

Д.Э. Ситников, П.Э. Ситникова, С.В. Титов, Е.В. Титова

Харьковский национальный университет радиоэлектроники, Харьков

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ОБОБЩЕННЫХ АССОЦИАТИВНЫХ ПРАВИЛ МЕТОДОМ ДЕКОМПОЗИЦИИ

Предлагается метод расчета параметров обобщенных ассоциативных правил: уровней поддержки, доверия и улучшения на основании параметров простых ассоциативных правил, которые являются составными частями агрегированной ассоциации. В отличие от метода поиска обобщенных ассоциаций с использованием таксономии признаков, предлагаемый метод не требует дополнительного сканирования БД для расчета характеристик правила, а позволяет рассчитывать их аналитическим путем с использованием построенного дерева покрытий. Объединение признаков объектов может быть осуществлено не для всей группы, а для некоторых значений атрибутов.

Ключевые слова: ассоциативные правила, параметры ассоциативной зависимости, поддержка, доверие, уровень улучшения.

Введение

Постановка проблемы. К основным задачам извлечения знаний из данных в виде нахождения регулярностей в больших массивах информации, (Data Mining) относятся:

- 1) построение классификатора;
- 2) кластеризация;
- 3) вывод правил ассоциации;
- 4) восстановление моделей зависимости.

Для решения реальной практической задачи, как правило, требуется применение нескольких подходов и методов.

Большинство методов решения было первоначально разработано в рамках теории искусственного интеллекта в 70-80-х годах, но получили распространение в последние десятилетия, когда проблема интеллектуализации обработки больших и быстро растущих объемов данных потребовала их использования в качестве надстройки над базами и хранилищами данных.

Поиск ассоциативных правил состоит в выводе правил имплицативного вида, где каждое правило отражает набор взаимно сопутствующих событий. Первоначально эта задача возникла как прикладная в сфере маркетинга (так называемая “корзина покупок”). Выявлялись товары, которые покупатели стремились приобретать вместе. Впоследствии задача поиска ассоциативных правил была сформулирована в более общем виде как задача нахождения регулярностей (логических закономерностей) в данных. Эти правила могут использоваться для анализа данных и принятия решений, а также после соответствующей оценки служить исходным материалом для построения баз знаний.

Пусть $L = I_1, I_2, \dots, I_m$ – множество признаков объектов. Пусть T – множество записей. Каждая

запись t представлена бинарным вектором $t[k] = 1$, если t содержит признак I_k и $t[k] = 0$, если t не содержит признак I_k ($k = \overline{1, m}$). Пусть X – подмножество некоторых признаков из L , т.е. $X \subseteq L$. Мы говорим, что запись t удовлетворяет X , если $\forall I_k \subseteq X, t[k] = 1$. Под ассоциативным правилом понимается выражение вида $X \rightarrow Y$, где $X \subseteq L, Y \subseteq L$. При этом $X \cap Y = \emptyset$.

Ассоциативное правило $X \rightarrow Y$ поддерживается во множестве записей с уровнем доверия (*Confidence* – сокр. *Conf*) c , если $c\%$ записей в T , которые содержат X , также содержат Y . Правило $X \rightarrow Y$ имеет поддержку (*Support* – сокр. *Supp*) s во множестве записей T , если $s\%$ записей в T содержат $X \cup Y$.

Проблема поиска ассоциативных правил состоит в генерации всех ассоциативных правил, которые преодолевают заданные пороги поддержки и доверия, т.е. имеют $Supp \geq \min Supp$ и $Cof \geq \min Conf$.

Разновидностью поиска ассоциативных правил является задача нахождения обобщенных ассоциативных зависимостей в базе данных (БД), которая решается или с использованием таксономии признаков или путем объединения ветвей дерева покрытий [1–2].

Метод, предлагаемый в [2], позволяет генерировать ассоциативные правила для случая, когда признаки объектов в БД представляют собой небинарные величины.

