряда или паттерна временного ряда (локальной нечеткой тенденции) к одному из нескольких predetermined классов $G\tau \in G\mathfrak{T}$, описывающих динамику поведения временного ряда [Ярушкина, 2004]. Общая (основная) нечеткая тенденция $G\tau$, введенная в главе 4, служит адекватной моделью представления классов поведения ВР.

Сформулируем постановку задачи. Дан нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}, i \in [1, n]$. И существует конечное множество predetermined классов $G\mathfrak{T} = \{G\tau\}$, таких, что каждая общая тенденция представима моделью $G\tau = \langle \tilde{v}, \tilde{a}, \Delta\tilde{t}, \mu \rangle$.

Требуется определить функционал $Gtend: N\mathfrak{T} \times N\mathfrak{T} \rightarrow G\mathfrak{T}$, вычисляющий параметры общей нечеткой тенденции, то есть идентифицировать в модели общей НТ $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta\tilde{t}_i, \mu_i \rangle$ параметры $\tilde{v}_i \in \tilde{V}$, $\tilde{a}_i \in \tilde{A}$ и степень принадлежности НВР этой нечеткой тенденции μ_i .

Так как локальная нечеткая тенденция представима в виде последовательности элементарных тенденций, то функционал $Gtend$ может быть выражен и через элементарные тенденции в виде $Gtend: \mathfrak{T} \times \mathfrak{T} \rightarrow G\mathfrak{T}$.

Решение сформулированной задачи будем строить в два этапа, на первом будем решать задачу классификации, для которой $\tilde{V} = \tilde{V}1$, $\tilde{V}1 = \{\tilde{v}0, \tilde{v}1, \tilde{v}2, \tilde{v}3\}$. На втором этапе, в случае необходимости, задачу классификации будем решать для $\tilde{V} = \tilde{V}2$, $\tilde{V}2 = \{\tilde{v}4, \tilde{v}5\}$. В соответствии с этим выделим следующие классы

(паттерны): ВР с общей тенденцией, тип которой относится к базовым типам НТ ($\tilde{v} 0$ =«Неопределенность», $\tilde{v} 1$ =«Стабилизация», $\tilde{v} 2$ =«Падение», $\tilde{v} 3$ =«Рост»), и ВР с общей нечеткой тенденции производного типа ($\tilde{v} 4$ =«Колебания периодические», $\tilde{v} 5$ =«Колебания Хаотические»).

5.5.1. Распознавание общей тенденции НВР базового типа

Дан временной ряд $Y = \{t_i, x_i\}$, $i \in [1, n]$, n – количество членов временного ряда. Требуется определить тип изменения переменной x , то есть определить общую тенденцию Gt (систематическую долгосрочную составляющую) временного ряда Y .

С этой целью для исходного временного ряда $Y = \{t_i, x_i\}$ спроектируем иерархическую гранулярную модель. Введем четыре уровня гранулирования, формирование которых соответствует подходу снизу-вверх (bottom-up). Нулевой уровень гранул будет представлен нечеткими метками уровней ВР, для образования гранул нулевого уровня воспользуемся «оценочной» операцией ACL-шкалы: $\tilde{x}_i = Fuzzy(x_i)$.

Используя введенные понятия, определим алгоритм классификации НВР:

1 этап. Определение лексики языка $LANG$: нечетких элементарных тенденций, идентификация их типов и параметров.

Таким образом, необходимо на основе нечетких уровней НВР определить нечеткие значения параметров тенденций. Эта процедура называется *идентификацией элементарных нечетких тенденций* исследуемого НВР. Для ее реализации используем операцию идентификации элементарной НТ: $\tau_i = ETend(\tilde{x}_i, \tilde{x}_{i+1})$.

2 этап. Определение синтаксиса (структуры) и семантики (параметров) модели НВР на основе операции объединения однородных элементарных тенденций в базовые группы. Введенный функционал $STend$ вычисляет результат объединения однородных тенденций, такой, что $N\tau_j = STend(\tau_i, \tau_s)$.

3 этап. Интерпретация модели в терминах типа и параметров общей тенденции НБР на основе анализа интегрированного показателя базовых групп и оценки степени принадлежности базового типа исходному временному ряду. Для идентификации общей тенденции НБР введем новый функционал $Gtend$.

Тогда цепочка $ETend(\tilde{x}_i, \tilde{x}_{i+1}) \rightarrow STend(\tau_i, \tau_s) \rightarrow GTend(N\tau_i, N\tau_s)$ позволяет получить значение общей тенденции НБР, что и будет являться результатом классификации НБР на основе нечетких тенденций.

Сопоставим нечетким тенденциям НБР информационные гранулы, обладающие структурной общностью. Определим операцию гранулирования первого уровня в виде функционала $ETend$, образующего гранулы элементарных нечетких тенденций: $\tau_i = ETend(\tilde{x}_i, \tilde{x}_{i+1})$, $i \in [1, n-1]$, n – количество членов НБР.

Функционал $ETend$ порождает гранулы элементарных нечетких тенденций на основе «оценочных» операций $TTend$ и $RTend$ ACL-шкалы.

Определим операцию гранулирования второго уровня в виде функционала $STend$, образующего гранулы локальных нечетких тенденций: $N\tau_j = STend(\tau_i, \tau_s)$, где τ_i, τ_s являются гранулами первого уровня.

Введенный функционал $STend$, вычисляется как результат объединения однотипных элементарных тенденций на основе «вычислительной» операции объединения $Union$ ACL-шкалы. Тогда объединение $N\tau_j = STend(\tau_i, \tau_s)$ есть такая нечеткая тенденция, для которой $v_j = v_i$, $\alpha_j = Union(\alpha_i, \alpha_s)$, $\mu_j = \mu_i \cup \mu_s$, длительность $\Delta t_j = \Delta t_i + \Delta t_s$.

Операция объединения однотипных тенденций определяет гранулы второго уровня.

Для порождения информационной гранулы, характеризующей общую нечеткую тенденцию базового типа, определим функционал $GTend$, $G\tau = GTend(N\tau_i, N\tau_s)$, где $N\tau_i, N\tau_s$ являются гранулами второго уровня. Функционал $GTend$ основан на применении «вычислительной» операции разности $Diff$ ACL-шкалы.

Третий уровень гранулирования реализуется нечеткими продукционными правилами:

$$R_1 := IF TInc \text{ is } A_{11} \text{ AND } TDec \text{ is } A_{12} \text{ THEN } G\tau \text{ is } B_1$$

.....

$$R_m := IF TInc \text{ is } A_{m1} \text{ AND } TDec \text{ is } A_{m2} \text{ THEN } G\tau \text{ is } B_m$$

Здесь использованы *TInc*, *TDec* для обозначения интенсивностей кластеров, образованных из НТ типа «Рост» и «Падение».

На основе введенных функционалов мы определили алгоритм классификации НБР как процедуру идентификации общей нечеткой тенденции *GTend* в терминах базовых типов нечетких тенденций. Эта процедура представляет собой последовательное порождение информационных гранул, моделирующих ВР на различных абстрактных уровнях. Результатом процедуры классификации НБР является гранула общей нечеткой тенденции, представляющая свертку элементарных тенденций в лингвистическую оценку поведения НБР:

$$ETend(\tilde{x}_i, \tilde{x}_{i+1}) \rightarrow STend(\tau_i, \tau_s) \rightarrow GTend(N\tau_i, N\tau_s).$$

Предлагаемый подход к решению задачи классификации НБР на основе модели нечеткой тенденции представлен на рисунках 5.5 и 5.6, на которых даны примеры выполнения процедуры классификации ВР в терминах базовых типов нечетких тенденций.

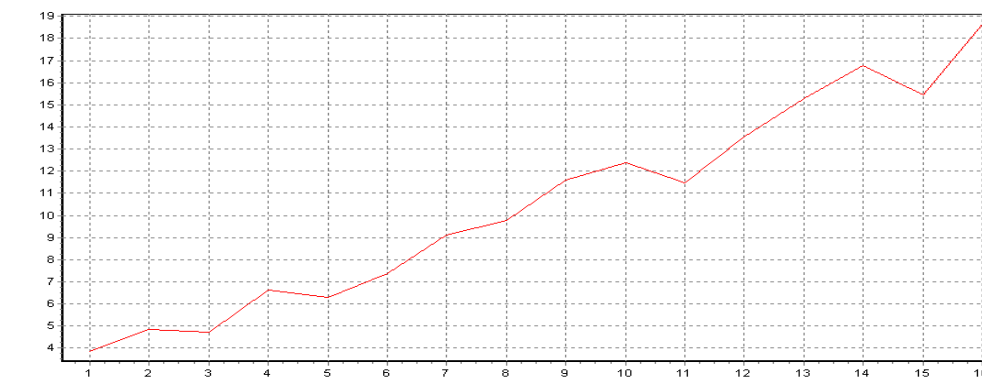


Рис. 5.5. Результат процедуры классификации: «Общая тенденция =Рост»

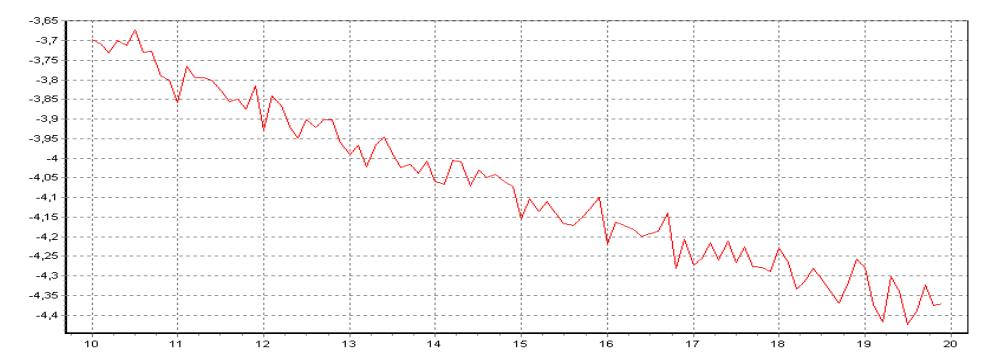


Рис. 5.6. Результат процедуры классификации: «Общая тенденция = Падение».

Реализация функционала, вычисляющего общую тенденцию НВР, может быть выполнена не только на основе лингвистически ясных «вычислительных» операций ACL-шкалы, но и на основе введения некоторого числового критерия доминирования того или другого типа общей тенденции. Одним из способов построения такого критерия может быть способ на основе вычисления нечеткой меры, выражающей интегральную характеристику «силы» нечеткой тенденции, например на основе нечетких интегралов [Яхьяева, 2006].

Безусловно, в качестве результата классификации НВР по приведенным нечетким правилам может формироваться и результат $\tilde{v} = \text{«Неопределенность»}$, свидетельствующий, что тенденции базового типа не обнаружены. В этом случае требуется второй этап процедуры классификации поведения НВР для выявления «периодических» или «хаотических» колебаний.

5.5.2. Распознавание периодического типа общей тенденции НВР

В том случае, когда невозможно определить общую тенденцию в виде нечеткой тенденции с базовым типом, то будем предполагать, что нечеткая тенденция анализируемого ВР может классифицироваться как «периодические колебания» или «хаотические колебания» [Афанасьева и др., 2009] .

Введем классифицирующее правило, согласно которому общая тенденция НВР может быть отнесена к типу «периодические колебания», если локальные

нечеткие тенденции нечеткого временного ряда периодически повторяются, период их повторения постоянный, состав локальных тенденций неизвестен.

Рассмотрим подход к решению задачи распознавания периодического типа общей тенденции НВР, в основе которого лежит понятие модели элементарной нечеткой тенденции.

Модель исходного нечеткого временного ряда представим в виде последовательности локальных нечетких тенденций (паттернов) с базовым типом $\tilde{V} 1 = \{\tilde{v} 1, \tilde{v} 2, \tilde{v} 3\}$, $\tilde{V} 2 = \{\tilde{v} 4, \tilde{v} 5\}$, где $\tilde{v} 1 = \text{«Стабилизация»}$, $\tilde{v} 2 = \text{«Падение»}$, $\tilde{v} 3 = \text{«Рост»}$.

Предположим, что тенденции, которые образуют период, заранее неизвестны. Введем ограничение. Пусть последовательности локальных НТ образованы элементарными тенденциями различных типов, но с постоянными значениями интенсивностей. Будем называть такие последовательности локальных НТ *паттернами*.

Определим следующие параметры паттерна:

$$sh = \langle Tsh, w, I \rangle,$$

где sh – наименование паттерна;

Tsh – тип паттерна;

w – длина паттерна;

I – интенсивность паттерна.

При решении проблемы распознавания общей тенденции НВР типа «периодические колебания» требуется решить подзадачи:

- Определение сходства паттернов локальных нечетких тенденций.
- Поиск повторяющихся паттернов локальных нечетких тенденций.
- Поиск длины периода повторяющихся паттернов локальных нечетких тенденций.

Предложим следующие этапы решения указанных подзадач:

1. Сегментация НВР на паттерны в виде локальных нечетких тенденций производного типа методом скользящего окна w , например, в случае $w = 2$, по-

лучим две ЛНТ $N\tau 1 = \{\tau_{11}, \tau_{21}, \dots, \tau_{m1}\}$ и $N\tau 2 = \{\tau_{12}, \tau_{22}, \dots, \tau_{m2}\}$, содержащие m упорядоченных по времени появления элементарных НТ, τ_{ij} – элементарная нечеткая тенденция.

2. Построение отношения сходства на основе вычисленной меры сходства. Для двух нечетких локальных тенденций отношение сходства вырождается в степень сходства $q(N\tau 1, N\tau 2)$ между полученными ЛТ. Меры сходства локальных нечетких тенденций, выраженные через меры различия $\rho(N\tau 1, N\tau 2)$, были определены в виде

$$q(N\tau 1, N\tau 2) = 1 - \rho(N\tau 1, N\tau 2),$$

где $\rho(\tau_{it}, \tau_{st}) = \rho(\tilde{v}_{it}, \tilde{v}_{st}) \cdot \sqrt{\rho(\tilde{a}_{it}, \tilde{a}_{st})^2 + \rho(\mu_{it}, \mu_{st})^2}$, согласно теореме о степени различия элементарных нечетких тенденций, рассмотренной в главе 4.

3. Анализ отношения сходства и вычисление интегральной степени сходства между всеми ЛТ, полученными в процессе сегментации, например, как интегральная мера сходства $Q(N\tau 1, N\tau 2)$.

В этом случае задача распознавания класса общей тенденции НВР типа «периодические колебания» заключается в следующем: найти такое w , для которого мера сходства нечетких тенденций $Q(N\tau 1, N\tau 2.. N\tau w)$ максимальна и/или превышает некоторый заданный порог λ , тогда w будет определять период, а тип паттерна $Tsh = N\tau 1$.

Для решения обозначенной задачи можно использовать переборные методы, в рамках которых исходный нечеткий временной ряд анализируется на наличие периода на основе перебора паттернов различной длины и оценки каждого как возможного периода.

Можно предложить и другие меры сходства паттернов, так, например, степень сходства паттернов можно вычислять на основе их классификации по базовым типам. И если анализируемые паттерны относятся к одному классу, то они сходны, а степень принадлежности классу задаст степень сходства. Интерес представляет рассмотрение меры, построенной на частотной характеристике

следования элементарных нечетких тенденций $N\tau = \{\tau_{11}, \tau_{21}, \dots, \tau_{m1}\}$, входящих в состав локальных тенденций.

Не вызывает сомнения, что применение совместно нескольких критериев, основанных как на мерах сходства по расстоянию между элементарными тенденциями, входящими в состав анализируемых локальных тенденций, так и на частотных мерах, позволит повысить достоверность алгоритмов классификации НВР.

В заключение отметим, что проблематика и подходы, рассмотренные в этом разделе, касающиеся определения сходства двух произвольных локальных нечетких тенденций, в состав которых входит одинаковое количество элементарных тенденций, могут быть расширены и на задачи вычисления степени сходства поведения различных временных рядов.

Контрольные вопросы

1. Сформулируйте задачу классификация НВР в терминах нечетких тенденций.
2. Дайте определение функционалу G_{tend} .
3. Опишите алгоритм распознавания общей тенденции НВР базового типа.
4. В каком случае нечеткая тенденция анализируемого ВР может классифицироваться как «периодические колебания» или «хаотические колебания»?
5. Какие задачи требуется рассмотреть при решении проблемы распознавания общей тенденции НВР типа «периодические колебания»?
6. Приведите алгоритм решения задачи распознавания общей тенденции ВР.

5.6. Извлечение правил

Извлечение правил из нечетких временных рядов – это задача автоматического построения базы знаний, которая может быть использована в явном виде в нечетких моделях и системах нечеткого логического вывода продукционного типа [Борисов и др., 2007] или в неявном виде в нечетких нейросетевых системах [Ярушкина, 2004] моделирования НВР.

Тип правил, включенных в базу знаний, определяет ее структуру. Введем следующие обозначения. Обозначим базу знаний $Rule = \{R_k^p\}$, где k – количество нечетких продукционных правил в базе знаний; p – количество переменных (нечетких высказываний) в каждом правиле. Каждое нечеткое продукционное правило ($i = [1, k]$) представляется в виде нечеткого продукционного правила:

$$R_i^p : \text{Если } x_1 \text{ есть } \tilde{x}_1 \text{ и } x \text{ есть } \tilde{x}_j \text{ и } \dots x_p \text{ есть } \tilde{x}_p \text{ то } y \text{ есть } \tilde{y}_i ,$$

где $\tilde{x}_j \in \tilde{X}$, $\tilde{y}_i \in \tilde{Y}$ – нечеткие множества.

Следовательно, методы генерации правил определяют процесс структурной идентификации базы знаний, в основе которых лежит процесс определения параметров p (порядок модели) и k (размерность модели). При параметрической идентификации производится настройка параметров функций принадлежности входных и выходных переменных $\tilde{x}_j \in \tilde{X}$, $\tilde{y}_i \in \tilde{Y}$ нечеткой модели, количество которых задается экспертно.

Алгоритмы на основе мер ассоциации выявляют паттерны ВР, которые появляются вместе в одной транзакции. В общем случае ассоциативные правила имеют форму правила импликации

$$\text{Если } A_{t-l} \text{ то } B_t(W) ,$$

где A_{t-l} – нечеткое ограничение, определенное в момент времени $t-l$ ряда x_t ;

B_t – нечеткое ограничение, определенное в момент времени t ряда y_t ;

W – мера ассоциации A и B .

Для каждого правила вычисляются две частотные характеристики: поддержка и достоверность. Меры ассоциации используются в том случае, если исследуется нечеткая импликация.

Используют следующие меры W :

– распространенность s для поддержки правила: частота одновременного проявления признаков A и B ;

– доверительность s для достоверности правила: взаимосвязь A и B , частота появления признака B по отношению к A .

Формулы для расчета мер:

$$s = \frac{\sum_{i=1}^N T(\mu_A(i-l), \mu_B(i))}{N},$$

$$c = \frac{\sum_{i=1}^N T(\mu_A(i-l), \mu_B(i))}{\sum_{i=1}^N \mu_A(i-l)},$$

где T – t -норма;

μ_A – функция принадлежности нечеткой переменной A ;

μ_B – функция принадлежности нечеткой переменной B .

Для анализа распространенности также используют матрицы частотности, в которых указываются частоты появления нечетких правил (рис. 5.7).

y_{t-l}

	B_1	B_2	...	B_i
A_1	z_{11}	z_{12}	...	z_{1i}
A_2	z_{21}	z_{22}	...	z_{2i}
...	...	...	...	...
A_j	z_{j1}	z_{j2}	...	z_{ij}

t

Рис. 5.7. Частоты появления нечетких правил.

A_i и B_i – нечеткие переменные рядов x_t

и y_t ; z_{ij} – частота срабатывания правила «ЕСЛИ A_j , ТО B_i »

Рассмотренные способы генерации правил базы знаний системы нечеткого логического вывода для нечеткого временного ряда могут быть дополнены правилами следования нечетких тенденций, которые строятся на базе ACL-шкалы в условиях предположения об упорядоченности нечетких уровней ряда (см. главу 4).

Основанием для извлечения правил базы знаний нечеткой модели ВР является последовательность элементарных тенденций, построенная на нечетком временном ряду.

Вид правил базы знаний $Rule_et = \{R_k^p\}$, используемый для генерации последовательности элементарных тенденций, образующей временной ряд элементарных тенденций, был рассмотрен в подразделе 5.2.

В этом случае количество входных, выходных переменных и их параметры для нечеткой модели определяются при генерации нечетких градаций ACL-шкалы S_x . Так количество входных переменных m может быть вычислено на основе количества нечетких градаций и теоремы о мощности ACL-шкалы (глава 4) $\tilde{x}_i \in \tilde{X}$, $i \in [1, m]$, а количество выходных переменных задается градациями $\tilde{V}$, $\tilde{A}$, то есть количество выходных переменных кратно двум. Тогда размерность нечеткой модели $k = 2 \cdot m \cdot m = 2 \cdot m^2$. Параметр структуры p базы правил $Rule_et = \{R_k^p\}$ в соответствии с определением элементарной тенденции как тенденции между двумя соседними уровнями НВР будет равен $p=2$. Следовательно, структурная идентификация базы знаний для генерации временного ряда элементарных тенденций определяется введенными формализмами элементарной тенденции и ACL-шкалой, что позволяет задать структуру базы знаний в виде $Rule_et = \{R_{2 \cdot m^2}^2\}$, где m – мощность ACL-шкалы, параметрическая идентификация которой также может быть реализована автоматическими процедурами.

Контрольные вопросы

1. Для чего нужно извлечение правил из нечетких временных рядов?
2. Перечислите параметры задачи базы знаний.
3. Какие меры ассоциации используются при нечеткой импликации?
4. Как генерируются правила базы знаний системы?

5.7. Моделирование ВР на основе нечетких тенденций

Моделирование ВР имеет целью решение задач определения и объяснения природы, а также прогнозирования временного ряда. Для этого модель ВР должна быть идентифицирована и формально описана. Основа модели – зависимость (формула, правила), связывающая элементы ряда. Построение зависимости есть процесс извлечения знаний.

