УДК 681.3.00:007

ПОЛНАЯ ВЗАИМНАЯ ИНФОРМАЦИЯ КАК ОБОБЩЕННЫЙ ПОКАЗАТЕЛЬ КАЧЕСТВА АССОЦИАТИВНЫХ ЗАВИСИМОСТЕЙ С БИНАРНЫМИ ПРИЗНАКАМИ

к.т.н. Д.Э. Ситников, Е.В. Титова
(представил д.т.н., проф. О.И. Сухаревский)

Проводится анализ характеристик ассоциативных правил при выделении знаний из данных (data mining). Доказывается необходимость и достаточность трех показателей: уровня поддержки, доверия и улучшения для определения взаимодействия двух частей ассоциативного правила. Предлагается интегральный показатель качества ассоциации – полная взаимная информация.

Введение. Обнаружение знаний в данных определяют как процесс выявления и идентификации в больших массивах информации новых, потенциально полезных и понятных зависимостей. В настоящее время это направление (англ. knowledge discovery in databases – KDD and data mining) интенсивно развивается, расширяется сфера его применений: в здравоохранении, биохимии, в социологии, финансовой, кредитной и маркетинговой деятельности, в правительственных службах, технической диагностике, в управлении производством и т.д.

Под термином data mining понимают центральный этап процесса выявления знаний в данных, т.е. собственно обнаружение зависимостей, регулярностей и отношений, который опирается на определенную форму повторяемости информации в базах данных. Операции этого этапа включают поиск и конструирование зависимостей, их оценку и отсеивание.

Одной из типичных задач data mining является задача открытия (генерации) правил ассоциации. Ассоциативное правило определяется как утверждение импликативного вида $A \to B$, где A и B – некоторые множества признаков объектов (записей в базе данных), причем данное утверждение должно иметь меру определенности не ниже заданного уровня.

Постановка проблемы. Одной из проблем, возникающих на этапе обработки полученных зависимостей – очистка результатов, их оценка, удаление частных и избыточных знаний. Иначе существует угроза астрономического изобилия обнаруживаемых зависимостей, подавляющая

часть которых оказывается ненужной [1].

Анализ литературы показывает, что предлагаемые меры качества ассоциативных правил: "интерес", корреляция [1], обобщенная мера ассоциации [2] и др. не всегда в достаточной мере характеризуют ассоциативную зависимость и имеют свои недостатки.

Цель статьи. Выработка интегрального показателя качества ассоциативных зависимостей при выделении знаний из данных, опирающегося на стандартные характеристики – уровень поддержки, доверие и улучшение.

Рассмотрим общепринятые характеристики ассоциативных зависимостей: поддержку (Support – сокр. Sup) и доверие (Confidence – сокр. Conf).

Поддержка определяется как отношение количества записей в базе данных (БД), удовлетворяющих этому правилу к общему количеству записей в БД (может определяться просто как количество записей в БД, удовлетворяющих данному правилу). С точки зрения теории вероятности поддержка определяет вероятность наличия у объекта (записи в БД) двух признаков A и B: P(AB). Доверие (вероятность правила) определяют как отношение количества записей, удовлетворяющих правилу к количеству записей, удовлетворяющих посылке (левой части). С точки зрения теории вероятности уровень доверия определяет условную вероятность $P_A(B) = \frac{P(AB)}{P(A)}$. Если $A \to B$ – является ассоциативным правилом, то должны выполняться следующие условия:

$$\mathrm{Sup}(A \to B) \geq \min\mathrm{Sup};$$

$$\frac{\mathrm{Sup}(A \to B)}{\mathrm{Sup}(A)} \geq \min\mathrm{Conf},$$

где Sup(A) = M(A) / M(БД) – отношение мощности множества объектов в БД, обладающих признаком A к количеству объектов в БД; $\mathrm{Sup}(A \to B) = \mathrm{Sup}(AB) = M(A \cap B) / M(БД)$ – отношение мощности множества объектов в БД, обладающих признаками A и B к количеству объектов в БД; minSup, minConf – установленные минимальные уровни поддержки и доверия.