Обобщенные ассоциативные правила (generalized association rules) нередко обладают лучшими показателями (*Supp, Conf*) что повышает их ценность для аналитика. Однако определение параметров обобщенных ассоциативных правил на основании параметров простых ассоциаций и наоборот,

представляет **проблему**, решение которой позволит исключить дополнительное сканирование БД. Это приведет к сокращению времени на реализацию наиболее трудоемкого этапа построения ассоциативных правил.

Анализ последних исследований и публикаций. Методы и алгоритмы построения ассоциативных правил, а также области применения широко освещены в литературе и в электронных источниках [3–6]. Метод построения обобщенных ассоциативных правил с учетом таксономии признаков описан в [1]. В [2] авторами предлагается метод генерации ассоциаций путем объединения ветвей дерева покрытий. Это позволяет получать обобщенные ассоциативные зависимости для признаков, не являющихся бинарными величинами. Для описания подобных величин в [2] использовался аппарат алгебры конечных предикатов.

Пусть признак X_j объекта может принимать значения из множества $A = \{a_1, a_2, \dots, a_n\}$, тогда такой признак может быть описан с помощью конечного предиката следующим образом (1):

$$X_j^{a_i} = \begin{cases} 1, & \text{если } X_j = a_i, \\ 0, & \text{если } X_j \neq a_i. \end{cases} \quad (1)$$

Предикат $X_j^{a_i}$ как бы “узнает” значение a_i признака X_j среди всех возможных значений $a \in A$.

Подобное описание позволяет не использовать отрицание признаков: если признак X_j не принимает значение a_i , то он принимает любое значение из оставшегося набора, т.е. $\overline{X_j^{a_i}} = X_j^{a_1} \vee \dots \vee X_j^{a_{i-1}} \vee X_j^{a_{i+1}} \vee \dots \vee X_j^{a_n}$. В случае бинарных признаков для их описания используется унарный предикат, различающий два значения: X^0 и X^1 .

Для оценки полученных правил в большинстве случаев недостаточно использовать такие параметры как *Supp* и *Conf*. В [7] описаны такие параметры как *Improvement (Lift)*, *Improve*, *Validity*, *Conviction* и др., а также предлагаются новые – *Bi-lift*, *Bi-improve*, и *Bi-confidence*. В [8–9] рассматривается фильтрация полученных ассоциативных зависимостей с точки зрения такой характеристики как “интересность” правила.

Зависимость параметров обобщенных ассоциативных правил от параметров простых ассоциаций рассмотрена в [10]. Предлагается метод расчета таких характеристик обобщенных ассоциативных правил как *Supp* и *Conf*.

Цель статьи – предложить метод расчета уровня улучшения обобщенных ассоциативных правил *Improvement* на основании характеристик простых ассоциативных правил.

Метод определения уровня улучшения обобщенных ассоциативных правил

Одним из наиболее распространенных параметров оценки ассоциативных правил, наряду с *Supp* и *Conf*, является уровень улучшения *Improvement*. Для правила $X \rightarrow Y$ он характеризует степень повышения вероятности события Y при условии наступления события X относительно простой вероятности события Y (2):

$$Imp(X \rightarrow Y) = \frac{Conf(X \rightarrow Y)}{Supp(Y)} = \frac{Supp(X \rightarrow Y)}{Supp(Y)Supp(X)}. \quad (2)$$

Отметим, что тройка параметров *Supp*, *Conf* и *Imp* позволяют дать исчерпывающую характеристику ассоциации и, таким образом, составляют систему показателей, характеризующую ассоциативную зависимость.

Система двух величин (в данном случае двух признаков, двух частей ассоциативного правила X и Y) с точки зрения теории вероятностей характеризуется следующими значениями вероятностей: P_{00} , P_{01} , P_{10} , P_{11} . Для их определения требуются как минимум три параметра.