В предыдущих разделах были рассмотрены задачи сегментации, кластеризации, классификации и извлечения знаний из ВР в терминах элементарной, локальной и общей нечетких тенденций, каждая из которых выражает поведение ВР на своем уровне иерархии. Исходя из этого, прогнозирование исходного ВР будем рассматривать как получение многоуровневого прогноза: в терминах общей, локальной, элементарной тенденций, при этом получение прогнозных значений нечетких и четких уровней ВР обеспечивается «оценочными» операциями ACL-шкалы. Отметим, что модель для прогноза в терминах общей тенденции основывается на распознавании и классификации общей НТ, рассмотренных в предыдущих разделах, и соответствует систематической долговременной компоненте в поведении ВР. В основу моделирования поведения ВР в терминах элементарной и локальной нечетких тенденций положим предположение о развитии системы как результата зависимости нечеткой тенденции от значений нечетких тенденций в предыдущие моменты времени. Таким образом, структуру модели ВР будем выражать нечетким разностным уравнением d -го порядка:

$$\tau_t = \tilde{f}(t, \tau_{t-1}, \dots, \tau_{t-d}),$$

где d – фиксированное число, параметр модели;

$\tau_t, \tau_{t-1}, \dots, \tau_{t-d}$ – последовательность нечетких тенденций;

$\tilde{f}$ – некоторая нечеткая зависимость.

5.7.1. Прогнозирование ВР в терминах элементарных тенденций

Предлагаемый подход к построению моделей ВР для целей прогноза [Афанасьева, 2009] подразумевает преобразование исходного ВР $Y = \{x_t\}$ в нечеткий временной ряд $\tilde{Y} = \{\tilde{x}_t\}$, где $\tilde{x}_t = Fuzzy(x_t)$, $\tilde{x}_t \in \tilde{X}, t \in [l, n]$, а затем – во временной ряд элементарных тенденций, где $\tau_t = ETend(\tilde{x}_t, \tilde{x}_{t+1})$ получается с использованием FT -преобразования и ACL-шкалы. В результате этих операций получается несколько временных рядов, четких числовых и нечетких лингвистических, например, временные ряды параметров нечетких тенденций $\tau_t = \langle \tilde{v}_t, \tilde{\alpha}_t, \Delta \tilde{t}_t, \mu_t \rangle$, ряд индексов нечетких уровней, ряд центров масс нечетких уровней и др. (рис. 5.8). Данные ряды являются потенциально информативными для построения моделей прогноза ВР четких x_t и нечетких $\tilde{x}_t$ уровней, типов $\tilde{v}_t$ и интенсивностей $\tilde{\alpha}_t$ нечетких тенденций.

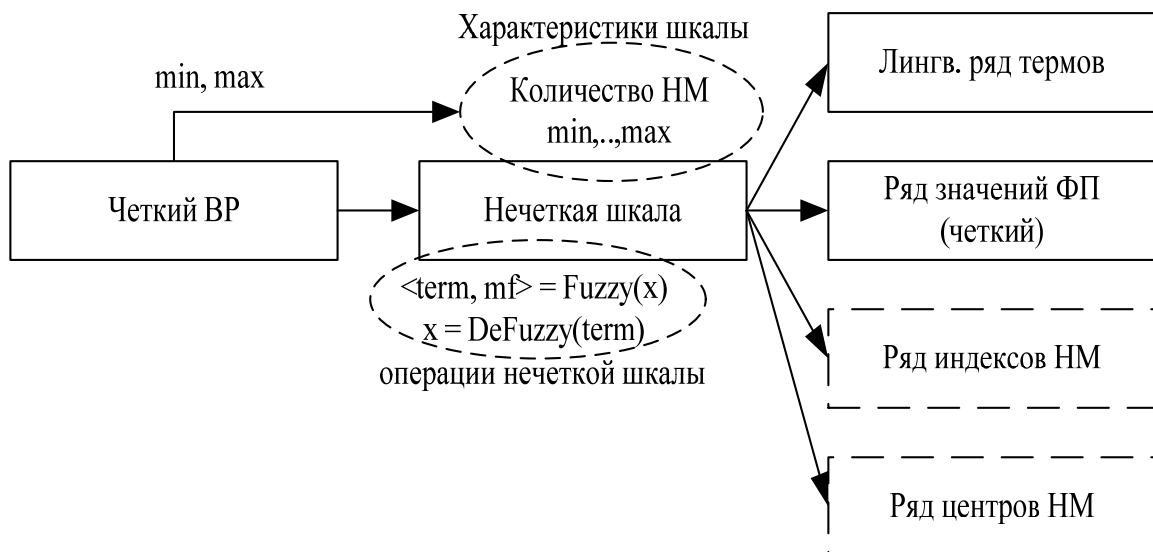


Рис. 5.8. Преобразование и декомпозиция временного ряда

Предлагается следующая модель ВР первого порядка для извлечения знаний о зависимостях параметров элементарных тенденций:

$$\begin{aligned}\tilde{x}_t &= Fuzzy(x_t), \\ \tilde{v}_t &= TTend(\tilde{x}_t, \tilde{x}_{t-1}), \\ \tilde{v}_{t+1} &= \tilde{f}_{\tilde{v}}(\tilde{v}_t), \\ \tilde{a}_t &= RTend(\tilde{a}_t, \tilde{a}_{t-1}), \\ \tilde{a}_{t+1} &= \tilde{f}_{\tilde{a}}(\tilde{a}_t), \\ \tilde{x}_{t+1} &= Comp(\tilde{x}_{t+1}, \tilde{v}_{t+1}, \tilde{a}_{t+1}), \\ x_{t+1} &= DeFuzzy(\tilde{x}_{t+1}) + \varepsilon_{t+1},\end{aligned}$$

где *Fuzzy*, *TTend*, *RTend*, *Comp*, *DeFuzzy* – операции ACL-шкалы;

$\tilde{f}_{\tilde{v}}, \tilde{f}_{\tilde{a}}$ – нечеткие зависимости, представимые на основе композиционного правила импликации в виде $\tilde{v}(t-1) \circ R_{\tilde{v}}(t, t-1) \rightarrow \tilde{v}(t)$ и $\tilde{a}(t-1) \circ R_{\tilde{a}}(t, t-1) \rightarrow \tilde{a}(t)$;

x_{t+1}, ε_t – числовая оценка и ошибка прогнозируемого уровня исходного ВР.

Предложенная модель ВР является частным случаем модели временного ряда p -порядка.

Форма представления прогнозных значений определяется моделью, и в случае нечетких значений результат представляется в лингвистической форме для нечетких уровней, типов и интенсивностей элементарной тенденции, а для получения числовой оценки прогноза применяется операция дефаззификация ACL-шкалы.

Таким образом, в предложенной модели имеется возможность получения многоуровневого прогноза, включающего оценку прогнозных типов и интенсивностей элементарных тенденций (изменений), что отличает ее от известных моделей.

Для решения задачи прогнозирования ВР в терминах элементарных НТ предлагается использовать схему, представленную на рис. 5.9, учитывая использование ACL-шкалы и различные представления параметров в модели элементарной тенденции.

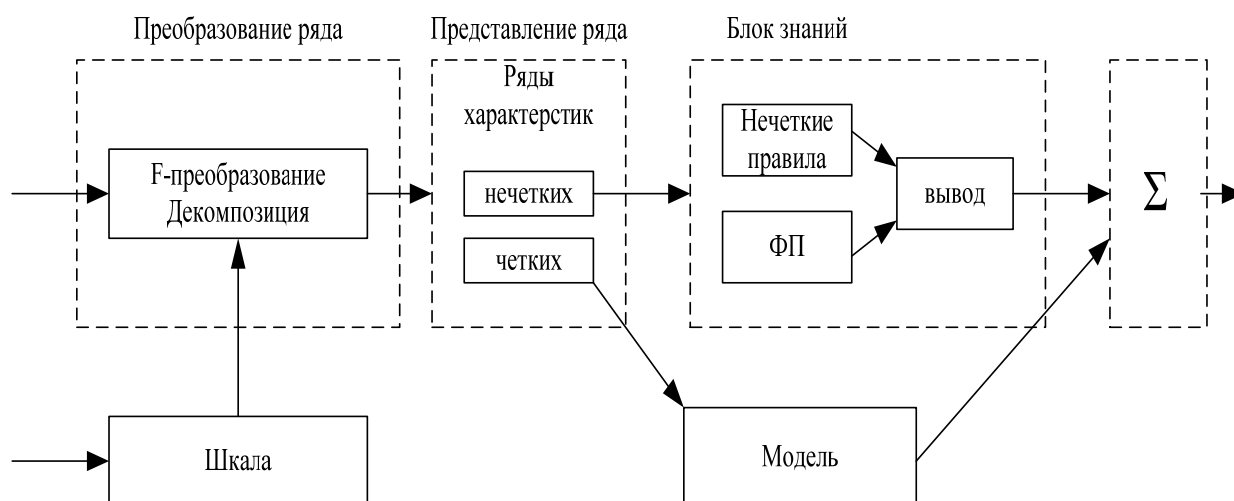


Рис. 5.9. Структура нечеткого моделирования ВР в терминах ЭТ

Общая схема этапов построения модели ВР для прогнозирования в терминах элементарных тенденций представлена на рисунке 5.10.



Рис. 5.10. Общая схема построения модели ВР в терминах ЭТ

Построение ACL-шкалы

На первом шаге по анализируемому ВР строится модель «псевдоинтервальной» ACL-шкалы. Параметры (нечеткие множества) этой шкалы могут устанавливаться пользователем или формироваться автоматически на основе методов кластерного анализа (в простейшем случае – равномерное разбиение шкалы). При автоматическом задании функций принадлежности (ФП) важное

значение имеет адекватное количество ФП. От этого количества зависит погрешность вычисления типа различия $TTend$ и интенсивности различия $RTend$, а значит, и ошибка модели. Заметим, что автоматическое задание ФП, в отличие от экспертного, затрудняет их предметно-зависимую интерпретацию, ФП рассматриваются как некоторые абстрактные функции.

В связи с этим возникает задача автоматического определения такого количества ФП, построенного для ACL-шкалы, при котором ошибка прогноза минимальна или хотя бы находится в допустимом интервале, длина которого не превышает определенного уровня ε . Фактически, это задача о минимальном разбиении диапазона (интервала) допустимых значений уровней ВР на покрывающие диапазон интервалы, при котором ошибка модели по модулю не превышает заданный уровень ε . Решение данной задачи может основываться на приведенной в главе 4 теореме о мощности ACL-шкалы.

Построение временных рядов элементарной тенденции

Для преобразования ВР во временной ряд нечетких элементарных тенденций применяется построенная на первом шаге ACL-шкала, которая преобразует сначала ВР в нечеткий временной ряд, а затем во временной ряд элементарных тенденций. В результате, получаем расширенную структурно-лингвистическую модель исходного временного ряда, проекции которой образуют несколько одновременных временных рядов, характеризующих различные свойства нечетких тенденций: тип, интенсивность, принадлежность (см. главу 4). Проекции структурно-лингвистической модели временного ряда и их вид (четкий или нечеткий) определяют представление модели ВР на основе объединения моделей временных рядов параметров элементарной тенденции.

Предлагаются три варианта представления модели временного ряда элементарных тенденций в зависимости от типа его компонент:

1. Модель F2S: $\tau = (\mu, \tilde{v}, \tilde{a})$.
2. Модель F1N: $\tau = (\mu, v, \alpha)$, где $v = DeFuzzy(\tilde{v})$, $\alpha = DeFuzzy(\tilde{a})$.

3. Модель F3N1S: $\tau = (\tilde{a}, \mu_1, \mu_2, \mu_3)$, где μ_1, μ_2, μ_3 – числовые временные ряды степеней принадлежности (функции принадлежности) тенденции «Рост», «Падение», «Стабильность».

Возможны и другие представления.

Выбор представления модели ВРЭТ определяет метод прогнозирования. Достоинством лингвистических временных рядов (модель F2S) является возможность использовать систему нечеткого логического вывода, прогноз числовых ВР (модель F1N) может быть основан на известных методах в этой области (статистические или нейросетевые методы). Интерес представляет применение повторного *FT*-преобразования полученных числовых рядов интенсивностей для анализа тенденций в них. В этом случае прогноз может строиться и в отношении изменения самих тенденций.

Извлечение знаний в виде продукционных правил

На третьем шаге из временного ряда нечетких тенденций извлекаются продукционные правила $Rule_et = \{R_k^p\}$, характеризующие поведение временного ряда в терминах элементарных нечетких тенденций. В учебном пособии продукционные правила, применительно к нечетким элементарным тенденциям, будут рассматриваться как ассоциативные правила следования ЭТ. В такой интерпретации эти правила образуют синтаксические правила грамматики языка *LANG*. Извлечение таких правил на основе лингвистических термов совместно с функциями принадлежности формирует и семантические правила языка *LANG*, которые будут в дальнейшем использованы для решения задач моделирования тенденций. Количество нечетких высказываний в каждом правиле p , количество правил k будут определять соответственно порядок и размерность нечеткой системы моделирования поведения ВР.

Общей проблемой как при прогнозировании в терминах НМ, так и в терминах нечетких тенденций, является проблема определения вида, порядка и ка-

чества модели прогноза, которая имеет много общего с задачей извлечения ассоциативных правил из БД.

Прогнозирование временного ряда

На четвертом шаге реализуется применение сгенерированной нечеткой системы для прогноза в лингвистических значениях тенденций, в нечетких значениях временного ряда и в числовых значениях анализируемого ВР, то есть выполняется прогнозирование отдельных ВР, полученных ранее. Затем производится свертка полученных значений. Результатом является кортеж:

$$\langle \hat{Y}, \hat{\nu}, \hat{\alpha}, \hat{\mu}, \hat{x} \rangle,$$

где $\hat{Y}, \hat{\nu}, \hat{\alpha}, \hat{\mu}, \hat{x}$ – прогнозные значения числового уровня ВР, типа нечеткой тенденции, интенсивности нечеткой тенденции, степени принадлежности и нечеткого уровня ВР соответственно. Полученные результаты генерируются алгоритмом поиска наилучшей модели из набора моделей нечетких тенденций.

5.7.2. Модели ВР в терминах элементарных тенденций

Модель F2S

Данная модель основана на представлении временного ряда как результата агрегации нечеткого временного ряда типов тенденций и нечеткого временного ряда интенсивностей $\tau = (\mu, \tilde{\nu}, \tilde{\alpha})$. Модель, используемая для целей прогноза, будет относиться к нечетким моделям логического вывода, вследствие нечеткого представления параметров ЭТ, типа нечеткой авторегрессии p -порядка ($AR(p)$):

$$Y(t-p) \circ R(t, t-p) \rightarrow Y(t).$$

Производя формальную подстановку вместо значений $Y(t)$, $Y(t-p)$ – значения $\tau(t)$, $\tau(t-p)$, а вместо $R(t, t-p)$ – набор извлеченных правил из НВР, получим следующее представление модели для прогноза элементарных тенденций:

$$\tau(t-p) \circ R(t, t-p) \rightarrow \tau(t),$$

$$\tilde{\nu}(t-p) \circ R_{\tilde{\nu}}(t, t-p) \rightarrow \tilde{\nu}(t),$$

$$\tilde{\alpha}(t-p) \circ R_{\tilde{\alpha}}(t, t-p) \rightarrow \tilde{\alpha}(t).$$

Результатом моделирования является база правил $\{R(t, t-p)\}$ нечеткой системы логического вывода, выражающая нечеткую зависимость следования элементарных тенденций, извлеченная из НВР, пример которой с частотной характеристикой каждого правила для $p=1$ представлен в таблице 5.2.

Таблица 5.2

Таблица правил для типов тенденций

Номер	Правило	Частота
0	if ((Input0 is Падение)) then (Output is Рост)	3
1	if ((Input0 is Рост)) then (Output is Рост)	1
2	if ((Input0 is Рост)) then (Output is Падение)	2
3	if ((Input0 is Падение)) then (Output is Падение)	2

Результатами прогноза являются параметры элементарной тенденции, выражающие тип и интенсивность будущих изменений, нечеткую и числовую оценки уровней ВР. Полученный результат может иметь лингвистическую, лингвистическо-графическую интерпретацию (см. рис. 5.11) и числовую при дефаззификации с помощью операции *DeFuzzy ACL*-шкалы.

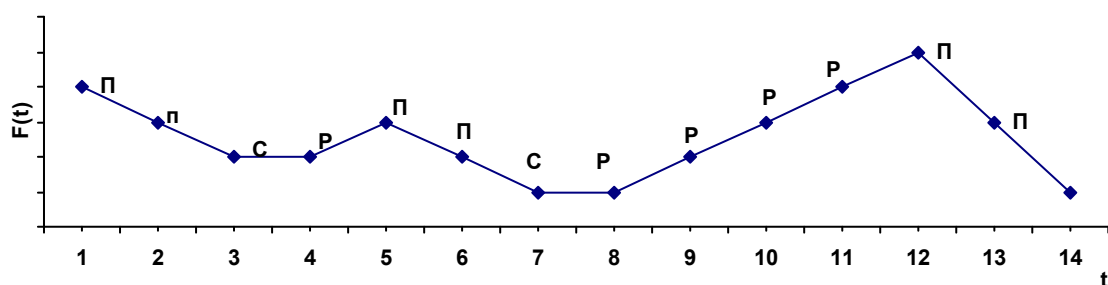


Рис. 5.11. Лингвистический ряд типов ЭТ

Структурная схема построения модели *F2S* приведена на рисунке 5.12.

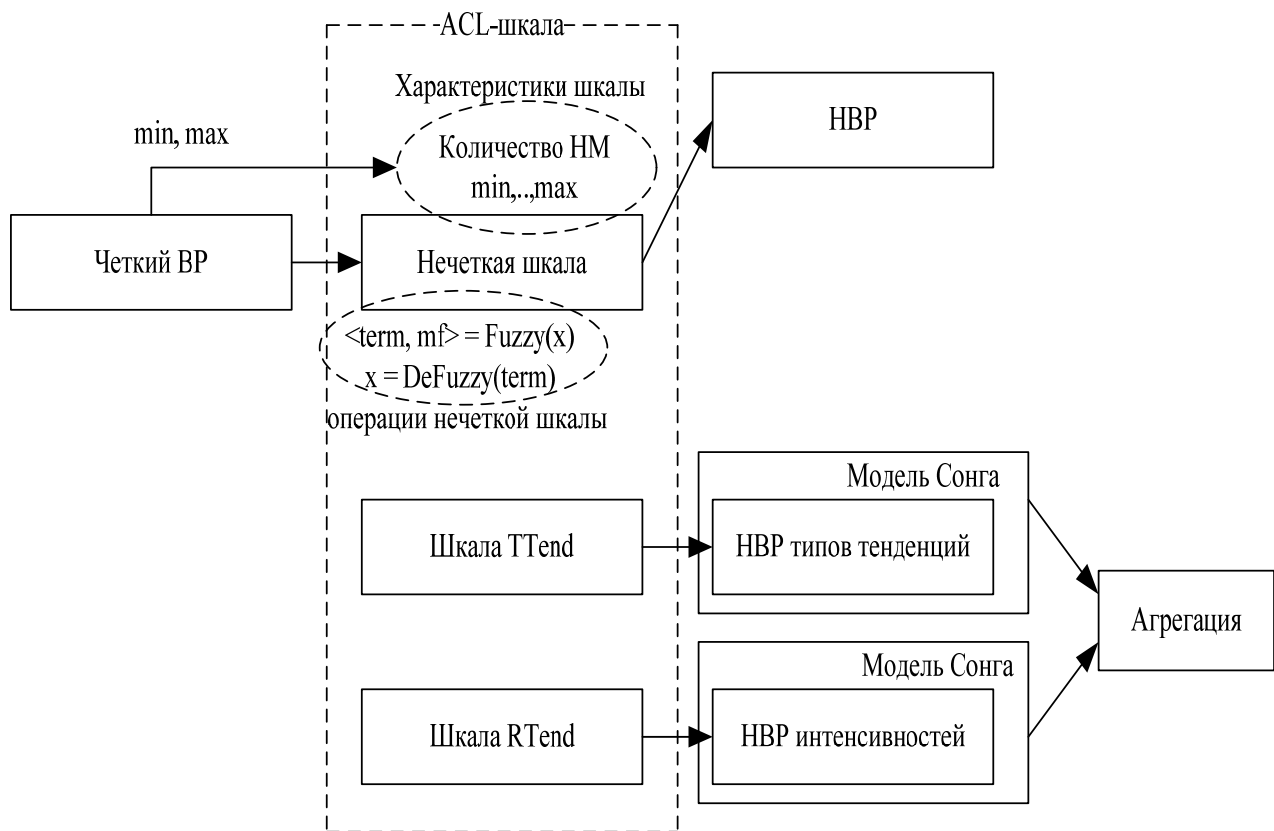


Рис. 5.12. Схема построения модели $F2S$

Модель $F1N$

Данная модель временного ряда имеет вид $\tau = (\mu, v, \alpha)$, где $v = DeFuzzy(\tilde{v})$, $\alpha = DeFuzzy(\tilde{\alpha})$, и основана на представлении модельного временного ряда ЭТ в виде числового ВР интенсивностей изменения элементарных нечетких тенденций. Четкий временной ряд интенсивностей получается с помощью процедуры дефаззификации по шкале интенсивности. Для прогнозирования значений этого ряда удобно использовать трехслойную нейросетевую модель с количеством входов, равным порядку модели p , и одним выходом:

$$\alpha' = F\left(\sum_{i=1}^p w_i \alpha_i\right),$$

где $(\alpha_1, \alpha_2, \dots, \alpha_p)^T$ – вектор входного сигнала;

$(w_1, w_2, \dots, w_p)$ – весовой вектор;

F – оператор нелинейного преобразования, неявно выражающий зависимость выхода от входа.

Алгоритм прогнозирования состоит в получении значения интенсивности со знаком, зависящим от типа ЭТ, и суммирования этой интенсивности и предыдущего значения ряда для получения следующего значения. Чтобы получить лингвистическую интерпретацию полученного прогноза, необходимо его оценить на основе оценочных операций *ACL*-шкалы.

Структурная схема построения модели *FIN* приведена на рисунке 5.13.

Модель F3N1S

Данная модель основана на представлении временного ряда в виде $\tau = (\tilde{a}, \mu_1, \mu_2, \mu_3)$, где μ_1, μ_2, μ_3 – числовые временные ряды степеней принадлежности (функции принадлежности) типов элементарных тенденций «Рост», «Падение», «Стабильность»; $\tilde{a}$ – временной ряд интенсивностей элементарных тенденций, обозначаемых $R0, R1$ и т. д.

