

АССОЦИАТИВНЫЕ ПРАВИЛА В ПРОГРАММАХ БАНКОВСКОЙ ЛОЯЛЬНОСТИ

Н.Б. ПАКЛИН,
кандидат технических наук,
доцент кафедры информационных технологий
Рязанский филиал Московского
государственного университета экономики,
статистики и информатики

С.В. УЛАНОВ,
кандидат экономических наук,
доцент кафедры «Экономика и гуманитарные науки»
Саранский политехнический институт

В статье рассказывается о подходе к обнаружению связанных событий, основанному на применении специального алгоритма машинного обучения – ассоциативных правил. В качестве событий выступают розничные банковские услуги, в различное время приобретенные клиентами банка. Приведен практический пример того, как можно выявлять, интерпретировать и использовать ассоциативные правила в сфере розничного банкинга при продаже дополнительных услуг и моделировании лояльности клиентов.

Ключевые слова: менеджмент, ассоциативные правила, банки, методы, моделирование, лояльность.

Одним из современных направлений количественных методов анализа в экономике и бизнесе, применяемых при рассмотрении транзакционных данных, является *анализ рыночной корзины*, главным инструментом которого служат *ассоциативные правила*. Они устанавливают наличие связей между фактами покупки определенных товаров. Иначе говоря, между товарами, которые в большинстве транзакций появляются совместно, предполагается наличие ассоциативных связей. Они устанавливаются и количественно описываются с помощью *ассоциативных правил* [1].

Ассоциативные правила впервые были опробованы в розничной торговле при поиске совместно покупаемых товаров. Но на этом их применение не ограничилось. Ассоциативные правила широко используются в сегментации покупателей по поведению при покупках, анализе предпочтений клиентов, кросс-маркетинге, адресной рассылке, в медицине при определении влияния побочных эффектов от лекарств, для анализа данных по пе-

реписи населения, в прогнозировании сбоев телекоммуникационного оборудования и т. д. [2].

Эффективно применение ассоциативных правил и в сфере розничных банковских услуг. Здесь анализу подвергается не рыночная корзина, а потребленные клиентами услуги в виде кредитов, депозитных вкладов, дебетовых карт и т. п. Выявленные ассоциации можно использовать при повышении рентабельности *программ лояльности* – продаже дополнительных услуг клиентам банка, а также для выделения целевого сегмента лиц для массовой рассылки рекламы.

Еще несколько лет назад основным препятствием на пути распространения количественных методов анализа было отсутствие как алгоритмов их поиска, так и соответствующего программного обеспечения. Но даже после появления такого эффективного алгоритма, как «А Ргіогі» (исследовательский центр ІВМ, 1993 г.), прошло еще несколько лет, прежде чем на рынке информационно-аналитических систем начали предлагать как коммерческие, так и свободно распространяемые пакеты программ с дружественным графическим интерфейсом и богатыми возможностями визуализации полученных результатов. Сегодня можно говорить о другой проблеме – недостаточной практической разработанности вопросов использования информационно-аналитических технологий в банковской сфере.

Основные понятия

Базовым понятием в теории ассоциативных правил является *транзакция* – некоторое мно-

Таблица 1

Пример набора транзакций у клиентов банка

№ п/п	Транзакция
1	автокредит, экспресс-кредит, дебетовая карта
2	аренда сейфа, срочный вклад
3	срочный вклад
4	кредитная карта, дебетовая карта, «металлический» вклад, срочный вклад
5	кредитная карта, мобильный банк, экспресс-кредит, срочный вклад, дебетовая карта
6	sms-банкинг, мобильный банк, аренда сейфа, дебетовая карта
7	ипотека, экспресс-кредит, дебетовая карта
8	мобильный банк, экспресс-кредит, дебетовая карта, срочный вклад
9	вклад до востребования, автокредит, дебетовая карта, доверительное управление, срочный вклад
10	вклад до востребования, «металлический» вклад

жество событий, происходящих совместно. Например, если в 90 из 100 транзакций наблюдается совместное появление товаров *A* и *B*, то можно с большой долей уверенности предположить, что появление в транзакции товара *A* влечет за собой появление товара *B*. Следовательно, приобретение покупателем товара *A* каким-то образом влияет на вероятность покупки товара *B*, а соответствующее правило формулируется в виде «Если *A*, то *B*» (записывается $A \rightarrow B$). Зная данное правило, несложно выработать решения, позволяющие повысить продажи и получить дополнительную прибыль: всегда иметь в ассортименте оба этих товара, предлагать их покупателям совместно, размещать в торговом зале рядом и т.д. Такое правило может отражать некоторый поведенческий аспект: при покупке чая клиент, увидевший рядом на прилавке конфеты, с большей вероятностью приобретет и их.