Зная значения параметров ассоциативного правила $X \rightarrow Y$: *Supp*, *Conf* и *Imp*, можно однозначно определить P_{00} , P_{01} , P_{10} , и P_{11} .

$$P_{11} = Supp(X \rightarrow Y);$$

$$P_{10} = \frac{Supp(X \rightarrow Y)}{Conf(X \rightarrow Y)} - Supp(X \rightarrow Y) = Supp(X \rightarrow \bar{Y});$$

$$P_{01} = \frac{Conf(X \rightarrow Y)}{Imp(X \rightarrow Y)} - Supp(X \rightarrow Y) = Supp(\bar{X} \rightarrow Y); \quad (3)$$

$$P_{00} = 1 - P_{11} - P_{10} - P_{01};$$

$$P_{00} = 1 + Supp(X \rightarrow Y) - \frac{Supp(X \rightarrow Y)}{Conf(X \rightarrow Y)} -$$

$$- \frac{Conf(X \rightarrow Y)}{Imp(X \rightarrow Y)} = Supp(\bar{X} \rightarrow \bar{Y}).$$

Таким образом, данная система характеристик удовлетворяет следующим условиям:

полнота – набор параметров достаточно полно характеризует ассоциативное правило;

минимальность – набор характеристик не может быть уменьшен;

не избыточность – различные характеристики учитывают разные показатели ассоциативной зависимости;

операциональность – каждая характеристика имеет понятную формулировку и характеризует определенные качества;

измеримость – каждая характеристика имеет количественную оценку.

Из трех вышеуказанных параметров два (*Supp* и *Imp*) являются симметричными, т.е. равны для правил $X \rightarrow Y$ и $Y \rightarrow X$ (4–5).

$$Supp(X \rightarrow Y) = Supp(Y \rightarrow X) = Supp(X \cap Y), \quad (4)$$

$$Imp(X \rightarrow Y) = Imp(Y \rightarrow X) = \frac{Supp(X \cap Y)}{Supp(Y)Supp(X)}. \quad (5)$$

Заметим, что знак \cap в данном случае означает, что рассматриваются объекты (записи в БД), которые содержат признаки X и Y одновременно.

Доверие *Conf* не является величиной симметричной. Однако вышеуказанные параметры составляют систему, удовлетворяющую требованиям полноты и минимальности, следовательно, может быть установлена следующая зависимость (4):

$$\begin{aligned} Conf(Y \rightarrow X) &= \frac{Supp(X \cap Y)}{Supp(Y)} = \\ &= Imp(X \rightarrow Y)Supp(X) = \\ &= \frac{Conf(X \rightarrow Y)}{Supp(Y)}Supp(X). \end{aligned} \quad (6)$$

Рассмотрим, как зависит такой параметр обобщенного ассоциативного правила как *Imp*, от параметров простых ассоциаций, входящих в его состав.

$$\begin{aligned} Imp(X \rightarrow Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k}) &= \\ &= Imp(Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k} \rightarrow X) = \\ &= \frac{Supp((Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k}) \cap X)}{Supp(Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k})Supp(Y)}. \end{aligned} \quad (7)$$

Но т.к. $Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k}$ описывают взаимно непересекающиеся множества объектов, то

$$\begin{aligned} Supp((Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k}) \cap X) &= \\ &= Supp(Y^{b_1} \cap X) + \dots + Supp(Y^{b_k} \cap X). \end{aligned} \quad (8)$$

Следовательно

$$\begin{aligned} Imp(X \rightarrow Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k}) &= \\ &= Imp(Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k} \rightarrow X) = \\ &= \frac{Supp(Y^{b_1} \cap X) + Supp(Y^{b_2} \cap X) + \dots + Supp(Y^{b_k} \cap X)}{Supp(Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k})Supp(X)} = \\ &= \frac{Conf(X \rightarrow Y^{b_1}) + \dots + Conf(X \rightarrow Y^{b_k})}{Supp(Y^{b_1} \vee Y^{b_2} \vee \dots \vee Y^{b_k})} = \\ &= \frac{Conf(X \rightarrow Y^{b_1}) + \dots + Conf(X \rightarrow Y^{b_k})}{Supp(Y^{b_1}) + Supp(Y^{b_2}) + \dots + Supp(Y^{b_k})}. \end{aligned} \quad (9)$$

Как видно из (9), уровень улучшения обобщенного ассоциативного правила может быть определен с использованием параметров простых ассоциаций, входящих в его состав.