Однако, эти две характеристики, будучи очень важными, не позволяют в достаточной степени оценить качество ассоциативного правила. Третьей характеристикой, предлагаемой для оценки ассоциации, является Improvement (сокр. Imp) – дословно “улучшение”. Imp определяется как отношение вероятности правила к вероятности результата

$$\mathrm{Imp}(A \to B) = \frac{P_A(B)}{P(B)} = \frac{\mathrm{Conf}(A \to B)}{\mathrm{Sup}(B)}.$$

Таким образом, Improvement характеризует степень повышения вероятности события B при условии наступления события A относительно простой

вероятности события B.

О важности Imp говорит хотя бы тот факт, что $\mathrm{Imp}(A \to B) = 1$ показывает, что $P_A(B) = P(B)$, т.е. признаки A и B независимы. В этом случае даже при хороших показателях уровней поддержки и доверия ассоциативное правило является абсолютно бессмысленным.

Однако этой характеристике не всегда уделяется достаточно внимания в отличие от Sup и Conf (например, в одной из известных систем выделения знаний из данных WizWhy).

Покажем, что для полной оценки взаимодействия двух признаков (двух частей ассоциативного правила) необходимы и достаточны все три характеристики.

Система двух величин (в данном случае двух признаков, двух частей ассоциативного правила) с точки зрения теории вероятности характеризуется следующими значениями вероятностей: P_{00}, P_{01}, P_{10}, P_{11}.

Зная три характеристики ассоциативного правила $A \to B$: Support, Confidence и Improvement, мы можем однозначно определить P_{00}, P_{01}, P_{10} и P_{11}.

Для начала выразим Support, Confidence и Improvement правила $A \to B$ через вероятности:

$$\mathrm{Sup}(A \to B) = P(AB) = P_{11}; \tag{1}$$

$$\mathrm{Conf}(A \to B) = \frac{\mathrm{Sup}(A \to B)}{\mathrm{Sup}(A)} = \frac{P(AB)}{P(A)} = \frac{P_{11}}{P_{11} + P_{10}}; \tag{2}$$

$$\mathrm{Imp}(A \to B) = \frac{\mathrm{Sup}(A \to B)}{\mathrm{Sup}(A)\mathrm{Sup}(B)} = \frac{P(AB)}{P(A)P(B)} = \frac{P_{11}}{(P_{11} + P_{10})(P_{11} + P_{01})}. \tag{3}$$

Отсюда:

$$P_{11} = \mathrm{Sup}(A \to B); \tag{4}$$

$$P_{10} = \frac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \mathrm{Sup}(A \to B) = \mathrm{Sup}(A \to \overline{B}); \tag{5}$$

$$P_{01} = \frac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)} - \mathrm{Sup}(A \to B) = \mathrm{Sup}(\overline{A} \to B); \tag{6}$$

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

$$= 1 + \mathrm{Sup}(A \to B) - \frac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \frac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)} = \mathrm{Sup}(\overline{A} \to \overline{B}). \tag{7}$$

Таким образом, для определения всех значений P_{00}, P_{01}, P_{10}, P_{11} требуются все три характеристики ассоциативного правила: Sup, Conf, Imp и при отсутствии хотя бы одной из них все вероятности P_{00}, P_{01}, P_{10}, P_{11} не могут быть получены.

Следует отметить, что на численные значения всех трех величин

Sup, Conf и Imp накладываются определенные ограничения:

1) $0 \le \mathrm{Sup}(A \to B) \le 1$.

2) $\mathrm{Sup}(A \to B) \le \mathrm{Conf}(A \to B) \le 1$.

Действительно, $\mathrm{Sup}(A \to B) = P_{11}$; $\mathrm{Conf}(A \to B) = \dfrac{P_{11}}{P_{11} + P_{10}}$.

Но $P_{11} \le \dfrac{P_{11}}{P_{11} + P_{10}}$, т.к. $P_{11} + P_{10} \le 1$.