Рассмотрим терминологию ассоциативных правил применительно к области банковских услуг. Типичная транзакция – приобретение клиентом банка какой-либо услуги. Часто он не ограничивается одной услугой, а сразу или спустя какое-то время обращается, например, за новым кредитом или оформляет кредитную карту. При этом возникает вопрос: связаны ли данные события? Это и устанавливают ассоциативные правила. Скажем, может быть обнаружено ассоциативное правило, утверждающее, что клиент, открывший срочный вклад, с вероятностью 65% оформит и дебетовую карту.

Современные банковские информационные системы, а также CRM-системы позволяют собирать и консолидировать данные о клиентах, их обращениях и потребленных услугах. В табл. 1 представлен простой пример, содержащий такие данные. В каждой строке указывается комбинация банковских услуг, приобретенных за определенный период времени. Хотя на практике приходится иметь дело с десятками и сотнями тысяч транзакций, для иллюстрации подхода ограничимся 10 транзакциями, содержащими 11 видов услуг.

Визуальный анализ таблицы показывает, что все четыре транзакции, в которых фигурирует экспресс-кредит, также включают дебетовую карту. А четыре из семи транзакций, содержащих дебетовую карту, также содержат экспресс-кредит. Услуги «экспресс-кредит» и «дебетовая карта» присутствуют у многих клиентов. Ассоциативные правила позволяют обнаруживать и количественно описывать такие совпадения.

Ассоциативное правило состоит из двух наборов предметов, называемых *условие* и *следствие*,

записываемых в виде $A \rightarrow B$, что читается следующим образом: «Из *A* следует *B*».

Условие может ограничиваться только одной услугой. Правила обычно отображаются с помощью стрелок, направленных от условия к следствию. Например, *срочный вклад* \rightarrow *дебетовая карта*. Ассоциативные правила описывают связь между наборами предметов, соответствующими условию и следствию. Она характеризуется двумя показателями – поддержкой (support) и достоверностью (confidence).

Обозначим базу данных транзакций как *D*, а число транзакций в этой базе как *N*. Каждая транзакция представляет собой некоторый набор предметов. Зададим, что *S* – поддержка, *C* – достоверность. *Поддержка ассоциативного правила* – это число транзакций, которые содержат как условие, так и следствие. Можно записать: $S(A \rightarrow B) = N_{AB} / N$, где N_{AB} – количество транзакций, содержащих *A* и *B*. *Достоверность ассоциативного правила* $A \rightarrow B$ представляет собой меру точности правила и определяется как отношение количества транзакций, содержащих и условие, и следствие (N_{AB}), к количеству транзакций, содержащих только условие (N_A): $C(A \rightarrow B) = N_{AB} / N_A$.

Если поддержка и достоверность достаточно высоки, можно с большой вероятностью утверждать, что любая будущая транзакция, которая включает условие, будет также содержать и следствие.

Вычислим поддержку и достоверность для ассоциаций из табл. 1. Возьмем ассоциацию *экспресс-кредит* \rightarrow *дебетовая карта*. Поскольку количество транзакций, содержащих как *экспресс-кредит*, так и *дебетовую карту*, равно 4, а общее число тран-

закций – 10, то поддержка данной ассоциации будет:

$$S (\text{экспресс-кредит} \rightarrow \text{дебетовая карта}) = 4 / 10 = 0,4.$$

Поскольку количество транзакций, содержащих только *экспресс-кредит* (условие), равно 4, то достоверность данной ассоциации будет:

$$C (\text{экспресс-кредит} \rightarrow \text{дебетовая карта}) = 4 / 4 = 1.$$

Иными словами, все наблюдения, содержащие *экспресс-кредит*, также содержат и *дебетовую карту*. Из чего делаем вывод: данная ассоциация может рассматриваться как правило. С точки зрения интуитивного поведения такое правило вполне объяснимо, поскольку оба банковских продукта часто у клиента фигурируют вместе.

