

АНАЛИЗ РЫНОЧНОЙ КОРЗИНЫ ПОКУПАТЕЛЯ

Наталья АПЕТРИЙ

Совершенствование информационных технологий хранения данных позволило фирмам накапливать огромные их объемы. Наряду с этим появились и технологии, выполняющие обработку этих данных. Технологией, позволяющей получать новые знания и использовать их в процессе принятия решений, является Data mining.

Одним из методов Data mining является поиск ассоциативных правил. Впервые задача поиска ассоциативных правил была предложена для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины покупателя.

Целью анализа рыночной корзины является выявление продукта или группы продуктов, которые часто продаются вместе, с тем чтобы выявить предпочтения потребителей.

Ассоциативное правило состоит из двух компонентов: условие (*antecedent*) и следствие (*consequent*) [1, 269]. Записывается в виде $A \rightarrow B$. Правило читается так: «Из А следует В». Таким образом, ассоциативное правило формулируется в виде: «Если условие, то следствие». Пример: правило *Молоко, хлеб* \rightarrow *Масло* означает, что если покупатель приобрел молоко и хлеб, он также купит и масло сливочное.

Для оценки полезности правил используются следующие показатели [1, 270]:

1. *Поддержка* – показывает, какой процент транзакций содержит данное правило;

2. *Достоверность* – определяется как отношение количества транзакций, содержащих и условие, и следствие, к количеству транзакций, содержащих только условие.

Н.Паклин, В.Орешков [1, 582] выделяют три вида правил:

1) *интересные (полезные) правила* – содержат информацию, которая ранее была неизвестна и может быть использована в процессе принятия решений;

2) *тривиальные правила* – содержат уже известную и легко объяснимую информацию. Редко используются в принятии решений.

3) *непонятные правила* – содержат информацию, которую невозможно объяснить. Не используются в принятии решений.

Наиболее популярным алгоритмом поиска ассоциативных правил является алгоритм *a priori*. Алгоритм *a priori* описан в 1994 г. С.Рамакришнан (Ramakrishnan Srikant) и Р. Агравалом (Rakesh Agrawal) [2].

Алгоритм для поиска ассоциативных правил, *a priori* выполняется в два этапа [3]:

1) поиск наиболее часто встречающихся наборов (*frequent itemsets*) и подсчет их поддержки;

2) генерация правил на основе часто встречающихся наборов.

Как только найдены все часто встречающиеся наборы элементов, происходит генерация правил. Показатели *поддержка* и *достоверность* используются при генерации правил. Аналитик при поиске ассоциативных правил задает минимальные и максимальные значения этих величин (пороги). Правила, которые не удовлетворяют заданному порогу, исключаются и не учитываются далее при решении задачи.

Правила, которые имеют поддержку и достоверность $> 80\%$, не интересны для анализа, так как в большинстве случаев они тривиальны. Высокая достоверность говорит только о том, что товары являются лидерами продаж.

Изменяя верхние и нижние пределы поддержки и достоверности, а также мощность набора, можно избежать очевидных и неинтересных правил. *Большинство интересных правил находится при низком пороге поддержки и достоверности.*

Для примера анализа рыночной корзины покупателя были использованы данные по транзакциям, взятые с сайта www.basegroup.ru. Целью анализа было выявить полезные правила для двух предметных популярных наборов. Для анализа данных была выбрана аналитическая платформа Deductor [4]. В качестве программного инструментария был использован обработчик *Ассоциативные правила*. Для более эффективной работы алгоритма были указаны следующие пороги: минимальная поддержка – 10%; максимальная поддержка – 60%; минимальная достоверность – 40%; максимальная достоверность – 90%. Сгенерированы обработчиком 14 правил (рис. 1).

В качестве примера интерпретации ассоциативных правил проинтерпретируем правило под номером 4: *если покупатель приобрел кетчуп, то он с вероятностью 86,96 % также приобретет макаронные изделия.*

Проанализируем полученные ассоциативные правила. Правила под номерами (4-5) имеют достоверность $> 80\%$, это говорит о том, что они тривиальны. Тривиальные правила не представляют интерес для анализа. Высокая достоверность говорит только, что эти товары являются лидерами продаж. Рассмотрим правила под номерами 2, 3, 6, 9, 10. Несмотря на достоверность выше 40% эти правила таковыми не являются, т.к. условие и следствие независимы (вафли, мед, сухари и макаронные изделия просто в данной торговой точке являются часто

покупаемыми товарами). Такие правила тоже имеет малую ценность для анализа. Интересными и полезными будем считать правила под номерами 1, 7, 8, 11, 12, 13.

Результаты анализа рыночной корзины чаще всего используют для решения таких задач, как размещение товаров в магазине, стимулирование продаж (различного рода акции), принятие решений о поставках товаров.

Ассоциативные правила по элементам транзакций Товар					
Фильтр: Без фильтрации					
Итого правил: 14					
№	Условие	Следствие	Поддержка		Достоверность, %
			%	Кол-во	
1	ВАФЛИ	МЕД	13,64	6	42,86
2	ВАФЛИ	СУХАРИ	22,73	10	71,43
3	СУХАРИ	ВАФЛИ	22,73	10	71,43
4	КЕТЧУПЫ, СОУСЫ, АДЖИКА	МАКАРОННЫЕ ИЗДЕЛИЯ	45,45	20	86,96
5	МАКАРОННЫЕ ИЗДЕЛИЯ	КЕТЧУПЫ, СОУСЫ, АДЖИКА	45,45	20	83,33
6	МЕД	КЕТЧУПЫ, СОУСЫ, АДЖИКА	20,45	9	40,91
7	КЕТЧУПЫ, СОУСЫ, АДЖИКА	СЫРЫ	22,73	10	43,48
8	СЫРЫ	КЕТЧУПЫ, СОУСЫ, АДЖИКА	22,73	10	52,63
9	МАКАРОННЫЕ ИЗДЕЛИЯ	МЕД	22,73	10	41,67
10	МЕД	МАКАРОННЫЕ ИЗДЕЛИЯ	22,73	10	45,45
11	МАКАРОННЫЕ ИЗДЕЛИЯ	СЫРЫ	22,73	10	41,67
12	СЫРЫ	МАКАРОННЫЕ ИЗДЕЛИЯ	22,73	10	52,63
13	СУХАРИ	МЕД	15,91	7	50,00
14	СЫРЫ	МЕД	18,18	8	42,11

Рис. 1. Визуализатор *Правил*

Выводы

Внедрение новейших информационных технологий, востребованных в бизнесе, в образовательный процесс будет только способствовать подготовке специалистов высокой квалификации.

На кафедре Бухгалтерский учет и экономическая информатика, факультет Экономические науки Молдавского государственного университета, разработан новый учебный курс “*Data Mining*” (интеллектуальный анализ данных) в системе Moodle. Экспериментальное обучение курса *Data Mining* было проведено с двумя группами по специальности «Бухгалтерский учет».

Литература:

1. ПАКЛИН, Н., ОРЕШКОВ, В. *Бизнес-аналитика: от данных к знаниям*. Санкт-Петербург: Питер, 2009.
2. ОРЕШКОВ, В. *Поиск последовательных шаблонов*. Часть 1. [Accesat 23.05.2013] Disponibil: http://www.basegroup.ru/library/analysis/association_rules/sequential_patterns_1/
3. ОРЕШКОВ, В. *Поиск последовательных шаблонов*. Часть 2. [Accesat 23.05.2013] Disponibil: http://www.basegroup.ru/library/analysis/association_rules/sequential_patterns_2/
4. <http://www.basegroup.ru/download/deductor/>