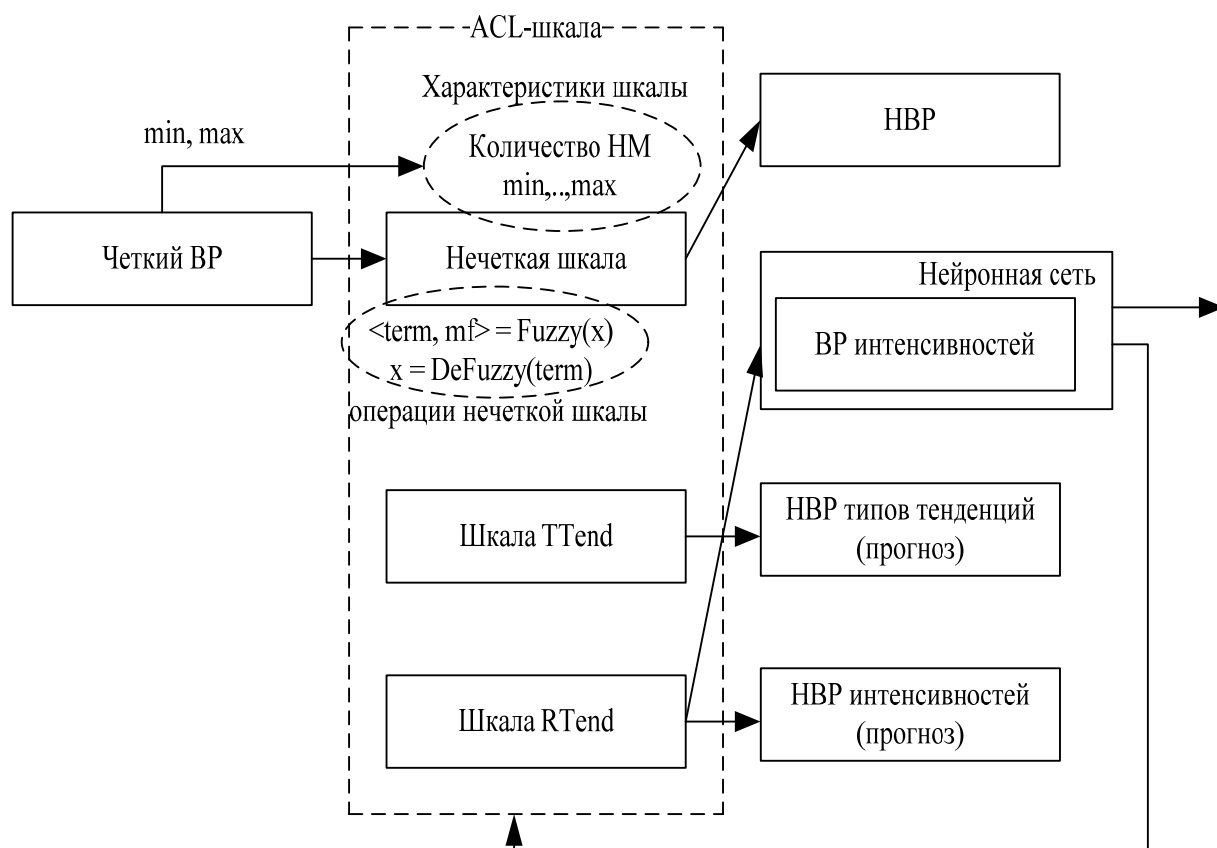


Рис. 5.13. Схема построения модели F1N

При моделировании ВР в модели *F3NIS* строится база правил нечеткой системы логического вывода интенсивностей элементарной тенденции (см. табл. 5.3).

Прогнозное значение интенсивности элементарной тенденции будет соответствовать такому типу тенденции, функция принадлежности которого максимальна.

Функции принадлежности типов элементарных тенденций моделирует нейронная сеть, входными данными нейронной сети служат соседние значения четкого ряда, сеть имеет три выхода для каждого типа тенденций. Значения на выходе – функции принадлежности тенденции соответствующего типа. Для получения однозначного результата о типе прогнозной тенденции выход нейронной сети объединяется для определения функции принадлежности с максимальным значением.

Таблица правил для интенсивностей тенденций

Номер	Правило	Частотность
0	if ((Input0 is R3)) then (Output is R2)	3
1	if ((Input0 is R2)) then (Output is R2)	3
2	if ((Input0 is R2)) then (Output is R1)	2
3	if ((Input0 is R1)) then (Output is R3)	2
4	if ((Input0 is R3)) then (Output is R1)	1
5	if ((Input0 is R1)) then (Output is R1)	5
6	if ((Input0 is R3)) then (Output is R5)	1
7	if ((Input0 is R5)) then (Output is R1)	1

Таким образом, модель ВР представлена следующими зависимостями;

$$\tilde{\alpha}(t - p) \circ R_{\tilde{\alpha}}(t, t - p) \rightarrow \tilde{\alpha}(t),$$

$$\mu'_1 = F_1 \left(\sum_{i=1}^p w_i \mu_{1i} \right),$$

$$\mu'_2 = F_2 \left(\sum_{i=1}^p w_i \mu_{2i} \right),$$

$$\mu'_3 = F_3 \left(\sum_{i=1}^p w_i \mu_{3i} \right),$$

$$r = \arg(\max(\mu'_1, \mu'_2, \mu'_3)),$$

где r – номер типа тенденции.

Структурная схема построения модели $F3NIS$ приведена на рисунке 5.14.

5.7.3. Критерии качества моделей

Оценка качества моделей прогнозирования основана на анализе расхождений между исходным ВР и его моделью. Используются внутренние и внешние меры качества моделей, основанные на разделении исходного ВР на три части (см. рис. 5.15).

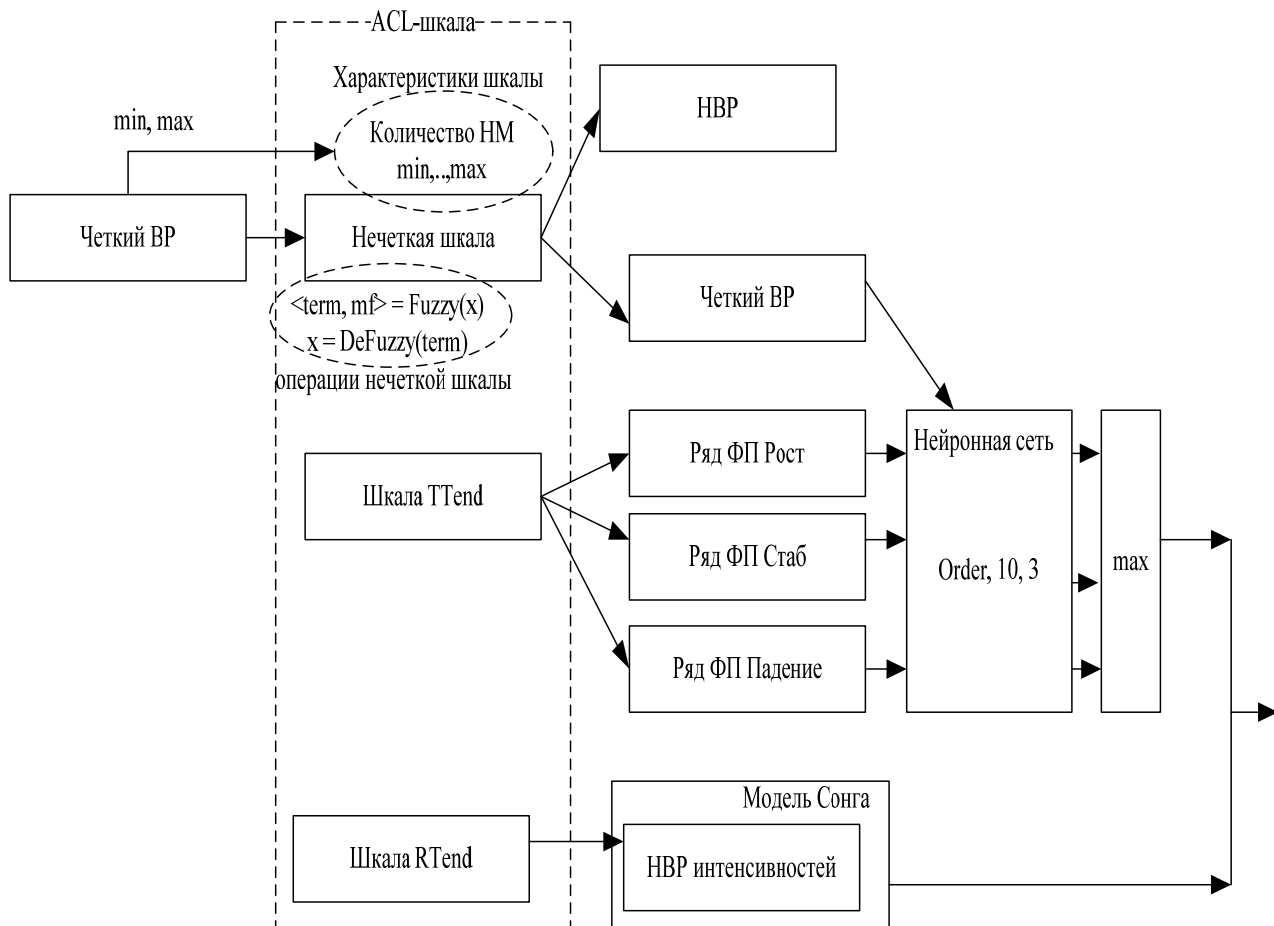


Рис. 5.14. Схема построения модели *F3NIS*

Ошибка на обучающей выборке называются внутренними ошибками модели, а ошибки на части экзаменационной выборке, не участвующей в обучении – внешними ошибками. Количество будущих значений задается глубиной прогноза. В качестве критериев точности нечеткого моделирования числовых уровней ВР будем использовать традиционные меры качества, приведенные в таблице (табл. 5.4).

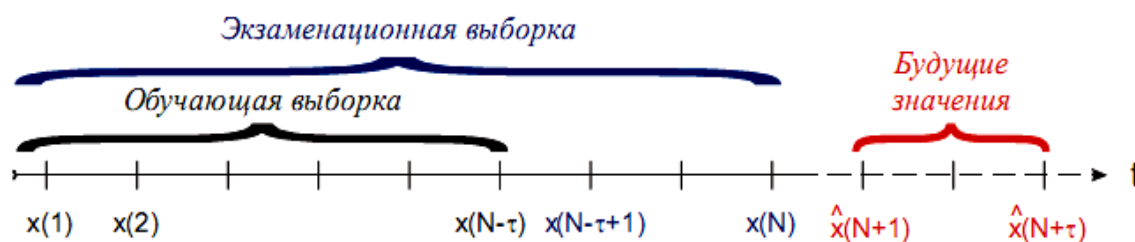


Рис.5.15. Подход к оценке качества моделей

Таблица 5.4

Критерии качества нечеткого моделирования числовых уровней ВР

Критерий	Формула расчета
Средняя квадратичная ошибка (СКО)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Квадрат из средней квадратичной ошибки	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Средняя относительная ошибка	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \cdot 100\%$
Симметричная средняя относительная ошибка	$SMAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{(y_i + \hat{y}_i) / 2} \right \cdot 100\%$

Для нечетких моделей в терминах тенденций необходимо разработать собственные критерии качества, характеризующие качество при моделировании нечетких уровней и параметров нечетких тенденций – типа и интенсивности. Предлагается дополнительно оценивать качество нечеткого моделирования нечетких объектов ВР новыми критериями *FPE*, *FCE* и *FME*. Первый критерий *FPE* показывает процент ошибок при моделировании нечетких объектов, например, 8.5%. Второй критерий *FCE* введен для удобства, он показывает количество неверно предсказанных нечетких объектов по отношению к общему количеству термов, например, 2/10. Третий критерий предназначен для вычисления средней ошибки при моделировании $n\tau$ нечетких элементарных тенденций на основе введенной меры различия $\rho(\tau_i, \tau'_i)$ в главе 3:

$$FPE = (ErrorCount / TotalCount) \cdot 100\%,$$

$$FCE = ErrorCount / TotalCount,$$

$$FME = \frac{1}{n\tau} \sum_{i=1}^{n\tau} \rho(\tau_i, \tau'_i).$$

В рассмотренных выше нечетких моделях ВР внутренние и внешние ошибки моделирования нечетких объектов рассчитываются для различных представлений:

- для нечетких уровней временного ряда FPE, FCE ;
- для элементарных тенденций FME ;
- для типов элементарных тенденций $TFPE, TFCE$;
- для интенсивностей элементарных тенденций $RFPE, RFCE$.

Проблемы и ограничения нечетких моделей ВР в терминах нечетких элементарных тенденций

Использование нечеткой продукционной модели для моделирования нечетких объектов временных рядов основано на реализации системы нечеткого логического вывода по схеме Мамдани. Нечеткая модель представляет собой совокупность нечетких множеств и продукционных правил типа

$$R_1: \text{ЕСЛИ } x_1 \text{ это } A_{11} \dots \text{И } \dots x_n \text{ это } A_{1n}, \text{ ТО } y \text{ это } B_1$$

$$R_i: \text{ЕСЛИ } x_1 \text{ это } A_{i1} \dots \text{И } \dots x_n \text{ это } A_{in}, \text{ ТО } y \text{ это } B_i$$

$$R_m: \text{ЕСЛИ } x_1 \text{ это } A_{m1} \dots \text{И } \dots x_n \text{ это } A_{mn}, \text{ ТО } y \text{ это } B_m$$

где $x_k, k=1..n$ – входные переменные;

y – выходная переменная;

A_{ik} – заданные нечеткие множества с функциями принадлежности.

Такие правила строятся для каждого нечеткого объекта временного ряда и затем в системе нечеткого логического вывода на основе алгоритма результаты применения этих правил объединяются (по максимуму) для получения агрегированного результата. В ходе исследований были выявлены следующие проблемы, возникающие при функционировании данной модели:

1. Проблема неполноты правил.
2. Проблема коллизий набора правил.

Проблема неполноты правил заключается в неопределенном поведении системы, если ни одно правило базы правил не сработало.

Частичное решение этой проблемы заключается в поиске наиболее подходящего правила из имеющихся в базе правил по критерию совпадения левых частей (предпосылок). Если существуют такие правила, то среди них выбирается правило с наименьшим уровнем срабатывания.

Проблема коллизий набора правил обнаруживается при моделировании типов нечетких элементарных тенденций. Она заключается в том, что в базе правил могут существовать правила с одинаковыми предпосылками и одинаковыми уровнями срабатывания, а заключения содержат противоположные типы элементарных нечетких тенденций, агрегация которых дает нечеткую тенденцию «Стабильность» (табл. 5.5).

Таблица 5.5

Коллизии правил для типов тенденций

Правило	Частотность
if ((Input0 is Рост) and (Input1 is Рост)) then (Output is Падение)	7
<i>if ((Input0 is Рост) and (Input1 is Падение)) then (Output is Падение)</i>	7
<i>if ((Input0 is Падение) and (Input1 is Падение)) then (Output is Падение)</i>	8
<i>if ((Input0 is Падение) and (Input1 is Рост)) then (Output is Рост)</i>	7
if ((Input0 is Рост) and (Input1 is Рост)) then (Output is Рост)	6

Для решения данной проблемы авторами предложен алгоритм отбора нечетких правил.

5.7.4. Алгоритм отбора нечетких правил

Идея алгоритма решения проблемы коллизии в правилах для типов нечетких элементарных тенденций заключается в сопоставлении каждому правилу числа – «степени принадлежности» правила. Данное число предлагается считать минимумом из степеней принадлежности нечетких объектов, входящих в данное правило (рис. 5.16).

То есть каждому продукционному правилу в нечеткой системе логического вывода приписывается дополнительно к частотной характеристике характе-

ристика, выражающая меру ассоциации распространенности правила, на основе которой будет производиться отбор правил. Можно использовать и другую меру ассоциации – доверительность правила.

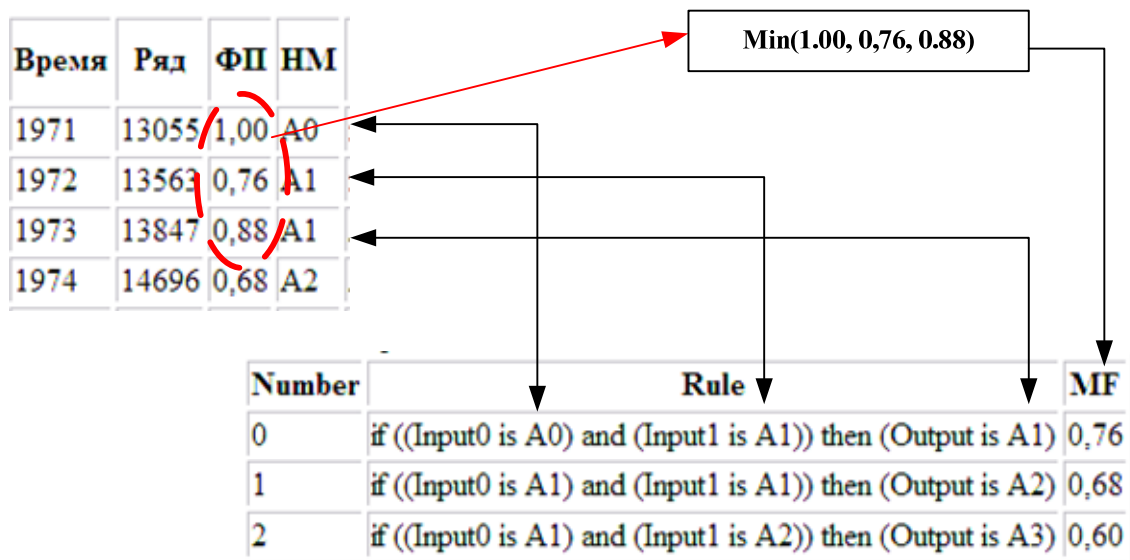


Рис. 5.16. Формирование «степени принадлежности» правила

Отбор правил производится непосредственно перед стандартной процедурой дефаззификации по следующей схеме (рис. 5.17):

1. Объединение лингвистически подобных правил: если правила отличаются только «степенью принадлежности», то выбирается правило с меньшей «степенью принадлежности».

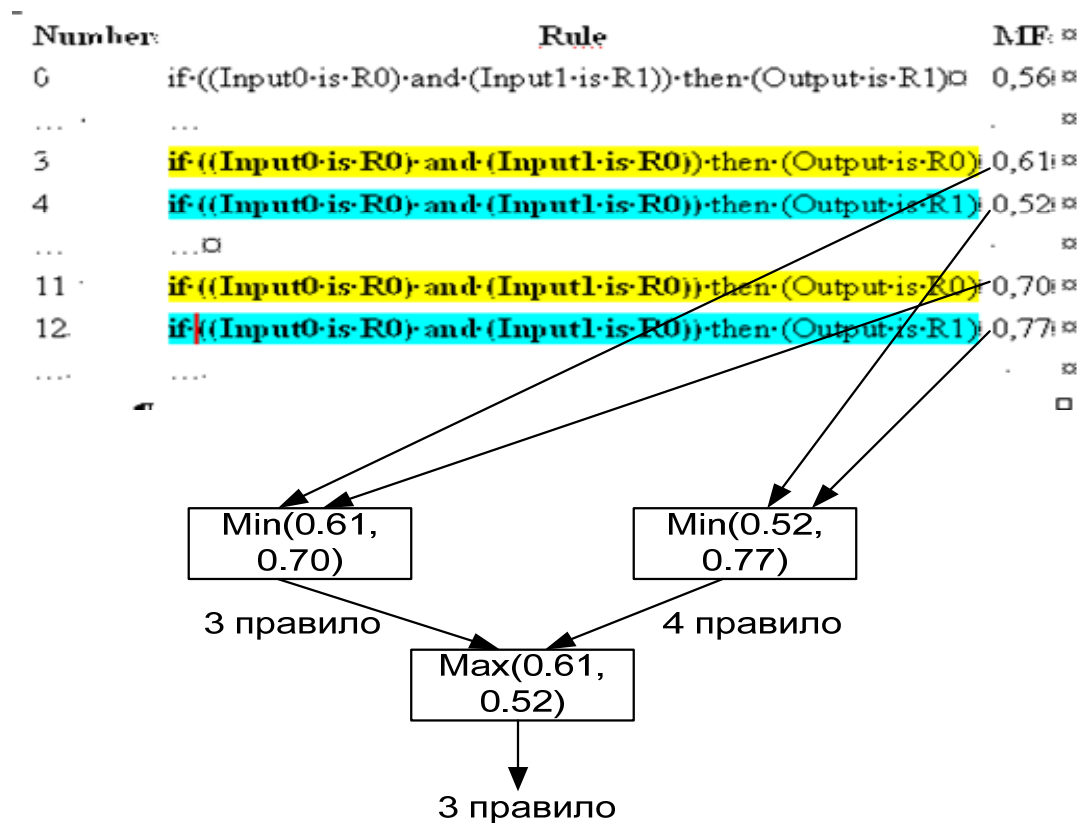


Рис. 5.17. Схема процесса отбора правил

- Выбор наиболее подходящего правила: среди правил, оставшихся после первого шага: выбирается продукционное правило с максимальной «степенью принадлежности».

5.7.5. Моделирование ВР в терминах локальных нечетких тенденций

Представим нечеткую тенденцию (НТ) в виде последовательности нечетких уровней временного ряда, формирующих правила распознавания, что позволит одновременно описать и дать формулу расчета степени принадлежности четкого ряда к тенденции. Таких правил может быть несколько на каждый вид НТ:

$$\text{ЕСЛИ } \tilde{y}_1^{a(i,1,1)} \text{ и } \dots \tilde{y}_k^{a(i,1,k)} \dots \text{и } \tilde{y}_m^{a(i,1,m)}, \text{ ТО } \tau \text{ есть } \tau^i,$$

...

ЕСЛИ $\tilde{y}_1^{a(i,j,1)}$ и ... $\tilde{y}_k^{a(i,j,k)}$... и $\tilde{y}_m^{a(i,j,m)}$, ТО τ есть τ^i .

Здесь $a(i, j, m)$ – индекс нечеткой переменной $\tilde{Y}$;

i – номер вида НТ;

j – номер правила;

k – момент времени.

Введем функционал $Tend$ для получения результата нечеткого логического вывода по данным правилам [Ярушкина, 2007]. Совокупность всех продукционных правил определения видов НТ составляет первый уровень общей системы логического вывода модели нечеткой тенденции ВР (МНТ), которая реализует функционал $Tend$ для получения временного ряда нечетких тенденций (в формализованном виде):

$$R_1 : \quad \tau_t = \{\tau_t^i = \bigcup_{j=1}^{m_j} (T_{t-k+1}(\tilde{y}_{t-k+1}^{a(i,j,k)}))\}.$$

Задача эксперта определить количество правил для каждого вида НТ и их состав.

Уравнение МНТ представим нечетким отношением:

$$\tau_t = (\tau_{t-1}, \dots, \tau_{t-l}) \circ R_2.$$

Конкретный вид НТ определяется следующими правилами:

ЕСЛИ $\tau_{t-1}^{c(i,1,1)}$ и ... $\tau_{t-k}^{c(i,1,k)}$... и $\tau_{t-l}^{c(i,1,l)}$, ТО τ есть τ^i ,

...