Теперь рассмотрим ассоциацию *экспресс-кредит* → *срочный вклад*, в которой содержатся, в общем-то, совершенно разные услуги: тот, кто решил взять *экспресс-кредит*, вряд ли станет открывать вклад, и наоборот. Поддержка данной ассоциации: $S = 2 / 10 = 0,2$, а достоверность $C = 2 / 4 = 0,5$. Таким образом, сравнительно невысокая достоверность данной ассоциации дает повод усомниться в том, что она является правилом.

Кроме поддержки и достоверности при анализе ассоциативных правил используются дополнительные показатели, позволяющие оценить значимость правила. Одной из таких мер является лифт (lift), которая связывает поддержку и достоверность, и есть отношение частоты появления условия в транзакциях, которые также содержат и следствие, к частоте появления следствия в целом: $L(A \rightarrow B) = C(A \rightarrow B) / S(B)$.

Значения лифта, большие, чем единица, показывают, что условие чаще появляется в транзакциях, содержащих следствие, чем в остальных.

Рассмотрим ассоциацию *дебетовая карта* → *экспресс-кредит* из табл. 1.

$$S (\text{экспресс-кредит}) = 4/10 = 0,4;$$

$$C (\text{дебетовая карта} \rightarrow \text{экспресс-кредит}) = 4/7 = 0,57.$$

Следовательно, $L (\text{дебетовая карта} \rightarrow \text{экспресс-кредит}) = 0,57/0,4 = 1,425$.

Теперь рассмотрим ассоциацию *дебетовая карта* → *срочный вклад*.

$$S (\text{срочный вклад}) = 0,6; C (\text{дебетовая карта} \rightarrow \text{срочный вклад}) = 4/7 = 0,57.$$

$$\text{Тогда } L (\text{дебетовая карта} \rightarrow \text{срочный вклад}) = 0,57/0,6 = 0,95.$$

Большее значение лифта для первой ассоциации показывает, что дебетовые карты больше

влияют на выбор *экспресс-кредита*, чем на выбор *срочного вклада*.

Алгоритмы поиска ассоциативных правил

В процессе поиска ассоциативных правил может производиться обнаружение всех ассоциаций, поддержка и достоверность для которых превышают заданный минимум. Простейший алгоритм поиска ассоциативных правил рассматривает все возможные комбинации условий и следствий, оценивает для них поддержку и достоверность, а затем исключает все ассоциации, которые не удовлетворяют заданным ограничениям. Однако число возможных ассоциаций с увеличением числа предметов (услуг) растет экспоненциально. Поэтому применяются методики, позволяющие уменьшить количество ассоциаций, которое требуется проанализировать. Одной из наиболее распространенных является методика, основанная на обнаружении так называемых часто встречающихся наборов. На этой концепции основан известный алгоритм поиска ассоциативных правил «A Priori» [1, 3]. Алгоритм состоит из двух шагов:

1. Ищутся все часто встречающиеся предметные наборы – наборы с поддержкой больше заданного порога либо равной ему. Под *частотой* понимается количество транзакций, в которых содержится данный предметный набор.

2. На основе часто встречающихся наборов генерируются ассоциативные правила, удовлетворяющие условиям минимальной поддержки и достоверности.

В своей работе алгоритм «A Priori» использует свойство антимонотонности. Оно утверждает, что если предметный набор P не является частым, то добавление некоторого нового предмета A к набору P не делает его более частым. Это свойство позволяет значительно уменьшить пространство поиска ассоциативных правил.