Продемонстрируем расчет параметров обобщенных ассоциативных правил на примере:

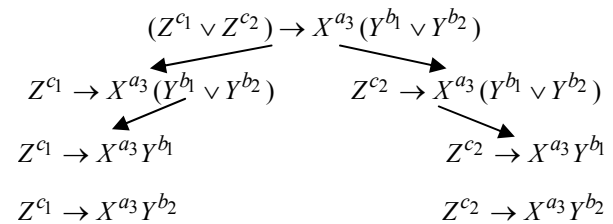
Пусть дан следующий фрагмент базы данных (табл. 1).

Таблица 1

Фрагмент базы данных

Запись Признак	1	2	3	4	5	6	7	8	9	10
X	a ₁	a ₂	a ₁	a ₂	a ₂	a ₃	a ₁	a ₃	a ₃	a ₁
Y	b ₁	b ₃	b ₁	b ₃	b ₃	b ₁	b ₂	b ₁	b ₂	b ₂
Z	c ₂	c ₁	c ₂	c ₃	c ₁	c ₁	c ₂	c ₂	c ₁	c ₂

Рассмотрим, например, покрытие $X^{a_3}(Y^{b_1} \vee Y^{b_2})(Z^{c_1} \vee Z^{c_2})$ с уровнем поддержки $Supp=3$. Из него получено ассоциативное правило $(Z^{c_1} \vee Z^{c_2}) \rightarrow X^{a_3}(Y^{b_1} \vee Y^{b_2})$. Для расчета уровня доверия данного правила декомпозируем его на простые ассоциации согласно методу, изложенному в [10].



Рассчитываем уровень доверия.

$$1) Conf(Z^{c_1} \rightarrow X^{a_3}Y^{b_1}) = \frac{1}{4};$$

$$2) Conf(Z^{c_1} \rightarrow X^{a_3}Y^{b_2}) = \frac{1}{4}.$$

Следовательно,

$$Conf(Z^{c_1} \rightarrow X^{a_3}(Y^{b_1} \vee Y^{b_2})) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}.$$

$$3) Conf(Z^{c_2} \rightarrow X^{a_3}Y^{b_1}) = \frac{1}{5};$$

$$4) Conf(Z^{c_2} \rightarrow X^{a_3}Y^{b_2}) = \frac{0}{5}.$$

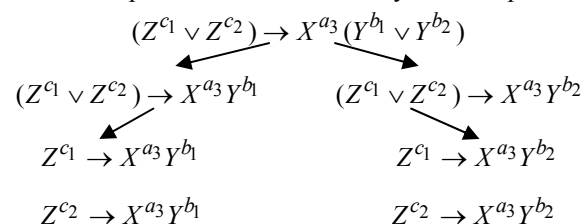
Следовательно,

$$Conf(Z^{c_2} \rightarrow X^{a_3}(Y^{b_1} \vee Y^{b_2})) = \frac{1}{5} + \frac{0}{5} = \frac{1}{5}.$$

Окончательно,

$$Conf((Z^{c_1} \vee Z^{c_2}) \rightarrow X^{a_3}(Y^{b_1} \vee Y^{b_2})) = \frac{2+1}{4+5} = \frac{3}{9}.$$

Отметим, что декомпозиция агрегированного ассоциативного правила может производиться как справа налево, так и наоборот. Для данного примера правило $(Z^{c_1} \vee Z^{c_2}) \rightarrow X^{a_3}(Y^{b_1} \vee Y^{b_2})$ может быть разложено на простые ассоциации следующим образом.



Покажем, что расчет уровня доверия для данного разложения приводит к тем же результатам.

$$1) \text{Conf}(Z^{c1} \rightarrow X^{a3} Y^{b1}) = \frac{1}{4};$$

$$2) \text{Conf}(Z^{c2} \rightarrow X^{a3} Y^{b1}) = \frac{1}{5}.$$

Следовательно,

$$\text{Conf}((Z^{c1} \vee Z^{c2}) \rightarrow X^{a3} Y^{b1}) = \frac{1+1}{4+5} = \frac{2}{9}.$$

$$3) \text{Conf}(Z^{c1} \rightarrow X^{a3} Y^{b2}) = \frac{1}{4};$$

$$4) \text{Conf}(Z^{c2} \rightarrow X^{a3} Y^{b2}) = \frac{0}{5}.$$

Следовательно,

$$\text{Conf}((Z^{c1} \vee Z^{c2}) \rightarrow X^{a3} Y^{b2}) = \frac{1+0}{4+5} = \frac{1}{9}.$$