3) при заданных значениях $\mathrm{Sup}(A \to B)$ и $\mathrm{Conf}(A \to B)$

$$\frac{(\mathrm{Conf}(A \to B))^2}{\mathrm{Sup}(A \to B) \cdot \mathrm{Conf}(A \to B) + \mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B)} \le$$

$$\le \mathrm{Imp}\,(A \to B) \le \frac{\mathrm{Conf}(A \to B)}{\mathrm{Sup}(A \to B)}.$$

Доказательство третьего неравенства будем вести, исходя из определения Improvement

$$\mathrm{Imp}\,(A \to B) = \frac{P_A(B)}{P(B)} = \frac{\mathrm{Conf}(A \to B)}{\mathrm{Sup}(B)} = \frac{\mathrm{Sup}(AB)}{\mathrm{Sup}(A)\mathrm{Sup}(B)}.$$

Если Sup(B) – минимальна, т.е. Sup(B) = Sup(AB), то $\mathrm{Imp}(A \to B)$ – максимален и равен

$$\mathrm{Imp}\,(A \to B) = \frac{1}{\mathrm{Sup}(A)} = \frac{\mathrm{Conf}(A \to B)}{\mathrm{Sup}(A \to B)}.$$

Если Sup(B) – максимальна, т.е. Sup(B) = 1 – Sup(A) + Sup(AB), то $\mathrm{Imp}(A \to B)$ – минимален и равен

$$\mathrm{Imp}\,(A \to B) = \frac{\mathrm{Sup}(AB)}{\mathrm{Sup}(A)(1 - \mathrm{Sup}(A) + \mathrm{Sup}(AB))}.$$

Но $\mathrm{Sup}(A) = \dfrac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)}$, откуда

$$\mathrm{Imp}(A \to B) = \frac{(\mathrm{Conf}(A \to B))^2}{\mathrm{Sup}(A \to B) \cdot \mathrm{Conf}(A \to B) + \mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B)}.$$

В процессе обработки полученных правил встает вопрос оценки качества ассоциаций. Каждая из приведенных выше величин, безусловно, характеризует полученную зависимость, однако необходима выработка некоторой интегральной характеристики, которая учитывала бы все три параметра. Как сравнить, например, два ассоциативных правила, у одного из которых больше Sup и Imp, а у другого – Conf? Какое из них “лучше” и насколько?

В [2] предложена обобщенная мера ассоциации

$$\Delta = \frac{P_{00} \cdot P_{11}}{P_{01} \cdot P_{10}}, \qquad (8)$$

которая характеризует отношение вероятности сходства признаков к вероятности их отличия.

Безусловно, эта мера может быть выражена через три характеристики ассоциативного правила. Подставляя в формулу (8) значения для P_{00}, P_{01}, P_{10}, P_{11} из формул (4 – 7), получим

$$\Delta = \frac{\mathrm{Sup}(A \to B) \cdot \left(1 + \mathrm{Sup}(A \to B) - \dfrac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \dfrac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)}\right)}{\left(\dfrac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \mathrm{Sup}(A \to B)\right) \cdot \left(\dfrac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)} - \mathrm{Sup}(A \to B)\right)}.$$

Однако, предлагаемая мера обладает существенным недостатком: она не всегда может быть рассчитана, например, если хотя бы одна из вероятностей P_{00}, P_{01}, P_{10}, P_{11} равна нулю. Действительно, пусть:

$P_{00} = 0{,}4$;
$P_{01} = 0{,}1$;
$P_{10} = 0{,}2$;
$P_{11} = 0{,}3$.

Рассчитываем характеристики Sup, Conf, Imp, Δ, пользуясь формулами (1, 2, 3, 8): Sup = 0,3; Conf = 0,6; Imp = 1,5.

$$\Delta = \frac{P_{00} \cdot P_{11}}{P_{01} \cdot P_{10}} = \frac{0{,}3 \cdot 0{,}4}{0{,}1 \cdot 0{,}2} = 6.$$

Пусть теперь:

$P_{00} = 0{,}4$;
$P_{01} = 0{,}1$;
$P_{10} = 0{,}1$;
$P_{11} = 0{,}4$.