Выявление наиболее интересных правил – это одна из главных задач при интерпретации результатов работы алгоритма по вычислению ассоциативных зависимостей. Дело в том, что ассоциативные правила сами по себе, как результат работы некоторого алгоритма, еще не готовы к использованию. Их нужно проинтерпретировать – т. е. понять, какие из ассоциативных правил представляют интерес, действительно ли правила отражают закономерности изучаемой области. Поэтому все множество ассоциативных правил принято делить на три вида: интересные, тривиальные и

непонятные. Первые содержат полезную информацию, которая ранее была неизвестна, но имеет логичное объяснение. Тривиальные правила – это легко объяснимые ассоциации, которые уже известны. Непонятные правила содержат информацию, которая не может быть объяснена. Они могут быть получены или на основе аномальных значений, или глубоко скрытых знаний. Напрямую такие правила для принятия решений использовать не рекомендуется. Для лучшего понимания требуется их дополнительный анализ.

Обычно количество правил ограничивают установкой параметров минимальной и максимальной поддержки (*minsupport*; *maxsupport*) и минимальной и максимальной достоверности (*minconfidence*; *maxconfidence*). Аналитики могут отдавать предпочтение правилам, которые имеют только высокую поддержку или только высокую достоверность либо, что является наиболее частым, оба этих показателя. Хотя отметим два важных аспекта. Правило со слишком большой поддержкой практически малоценно, так как оно либо всем известно, либо услуги, присутствующие в нем, являются самыми популярными. А правило со слишком большой достоверностью тоже не имеет практической ценности, так как услугами, входящими в следствие, клиент, скорее всего, уже воспользовался. Интерес представляют правила с максимальным лифтом. Поэтому в первую очередь нужно анализировать именно их.

Пример поиска ассоциаций в розничных банковских услугах

Для демонстрации предлагаемого подхода будет использоваться база данных, содержащая транзакции по розничным банковским услугам клиентам. Набор данных содержит 7 000 записей по следующим 13 услугам: sms-банкинг, автокредит; депозитный сертификат, доверительное управление, индивидуальный пенсионный счет, кобрендинговая карта; кредитная карта; потребительский кредит, сберегательный счет, срочный вклад, счет до востребования, ценные бумаги, экспресс-кредит. Сформулируем следующую задачу: банк сделал эмиссию новых кобрендинговых карт, существенно изменив условия и снизив ставки. Требуется сделать персонифицированную рекламу новой программы кобрендинга существующим и новым клиентам банка.

Одним из вариантов решения вышеописанной задачи является применение алгоритма выявле-

ния ассоциативных правил. Результаты работы алгоритма «A PriogI» с параметрами *minsupport* = 1%, *maxsupport* = 20%, *minconfidence* = 40%, *maxconfidence* = 90% приведены в табл. 2 (использовалось ПО «SPSS Clementine 12»).

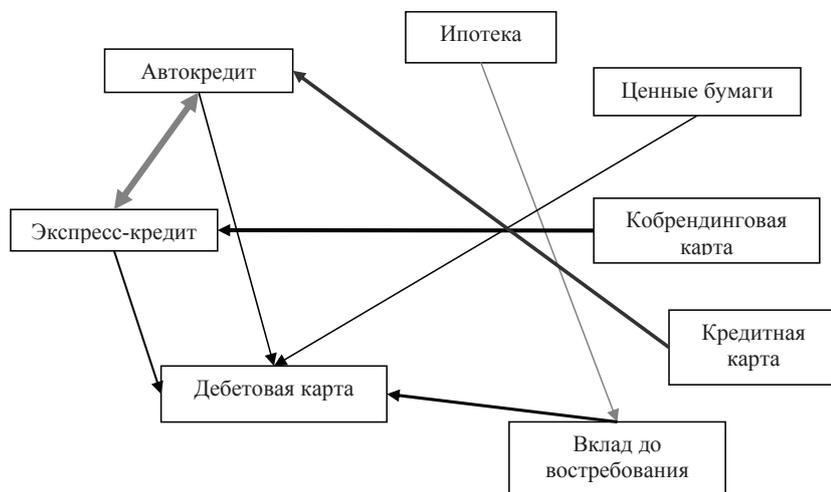
Так, в табл. 2 можно выделить интересную ассоциацию *доверительное управление* → *кобрендинговая карта*. Кобрендинговые карты одновременно включают в себя преимущества как кредитных, так и дисконтных карт. Анализ базы транзакций выявил их сильную ассоциацию с доверительным управлением. Это, видимо, объясняется следующим образом: в первую очередь менеджеры банка стараются предложить свою новую программу проверенным, лояльным клиентам (предположим, что клиенты, осуществляющие в банке доверительное управление своими активами, являются более чем лояльными).