Окончательно,

$$\text{Conf}((Z^{c1} \vee Z^{c2}) \rightarrow X^{a3} (Y^{b1} \vee Y^{b2})) = \frac{2}{9} + \frac{1}{9} = \frac{3}{9}.$$

Рассмотрим на примере расчет такой характеристики ассоциативного правила как уровень улучшения.

$$\begin{aligned} \text{Imp}((Z^{c1} \vee Z^{c2}) \rightarrow X^{a3} (Y^{b1} \vee Y^{b2})) &= \\ &= \frac{\text{Conf}((Z^{c1} \vee Z^{c2}) \rightarrow X^{a3} (Y^{b1} \vee Y^{b2}))}{\text{Supp}(X^{a3} Y^{b1}) + \text{Supp}(X^{a3} Y^{b2})}. \end{aligned}$$

В числителе дроби находится величина, посчитанная на предыдущем шаге. В знаменателе

$$\text{Supp}(X^{a3} Y^{b1}) + \text{Supp}(X^{a3} Y^{b2}) = 2 + 1 = 3.$$

Уровень доверия, таким образом, равен $\frac{1}{9}$.

Предлагаемый метод позволяет более гибко подходить к построению обобщенных ассоциатив-

ных правил. Объединение признаков (атрибутов) объектов может быть осуществлено не для всей группы, как предлагается в [1; 10], а для некоторых значений атрибутов.

Выводы

Предлагается метод расчета параметров обобщенных ассоциативных правил на основании параметров простых ассоциативных правил, которые являются составными частями агрегированной ассоциации. Уровень поддержки обобщенного покрытия определяется суммированием уровней поддержки простых покрытий, а для нахождения уровня доверия обобщенного правила применяется декомпозиция его на простые. В отличие от метода поиска обобщенных ассоциаций с использованием таксономии признаков, предлагаемый метод не требует дополнительного сканирования БД для расчета характеристик правила, а позволяет рассчитывать их аналитическим путем с использованием построенного дерева покрытий.

Использование объединения значений атрибутов позволяет определить связи, входящие в более высокие уровни обобщения, поскольку поддержка набора увеличивается, если подсчитывается вхождение группы значений атрибутов. Главным недостатком метода объединения значений атрибутов с помощью таксономии [1; 10] является тот факт, что объединение возможно только путем перехода на более высокий уровень иерархии. Это ведет к меньшей “интересности” (“полезности”) правил, т.к. они в этом случае относятся к группам значений атрибутов [10]. Метод нахождения обобщенных ассоциативных правил с помощью объединения ветвей дерева покрытий позволяет избежать этого недостатка.

Список літератури

1. Srikant R. Mining generalized association rules / R. Srikant, R. Agrawal // Future Generation Computer Systems. – 1997. – Vol. 13, Issues 2–3. – P. 161-180.
2. A method for generating aggregated association between discrete data features / E. Titova, D. Sitnikov, O. Ryabov, B. D’Cruz // WIT Transactions on Information and Communication Technologies. – 2005. – Vol. 35. – P. 25-34.
3. Марченко О.О. Актуальні проблеми Data Mining / О.О. Марченко, Т.В. Россада. – Київ: КПІ, 2017. – 150 с.
4. Корнилков А. О реализации поиска ассоциативных правил средствами языка программирования PHP [Электронный ресурс] / А.П. Корнилков, Т.В. Хабибулина // Современная техника и технологии – 2014. – № 5. – Режим доступа к журн.: <http://technology.snauka.ru/2014/05/3659>.
5. Фам К.Х. Применение ассоциативных правил в информационно-аналитической системе оценки качества предоставления телекоммуникационных услуг / К.Х. Фам, И.Ю. Квятковская // Научный вестник Новосибирского государственного технического университета. – 2015. – № 2 (59). – С. 33-42.
6. Желізняк І.Й. Правила побудови асоціативних правил на прикладі фізичних показників пацієнта / І.Й. Желізняк // Науковий вісник НЛТУ України. – 2017. – Т. 27, No. 9. – С. 107-110.
7. A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit / C. Ju, F. Bao, C. Xu, X. Fu // Discrete Dynamics in Nature and Society. – Vol. 2015. – Article ID 868634. – 10 p.
8. Using information-theoretic measures to assess association rule interestingness / J. Blanchard, F. Guillet, R. Gras, H. Briand // 5th IEEE International Conference on Data Mining ICDM’05, 2005. – United States. IEEE Computer Society. – P. 66-73.