Отметим, что ассоциация усилилась: вероятность сходства признаков возросла, соответственно возросли все три характеристики.

Sup = 0,4; Conf = 0,8; Imp = 1,6.

$$\Delta = \frac{P_{00} \cdot P_{11}}{P_{01} \cdot P_{10}} = \frac{0{,}4 \cdot 0{,}4}{0{,}1 \cdot 0{,}1} = 16.$$

До сих пор предложенная мера ассоциации работает нормально: она увеличивается.

Теперь еще усиливаем ассоциацию:

$P_{00} = 0{,}4$;
$P_{01} = 0$;
$P_{10} = 0{,}1$;

$P_{11} = 0{,}5$.

Sup = 0,5; Conf = 0,8(3); Imp = 1,0(6).

Теперь обобщенная мера ассоциации Δ не может быть рассчитана из-за равенства нулю знаменателя. Простое отбрасывание нулевой вероятности приводит к абсурдным результатам: $\Delta = \frac{P_{00} \cdot P_{11}}{P_{10}} = \frac{0{,}4 \cdot 0{,}5}{0{,}1} = 2$, т.е. при усилении ассоциации обобщенная мера уменьшилась.

Таким образом, необходим такой показатель качества ассоциации, который, во-первых, включал бы в себя все три характеристики ассоциативного правила, во-вторых, мог бы быть рассчитан при любых значениях вероятностей P_{00}, P_{01}, P_{10}, P_{11}, в-третьих, его численное значение должно возрастать при усилении ассоциации и, соответственно, уменьшаться при ослаблении.

Показатель, который отвечает всем поставленным требованиям – взаимная информация. Действительно, ассоциативное правило $A \to B$ с точки зрения теории информации можно рассматривать, как информацию о событии B, получаемую в результате сообщения о событии A.

Информация “от события к событию” определяется как $I_{A \to B} = \log \frac{P(AB)}{P(A)P(B)}$ [3], что есть не что иное, как $\log(\mathrm{Imp}(A \to B))$.

Полная взаимная информация определяется [3]

$$I_{A \leftrightarrow B} = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log \frac{P_{ij}}{p_i r_j},$$

где $P_{ij} = P((A \sim a_i)(B \sim b_j))$ – вероятность того, что A находится в состоянии a_i, а B – в состоянии b_j; $p_i = P(A \sim a_i)$ – вероятность того, что A находится в состоянии a_i; $r_j = P(B \sim b_j)$ – вероятность того, что B находится в состоянии b_j.

Таким образом, для нашего случая формула полной взаимной информации примет вид

$$\begin{aligned} I = {} & \mathrm{Sup}(A \to B) \cdot \log(\mathrm{Imp}(A \to B)) + \\ & + \mathrm{Sup}(A \to \overline{B}) \cdot \log(\mathrm{Imp}(A \to \overline{B})) + \\ & + \mathrm{Sup}(\overline{A} \to B) \cdot \log(\mathrm{Imp}(\overline{A} \to B)) + \mathrm{Sup}(\overline{A} \to \overline{B}) \cdot \log(\mathrm{Imp}(\overline{A} \to \overline{B})). \end{aligned} \quad (9)$$

Как было показано ранее, три характеристики ассоциативного правила Sup, Conf и Imp полностью определяют систему двух величин. Следовательно, все параметры в формуле (9) могут быть выражены че-

рез них. Выражения для $\mathrm{Sup}(\overline{A} \to \overline{B})$, $\mathrm{Sup}(\overline{A} \to B)$, $\mathrm{Sup}(A \to \overline{B})$ были получены в формулах (5 – 7). Выразим величины $\mathrm{Imp}(\overline{A} \to \overline{B})$, $\mathrm{Imp}(\overline{A} \to B)$, $\mathrm{Imp}(A \to \overline{B})$ через известные характеристики ассоциативного правила:

$$\mathrm{Imp}(\overline{A} \to B) = \frac{\mathrm{Sup}(\overline{A} \to B)}{\mathrm{Sup}(\overline{A})\mathrm{Sup}(B)} = \frac{P_{01}}{(P_{01} + P_{00})(P_{01} + P_{11})}; \quad (10)$$