Для визуализации связей получила распространение специальная диаграмма, которая в случае ассоциативных правил выглядит, как это показано на рисунке. Каждая услуга представляется в виде узла, а транзакция отображается в виде стрелки, направленной от условия к следствию. Сила связи между условием и следствием будет пропорциональна поддержке соответствующей ассоциации (для ее обозначения используют цветовую гамму – чем темнее оттенок линии, тем больше значение), а толщина линий пропорциональна достоверности соответствующего правила. Диаграмма удобна тем, что достаточно беглого взгляда на нее, чтобы выделить правила с наибольшей и наименьшей достоверностью или поддержкой.

Посмотрим, что изменится, если снизить порог достоверности правила. В табл. 3 приведены резуль-

Таблица 2
Ассоциативные правила для банковских услуг

№ правила	Условие	Следствие	S, %	C, %	L
1	доверительное управление	кобрендинговая карта	2,49	51,0	2,9
2	срочный вклад	кредитная карта	5,58	49,4	3,2
3	кобрендинговая карта	срочный вклад	1,78	51,1	4,5
4	кредитная карта	срочный вклад	1,78	67,6	4,4
5	кобрендинговая карта	кредитная карта	2,00	43,2	3,8
6	срочный вклад	кредитная карта	2,00	58,0	3,7



Визуализация связей для ассоциативных правил

Таблица 3

Правила, полученные при сниженном пороге достоверности

№ правила	Условие	Следствие	S, %	C, %	L
1	кредитная карта	кобрендинговая карта	3,5	22,5	1,29
2	срочный вклад	кобрендинговая карта	2,6	23,3	1,33
3	кредитная карта срочный вклад	кобрендинговая карта	1,8	31,8	1,83
4	кредитная карта экспресс-кредит	кобрендинговая карта	1,1	22,7	1,3

таты работы алгоритма при следующих параметрах: $minsupport = 1\%$; $maxsupport = 20\%$; $minconfidence = 20\%$; $maxconfidence = 40\%$.

Уменьшение порога достоверности позволило выявить дополнительные правила. Например, ассоциация (кредитная карта и экспресс-кредит) → кобрендинговая карта представляется достаточно интересной. Очевидно, что люди, пользующиеся обычными кредитными картами и краткосрочными экспресс-кредитами, с радостью воспользуются их улучшенной комбинацией. Поэтому можно предположить, что новые кобрендинговые карты банка в первую очередь заинтересуют проверенных клиентов, имеющих вклады в этом банке, а

во вторую – владельцев обычных кредитных карт и тех, кто пользовался краткосрочными экспресс-кредитами. Эта информация может учитываться при выделении целевого сегмента лиц для массовой рассылки рекламы новой услуги.

Отечественные банки в своих корпоративных информационных системах еще только накапливают информацию, на основе которой возможно выявлять ассоциативные правила. Нередко большинство клиентов, на привлечение которых банк уже затратил значительные средства, ограничивается одной банковской услугой. Но рынок розничных банковских услуг активно развивается, и потребность в подобных технологиях количественного анализа в России будет только расти.

В рассмотренном примере не учитывалась временная компонента, т. е. не обращалось внимания на последовательность приобретения услуг. Тем не менее эту информацию полезно использовать. Для этого существуют *последовательные шаблоны*, выявляющие зависимости между связанными во времени событиями (например, после погашения автокредита клиент с вероятностью 60% через полгода оформит кредитную карту).

В основе последовательных шаблонов лежат ассоциативные правила и исследование возможностей их применения в сфере розничных банковских услуг также представляет собой актуальную задачу.

Список литературы

1. Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A. I. Fast discovery of associations rules // In: «Advances in Knowledge Discovery and Data Mining», American Association for Artificial Intelligence, 1996. P. 307 – 328.
2. Количественные методы анализа в маркетинге / Под ред. Т. П. Данько, И. И. Скоробогатых. СПб.: Питер, 2005. 384 с.
3. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2009. 624 с.