9. Prajapati D.J. Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment / Dinesh J. Prajapati, Sanjay Garg, N.C. Chauhan // *Future Computing and Informatics Journal*. – 2017. – Vol. 2, Issue 1. – P. 19-30.

10. Assessment of extended aggregated association rules / D. Sitnikov, O. Ryabov, O. Titova, A. Kovalenko // *The 9th IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2018*. 24-27 May, 2018. – Kyiv, Ukraine.

11. Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009.

References

1. Srikant, R. and Agrawal, R. (1997), Mining generalized association rules, *Future Generation Computer Systems*, Vol. 13, Issues 2–3, pp. 161-180.

2. Sitnikov, D., Titova, E., Ryabov, O. and D'Cruz, B. (2005), A method for generating aggregated association between discrete data features, *WIT Transactions on Information and Communication Technologies*, Vol. 35, pp. 25-34.

3. Marchenko, O.O. and Rossada, T.V. (2017), “Aktual'ni problemy Data Mining” [*Actual problems of Data Mining*], KPI, Kyiv, 150 p.

4. Kornilkov, A.P. and Khabibulina, T.V. (2014), “O realizatsii poiska assotsiativnykh pravil sredstvami yazyka programirovaniya PHP” [On the implementation for association rule mining in the programming language PHP], *Modern technology and technology*, No. 5, available at: www.technology.snauka.ru/2014/05/3659.

5. Fam, K.Kh. and Kvyatkovskaya, I.Yu. (2015), “Primeneniye assotsiativnykh pravil v informatsionno-analiticheskoy sisteme otsenki kachestva predostavleniya telekommunikatsionnykh uslug” [Application of associative rules in the information and analytical system for assessing the quality of telecommunications services], *Scientific Bulletin of the Novosibirsk State Technical University*, No. 2 (59), pp. 33-42.

6. Zheliznyak, I.Y. (2017), “Pravyla pobudovy asotsiativnykh pravyl na prykladi fizychnykh pokaznykiv patsiyenta” [Rules of construction of associative rules on the example of physical indicators of the patient], *Scientific Bulletin of NLTU of Ukraine*, Vol. 27, No. 9, pp. 107-110.

7. Ju, C., Bao, F., Xu, C. and Fu, X. (2015), A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit, *Discrete Dynamics in Nature and Society*, Article ID 868634, 10 p.

8. Blanchard, J., Guillet, F., Gras, R. and Briand, H. (2005), Using information-theoretic measures to assess association rule interestingness, *5th IEEE International Conference on Data Mining ICDM'05*, IEEE Computer Society, United States, pp. 66-73.

9. Prajapati, D.J., Sanjay, Garg and Chauhan, N.C. (2017), Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment, *Future Computing and Informatics Journal*, Vol. 2, Issue 1, pp. 19-30.

10. Titova, O., Sitnikov, D., Ryabov, O. and Kovalenko, A. (2018), Assessment of extended aggregated association rules, *The 9th IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2018*, Kyiv, Ukraine.

11. Barsегян, A.A. Kupриянов, M.S., Kholod, I.I., Tess, M.D. and Yelizarov, S.I. (2009), “Analiz dannykh i protsessov” [*Data and process analysis*], BHV-Petersburg, Sankt-Petersburg, 512 p.

