

基于兴趣度含正负项目的关联规则挖掘方法

张玉芳¹, 熊忠阳², 彭燕³, 刘君¹

(1. 重庆大学计算机学院 重庆 沙坪坝区 400030; 2. 重庆大学电气工程博士后流动站 重庆沙坪坝区 400030;
3. 华为深圳技术有限公司 广东 深圳 518129)

【摘要】项目的引入使得挖掘出的频繁项集成倍增加,同时生成的关联规则数量更加庞大,引入兴趣度来约束从频繁项集中提取关联规则的数量。分析现有的兴趣度模型,从中选择了一种适合于含正负项目的关联规则挖掘的兴趣度方法,并且提出了置信度的一个性质,描述了含正负项目的频繁项集挖掘关联规则的算法,并对矛盾关联规则进行了分析。实验结果表明,该算法是有效和可行的。

关键词 关联规则; 置信度; 兴趣度; 负项目

中图分类号 TP311

文献标识码 A

doi:10.3969/j.issn.1001-0548.2010.03.018

Association Rule Mining Method Based on Interest Measure with Positive and Negative Items

ZHANG Yu-fang¹, XIONG Zhong-yang², PENG Yan³, and LIU Jun¹

(1. College of Computer Science, Chongqing University Shapingba Chongqing 400030;

2. Post-Doctoral Research Station of Electrical Engineering, Chongqing University Shapingba Chongqing 400030;

3. Huawei Technologies Co., Ltd. Shenzhen Guangdong 518129)

Abstract Negative item brings the increase of frequent items and makes association rules doubled. Interest measure is adopted to restrict the amount. By analyzing current interest measure models, the deviation-based interest measure is chosen and the confidence property is presented. An association rule mining algorithm based on interest measure with positive and negative items is described. And the analysis of conflict association rules is given. The experimental results indicate the given algorithm is efficient and feasible.

Key words association rule; confidence measure; interest measure; negative item

现有的关联规则挖掘算法普遍采用“支持度-置信度”的评价标准,但有时“支持度”和“置信度”都很高的强关联规则却没有应用价值。如何评价强关联规则,以剪切没有应用价值的规则,引起了人们的关注。有些学者在“支持度”和“置信度”的框架下,引入“兴趣度”修剪无用的规则,即避免生成“干扰性”的关联规则。

由于引入了负项目,使得挖掘出的频繁项集成倍增加,由此生成的关联规则数量更加庞大^[1],其中可能存在矛盾、冗余的关联规则,所以更有必要为含正负项目的关联规则引入“兴趣度”这一评价参数,作为对无价值规则的修剪工具。

关联规则挖掘算法主要考虑置信度和支持度指标,存在一定的局限性。如果人们仍把支持度作为最初项集产生的主要决定因素,把最小支持度设得

足够低,就可能产生一些对用户不具有实际应用价值甚至具有误导性的规则;把最小支持度设得过高,就有可能丢失对用户有价值的规则。为了避免生成令人产生“错觉”的关联规则,可引入兴趣度进一步约束挖掘出的关联规则。

1 兴趣度的相关介绍

1.1 几种兴趣度模型

兴趣度是表征用户对规则关注程度的度量,是用户对挖掘出的知识的新颖性、可用性和可理解性的综合考虑。不同挖掘任务和应用环境,用户对所挖掘出的规则的关注程度不同。

1.1.1 基于模板的兴趣度模型

文献[2]等定义了模板的概念。模板描述的是一组规则,用于限定哪些属性可以出现在规则前件,哪些出现在规则后件。如果一条规则匹配一个包含

收稿日期: 2008-11-12; 修回日期: 2009-09-03

基金项目: 教育部留学回国人员启动基金(教外司留[2007]1108-10); 中国博士后科学基金(20070420711)