$$\mathrm{Imp}(A \to \overline{B}) = \frac{\mathrm{Sup}(A \to \overline{B})}{\mathrm{Sup}(A)\mathrm{Sup}(\overline{B})} = \frac{P_{10}}{(P_{10} + P_{11})(P_{10} + P_{00})}; \quad (11)$$

$$\mathrm{Imp}(\overline{A} \to \overline{B}) = \frac{\mathrm{Sup}(\overline{A} \to \overline{B})}{\mathrm{Sup}(\overline{A})\mathrm{Sup}(\overline{B})} = \frac{P_{00}}{(P_{00} + P_{01})(P_{00} + P_{10})}. \quad (12)$$

Подставим в формулы (10 – 12) выражения для P_{01}, P_{00}, P_{10}, P_{11} из формул (4 – 7) и проведем алгебраические преобразования. Получим:

$$\mathrm{Imp}(\overline{A} \to B) = \frac{\mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B) \cdot \mathrm{Imp}(A \to B)}{\mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B)}; \quad (13)$$

$$\mathrm{Imp}(A \to \overline{B}) = \frac{\mathrm{Imp}(A \to B)(1 - \mathrm{Conf}(A \to B))}{\mathrm{Imp}(A \to B) - \mathrm{Conf}(A \to B)}; \quad (14)$$

$$\mathrm{Imp}(\overline{A} \to \overline{B}) = \frac{1 + \mathrm{Sup}(A \to B) - \frac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \frac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)}}{\left(1 - \frac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)}\right)\left(1 - \frac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)}\right)}. \quad (15)$$

Теперь можно записать несколько громоздкую формулу для полной взаимной информации, которая выражается через три характеристики ассоциативного правила: Sup, Conf и Imp.

$$I = \mathrm{Sup}(A \to B) \cdot \log(\mathrm{Imp}(A \to B)) + \left(\frac{\mathrm{Sup}(A \to B)}{\mathrm{Conf}(A \to B)} - \mathrm{Sup}(A \to B)\right) \times$$

$$\times \log\left(\frac{\mathrm{Imp}(A \to B)(1 - \mathrm{Conf}(A \to B))}{\mathrm{Imp}(A \to B) - \mathrm{Conf}(A \to B)}\right) + \left(\frac{\mathrm{Conf}(A \to B)}{\mathrm{Imp}(A \to B)} - \mathrm{Sup}(A \to B)\right) \times$$

$$\times \log\left(\frac{\mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B) \cdot \mathrm{Imp}(A \to B)}{\mathrm{Conf}(A \to B) - \mathrm{Sup}(A \to B)}\right) + \quad (16)$$

$$+\left(1+\mathrm{Sup}(A\to B)-\frac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}-\frac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}\right)\times$$

$$\times\log\left(\frac{1+\mathrm{Sup}(A\to B)-\dfrac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}-\dfrac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}}{\left(1-\dfrac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}\right)\left(1-\dfrac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}\right)}\right).$$

Проверим, как ведет себя полная взаимная информация при разных значениях P_{01}, P_{00}, P_{10}, P_{11}. Обратимся к нашему примеру.

При Sup = 0,3; Conf = 0,6; Imp = 1,5 полная взаимная информация, рассчитанная по формуле (16), равна I = 0,125.

При Sup = 0,4; Conf = 0,8; Imp=1,6 информация растет: I = 0,278.