ЕСЛИ $\tau_{t-1}^{c(i,j,1)}$ и ... $\tau_{t-k}^{c(i,j,k)}$... и $\tau_{t-l}^{c(i,j,l)}$, ТО τ есть τ^i ,

где $c(i, j, l)$ – индекс вида НТ;

i – номер вида НТ;

j – номер правила;

l – временной лаг.

Множество следующих правил составляет второй уровень вывода:

$$R_2 : \quad \tau_t = \{ \tau_t^i = \bigcup_{j=1}^l \left(\bigcap_{k=1}^l (\tau_{t-k}^{c(i,j,k)}) \right) \}.$$

Для анализа и построения четкого временного ряда по нечеткой модели (функционал deTend) каждому виду нечеткой тенденции сопоставим нечеткий временной ряд, имеющий наибольшую степень принадлежности, в виде правил:

$$\text{ЕСЛИ } \tau^i, \text{ ТО } \tilde{y}_1 \text{ есть } \tilde{y}^{b(i,1)},$$

...

$$\text{ЕСЛИ } \tau^i, \text{ ТО } \tilde{y}_m \text{ есть } \tilde{y}^{b(i,m)},$$

где $b(i, m)$ – индекс нечеткой метки НВР;

i – номер вида НТ.

Множество таких правил, реализующее функционал преобразования временного ряда нечеткой тенденции в НВР *deTend*, формирует третий уровень правил:

$$R_3 : \quad \tilde{y}_t = \{ \tilde{y}_t^k = \bigcup_{i=1}^p \left(\bigcap_{\substack{j=1, \\ b(i,j)=k}}^{m_i} \tau_{t+j-1}^i \right) \}.$$

Задача эксперта определить состав видов НТ и временные моменты, определяющие нечеткую переменную НВР.

Таким образом, модель НТ полностью реализуется многоуровневой системой логических отношений [Ярушкина, 2007]:

$$R_1 \Rightarrow R_2 \Rightarrow R_3,$$

где выходы в виде нечетких переменных одного набора правил подаются на входы следующего набора правил без дефаззификации и фаззификации. Преобразования в нечеткие и четкие значения происходят только в отношениях R_1 и R_3 соответственно.

В качестве инструмента построения функциональной зависимости (уровень R_2) в виде нечетких отношений воспользуемся аппаратом нечетких нейронных сетей. При этом необходимо учитывать: входами и выходами сети яв-

ляются нечеткие значения. Таким образом, будут отсутствовать слои фаззификации и дефаззификации, что сохранит от изменения первоначальных знаний в виде экспертных оценок.

Введем классические нечеткие нейроны, в которых операции сложения и умножения заменяются триангулярными нормами:

- И-нейрон $\beta = T(S(x_1, w_1), S(x_2, w_2))$: импликация предпосылок правила;
- ИЛИ-нейрон $\varphi = S(T(\beta_1, z_1), T(\beta_2, z_2))$: агрегация правил.

Интерпретация этого в лингвистической форме приводит к записям:

ЕСЛИ $(x_1$ или w_1) и $(x_2$ или w_2), ТО β ,

ЕСЛИ $(\beta_1$ и z_1) или $(\beta_2$ и z_2), ТО φ .

Веса w и z можно рассматривать как степени влияния соответствующего входа на выход. Так веса в И-нейроне – это нечеткая переменная, выражающая степень невлипания:

- при $w_1=0$: $\beta = T(S(x_1, 0), S(x_2, w_2)) = T(x_1, S(x_2, w_2))$ – вход x_1 влияет на результат правила;
- при $w_1=1$: $\beta = T(S(x_1, 1), S(x_2, w_2)) = T(1, S(x_2, w_2)) = S(x_2, w_2)$ – вход x_1 не влияет на результат правила.

А вес в ИЛИ-нейроне – это нечеткая переменная, выражающая степень влияния:

- при $z_1=0$: $\varphi = S(T(\beta_1, 0), T(\beta_2, z_2)) = S(0, T(\beta_2, z_2)) = T(\beta_2, z_2)$ – правило β_1 не влияет на результат;
- при $z_1=1$: $\varphi = S(T(\beta_1, 1), T(\beta_2, z_2)) = S(\beta_1, T(\beta_2, z_2))$ – правило β_1 влияет на результат.

На основе таких нейронов можно построить сеть логического вывода по Мамдани для нахождения нечеткой тенденции, добавив в R_2 весовые коэффици-

циенты. Каждому виду нечеткой локальной тенденции соответствует своя нейронная сеть.

Контрольные вопросы

1. В чем заключается подход к прогнозированию ВР в терминах элементарных тенденций?
2. Приведите три варианта представления модели временного ряда элементарных тенденций.
3. Какие зависимости определяются в модели F2S?
4. Чем отличаются модели F1N, F3N1S и F2S?
5. На чем основана оценка качества моделей прогнозирования?
6. Приведите критерии качества нечеткого моделирования числовых уровней ВР.
7. Какие существуют проблемы и ограничения нечетких моделей ВР?
8. Для каких целей используется алгоритм отбора нечетких правил?
9. Опишите алгоритм моделирования ВР в терминах локальных нечетких тенденций.

5.8. Поиск аномалий во временных рядах

При решении задач экспертной деятельности с целью диагностики процессов, интерпретированных нечеткими временными рядами, целесообразно применять методы контроля, основанные на поиске аномалий. Эти методы, как представляется, могут включать сопоставление НВР, отражающего реализованную динамику процесса с НВР, с ожидаемой, требуемой динамикой.

Если имеется эталонный НВР, можно построить новый НВР, выражающий отклонения реализованного НВР от эталонного НВР. Возможны следующие значения отклонений, представленные нечеткими термами, такими как «Без отклонений», «Значительные отклонения», «Незначительные отклонения» и др. Тогда поиск аномалий заключается в формировании и анализе состава кластеров уровней нечеткого временного ряда отклонений. Так, например, кла-

стер уровней «Значительные отклонения» пуст, то можно классифицировать исходный ВР как временной ряд без аномалий.

Уточнение задачи поиска аномалий применительно к нечеткому временному ряду, обозначенной в работе [Батыршин, 2007], в рамках структурно-лингвистического подхода можно связать с выявлением лексических (нетипичных локальных нечетких тенденций), синтаксических (выявление нетипичных правил следования нечетких тенденций, выявление рассогласований в общих нечетких тенденциях) ошибок в НВР на основе грамматики языка *LANG*. Грамматика языка *LANG* может быть задана экспертно или сгенерирована по исходному временному ряду при решении задач сегментации, кластеризации и классификации, рассмотренных в предыдущих разделах, и формировании временных рядов нечетких уровней, элементарных и локальных тенденций.

Решение задачи поиска аномалий может быть основано на предположении, что аномальным является поведение ВР, выраженное в терминах редко встречающихся значений или недопустимых значений.

В связи с этим поиск аномалий – задача, которую можно решать на различных уровнях представления исходного ВР в зависимости от поставленных целей.

В качестве аномальных значений могут быть рассмотрены следующие типы значений ВР:

- 1) уровни исходного временного ряда;
- 2) нечеткие уровни нечеткого временного ряда;
- 3) нечеткие элементарные тенденций (аномальные термы) временного ряда элементарных тенденций;
- 4) нечеткие локальные тенденции, то есть паттерны, составленные из элементарных нечетких тенденций (аномальные лексемы);
- 5) правила следования локальных нечетких тенденций, то есть аномальные синтаксические правила;
- 6) общая нечеткая тенденция.

Для обнаружения редко встречающихся нечетких тенденций ВР целесообразно использовать частотный анализ НТ. При поиске недопустимых значе-

ний необходимо экспертно идентифицировать, какие значения (например, нечетких уровней, нечетких тенденций) являются недопустимыми. Тогда поиск аномалий можно свести к задаче сегментирования, кластеризации, классификации ВР в базисе нечетких тенденций с последующим выполнением частотного анализа.

Подход к обнаружению недопустимых правил следования НТ может использовать не частотный портрет тенденций, а синтаксический. Обнаружение недопустимых правил следования локальных нечетких тенденций связано с отсутствием обнаруженных правил в имеющейся базе правил нечеткой модели ВР. Одним из вариантов решения указанной задачи может служить применение нечетких грамматик, в которых фиксируется допустимое поведение. Тогда, если последовательность правил выводима в этой грамматике, то аномалии не обнаружены, иначе существует их потенциальная возможность.

Ниже рассмотрим решение задачи поиска аномалии поведения во ВР в рамках второго подхода, рассматривая аномальности как недопустимые синтаксические конструкции грамматики языка тенденций, построенного на основе эталонного ВР.

5.8.1. Нечеткие грамматики для поиска аномалий

Нечетким языком L в конечном алфавите V_t является нечеткое подмножество множества всех конечных цепочек $V_t^* = \{x\}$, полученных с помощью конкатенации элементов V_t :

$$\mu_L : V_t^* \rightarrow [0,1],$$

где $\mu_L(x)$ является степенью принадлежности x языку L и может быть интерпретирована как степень правильности цепочки x или степень возможности ее использования. По аналогии с формальным языком возможны, по крайней мере, три описания нечетких языков: нечеткие грамматики, нечеткие автоматы и нечеткие алгебраические выражения.

Нечеткой грамматикой называют шестерку [Аверкин, 1986]

$$G = \langle V_N, V_T, P, S, L, \varphi \rangle,$$

где V_N – множество нетерминальных символов;

V_T – множество терминальных символов;

S – начальный символ $S \in V_N$;

P – конечное множество правил подстановки вида
 $u \xrightarrow{p} v; u, v \in (V_N \cup V_T)^*, p \in P$;

L – множество весов (например, дистрибутивная решетка с 0 и 1);

$\varphi: P \rightarrow L, \varphi(p)$ – степень принадлежности выводу правила $p \in P$.

Пусть задана грамматика G ; $u, v \in V^*$. Говорят, что u непосредственно порождает v со степенью φ_1 , если найдутся такие $u_1, u_2, x, y \in V^*$, $p \in P: \varphi(p) = \varphi_1$, что $u = u_1 x u_2$, $v = u_1 y u_2$, $x \xrightarrow{p} y$. Это можно обозначить как $u \xrightarrow{\varphi_1, p} v$.

Пусть $u, v, z_0, \dots, z_m \in V^*$. Последовательность $z_0, \dots, z_m$ называется выводом u из v , а v – выводимой из u в грамматике G , если существует последовательность подстановок из P :

$$u = z_0 \xrightarrow{\varphi_1, p_1} z_1 \rightarrow \dots \rightarrow z_{m-1} \xrightarrow{\varphi_{mpm}} z_m = v.$$

Для лучшего понимания структуры предложения, являющегося образом НВР, можно построить дерево грамматического разбора. Узлы дерева помечаются терминалами и нетерминалами. Самые нижние узлы – терминалы, самый верхний – начальный символ. Обычно дерево строят снизу вверх.

Одним из способов описания нечетких языков являются *нечеткие распознающие автоматы*. Считается, что автомат допускает слово некоторого нечеткого языка, если при подаче на вход последовательности сигналов, соответствующих заданному слову, нечеткий автомат генерирует выходной сигнал, указывающий на степень принадлежности данного слова нечеткому языку.

Нечетким конечным автоматом называется упорядоченная шестерка [Аверкин, 1986]

$$A = \langle U, X, Y, s_0, \delta, \sigma \rangle,$$

где $U = \{a_1, \dots, a_2\}$ – конечное множество входов;

$X = \{x_1, \dots, x_n\}$ – конечное множество состояний;

$Y = \{y_1, \dots, y_p\}$ – конечное множество выходов;

$\delta: X \times U \times X \rightarrow L$ – функция переходов;

$\sigma: X \times Y \rightarrow L$ – функция выходов;

s_0 – нечеткое начальное состояние.

Функция δ порождает множество нечетких матриц переходов, функция σ порождает нечеткую матрицу выхода $\sigma = (\sigma_{xij})$, $1 \leq i \leq n, 1 \leq j \leq p$.

5.8.2. Применение нечетких грамматик для описания аномалий НВР

Рассмотрим возможность описания НВР при помощи нечетких грамматик с дальнейшей генерацией нечеткого автомата, распознающего допустимые и аномальные последовательности нечетких тенденций НВР. Обозначим процедуру генерации нечеткого автомата функционалом R_F_A .

Временной ряд, выраженный метками типов элементарных нечетких тенденций «Рост», «Стабилизация», «Падение», можно рассматривать как предложения некоторого языка.

Рассмотрим лингвистический ряд типов ЭТ на рис. 5.18 и сгенерируем для него грамматику:

$$V_N = \{ \langle n \rangle, \langle c \rangle, \langle p \rangle, \langle n \rangle \}, V_T = \{ n, c, p \},$$

P – конечное множество правил подстановки (см. табл. 5.6):

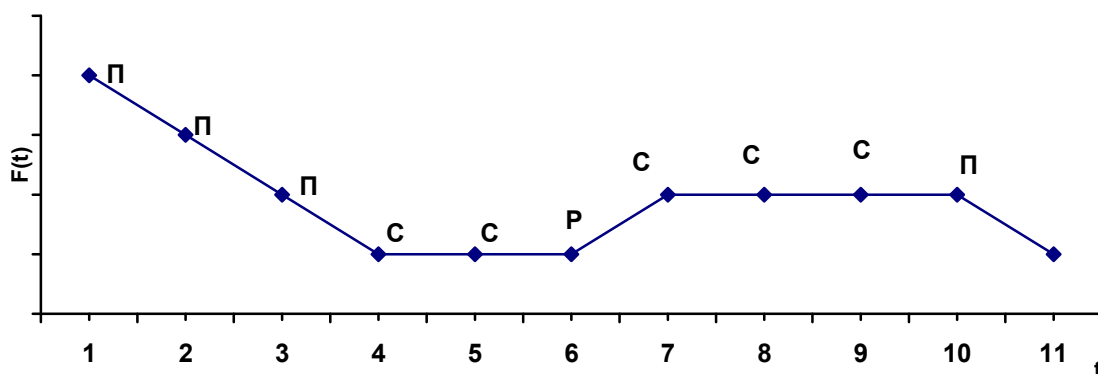


Рис. 5.18. Пример лингвистического ВР из типов элементарных тенденций

Таблица 5.6

Правила нечеткой грамматики			
Правила	Степени принадлежности	Правила	Степени принадлежности
p1: $\langle n \rangle ::= \langle p \rangle$	1	p8: $\langle c \rangle ::= c \langle p \rangle$	0,8
p2: $\langle n \rangle ::= \langle c \rangle$	1	p9: $\langle c \rangle ::= c \langle p \rangle$	0,7
p3: $\langle n \rangle ::= \langle p \rangle$	1	p10: $\langle c \rangle ::= c$	0,9
p4: $\langle p \rangle ::= p \langle p \rangle$	0,8	p11: $\langle p \rangle ::= p \langle p \rangle$	0,8
p5: $\langle p \rangle ::= p \langle c \rangle$	0,7	p12: $\langle p \rangle ::= p \langle c \rangle$	0,7
p6: $\langle p \rangle ::= p$	0,9	p13: $\langle p \rangle ::= p$	0,9
p7: $\langle c \rangle ::= c \langle c \rangle$	0,6	—	—

Построим дерево грамматического разбора для строки «пппср» (рис. 5.19).

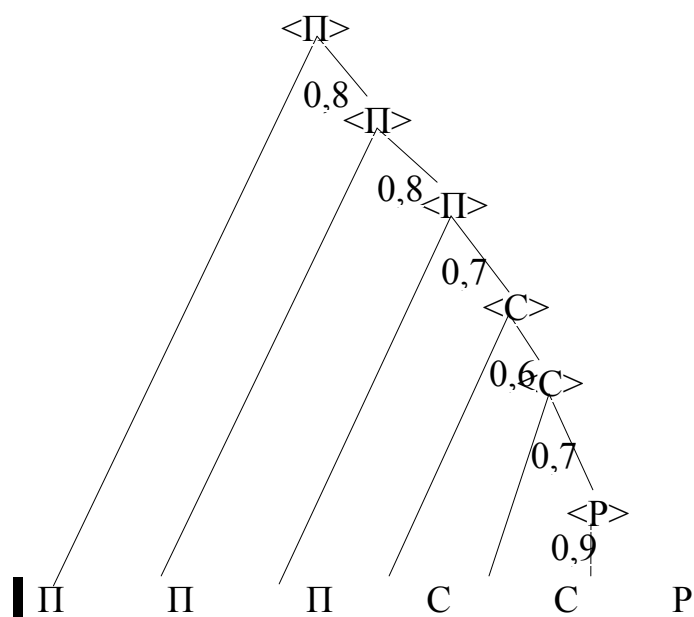


Рис. 5.19. Дерево грамматического разбора

Если принять грамматику за пессимистическую, то степень принадлежности выводу данной последовательности типов элементарных тенденций «пппср» равна 0,6.

Построим нечеткий автомат, распознающий данный язык, сгенерированной грамматики. Матрица переходов между состояниями имеет вид, представленный в табл. 5.7.

Таблица 5.7

Матрица переходов нечеткого автомата

	1	2	3	4	5	6	7	8	9	10	11	12	13
0	п, 1	с, 1	р, 1										
1				п, 0,8	п, 0,7	п, 0,9							
2							с, 0,6	с, 0,8	с, 0,7	с, 0,9			
3											р, 0,8	р, 0,7	р, 0,9
4				п, 0,8	п, 0,7	п, 0,9							
5							с, 0,6	с, 0,8	с, 0,7	с, 0,9			

6													
7							с, 0.6	с, 0.8	с, 0.7	с, 0.9			
8				п, 0.8	п, 0.7	п, 0.9							
9											р, 0.8	р, 0.7	р, 0.9
10													
11											р, 0.8	р, 0.7	р, 0.9
12							с, 0.6	с, 0.8	с, 0.7	с, 0.9			
13													

Для нечеткого автомата множество входов $U=V_T=\{п,с,р\}$, множество состояний $X=\{0,1,2,3,4,5,6,7,8,9,10,11,12,13\}$, начальные состояния $S_0=\{0\}$, $\sigma_0=(1,0,0,0,0,0,0,0,0,0,0,0,0,0)$, конечные состояния $\sigma_1=(0,0,0,0,0,1,0,0,0,1,0,0,1)$, $X_f=\{6,10,13\}$.

Если на вход нечеткого автомата подать слово «пппсрр», то выход автомата вычисляется следующим образом:

$$\mu(\text{пппсрр}) = \sigma_0 \circ T_p \circ T_p \circ T_p \circ T_c \circ T_c \circ T_r \circ \sigma_1 = 0.6.$$

Если попробовать вывести цепочку, не принадлежащую данному языку (например, «ппспр»), то на выходе получим 0, что является признаком аномальной последовательности типов тенденций.

Контрольные вопросы

1. Какие методы контроля, основанные на поиске аномалий, вы знаете?
2. Какие типы значений могут быть рассмотрены в качестве аномальных значений?
3. Дайте определение нечеткому языку L .
4. Дайте определение нечеткой грамматики
5. Что такое нечеткие распознающие автоматы?

5.9. Резюмирование нечеткого временного ряда

Резюмирование предполагает краткое описание ключевых тенденций нечеткого временного ряда в краткой форме. Для числовых временных рядов обычно для этого используется описательная статистика. Для нечетких временных рядов, анализ которых основывается на нечетких множествах, имеющих лингвистическую интерпретацию, целесообразно предложить лингвистическое резюмирование, кратко выражающее наиболее существенные паттерны поведения НВР.

Для этих целей введем дескрипторное описание НВР в терминах результатов решения задач извлечения знаний из НВР, рассмотренных выше.

Дескрипторное описание НВР будет соответствовать наиболее типичным паттернам НВР:

$$D_ \tilde{Y} = \langle G\tau, \max_ \tau, \max_ N\tau, \max_ Rule_ e\tau \rangle ,$$

где $G\tau$ – общая тенденция НВР, полученная на основе функционала $GTend$;

$\max_ \tau$ – элементарная тенденция с максимальной частотной характеристикой, вычисленной с помощью функционала $g_Clus(E\tau)$;

$\max_ N\tau$ – локальная нечеткая тенденция с максимальной частотной характеристикой, вычисленной с помощью функционала $g_Clus(N\tau)$;

$\max_ Rule_ e\tau$ – правило следования элементарных тенденций с максимальной частотной характеристикой, полученных на основе функционала $g_Clus(Rule_ e\tau)$.

На основе дескрипторного описания лингвистическая трансляция результатов моделирования нечетких тенденций как процесса резюмирования ВР может генерироваться функционалом SMR в виде заполнения некоторого лингвистического шаблона, типа

«Временной ряд» $is_name=Name$ «характеризуется общей тенденцией: tin »
 $is_Ttend=\tilde{v}$ «интенсивность» $is_Rtend=\tilde{\alpha}$.

«Типичные элементарные тенденции: tin » $is_Ttend=\tilde{v}$ «интенсивность»
 $is_Rtend=\tilde{\alpha}$.

«Типичные локальные тенденции: tin » $is_Ttend=\tilde{v}$ «интенсивность»
 $is_Rtend=\tilde{\alpha}$ «длительность» $is_Dt=\Delta\tilde{t}$.

«Типичные правила следования элементарных тенденций» $iz_Rule_et=R_{max}$.

Другой подход к резюмированию может быть основан на построении нечетких оценок моментов и интервалов времени: pw -ограничений и ww -ограничений. Эти ограничения могут задаваться экспертно, на основе хронологии и естественной классификации временных интервалов, или вычисляться на основе сегментации НВР на локальные или элементарные нечеткие тенденции и сопоставлении временным участкам общей тенденции.

Задача резюмирования, таким образом, имеет несколько аспектов и уровней ее представления: лингвистическое, числовое, временное, поуровневое. Указанная особенность позволяет сделать вывод о целесообразности введения некоторого языка запросов, с помощью которого можно будет специфицировать параметры процедуры резюмирования SMR .

Контрольные вопросы

1. Что значит термин «резюмирование временного ряда»?
2. Приведите пример дескрипторного описания ВР.
3. Опишите различные подходы к резюмированию ВР.

5.10. Принципы построения интеллектуальной системы анализа нечетких тенденций временных рядов

Временные ряды, получившие название нечетких временных рядов (НВР), обладают специфическим свойством: в значениях уровней НВР отобра-

жена контекстно-зависимая лингвистическая трихотомия «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ», в отличие от традиционной лингвистической дихотомии «ЗНАК-ЗНАЧЕНИЕ». Лингвистическая трихотомия расширяет семантическое пространство «ЗНАКА» за счет возможности использования четких и нечетких значений. Согласно концепции точности/неточности понятий Заде, каждое понятие, знак, имеет содержание (*value*), которое может быть задано точно или неточно (*v-precise*, *v-imprecise*).