Надійшла до редколегії 28.12.2018

Схвалена до друку 22.01.2019

Відомості про авторів:

Ситніков Дмитро Едуардович

кандидат технічних наук доцент
професор кафедри Харківського національного
університету радіоелектроніки,
Харків, Україна
<https://orcid.org/0000-0003-1240-7900>

Ситнікова Поліна Едуардівна

кандидат технічних наук доцент
доцент кафедри Харківського
національного університету радіоелектроніки,
Харків, Україна
<https://orcid.org/0000-0002-6688-4641>

Тітов Сергій Володимирович

кандидат технічних наук доцент
доцент кафедри Харківського
національного університету радіоелектроніки,
Харків, Україна
<http://orcid.org/0000-0003-0910-4415>

Information about the authors:

Dmitro Sitnikov

Candidate of Technical Sciences Associate Professor
Professor of Department of Kharkiv National University
of Radio Electronics,
Kharkiv, Ukraine
<https://orcid.org/0000-0003-1240-7900>

Polina Sitnikova

Candidate of Technical Sciences Associate Professor
Senior Lecturer of Kharkiv National
University of Radio Electronics,
Kharkiv, Ukraine
<https://orcid.org/0000-0002-6688-4641>

Serhii Titov

Candidate of Technical Sciences Associate Professor
Senior Lecturer of Kharkiv National
University of Radio Electronics,
Kharkiv, Ukraine
<http://orcid.org/0000-0003-0910-4415>

Тітова Олена Вітольдівна

кандидат технічних наук доцент
доцент кафедри Харківського
національного університету радіоелектроніки,
Харків, Україна
<https://orcid.org/0000-0001-8894-2040>

Olena Titova

Candidate of Technical Sciences Associate Professor
Senior Lecturer of Kharkiv National
University of Radio Electronics,
Kharkiv, Ukraine
<https://orcid.org/0000-0001-8894-2040>

ВИЗНАЧЕННЯ ПАРАМЕТРІВ УЗАГАЛЬНЕНИХ АСОЦІАТИВНИХ ПРАВИЛ МЕТОДОМ ДЕКОМПОЗИЦІЇ

Д.Е. Ситніков, П.Е. Ситнікова, С.В. Тітов, О.В. Тітова

Запропоновано метод розрахунку параметрів узагальнених асоціативних правил на основі параметрів простих асоціативних правил. Рівень підтримки узагальненого покриття визначається підсумовуванням рівнів підтримки простих покриттів, а для знаходження рівня довіри узагальненого правила застосовується декомпозиція на прості. На відміну від методу з використанням таксономії ознак, даний метод не вимагає додаткового сканування БД та дозволяє отримувати параметри аналітичним шляхом з використанням побудованого дерева покриттів.

Ключові слова: асоціативні правила, параметри асоціативної залежності, підтримка, довіра, рівень поліпшення.

DEFINING PARAMETERS OF GENERALIZED ASSOCIATIVE RULES BY MEANS OF DECOMPOSITION

D. Sitnikov, P. Sitnikova, S. Titov, O. Titova

A method is proposed for calculating parameters of generalized association rules based on the parameters of simple association rules that are part of an aggregated association. Under the generalized associative rules we mean the logical dependencies between the attributes of objects in databases, where attributes can take values from some set. The apparatus of the algebra of finite predicates was used to describe such attributes. The level of support for a generalized cover is determined by summing the levels of support for simple covers. In order to find the level of confidence of a generalized rule, we decompose it into simple ones. We show that the decomposition of an aggregated association rule can be made both from right to left and vice versa. In contrast to the method of searching for generalized associations using the taxonomy of features, the proposed method does not require additional scanning of the database to calculate the characteristics of the rule, but allows calculating them analytically using the constructed tree of covers. Using the union of attribute values allows defining relationships that are included in higher levels of generalization, since the set support increases as the occurrence of the group of attribute values is counted. The main disadvantage of the method of combining attribute values using taxonomy is the fact that merging is possible only by moving to a higher level of hierarchy. This leads to less "interesting" ("utility") rules, since in this case they belong to groups of attribute values. The method of finding generalized association rules by combining the branches of the cover tree avoids this disadvantage. The proposed method allows a more flexible approach to the construction of generalized associative rules. Association of attributes of objects can be carried out not for the whole group, but for some attribute values.

Keywords: associative rules, parameters of associative dependency, support, confidence, improvement.