Когда $P_{01} = 0$ в формуле (16) слагаемое, соответствующее этой вероятности, перестает работать (отбрасывается). Формула принимает вид:

$$I=\mathrm{Sup}(A\to B)\cdot\log(\mathrm{Imp}(A\to B))+\mathrm{Sup}(A\to\overline{B})\cdot\log(\mathrm{Imp}(A\to\overline{B}))+$$

$$+\mathrm{Sup}(\overline{A}\to\overline{B})\cdot\log(\mathrm{Imp}(\overline{A}\to\overline{B}))$$

или

$$I=\mathrm{Sup}(A\to B)\cdot\log(\mathrm{Imp}(A\to B))+\left(\frac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}-\mathrm{Sup}(A\to B)\right)\times$$

$$\times\log\left(\frac{\mathrm{Imp}(A\to B)\,(1-\mathrm{Conf}(A\to B))}{\mathrm{Imp}(A\to B)\;-\mathrm{Conf}(A\to B)}\right)+$$

$$+\left(1+\mathrm{Sup}(A\to B)-\frac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}-\frac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}\right)\times$$

$$\times\log\left(\frac{1+\mathrm{Sup}(A\to B)-\dfrac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}-\dfrac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}}{\left(1-\dfrac{\mathrm{Sup}(A\to B)}{\mathrm{Conf}(A\to B)}\right)\left(1-\dfrac{\mathrm{Conf}(A\to B)}{\mathrm{Imp}(A\to B)}\right)}\right).$$

Для нашего примера при $P_{01} = 0$: Sup = 0,5; Conf = 0,8(3); Imp = 1,0(6).

I = 0,612 – взаимная информация, как показатель качества ассоциации, выросла.

Отметим, что при $\mathrm{Imp}(A\to B) = 1$ (т.е. при независимости A и B) полная взаимная информация равна нулю, что является вполне логичным результатом.

Своего максимума I = 1 достигает при $P_{00} = 0{,}5$; $P_{01} = 0$; $P_{10} = 0$; $P_{11} = 0{,}5$: вероятность сходства максимальна Sup = 0,5; Conf = 1; Imp = 2. Безусловно,

в реальных базах данных получение таких показателей для ассоциативного правила вряд ли возможно.

Хорошими показателями поддержки и доверия, как показывает опыт, являются Sup = 0,1 ÷ 0,3; Conf = 0,7 ÷ 0,8. При Sup = 0,3; Conf = 0,8 и максимальном Imp = 8/3, полная взаимная информация достигает I ≈ 0,58.

При среднем $\mathrm{Imp}(A \to B) = \frac{\max \mathrm{Imp}(A \to B) + 1}{2}$, где единица соответствует I = 0, значения полной взаимной информации находятся в пределах I ≈ 0,2 ÷ 04.

Надо отметить, что вместо задания трех параметров при выявлении ассоциации в БД, можно выставлять требования на значения полной взаимной информации (т.е. производить отсев правил, информативность которых меньше заданного порога).

Выводы. В статье проанализированы общепринятые характеристики ассоциативных правил: уровень поддержки, доверие и улучшение. Показана важность третьей характеристики, которой не всегда уделяется достаточно внимания. Доказана необходимость и достаточность этих трех показателей для описания взаимодействия двух частей ассоциативной зависимости. Предложен интегральный показатель качества ассоциации – полная взаимная информация, который рассчитывается при любых показателях системы двух величин и позволяет количественно сравнивать полученные правила. Кроме этого, информативность правила (полная взаимная информация) может задаваться в качестве ограничения при генерации ассоциативных зависимостей во время выделения знаний из данных.

ЛИТЕРАТУРА

1. *Балабанов А.С. Выделение знаний из баз данных – передовые компьютерные технологии интеллектуального анализа данных // Математичні машини і системи. – 2001. – № 1, 2. – С. 40 – 54.*
2. *Edwards A.W.F. The measure of association in 2×2 table // Journal of the Royal Statistical Society. ser. A29. – P. 109 – 114.*
3. *Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964. – 576 с.*

Поступила 17.12.2003

***СИТНИКОВ Дмитрий Эдуардович**, канд. техн. наук, доцент, зав. кафедрой информационно-документных систем ХГАК. В 1988 году окончил Харьковский институт радиоэлектроники. Область научных интересов – выделение знаний из данных, нечеткие множества.*

***ТИТОВА Елена Витольдиевна**, младший научный сотрудник НЦ Войск ПВО, аспирантка кафедры информационно-документных систем ХГАК. В 1988 году окончила*

Харьковский институт радиоэлектроники. Область научных интересов – выделение знаний из данных.