Термин «ЗНАЧЕНИЕ» применительно к НВР будем трактовать как точную числовую оценку уровня временного ряда (*v-precise*). Для указанного аспекта «ЗНАКА» разработаны эффективные методы анализа ВР. «ОБОЗНАЧЕНИЕ» – это лингвистическая оценка, выражающая уровень ВР в контекстно-зависимых терминах конечного пользователя (*v-imprecise*). Наиболее удобными математическими средствами моделирования и описания лингвистических оценок могут служить средства нечетких множеств и нечетких шкал.

Следует отметить ощутимый разрыв между возможностями, предоставляемыми традиционными инструментальными системами поддержки задач анализа ВР, и практическими потребностями конечных пользователей в автоматизации анализа тенденций нечетких временных рядов. Этот разрыв обусловлен, с одной стороны, ориентацией традиционных систем на использование лингвистической дихотомии, а с другой – недостаточным развитием формального описания моделей и методов анализа нечетких временных рядов.

Одним из способов преодоления указанной проблемы является разработка и развитие интеллектуальных систем анализа НВР, в которых исходные данные, выходные результаты определены на понятном пользователю языке, выражающем лингвистическую трихотомию в семантике состояний и изменений объектов предметной области, а модели и методы анализа и прогноза основываются на технологиях быстро развивающегося направления Time Series Data Mining (TSDM).

Можно обозначить в соответствии с общими методологическими положениями моделирования ВР, сформулированными в главе 3, следующие особенности интеллектуальных систем анализа нечетких временных рядов (ИСА-НВР) [Афанасьева и др., 2008]:

1. Ориентация в представлении входных и выходных данных на лингвистическую трихотомию «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ».

Например, «ЗНАК» – это уровень временного ряда, «ЗНАЧЕНИЕ» – числовое значение уровня конкретного временного ряда, «ОБОЗНАЧЕНИЕ» – это нечеткая метка (нечеткое множество), соответствующая числовому уровню временного ряда. Для перехода от «ЗНАКА» к «ЗНАЧЕНИЮ» необходимы процедуры измерения, вычисления или извлечения, от «ЗНАКА» к «ОБОЗНАЧЕНИЮ» – процедуры построения нечеткой шкалы, от «ЗНАЧЕНИЯ» к «ОБОЗНАЧЕНИЮ» – процедуры оценки по нечеткой шкале (фаззификация).

Например, «ЗНАК» – это временной ряд, «ЗНАЧЕНИЕ» – реализации временного ряда, конкретный временной ряд (числовой или нечеткий), «ОБОЗНАЧЕНИЕ 1» – это модель тенденции поведения в некоторой прикладной области, построенная на конкретном временном ряду, «ОБОЗНАЧЕНИЕ 2» – это аддитивная модель в виде временных рядов, определяющих трендовую, сезонную, периодическую и случайную компоненты, «ОБОЗНАЧЕНИЕ 3» – это конкретная функция, моделирующая поведение временного ряда. Для перехода от «ЗНАКА» к «ЗНАЧЕНИЮ» необходимы процедуры измерения и оценки, от «ЗНАЧЕНИЯ» к «ОБОЗНАЧЕНИЮ» – процедуры идентификации поведения, моделирования и вычисления новых «ЗНАЧЕНИЙ».

2. Применение множеств нечетких оценок, задающих семантику предметной области и ориентированных на задачи конечного пользователя.

3. Использование нечетких лингвистических шкал для генерации нечеткой оценки.

4. Извлечение знаний из ВР в лингвистической форме продукционных правил, содержащих зависимости понятные конечному пользователю.

5. Способность к обучению и развитию.

Указанные особенности определяются тем фактом, что целью создания интеллектуальных систем анализа НВР является обработка и управление новыми типами данных, образованными сложными слабо-формализованными

структурами представления семантических понятий, для извлечения новых знаний о состоянии и тенденциях развития объектов предметной области.

Поэтому проектировщику ИСА-НВР необходимо определить в первую очередь семантику системы, выраженную следующими стратами знаний:

1. *Где-знания*. Область применения системы, пространственные ограничения, условия, соотношение с другими системами.
2. *Кто-знания*. Пользователи системы, компетенции и требования пользователей.
3. *Зачем-знания*. Цель применения системы и совокупность решаемых задач.
4. *Что-знания*. Входные и выходные данные системы. Их типы, структуры, модели.
5. *Как-знания*. Функции, операции, алгоритмы, определяющие функционирование системы.
6. *Почему-знания*. Значимость, актуальность применения системы.
7. *Когда-знания*. Временные ограничения.
8. *Сколько-знания*. Количественные показатели, постоянные, затраты, показатели эффективности.
9. *Какие-знания*. Качественные оценки системы.

Заметим, что страты знаний могут быть укрупнены в такие объекты системы, как вход-выход с возможными ограничениями (страты 4, 1, 2, 7), функции (страты 3, 5), критерии (страты 6, 8, 9). Эти укрупненные страты знаний могут быть взяты за основу при концептуальном проектировании ИСА-НВР.

Рассмотрим в рамках концептуального проектирования абстрактной интеллектуальной системы анализа нечетких временных рядов S определение ее модели, ориентированной на лингвистическую трихотомию.

Обобщенная компонентная модель системы S , включающая вход Y , выход $\tilde{Y}$, множество моделей преобразования входа в выход F , множество критериев K , используемых для идентификации модели, оценки ее параметров и критериев, определяющих качество моделей F , представима в виде $Cs = \langle Y, \hat{Y}, F, K \rangle$.

Эта традиционная модель системы может иметь, по крайней мере, двухуровневое представление для задач в конкретной предметной области: на одном уровне – в терминах конечного пользователя, на другом – в терминах математических моделей и методов.

Каждая компонента модели C_s характеризуется своей семантикой, имеющей внутреннюю и внешнюю интерпретации. Внешняя интерпретация семантики задает интенсификацию модели S в форме лингвистического представления каждой компоненты, выраженных в терминах конечного пользователя, а внутренняя – определяет экстенсификацию модели S в виде правил выполнения допустимых операций для каждой компоненты, совокупности отношений и математических зависимостей между компонентами модели. Таким образом, внутренняя семантика определяется моделями решения задач анализа НВР и используется для генерации «ЗНАЧЕНИЙ», а внешняя семантика задает субъектно-объектное отображение предметной области и объектов известных исследователю формальных систем в компоненты модели S и определяет «ОБОЗНАЧЕНИЕ».

Определим для модели C_s системы S функционал $R_s = R(Y, \hat{Y}, F, K)$, задающий внутреннюю семантику, и функционал $P_s = P(Y, \hat{Y}, F, K)$ для представления внешней семантики.

Тогда концептуальная модель абстрактной интеллектуальной системы анализа НВР S может быть определена в виде $M_s = \langle C_s, R_s, P_s \rangle$, и в этой модели представлена лингвистическая трихотомия «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ». Сопоставление компонент и семантических моделей систе-

мы S при анализе конкретного класса НВР обычно осуществляет проектировщик системы, который принимает наиболее адекватное решение исходя из контекста среды, в которой наблюдается НВР, и своего опыта.

В предыдущих разделах настоящей главы в рамках структурно-лингвистического подхода были обозначены функционалы, реализующие задачи $TSDM$ (см. табл. 5.8), совокупность которых является, как нам представляется, основой для формализации функционала P_s , представляющего внешнюю семантику системы S .

Функционалы структурно-лингвистического подхода для решения задач интеллектуального анализа временных рядов

Задачи	Функционалы
Сегментация ВР	<i>Etend</i> <i>Stend</i>
Резюмирование	<i>SMR</i>
Прогноз	<i>F2S</i> <i>FIN</i> <i>F3NIS</i> $R_1 \Rightarrow R_2 \Rightarrow R_3$
Кластеризация	<i>f_Clus</i>
Классификация	<i>G_Tend</i>
Частотный анализ	<i>g_Clus</i>
Поиск аномалий	<i>R_F_A</i>
Извлечение ассоциативных правил	<i>Rule</i>

Приведенные функционалы могут быть использованы и на случай многомерных ВР. Так, в пространстве однородных объектов по отдельной временной координате могут быть сформированы пространственные ВР общих свойств объектов.

С помощью ACL-шкалы могут быть оценены и сформированы пространственные элементарные тенденции, временной анализ которых в рамках структурно-лингвистического подхода позволит выявлять тенденции многомерных объектов.

Контрольные вопросы

1. Каким специфическим свойством обладают ВР?
2. Для каких целей разрабатывают интеллектуальные системы анализа ВР?
3. Какие особенности имеют интеллектуальных систем анализа ВР?
4. Какие страты знаний должен выделить проектировщик ИСА-НВР?
5. Выделите функционалы структурно-лингвистического подхода для решения задач интеллектуального анализа временных рядов.

Выводы

Изложенный в настоящей главе структурно-лингвистический подход, основанный на технологии извлечения знаний о тенденциях ВР, не противоречит основным методологическим принципам моделирования временных рядов (см. главу 3). Он расширяет сферу применимости систем моделирования ВР на данные, обладающие объективной нечеткостью, и нацелен в первую очередь на распознавание различных тенденций и их параметров, происходящих в исследуемом процессе. В то же время использование, автоматическая обработка и вывод формализованных нечетких лингвистических высказываний обеспечат снижение времени на разведывательный анализ данных, который значительно превосходит время, затрачиваемое на моделирование ВР, и вносит весомую долю в неточность результатов при принятии решений.

Отметим новые принципы интеллектуальных систем нечеткого моделирования ВР на основе структурно-лингвистического подхода: снижение требований к квалификации конечных пользователей, высокая степень интерпретируемости, понятности для конечного пользователя. Реализация этих принципов позволяет надеяться, что структурно-лингвистический подход в нечетком моделировании и анализе нечетких тенденций временных рядов найдет применение в первую очередь как инструментальное средство, дополняющее арсенал методов моделирования ВР, ориентированное на массового пользователя, для которых допустимы интервальные числовые оценки уровней.

Библиографический список

1. [Аверкин и др., 1986] Нечеткие множества в моделях управления и искусственного интеллекта / А. Н. Аверкин, И. З. Батыршин, А. Ф. Блишун и др. ; под ред. Д. А. Поспелова. – М. : Наука. Гл. ред. физ.-мат. лит., 1986. – 312 с.
2. [Афанасьева, 2008] Афанасьева, Т. В. Структурно-лингвистический подход в анализе нечетких временных рядов / Т. В. Афанасьева // Программные продукты и системы. – 2008. – №4(84). – С. 61-65.
3. [Афанасьева и др., 2008] Афанасьева, Т. В. Концептуальная модель гибридной системы анализа нечетких временных рядов / Т. В. Афанасьева, С. Г. Валеев, Н. Г. Ярушкина // Вопросы современной науки и практики. Университет им. В. И. Вернадского. Т.2. Серия «Технические науки». – Тамбов: ТГТУ, 2008. – № 4(14). – С. 85-91.
4. [Афанасьева, 2009] Афанасьева, Т. В. Моделирование нечеткого временного ряда на основе элементарных тенденций/ Т. В. Афанасьева // Труды V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). – Т.2. – М. : Физматлит, 2009. – С. 291-297.
5. [Афанасьева, 2009] Афанасьева, Т. В. Метод прогнозирования временных рядов на основе нечетких тенденций / Т. В. Афанасьева ; под ред. д. т. н., проф. Ю. В. Полянского, д. ф.-м. н., проф. В. Л. Леонтьева // Труды Седьмой Международной конференции «Математическое моделирование физических, экономических, технических, социальных систем и процессов», 2-5 февраля 2009 года, г. Ульяновск. – Ульяновск : УлГУ, 2009. – С. 33-35.
6. [Батыршин и др., 2007] Батыршин, И. З. Модели и методы перцептивного дата майнинга временных рядов для систем поддержки принятия решений / И. З. Батыршин, Л. Б. Шереметов // Нечеткие системы и мягкие вычисления. –2007. – Т. 2. – №1.

7. [Борисов и др., 2007] Борисов, В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия – Телеком, 2007. – 284 с.
8. [Ярушкина, 2004] Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
9. [Яхьяева, 2006] Яхьяева, Г. Э. Нечеткие множества и нейронные сети: учебное пособие / Г. Э. Яхьяева. – М. : Интернет-Университет Информационных технологий: БИНОМ. Лаборатория знаний, 2006. – 316 с.

ГЛАВА 6. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ НЕЧЕТКОГО МОДЕЛИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ И АНАЛИЗА НЕЧЕТКИХ ТЕНДЕНЦИЙ

Введение

Для исследования результативности предложенного структурно-лингвистического подхода моделирования ВР разработана программная система *FuzzyTend* для нечеткого моделирования и анализа нечетких тенденций временных рядов [Афанасьева и др., 2009], реализующая принципы построения интеллектуальных систем анализа ВР, изложенные в главе 5.

Отличительными особенностями программной системы *FuzzyTend* являются:

1. Решение новых задач анализа нового объекта ВР – нечеткой тенденции. Выявление новых закономерностей поведения ВР в форме продукционных правил зависимостей между элементарными нечеткими тенденциями:
2. Автоматизация построения и поиска наилучшей модели ВР по комплексу критериев при допустимом уровне погрешности.
3. Высокая интерпретируемость результатов и отсутствие требований к математической подготовке конечных пользователей в области анализа ВР.
4. Решение комплекса задач Time Series Data Mining (TSDM).

6.1. Программная система моделирования временных рядов на основе элементарных нечетких тенденций *FuzzyTend*

Программная система *FuzzyTend* реализована в виде пакета прикладных программ. Данный пакет содержит в себе шесть параметрических нечетких моделей ВР и одну нейросетевую модель, каждая из которых имеет две модификации на основе включения/выключения алгоритма отбора правил. Совокупность нечетких моделей включает авторские нечеткие модели временных рядов на основе элементарных нечетких тенденций **F2S**, **F1N**, **F3N1S**, а также нечет-

кое F-преобразование [Perfilieva, 2006]. Кроме перечисленных в программной системе *FuzzyTend* реализованы нечеткие S-модель [Song, 1993] и D-модель [Şah, 2004], которые будут использованы для сравнения результатов моделирования ВР.

Интеллектуальные модели могут быть использованы автономно или в комплексе под управлением автоматической процедуры выбора наилучшей из набора конкурирующих интеллектуальных моделей ВР по нескольким критериям.

Как правило, специфика разработки математического пакета определяет жесткие требования к интерфейсу таких программ, что также было учтено при проектировании:

- Удобство ввода исходных данных: поддержка различных форматов файлов с данными, а также автоматическое генерирование наилучшей модели и многоуровневого представления результатов моделирования при задании минимального количества параметров, зависящих от предметной области.
- Наглядное отображение графиков.
- Удобство одновременной работы с множеством моделей в проекте.
- Просмотр графиков различных компонент, на которые раскладывается исходный временной ряд в рамках используемой модели.
- Возможность применения к одному временному ряду несколько моделей.
- Интерфейс для настройки параметров моделей временных рядов и шкалы.
- Возможность дополнительного анализа параметров нечетких тенденций.
- Возможность анализа остатков моделей.

- Отображение результатов моделирования и прогнозирования в табличной, графической, числовой и лингвистической формах, понятных конечному пользователю.
- Сохранение полученных графиков в файл.
- Генерация html-отчета применения модели с возможностью его просмотра и сохранения на диске.
- Сохранение проекта в файл и загрузка проекта из файла.

На рис. 6.1 приведено главное окно программы, в котором представлена вся необходимая информация для исследования поведения ВР.

Библиотека моделей временных рядов представляет собой объектно-ориентированные модели, реализующие модели временных рядов: нейросетевые модели для моделирования числовых представлений и нечеткие модели для моделирования нечеткого представления параметров элементарных тенденций, описанные в главе 5.

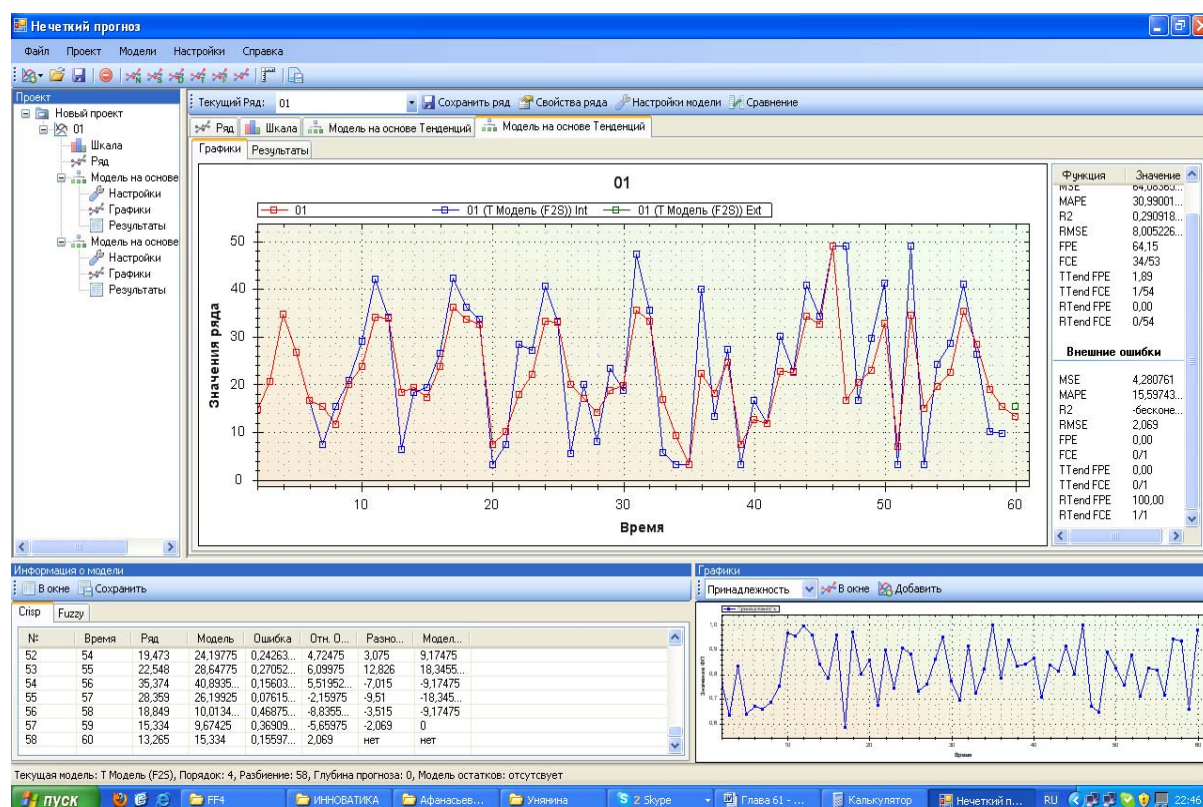


Рис. 6.1. Интерфейс программной системы

На рисунке 6.2 приведена структурная организация программного обеспечения нечеткого моделирования и анализа нечетких тенденций временных рядов и библиотеки моделей ВР. Структурно библиотека моделей ВР представляет собой набор модулей, реализующих интерфейс моделей прогнозирования и использующих библиотеки более низкого уровня.

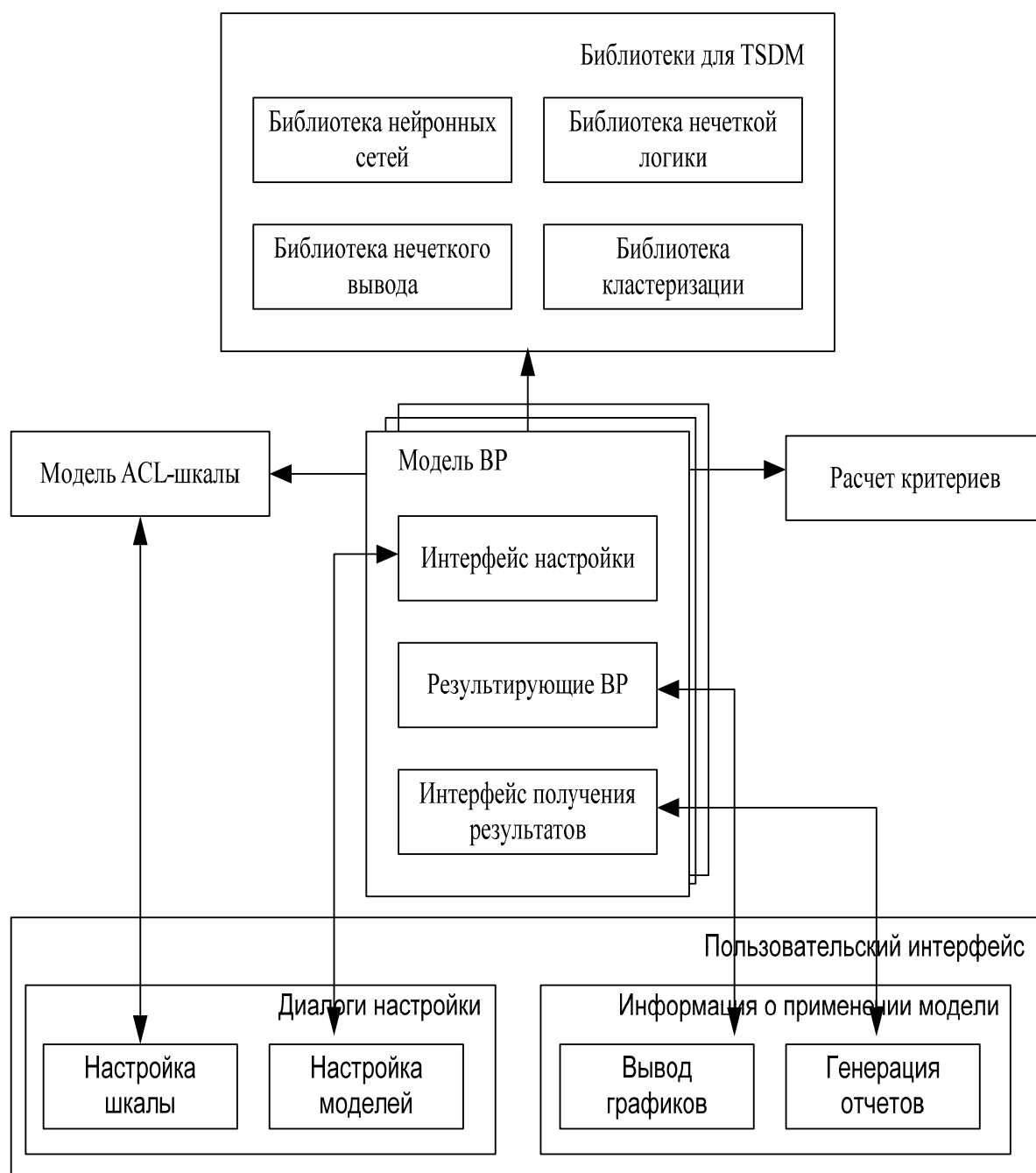


Рис. 6.2. Структура пакета прикладных программ нечеткого моделирования

После загрузки временного ряда программа предлагает настроить лингвистическую ACL-шкалу. Диалоговое окно настройки шкалы показано на рисунке 6.3. В этом окне отображаются графики функций принадлежности для составляющих шкалы – базовая шкала, шкала типов, шкала интенсивностей, а также шкала разностей.

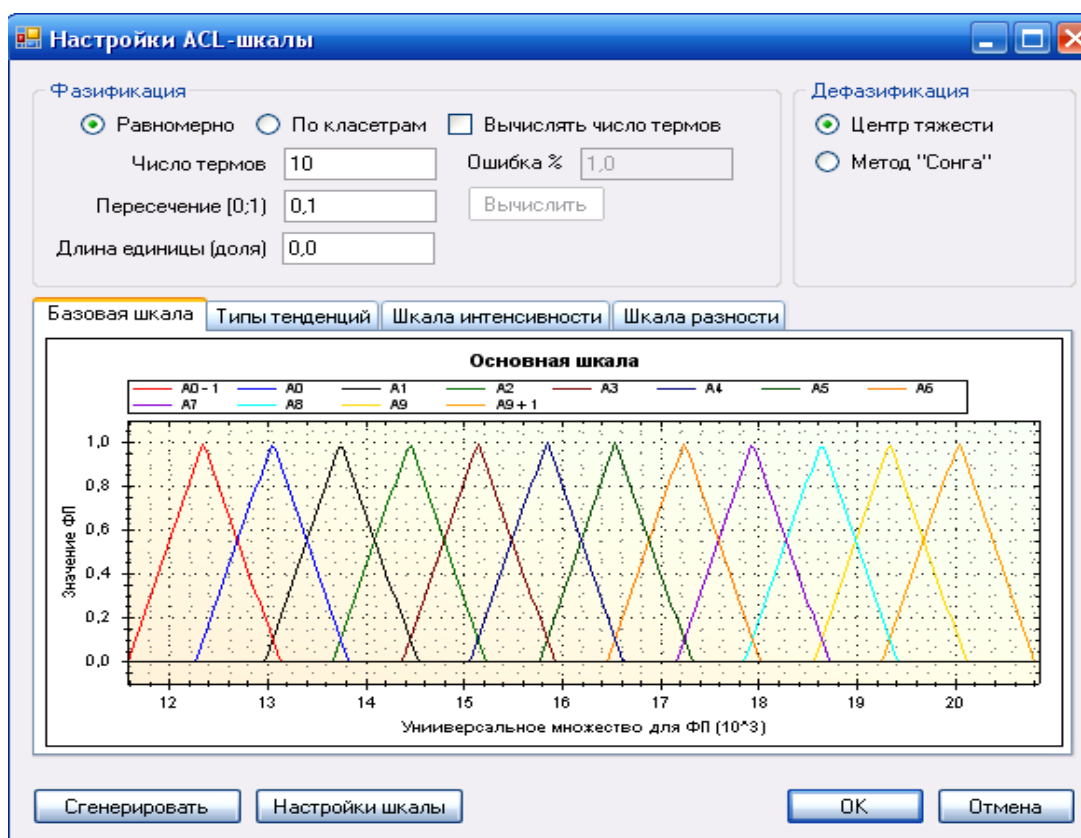


Рис. 6.3. Настройка ACL-шкалы

Параметры функций принадлежности можно изменять в зависимости от экспертной оценки качественных интервалов предметной области или вычислять в зависимости от ожидаемого экспертом значения ошибки моделирования MAPE (расчетная формула приведена в главе 4). Последний вариант настройки шкалы предпочтителен, если экспертные оценки отсутствуют, так как в результате происходит автоматическая генерация параметров функций принадлежности, моделирующих нечеткие градации шкалы. Настройка параметров отдельных функций принадлежности вручную так же возможна, но требует дополнительных экспертных знаний и реализуется в дополнительном окне настройки ACL-шкалы (рис. 6.4).

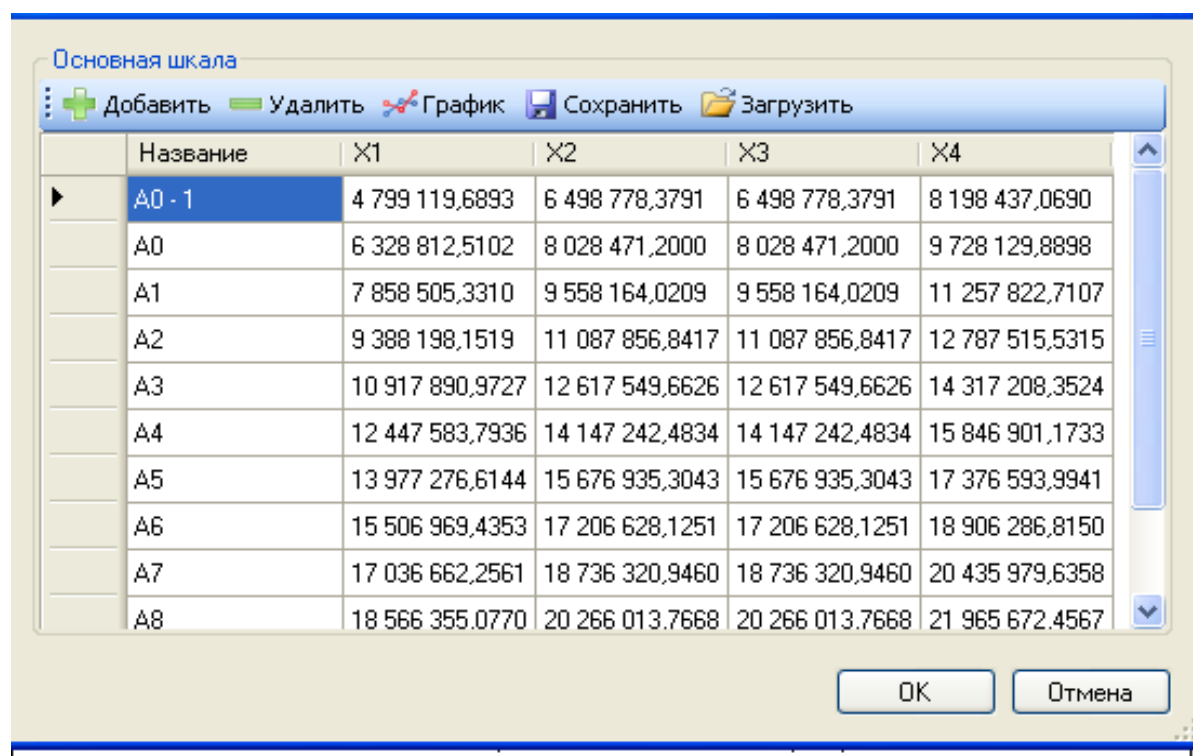


Рис. 6.4. Настройка параметров ACL-шкалы

Настройка параметров и выбор модели

При работе с моделями пользователь, задавая порядок модели, количество точек для тестирования модели и глубину прогноза, может работать с каждой моделью отдельно (рис. 6.5), при этом имеется возможность использовать модель остатков, а также использовать режим выбора наилучшей модели из набора моделей интеллектуального анализа данных по одному или совокупности выбранных критериев качества (рис. 6.6).

Таким образом, обеспечивается исследование модели временного ряда, анализ его способности к прогнозированию на выбранное количество точек (разбиение ряда) и генерация прогноза на будущее.

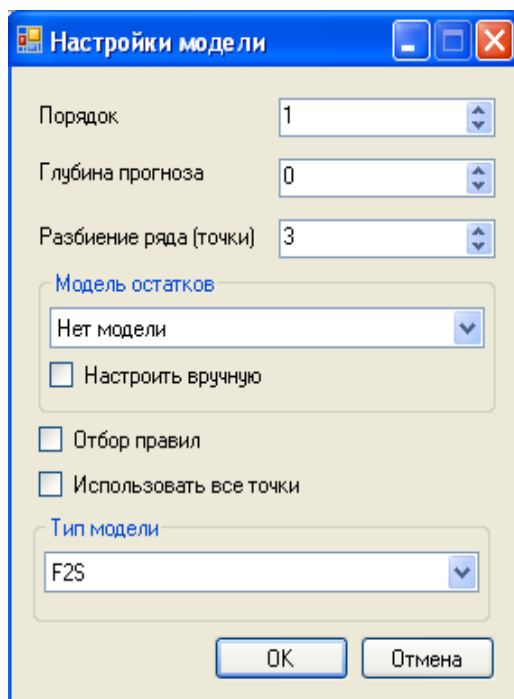


Рис. 6.5. Настройка модели

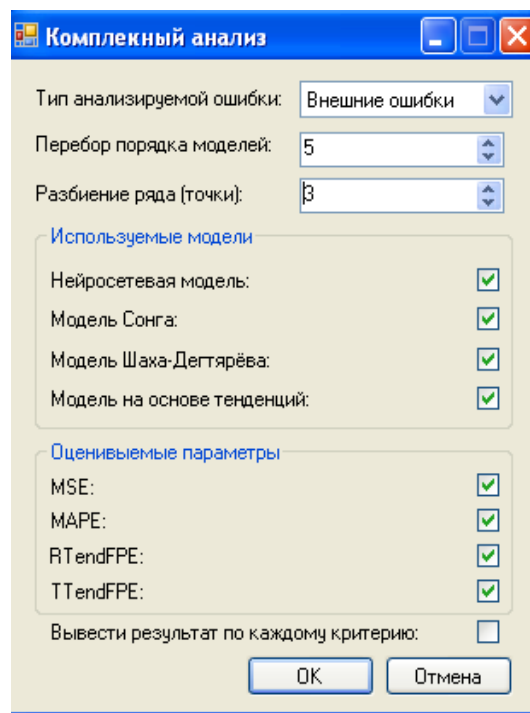


Рис. 6.6. Выбор наилучшей модели

Просмотр результатов моделирования

После применения модели к временному ряду будет отображаться информация о применении модели к данному ряду: в четком (рис. 6.7), в нечетком (рис. 6.8), в графическом (рис. 6.9, 6.10).

Информация о модели								
В окне Сохранить								
Crisp Fuzzy								
№	Время	Ряд	Модель	Ошибка	Отн. О...	Разность	Модел...	
0	1971	13055	13055	0	0	508	508	
1	1972	13563	13563	0	0	284	475,900...	
2	1973	13847	14038,9...	0,01385...	191,900...	849	436,707...	
3	1974	14696	14475,6...	0,01499...	-220,39...	764	872,412...	
4	1975	15460	15348,0...	0,00724...	-111,98...	-149	709,805...	
5	1976	15311	16057,8...	0,04877...	746,825...	292	-534,17...	
6	1977	15603	15523,6...	0,00508...	-79,344...	258	339,429...	
7	1978	15861	15863,0...	0,00013...	2,08475...	946	127,370...	
8	1979	16807	15990,4...	0,04858...	-816,54...	112	1521,70...	
9	1980	16919	17512,1...	0,03505...	593,160...	-531	-943,31...	
10	1981	16388	16568,8...	0,01103...	180,843...	-955	-827,83...	
11	1982	15433	15741,0...	0,01995...	308,003...	64	-414,01...	
12	1983	15497	15326,9...	0,01097...	-170,01...	-352	362,494...	
13	1984	15145	15689,4...	0,03595...	544,479...	18	-287,72...	
14	1985	15163	15401,7...	0,01574...	238,758...	821	111,785...	

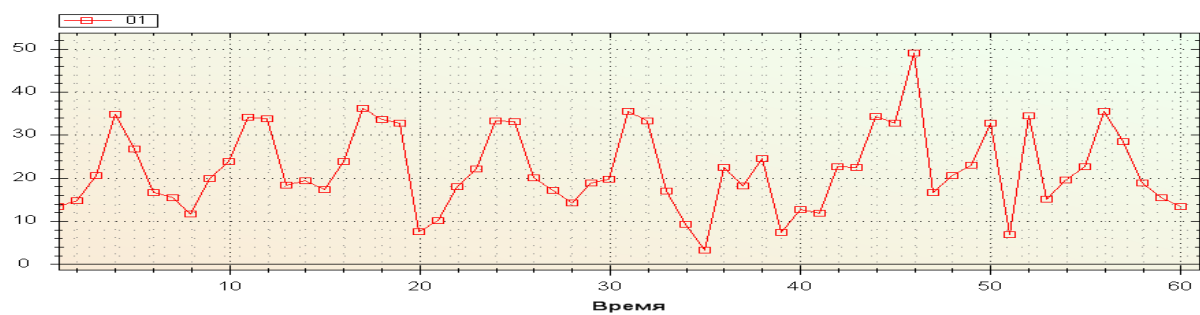
Рис. 6.7. Информация о модели (четкая)

Информация о модели											
В окне Сохранить Crisp Fuzzy											
№	Время	Ряд	ФП	НМ	Модел...	Ошибка	TTend	Модел...	Ошибка	RTend	Мод...
0	0	0	0,55	A5	не выч...	0	Рост	Рост	0	R3	R1
1	1	0,84147...	0,74	A8	A8(0,55...	1	Рост	Рост	0	R1	R1
2	2	0,90929...	0,63	A9	A9(0,74...	1	Падение	Падение	0	R4	R1
3	3	0,14112...	0,88	A5	A5(0,63...	1	Падение	Стабил...	0,5	R4	R0
4	4	-0,7568...	0,92	A1	A8(0,88...	1	Падение	Падение	0	R1	R3
5	5	-0,9589...	0,83	A0	A0(0,92...	1	Рост	Стабил...	0,5	R3	R0
6	6	-0,2794...	0,78	A3	A3(0,83...	0	Рост	Рост	0	R4	R1
7	7	0,65698...	0,59	A7	A7(0,78...	1	Рост	Рост	0	R2	R2
8	8	0,98935...	0,96	A9	A9(0,59...	1	Падение	Рост	1	R3	R1
9	9	0,41211...	0,68	A6	A5(0,96...	1	Падение	Падение	0	R4	R2
10	10	-0,5440...	0,95	A2	A2(0,68...	1	Падение	Падение	0	R2	R2
11	11	-0,9999...	1,00	A0	A0(0,95...	1	Рост	Падение	1	R2	R1
12	12	-0,5365...	0,92	A2	A3(1,00...	0	Рост	Рост	0	R4	R1
13	13	0,42016	0,65	A6	A0(0,92...	1	Рост	Рост	0	R3	R2

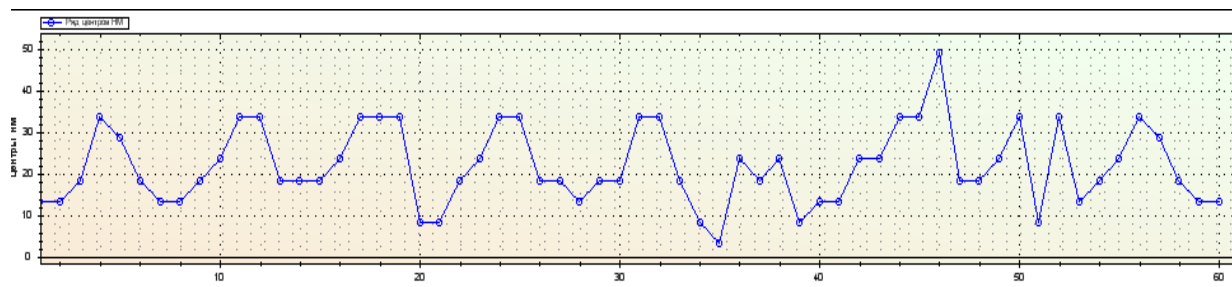
Рис. 6.8. Информация о модели (нечеткая)

Функция	Значение
MSE	64,00363...
MAPE	30,99001...
R2	0,290918...
RMSE	8,005226...
FPE	64,15
FCE	34/53
TTend FPE	1,89
TTend FCE	1/54
RTend FPE	0,00
RTend FCE	0/54
Внешние ошибки	
MSE	4,280761
MAPE	15,59743...
R2	-бесконе...
RMSE	2,069
FPE	0,00
FCE	0/1
TTend FPE	0,00
TTend FCE	0/1
RTend FPE	100,00
RTend FCE	1/1

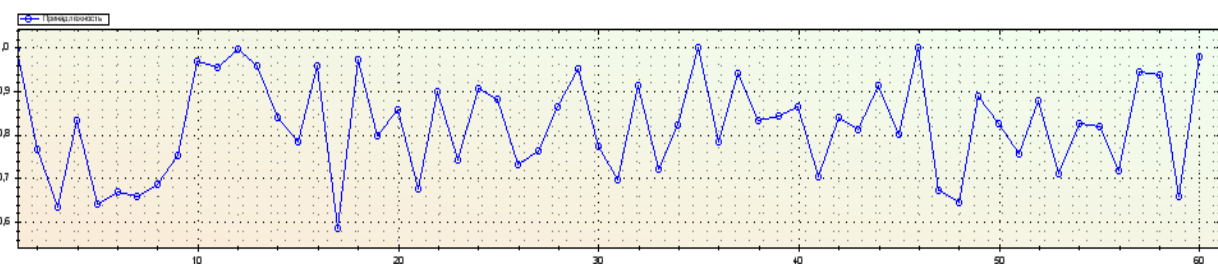
Рис. 6.9. Результаты моделирования и критерии оценивания ошибок



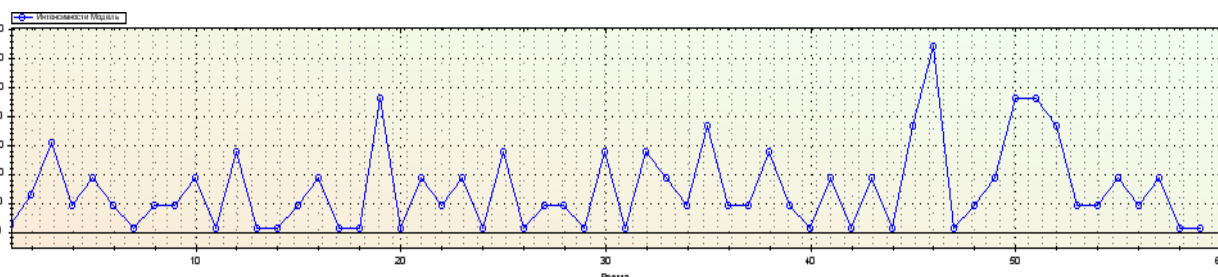
а) исходный временной ряд



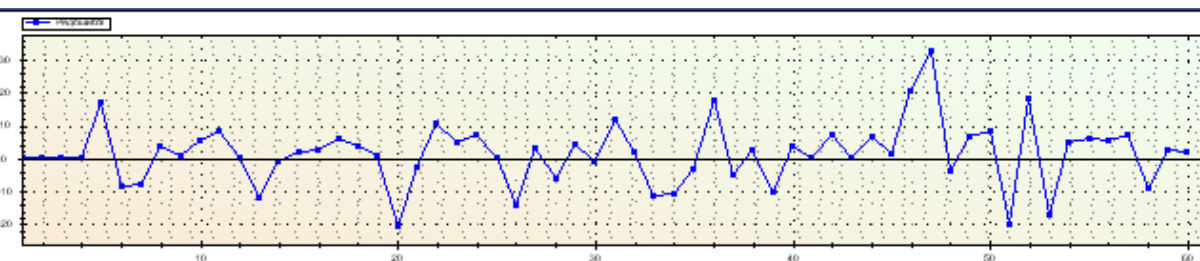
б) исходный временной ряд после нечеткого сглаживания



в) функция принадлежности нечеткого временного ряда элементарным тенденциям



г) функция интенсивностей элементарных тенденций



д) временной ряд остатков

Рис. 6.10. Графики компонент модели нечеткой тенденции

Внутренние и внешние ошибки модели отображаются в правой части окна совместно с графическим отображением исходного и моделированного временных рядов (рис. 6.10).

В программной системе использованы следующие показатели качества (y_i – реальные значения, $\hat{y}_i$ – смоделированные значения временного ряда):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad - \quad \text{среднеквадратическое отклонение;}$$

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)} \quad - \quad \text{квадрат среднеквадратического отклонения;}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% \quad - \quad \text{средняя относительная ошибка;}$$

$TTend = (ErrorCount/TotalCount) \cdot 100\%$ – коэффициент ошибок в моделировании нечетких тенденций временного ряда в процентах.

Лингвистическая характеристика поведения ВР представлена на рис. 6.11. Для нечетких моделей ВР, в том числе на основе нечетких тенденций, на вкладке «Результаты» отображаются лингвистические нечеткие правила (Rule), извлеченные из временного ряда (табл. 6.1) вместе с частотной (Count) характеристикой и степенью принадлежности (MF).

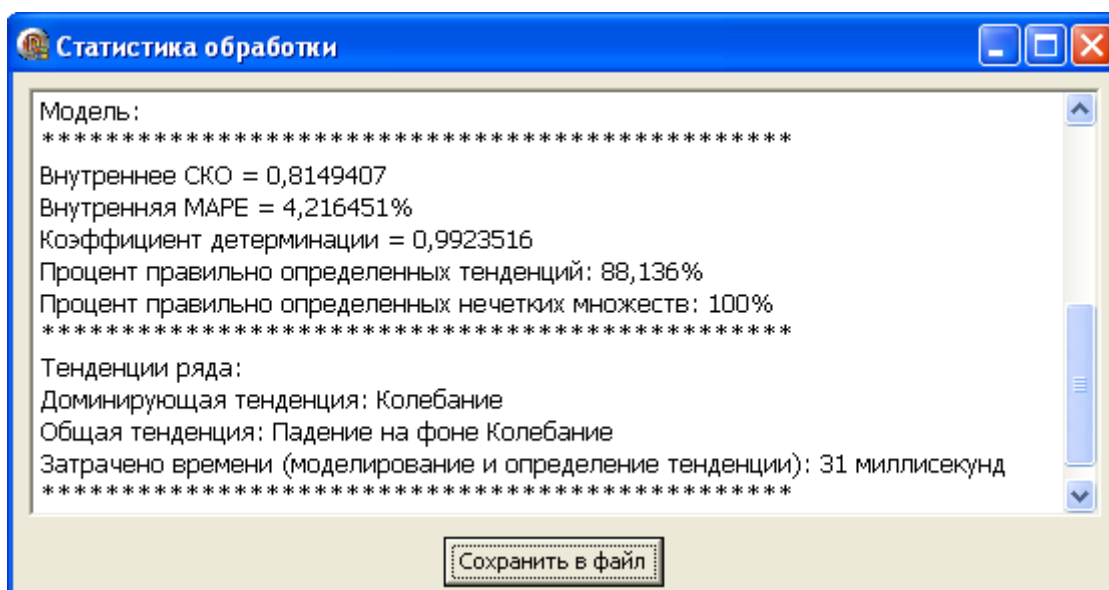


Рис. 6.11. Лингвистическое представление результатов моделирования

Нечеткая модель типов нечетких тенденций

Rule	Count	MF
if ((Input0 is Стабильность) and (Input1 is Рост)) then (Output is Рост)	5	0,76
if ((Input0 is Рост) and (Input1 is Рост)) then (Output is Падение)	3	0,64
if ((Input0 is Рост) and (Input1 is Падение)) then (Output is Падение)	2	0,64
if ((Input0 is Падение) and (Input1 is Падение)) then (Output is Падение)	3	0,64
if ((Input0 is Падение) and (Input1 is Падение)) then (Output is Стабильность)	2	0,64
if ((Input0 is Падение) and (Input1 is Стабиль- ность)) then (Output is Рост)	3	0,66
if ((Input0 is Рост) and (Input1 is Рост)) then (Out- put is Рост)	3	0,66
if ((Input0 is Рост) and (Input1 is Рост)) then (Out- put is Стабильность)	3	1,00

6.2. Вычислительный эксперимент

Целью вычислительного эксперимента являлось исследование продуктивности структурно-лингвистического подхода для генерации авторских моделей временных рядов, позволяющих извлекать знания о нечетких тенденциях в форме продукционных правил «ЕСЛИ-ТО» и использовать их для прогнозирования нечетких тенденций и числовых уровней ВР. Для этого предлагаемые модели ВР были исследованы по критериям точности моделей в терминах нечетких элементарных тенденций TTend, а также числовой оценки прогноза по критериям MAPE, MSE и RMSE, применяемых в работах зарубежных авторов, результаты которых были использованы для сравнения. Вычислительный эксперимент показал, что авторские нечеткие моде-

ли ВР в терминах нечетких тенденций имеют для краткосрочного прогноза хорошие результаты, как по критериям качества моделирования и оценки прогноза нечетких тенденций, так и для оценки числовых уровней ВР. Полагая, что лиц, принимающих решения, часто интересует знания о будущих тенденциях и закономерностях изменения в них в самом ближайшем будущем, ниже приведем именно такие примеры, связанные с получением краткосрочных прогнозов нечетких элементарных тенденций, а также и оценки числовых уровней временных рядов.

В эксперименте были использованы следующие авторские нечеткие модели тенденций: $F2S(X,Y,Z)$, $F1N(X,Y,Z)$, $F3N1S(X,Y,Z)$ и для сравнения программная реализация нечетких моделей $S(X,Y,Z)$ [Song, 1993], $D(X,Y,Z)$ [Şah, 2004]. В этих моделях X обозначает мощность генерируемой ACL-шкалы, Y – порядок модели, Z – разбиение ряда (количество тестовых интервалов для прогноза). В некоторых моделях будет использоваться алгоритм отбора нечетких правил, в этом случае к модели будет добавлен текст «+ отбор».

В вычислительном эксперименте были использованы следующие временные ряды:

№1 – временной ряд цены акций одного предприятия (данные взяты из работы [Кремер, 2007]).

№2 – временной ряд статьи 211 «Доходов и расходов» одной бюджетной организации.

№3 – временной ряд, содержащий данные о поступающих в Алабамский университет за 22 года, начиная с 1971 года (данные взяты из работы [Song, 1993]).

№4 – временной ряд статьи 223 «Доходов и расходов» одной бюджетной организации.

№5 – временной ряд (приведен не полностью), взятый с сайта <http://www.neural-forecasting-competition.com/NN5/index.htm>, посвященного тестированию новых интеллектуальных моделей временных рядов и их сравнению со статистическими моделями (NN5-2005).

№6 – временной ряд среднегодовой температуры в городе Мюнхене (по материалам [Корнеев, 2000]).

Временной ряд №1

Временной ряд № 1 отражает динамику цены акций одного предприятия (данные взяты из работы [Кремер, 2007]). Данный временной ряд относится к нестационарным временным рядам, количество значений – 22 (рис.6.12), для него нецелесообразно использовать модели с линейным или полиномиальным трендом. Результаты моделирования различными моделями представлены в таблице 6.2.



Рис. 6.12. Временной ряд акций предприятия

Таблица 6.2

Извлеченные правила

Rule	Count
if ((Input0 is Стабильность)) then (Output is Стабильность)	4
if ((Input0 is Стабильность)) then (Output is Падение)	4
if ((Input0 is Падение)) then (Output is Стабильность)	4
if ((Input0 is Стабильность)) then (Output is Рост)	3
if ((Input0 is Рост)) then (Output is Стабильность)	3
if ((Input0 is Рост)) then (Output is Рост)	1

В таблице 6.3 представлено сравнение авторских моделей ВР №1 с моделями для различного количества тестовых значений. Анализ продуктив-

ности авторских нечетких моделей для ACL-шкалы с тремя нечеткими градациями для краткосрочного прогноза акций предприятия (BP №1) показал, что они имеют лучшие показатели качества при прогнозе на 2 интервала, в то же время при прогнозировании на одно и на три интервала процедура поиска наилучшей модели определила D-модель.

Интерес представляет, что качество моделей временного ряда №1 для прогноза меняется в сторону улучшения в пользу авторских моделей при использовании менее грубой ACL-шкалы, например, с десятью нечеткими градациями (см. табл. 6.4).

Таблица 6.3

Сравнение моделей для прогноза BP №1 (мощность шкалы =3)

Модель	MAPE	TTend	MSE
F2S(3,1,1)	1,9	0	576
S(3,1,1)	11	0	18360
D(3,1,1)	0,45	0	30
ARIMA(1,0,1) на 1 интервал	12,6	нет	23104
F2S(3,1,2)	2	0	769
S(3,1,2)	10	25	15396
D(3,1,2)	5,7	25	4802
ARIMA(1,0,1) на 2 интервала	8,715568	нет	6400
F2S(3,1,3)	9	16,6	13613
S(3,1,3)	6,2	50	7713
D(3,2,3)	3,4	0	2547

Таблица 6.4

Сравнение нечетких моделей для прогноза ВР №1 (мощность шкалы =10)

Модель	MAPE	TTend	MSE
F2S(10,1,1)	1,9	0	576
S(10,1,1)	11	0	18360
D(10,2,1)	0,45	0	30,5
F2S(10,1,2)	2	0	769
F1N(10,1,2))	1,5	0	353
S(10,1,2)	10	25	15396
D(10,1,2)	6,9	75	7138
F2S(10,1,3)	9	16,6	13613
S(10,1,3)	6,2	33	7713
D(10,2,3)	14	33	31579

Временной ряд №2

Временной ряд №2 извлечен из статьи 211 «Доходы и расходы» одной бюджетной организации, данные характеризуют месячные расходы на заработную плату за четыре с половиной года.

Для этого нестационарного ВР (рис. 6.13) проводился прогноз на один интервал сначала (а) для усеченного ВР на 3 точки, (б) затем – на две точки, (в) затем на одну точку, в заключении (с) исследован весь временной ряд. Цель такого исследования одного и того же ряда – проанализировать устойчивость краткосрочного прогноза предлагаемых нечетких моделей по сравнению с известными S- и D-моделями. Результаты лучших моделей приведены в табл. 6.5.

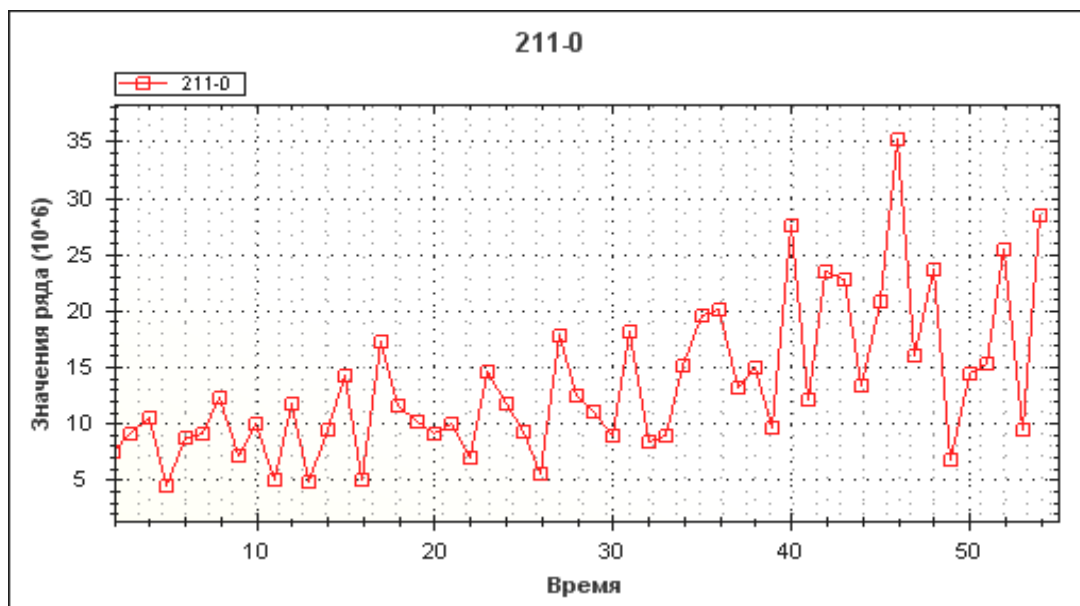


Рис. 6.13. Временной ряд статьи 211

Проведенные исследования показывают, что авторские модели при прогнозе на один интервал характеризуются стабильными показателями качества ($MAPE < 10\%$, $TTend=0$), что нельзя сказать о выбранных по наилучшим показателям качества S- и D-моделях.

Таблица 6.5

Сравнение нечетких моделей для прогноза ВР №2 (мощность шкалы =10)

Временной ряд	Модель	MAPE	TTend	MSE
(a)211-3	F2S(10,1,1)	5,4	0	$0,69 \cdot 10^{12}$
	S(10,1,1)	0,6	50	$0,8 \cdot 10^{10}$
	D(10,3,1)	9,4	0	$0,2 \cdot 10^{13}$
(б)211-2	F2S(10,1,1)	8,2	0	$0,43 \cdot 10^{13}$
	S(10,3,1)	22	0	$0,3 \cdot 10^{14}$
	D(10,1,1)	42	50	$0,1 \cdot 10^{15}$
(в)211-1	F1N(10,1,1)	5,9	0	$0,1 \cdot 10^{13}$
	S(10,4,1)	17	0	$0,28 \cdot 10^{13}$
	D(10,1,1)+отбор	52	0	$0,24 \cdot 10^{15}$
(с)211-0	F1N(10,2,1)	4	0	$0,1 \cdot 10^{13}$
	S(10,4,1)	30	0	$0,78 \cdot 10^{14}$
	D(10,1,1)+отбор	29	0	$0,7 \cdot 10^{14}$

Проведенное статистическое моделирование ВР №2 на три интервала в пакете «Статистика» показало, что наименьшее RMSE соответствует модели SARIMA(1,1,1)x(1,0,0): MAPE=49%, RMSE=88*10⁵. При нечетком моделировании в рассматриваемом авторском пакете были получены сравнимые результаты, приведенные в табл. 6.6.

Таблица 6.6

Сравнение нечетких моделей для прогноза ВР №1
и статистической модели

Модель	MAPE	RMSE	TTend
F1N(3,3,3)+отбор	28	70*10 ⁵	16,6
F3N1S(3,3,3) + отбор	36	122*10 ⁵	16,6
F2S(3,4,3) + отбор	42	80*10 ⁵	16,6
SARIMA(1,1,1)x(1,0,0)	49	88*10 ⁵	50
F3N1S(3,4,3) + отбор	49	101*10 ⁵	0
F2S(3,1,3) + отбор	55	88*10 ⁵	0

По сравнению с моделью SARIMA(1,1,1)x(1,0,0) нечеткая авторская модель F1N(3,3,3) показала лучшие результаты в оценке числового прогноза и в оценке ожидаемой нечеткой тенденции, а модели F3N1S(3,4,3) и F2S(3,1,3) вычислили оценку числового прогноза на уровне SARIMA(1,1,1)x(1,0,0), но к тому же безошибочно распознали будущую тенденцию.

В программной системе предусмотрена возможность после моделирования временного ряда провести моделирование временного ряда остатков. В результате моделирования остатков наилучшей модели F1N(3,3,3) (табл. 6.6) точность оценки прогноза по критерию RMSE улучшилась более чем в 2 раз и составила 30*10⁵. При этом для остатков модели F1N(3,3,3) была автоматически выбрана другая наилучшая модель F3N1S(3,1,3) + отбор.

Временной ряд №3

Тестирование авторских нечетких моделей проводилось на временном ряду, содержащем данные о поступающих абитуриентах в университет штата

Алабама, который стал уже, де-факто, стандартом для тестирования систем, связанных с моделированием и прогнозированием НВР (рис. 6.14).

Результаты моделирования на основе авторских нечетких моделей в таблице 6.7 сравниваются с моделями Сонга [Song, 1993] и Хуарга [Huarng, 2006], а в табл. 6.8 приводятся результаты прогноза моделей, которые сопоставляются с S-и D-моделями.

Таблица 6.7

Сравнение точности моделирования ВР №3

MAPE	ARIMA(1,0,1)	Метод Song	Метод Huarng	Предлагаемый метод нечетких тенденций	
		[Song, 1993]	[Huarng, 2006]	F2S(33,5,0) +отбор	F1N(33,4,0) +отбор
	5,49	3,11	1,5294	1,2 TTend =0	0,96 TTend =0

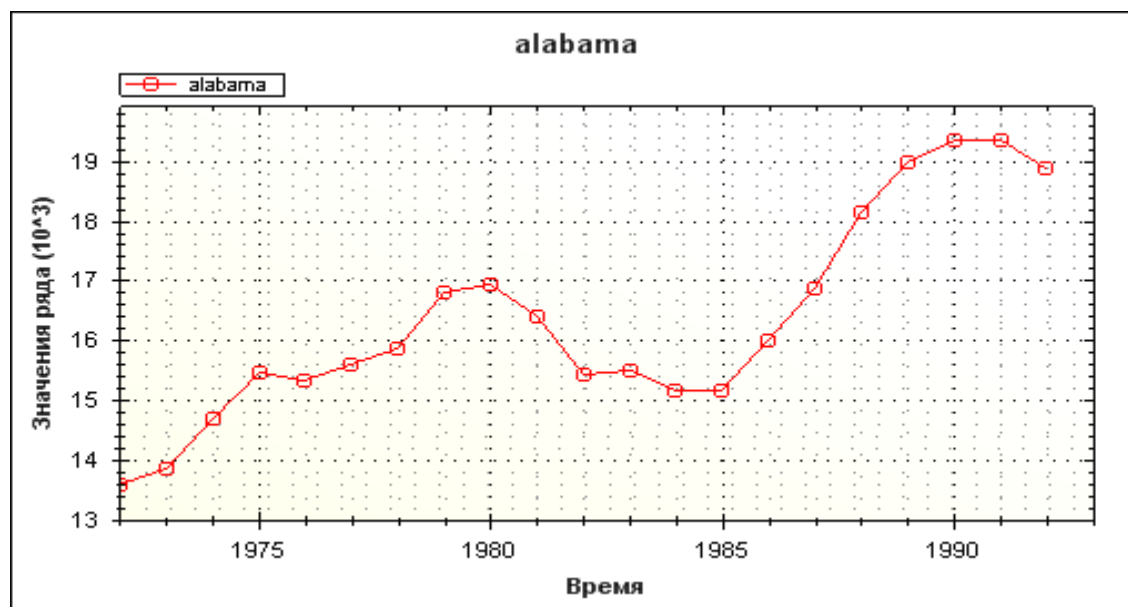


Рис. 6.14. График временного ряда «Алабама»

Таблица 6.8

Сравнение нечетких моделей для прогноза ВР №3

Модель	MAPE	TTend	MSE	RMSE
F2S(19,2,1)	0,2	0	1,9*10 ³	44
S(19,1,1)	2,4	50	212*10 ³	461
D(19,2,1)	3,3	100	395*10 ³	629
F2S(19,2,2)	0,11	0	666	25
S(19,2,2)	13	50	7110 786	2666
D(19,1,2)	0,5	25	13 416	115

Анализ таблиц 6.7 и 6.8 показывает, что авторские модели на основе нечетких тенденций по сравнению с известными нечеткими моделями при моделировании ВР «Алабама» (табл. 6.7) и при краткосрочном прогнозе (табл. 6.8) демонстрируют лучшие результаты по всем исследуемым критериям.

Временной ряд №4

Предлагаемый метод нечеткого моделирования был апробирован на временном ряду месячных данных показателя «Коммунальные расходы», извлеченном из статьи 223 «Доходы и расходы» бюджетной организации, начиная с января 2005 года до середины 2009, всего 53 значения (рис. 6.15). Управляющий алгоритм выбрал наилучшую нечеткую модель класса F2S(6,4,3), где цифра 6 обозначает количество нечетких термов модели, 4 – порядок модели, 3 – количество прогнозируемых точек, используемых для вычисления внешних критериев качества RMSE и MAPE.

Полученные результаты сравнивались с результатами статистической модели, выбранной из нескольких моделей класса АРПСС (ARIMA) в программной системе «Статистика». Выбор производился на основе визуального анализа

периодограммы и автокорреляционной функции, в результате была выбрана модель SARIMA (1,1,1)x(1,0,0).

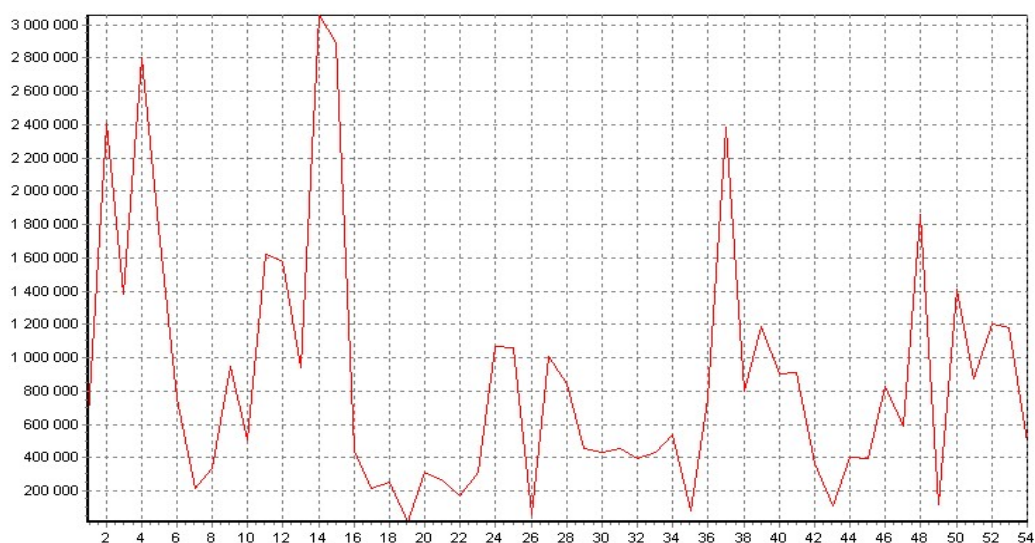


Рис. 6.15. Временной ряд «Коммунальные расходы»

Данные внешних критериев качества при прогнозе на три тестовых интервала показывают (см. табл. 6.9), что предлагаемый подход к нечеткому моделированию исследуемого ВР на основе нечетких тенденций не уступает, даже превосходит по точности модель SARIMA (1,1,1)x(1,0,0), а также лучше по сравнению с нечеткими S- и D- моделями (табл. 6.10).

Таблица 6.9

Сравнение нечетких моделей ВР №4 со статистической моделью

Модель	MAPE	TTend	MSE
F1N(6,2,3)+отбор	28,2	0	18*10 ¹⁰
F2S(6,4,3)	33,9	16,6	8*10 ¹⁰
SARIMA(1,1,1)x(1,0,0)	78	50	49*10 ¹⁰

Сравнение нечетких моделей для прогноза ВР №4

Название	MAPE	TTend	MSE
F1N(6,2,3)+отбор	28,2	0	$18 \cdot 10^{10}$
F2S (6,4,3)	33,9	16,6	$8 \cdot 10^{10}$
S(6,3,3)	58	33	$26 \cdot 10^{10}$
D(6,5,3)	39	33	$22 \cdot 10^{10}$

Временной ряд №5

Временной ряд №5 (рис. 6.16) отражает уровень изъятия денежных средств в одном из банкоматов за 113 дней. Этот временной ряд был предложен для тестирования в европейском соревновании моделирования временных рядов NN5 (см. <http://www.neural-forecasting-competition.com/NN5/index.htm>).

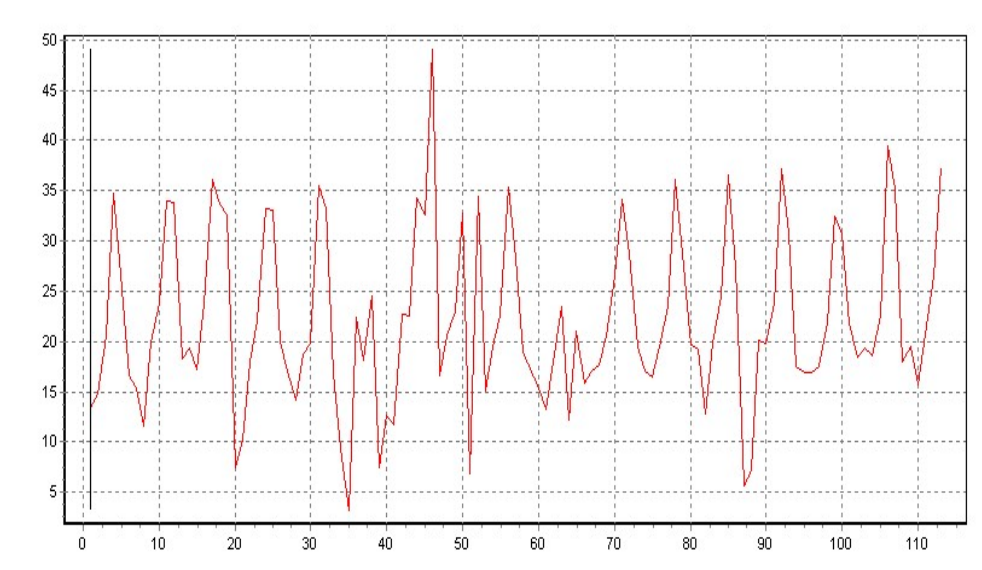


Рис. 6.16. Временной ряд «Данные банкомата»

Результаты по критериям числового прогноза MAPE и MSE, а также по критерию качества модели нечетких тенденций TTend для различных интервалов прогнозирования представлены в табл. 6.11.

Сравнение моделей для прогноза ВР №5

1. Модель	MAPE	TTend	MSE
2. F1N(13,1,2)	1	0	0,164
3. ARIMA(1,0,0)	30	нет	121
4. на 2 интервала			
5. F3N1S (6,2,5)+отбор	15	10	40
6. F2S(10,1,10)	22	5	105,5
7. S(10,1,10)	25	40	73
8. D(10,1,10)	23	30	67

Как видно из табл. 6.11, метод нечеткого моделирования ВР в терминах нечетких тенденций в сравнении с результатами лучших методов, заявленных на европейском соревновании моделей ВР NN5, показал удовлетворительные результаты числового прогноза. Отметим, что модель, основанная на нечетких правилах следования нечетких тенденций, не генерирует правила для сезонных компонент ВР, поэтому получаемые результаты здесь не достаточно точны.

Использование нечеткого сглаживания временного ряда

При использовании моделей ВР на основе нечетких элементарных тенденций для краткосрочного прогнозирования временных рядов использовался подход, при котором исходный ВР не подвергался предварительной обработке. На практике часто интерес представляет выявление основной тенденции и прогноз именно этой тенденции. Для построения моделей основной тенденции применяют методы сглаживания временного ряда. В реализованном пакете прикладных программ также предусмотрен метод нечеткого сглаживания временных рядов на основе метода F-преобразований [Perfilieva, 2006].

На рис. 6.17 приведены графики исходного ВР №2 и результат его сглаживания на основе метода F-преобразований с параметром, равным 5.

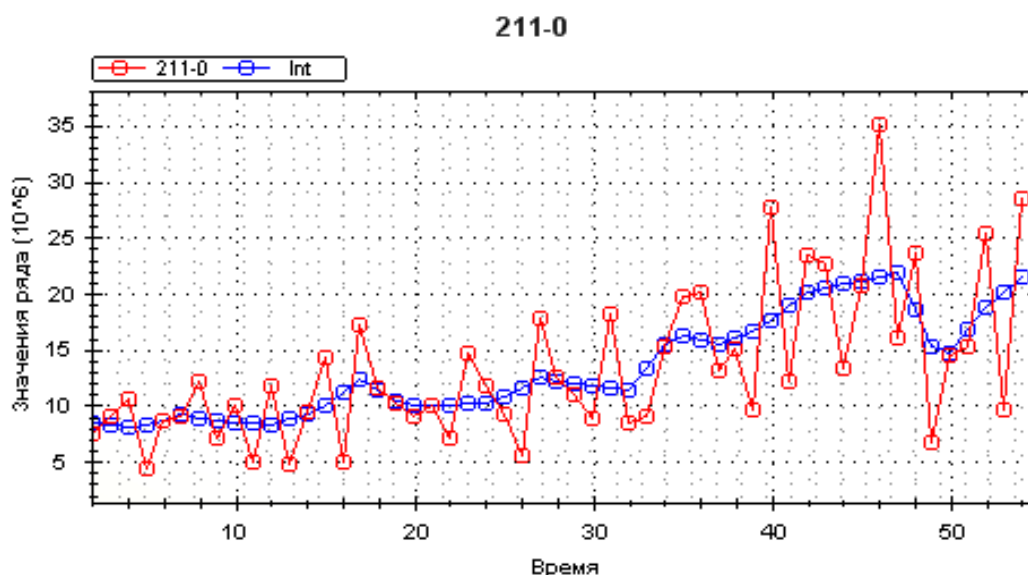


Рис. 6.17. Исходный и сглаженный ВР №2

Для сглаженного временного ряда были применены авторские нечеткие модели для прогноза на один интервал, которые показали лучшие результаты по сравнению с результатами прогноза исходного ВР: $MAPE = 2,9\%$, $TTend = 0$, модель $F1N(10,1,1)$. Данный факт позволяет сделать вывод, что структурно-лингвистический подход для построения нечеткой модели ВР в терминах нечетких элементарных тенденций может быть успешно применен для моделирования сглаженных временных рядов, для которых помимо числового прогноза формируется база правил в терминах нечетких тенденций и прогнозируются элементарные тенденции. Таким образом, имеется возможность рассматривать ВР в традиционном представлении как совокупность регулярной (трендовой) и нерегулярной компонент и моделировать для каждой компоненты отдельно элементарные тенденции.

Временной ряд №6

Временной ряд №6 (рис. 6.18) содержит данные о среднегодовой температуре в г. Мюнхене с 1925 по 1985 гг. [Корнеев и др., 2000]. Результаты прогноза по авторским нечетким моделям в сравнении с другими нечеткими моделями представлены в табл. 6.12.

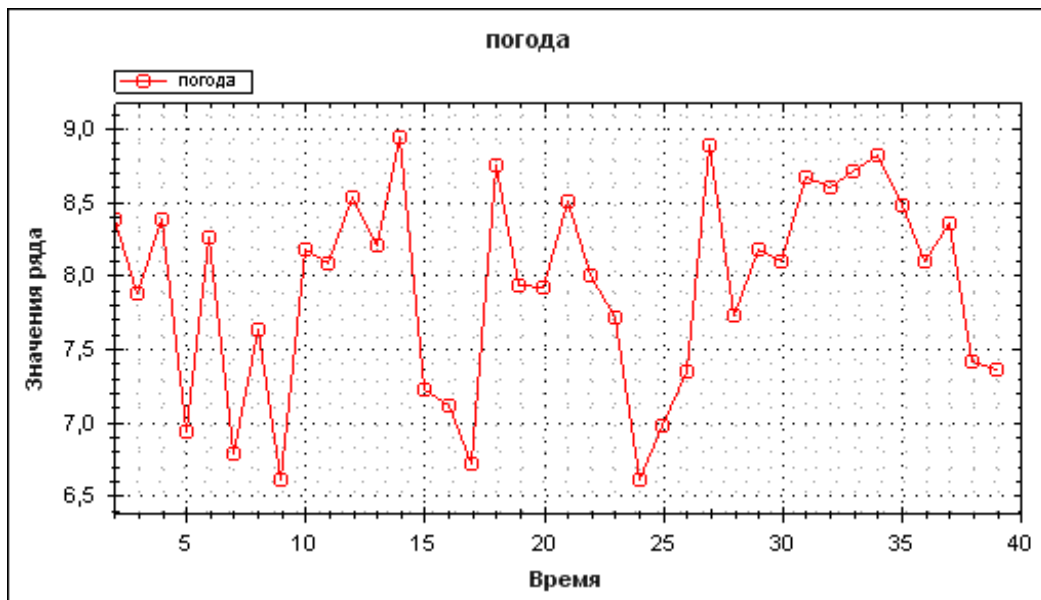


Рис. 6.18. Временной ряд «Погода в Мюнхене»

Таблица 6.12

Сравнение нечетких моделей для прогноза ВР №6

9. Модель	MAPE	TTend	MSE
10. F2S(10,1,1)	0,67	0	0,002
11. D(10,3,1)	2,7	0	0,04
12. S (10,3,1)	5,6	50	0,17

Контрольные вопросы

1. Какие модели включает в себя программная система *FuzzyTend*?
2. Какие показатели качества использованы в программе *FuzzyTend*?
3. В чем особенности и отличия от аналогов программной системы *FuzzyTend*?
4. Как могут в программе быть использованы интеллектуальные модели?
5. Что представляется собой библиотека моделей временных рядов?
6. Для чего предназначен пакет программ *FuzzyTend*.

Выводы

Программная система нечеткого моделирования и анализа нечетких тенденций *FuzzyTend*, рассмотренная в настоящей главе, является экспериментальной реализацией структурно-лингвистического подхода к нечеткому моделированию временных рядов. Она может быть использована для исследования применимости предложенных нечетких моделей, а также и для моделирования временных рядов и извлечение знаний о параметрах элементарных тенденций.

Исследование разработанных моделей и методов в программной системе нечеткого моделирования временных рядов в терминах нечетких элементарных тенденций показало, что чем выше порядок модели, тем выше качество модели. На основе проведенного анализа можно сделать вывод не только о полезности представленных нечетких моделей ВР, но и о результативности использования алгоритма отбора правил, который во многих случаях повышает точность моделирования (и прогноза) в числовом выражении и в терминах нечетких элементарных тенденций (по критерию *TTend*).

Отличием данной программной системы моделирования ВР по сравнению с аналогами является то, что в ней реализованы автоматический подбор интеллектуальных моделей, извлечение дополнительных знаний из ВР о зависимостях между нечеткими элементарными тенденциями и естественно-языковая интерпретация результатов. Проведенный сравнительный анализ показал, что предлагаемый метод нечеткого моделирования ВР может конкурировать с известными методами нечеткого моделирования, а также в краткосрочном прогнозе может генерировать более точные числовые оценки будущих тенденций по сравнению со статистическими моделями для некоторых нестационарных временных рядов. Кроме моделирования и прогнозирования ВР в терминах нечетких тенденций и числовых оценок, в программной системе реализован алгоритм извлечения ассоциативных правил, для каждого из которых вычислено соответствующее значение коэффициента поддержки. Данные знания, выраженные в лингвистической форме «ЕСЛИ-ТО», могут быть исполь-

зованы для поддержки принятия решений в экспертной системе или лицом, принимающим решения.

Библиографический список

1. [Huarng, 2006] Huarng, K. The application of neural networks to forecast fuzzy time series / K. Huarng // *Physica*. – 2006. – A 336. – P. 481-491.
2. [Perfilieva, 2006] Perfilieva, I. Fuzzy transforms: Theory and applications / I. Perfilieva // *Fuzzy Sets and Systems*. – 2006. – №157.
3. [Song, 1993] Song, Q. Forecasting enrollments with fuzzy time series – Part I / Q. Song, B. Chissom // *Fuzzy Sets and Systems*. – №54 (1993). – P. 1-9.
4. [Şah, 2004] Şah, M. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series / M. Şah, K. Y. Degtiarev // *Proc. Int. Conf. Computational Intelligence (ICCI) (2004)*. – P. 375-378.
5. [Афанасьева и др., 2009] Афанасьева, Т. В. Программная реализация интегрального метода нечеткого моделирования и анализа нечетких тенденций временных рядов / Т. В. Афанасьева, А. Г. Чекмарев, Д. Е. Савельев // *Сборник трудов всероссийской конференции «Проведение научных исследований в области обработки, хранения, передачи и защиты информации»*. Т.2. – Ульяновск : УлГТУ, 2009. – С. 542-549.
6. [Корнеев, 2000] Корнеев, В. В. Базы данных. Интеллектуальная обработка информации / В. В. Корнеев. – М. : «Нолидж», 2000. – С. 191-192.
7. [Кремер, 2007] Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям / Н. Ш. Кремер. – 3-е изд., перераб. и доп. – М. : ЮНИТИ-ДАНА, 2007. – С. 494-495.

ЗАКЛЮЧЕНИЕ

Учебное пособие посвящено проблеме моделирования временных рядов. Данная проблема порождена, с одной стороны, ростом информации, хранимой в БД, которая может быть организована в виде ВР, а с другой стороны, неготовностью специалистов различного профиля выполнять анализ таких объемов данных. Новые требования пользователей к системам моделирования ВР, извлеченных из БД, определили и новые задачи, среди которых обнаружение тенденций, их качественная оценка и прогноз выступают как отдельная задача анализа ВР.

На основе обобщения методологических положений различных подходов к моделированию ВР, удовлетворяющих новым требованиям пользователей, определены перспективы развития систем моделирования и анализа ВР в направлении Time Series Data Mining в виде интеграции интеллектуальных моделей и анализа нового объекта ВР – нечеткой тенденции.

Рассмотренный структурно-лингвистический подход представляет собой интеграцию нечетких шкал, нечетких моделей и нейронных сетей, ориентирован на анализ новых нечетких объектов временного ряда – нечетких тенденций. По результатам исследования моделей нечетких тенденций в авторской программной системе *FuzzyTend* можно сделать вывод об их полезности и продуктивности в качестве метода извлечения из ВР новых знаний о качественной оценке текущего и будущего состояний, о типичных и аномальных тенденциях, об имевших место качественных изменениях и их прогнозе.

Дальнейшее развитие структурно-лингвистического подхода к моделированию ВР связано с разработкой нового интегрального метода нечеткого моделирования временных рядов и анализа тенденций и исследованием его возможностей и ограничений применительно к различным классам временных рядов.

ГЛОССАРИЙ

Временной ряд – это последовательность дискретных упорядоченных в неслучайные равноотстоящие моменты времени измерений (показателей, наблюдений) $y(t_1), y(t_2), \dots, y(t_N)$, характеризующих уровни состояний изучаемого процесса, протекающего в условиях неопределенности.

Диагностика – это процесс поиска неисправностей, проблем, дефектов, аномалий или их отсутствия.

Интерпретация – это процесс обработки данных для описания состояния и динамики процесса с целью определения их смыслового значения.

Искусственные нейронные сети – это вычислительные модели, построенные по принципу организации и функционирования биологических нейронных сетей.

Классификация – распознавание общей нечеткой тенденции НВР как вывод нечеткой аксиомы грамматики языка *LANG*.

Кластеризация – построение лексики языка *LANG*: образование локальных нечетких тенденций.

Лингвистическая оценка – средство качественного оценивания и сравнения характеристик элементов систем, используемые проектировщиками, менеджерами, лицами, принимающими решения (ЛПР), экспертами.

Лингвистическая переменная – это переменная, значениями которой являются слова или высказывания естественного или искусственного языка.

Локальная нечеткая тенденция – тенденция, которая определяется между двумя выбранными значениями нечеткого временного ряда $\tilde{x}_i$ и $\tilde{x}_j$, при $i < j$ может быть сопоставлена с

моментом времени t_i .

Локальные нечеткие тенденции (ЛНТ) – результат соединения элементарных НТ.

Мощность ACL-шкалы – количество нечетких множеств (функций принадлежности), задающих количество ее градаций m .

Начальная элементарная тенденция $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$ – элементарная тенденция, которая образована на первой паре соседних значений $\tilde{x}_{i+1}, x_i$ НВР.

Неоднородные нечеткие тенденции – нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{v}_i \neq \tilde{v}_s$.

Непротивоположные элементарные тенденции – неоднородные элементарные тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, такие, что ($\tilde{v}_i = C$ и $\tilde{v}_s \neq C$) или ($\tilde{v}_i \neq C$ и $\tilde{v}_s = C$).

Нечеткая логика – это набор стандартной (Булевой) логики, которая была расширена до обработки понятий частичной истинности для оперирования нечеткими высказываниями и рассуждениями, так как человеческие рассуждения носят приблизительный и нечеткий характер.

Нечеткая метка $\tilde{x}_i$ может быть сформирована непосредственно экспертом или получена на основе некоторого преобразования исходного временного ряда.

Нечеткая нейронная сеть – четкая нейросеть, которая построена на основе многослойной архитектуры с использованием специальных «И»-, «ИЛИ»-нейронов.

Нечеткая тенденция НВР – нечеткая метка, выражающая характер изменения (систематическое движение) последовательности нечетких уровней НВР в заданном интервале времени.
Нечеткие модели или нечеткие системы – класс моделей, предназначенных для манипулирования со значениями, представленными нечеткими множествами на основе операций нечеткой логики.

Нечеткие распознающие автоматы – один из способов описания нечетких языков.

Нечеткий временной ряд – упорядоченная в равноотстоящие моменты времени последовательность наблюдений над некоторым процессом, состояния которого изменяются во времени, если значение состояния процесса в момент t_i может быть выражено с помощью нечеткой метки $\tilde{x}_i$.

Нечеткий конечный автомат – упорядоченная шестерка [Аверкин, 1986]
 $A = \langle U, X, Y, s_0, \delta, \sigma \rangle$.

Нечетким языком L в конечном алфавите V_t является нечеткое подмножество множества всех конечных цепочек $V_t^* = \{x\}$, полученных с помощью конкатенации элементов V_t :
 $\mu_L : V_t^* \rightarrow [0,1]$.

Нечеткое множество – функциональная модель лингвистической неопределенности.

Нечеткой грамматикой называют шестерку [Аверкин, 1986] $G = \langle V_N, V_T, P, S, L, \varphi \rangle$.

Носитель нечеткой метки $\tilde{x}_i$ – это четкое множество $w \subseteq B$ таких точек $x_i \in W$, для которых $\mu_{\tilde{x}_i}(w) > 0$, где $B \subset X$ – базовое множество нечетких меток $\tilde{X}$.

Общая (основная) тенденция НВР – нечеткая тенденция, модель которой определяется между начальным и конечным уровнем НВР.

Объединение непротивоположных элементарных тенденций – $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, $\tilde{v}_i \neq C$ и $\tilde{v}_s = C$ является тенденция $\tau_j = \tau_i \cup \tau_s$, такая, что $\tau_j = \langle \tilde{v}_j, \tilde{a}_j, \Delta \tilde{t}_j, \mu_j \rangle$, для которой $\tilde{v}_j = \tilde{v}_s$, $\tilde{a}_j = Union(\tilde{a}_i, \tilde{a}_s)$, $\mu_j = \mu_i \cup \mu_s$, длительность тенденции $\Delta \tilde{t}_j = Union(\Delta \tilde{t}_i, \Delta \tilde{t}_s)$.

Объединение однородных нечетких тенденций $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, является тенденция $\tau_j = \tau_i \cup \tau_s$, такая, что $\tau_j = \langle \tilde{v}_j, \tilde{a}_j, \Delta \tilde{t}_j, \mu_j \rangle$, для которой $\tilde{v}_j = \tilde{v}_s$, $\tilde{a}_j = Union(\tilde{a}_i, \tilde{a}_s)$, $\mu_j = \mu_i \cup \mu_s$, длительность тенденции $\Delta \tilde{t}_j = Union(\Delta \tilde{t}_i, \Delta \tilde{t}_s)$.

Однородные нечеткие тенденции – нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta \tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta \tilde{t}_s, \mu_s \rangle$, для которых верно $\tilde{v}_i = \tilde{v}_s$.

Определение нечетких переменных – разбиение данных на множество интервалов (носителей нечетких множеств), определение лингвистических значений нечетких множеств и их функций принадлежности.

Оценка изменения нечетких уровней (нечеткая тенденция) – сравнительная нечеткая оценка.

Оценка нечеткого уровня ВР – абсолютная нечеткая оценка.

Подобные нечеткие тенденции – однородные нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta\tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta\tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{a}_i = \tilde{a}_s$, $\Delta\tilde{t}_i = \Delta\tilde{t}_s$.

Поиск аномалий – выявление лексических, синтаксических и семантических ошибок и нетипичных нечетких объектов и правил на основе грамматики языка *LANG*.

Поиск ассоциативных правил – определение синтаксиса языка *LANG* на основе извлечения синтаксических правил следования нечетких тенденций.

Прогноз – это приближенная оценка будущих изменений, хода событий, поведения на основе модели динамики в прошлом и настоящем.

Прогнозирование – получение многоуровневого прогноза: в терминах общей, локальной, элементарной тенденций, в терминах нечетких и числовых уровней ВР.

Производные тенденции – это такие нечеткие тенденции, которые представлены паттернами неоднородных локальных тенденций $N\tau$ и для грамматики языка *LANG* представляются нетерминальными символами.

Противоположные элементарным тенденциям – неоднородные элементарные тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta\tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta\tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{v}_i \neq C$ и $\tilde{v}_s \neq C$.

Резюмирование – генерация гранулярного описания поведения НВР, определение семантики гранул и трансляция гранулярного описания НВР в виде предложения на естественном языке.

Сегментация – представление НВР в виде последовательности терминальных символов (нечетких тенденций) грамматики языка *LANG*.

Система нечеткого логического вывода – модель описания поведения систем на естественном (или близком к естественному) языке в виде приближенных рассуждений в теории нечетких множеств и нечеткой логики, основанная на композиционном правиле вывода.

Теория нечетких множеств – это раздел прикладной математики, посвященный методам анализа неопределенных данных, в которых описание неопределенностей реальных явлений и процессов проводится с помощью понятия о множествах, не имеющих четких границ.

Универсальное множество (или универсум) – множеством значений, произвольной природы, которым задается область определения.

Фаззификация входных данных – определение степени принадлежности входных данных входным нечетким переменным.

Финальная элементарная тенденция $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta\tilde{t}_s, \mu_s \rangle$ – элементарная тенденция, которая образована на последней паре соседних значений $\tilde{x}_{i+1}, x_i$ НВР.

Функция принадлежности – параметрическая функция особого класса, которая моделирует нечеткое множество.

Частотный анализ – формирование компонент гранулярного описания НВР в виде часто встречающихся нечетких тенденций и их паттернов.

Эквивалентные нечеткие тенденции – однородные нечеткие тенденции $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \Delta\tilde{t}_i, \mu_i \rangle$ и $\tau_s = \langle \tilde{v}_s, \tilde{a}_s, \Delta\tilde{t}_s, \mu_s \rangle$, такие, что $\tilde{a}_i = \tilde{a}_s$, $\Delta\tilde{t}_i = \Delta\tilde{t}_s$, $\mu_i = \mu_s$.

Экспертная оценка – оценка, полученная в результате экспертного оценивания. Качественный аспект оцениваемого элемента системы и самой системы выражаются в лингвистической форме.

Экспертное оценивание – интеллектуальная деятельность, содержанием которой является процесс оценивания параметров, моделей, объектов, состояний, процессов, явлений в условиях «нестохастической» неопределенности.

Элементарная тенденция (ЭТ) – это такая нечеткая тенденция НВР $\tau_k \in \mathfrak{Z}$, выражающая характер изменения на участке НВР между двумя соседними нечеткими метками НВР $\tilde{x}_i, \tilde{x}_{i+1}$.

Учебное издание

ЯРУШКИНА Надежда Глебовна
АФАНАСЬЕВА Татьяна Васильевна
ПЕРФИЛЬЕВА Ирина Григорьевна

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Учебное пособие

Редактор М. В. Штаева

ЛР № 020640 от 22.10.97.

Подписано в печать 09.08.2010. Формат 60×84/16.

Усл. печ. л. 18,60. Тираж 100 экз. Заказ 814.

Ульяновский государственный технический университет
432027, г. Ульяновск, ул. Северный Венец, д. 32.

Типография УлГТУ. 432027, г. Ульяновск, ул. Северный Венец, д. 32.