作者简介: 张玉芳(1965-),女,博士,副教授,主要从事数据挖掘方面的研究。

的模板,则是令人感兴趣的;如果一条规则匹配一个限制的模板,则是令人缺乏兴趣的。

1.1.2 基于概率相关性的兴趣度模型

给定事务集 D ,该事务集上的关联规则 $X \Rightarrow Y$ 的兴趣度为^[3]:

$$\text{Intr}(X \Rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)\text{Supp}(Y)} \quad (1)$$

从兴趣度的度量以及支持度和置信度的度量可以看出,关联规则的评价体系属于概率论范畴。从概率论角度来讲,兴趣度反映关联规则中 X 和 Y 的关系,体现 X 和 Y 的密切程度;置信度反映规则的依赖方向;支持度反映规则在事务集中出现的频率。当 $\text{Intr}(X \Rightarrow Y) = 1$,说明 X 和 Y 相互独立,即 X 和 Y 同时发生纯属偶然,此时 $\text{Supp}(X \cup Y) = \text{Supp}(Y)\text{Supp}(X)$,所对应的关联规则是没有任何意义的。当规则的兴趣度 $\text{Intr}(X \Rightarrow Y)$ 大于1,说明对该规则越感兴趣;当 $\text{Intr}(X \Rightarrow Y)$ 大于0,但小于1时,说明对反面规则越感兴趣。

1.1.3 基于信息量的兴趣度模型

文献[4]提出计算关联规则 $X \Rightarrow Y$ 的兴趣度的函数^[4]:

$$\text{Intr}(X \Rightarrow Y) = P(X) \left[P(Y/X) \lg \left(\frac{P(Y/X)}{P(Y)} \right) + (1 - P(Y/X)) \lg \left(\frac{1 - P(Y/X)}{P(Y/X)} \right) \right] \quad (2)$$

该计算方法主要对规则的简洁性和包含的信息量进行综合度量,综合考虑规则的前件 X 和后件 Y 的概率分布的相似程度,以及用 X 的出现概率 $P(X)$ 作为规则前项 X 的简洁性的度量。 X 的属性数目越少,规则越简洁,客观感兴趣度越高。一般地, X 包含的属性越少, $P(X)$ 越大。但是,该方法忽略了 $P(Y)$ 的作用。兴趣度的这种判定方法,在一定程度上考虑了规则的前项和后项的耦合程度,耦合程度越高,用户对规则感兴趣的可能越大。但该方法忽略了 $P(Y)$ 的作用,没有考虑到紧密耦合的事件之间的关系通常是明显的和已知的^[4]。

1.1.4 基于差异思想的兴趣度模型

文献[5-6]提出了一种基于差异思想的兴趣度模型,用以指导关联规则的发现。定义规则 $X \Rightarrow Y$ 的兴趣度为:

$$\text{Intr}(X \Rightarrow Y) = \frac{P(X/Y) - P(Y)}{\max\{P(X/Y), P(Y)\}} = \frac{\text{Conf}(X \Rightarrow Y) - \text{Supp}(Y)}{\max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} \quad (3)$$

显然,规则置信度与后件的支持度之间不存在任何数量关系,然而计算出来的兴趣度可能大于0,也可能小于0。引入的分母 $\max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}$ 是一个标准化因子,使得 $|\text{Intr}(X \Rightarrow Y)| < 1$,以方便对其问题的讨论。该兴趣度的定义把规则的支持度与信任度联系起来,很好地反映了 Y 在 X 影响下发生的概率与自身发生的概率的差异,以此判断规则是否有意义。并且该定义对于由同一项集 $\{X, Y\}$ 挖掘不同规则($X \Rightarrow Y$ 和 $Y \Rightarrow X$)赋予不同的兴趣度,更符合实际。 $\text{Intr}(X \Rightarrow Y)$ 绝对值大于最小兴趣度阈值,说明 Y 的支持度与规则 $X \Rightarrow Y$ 的置信度的差异较大,则规则 $X \Rightarrow Y$ 越是新奇,用户对该规则越感兴趣;若 $\text{Intr}(X \Rightarrow Y)$ 的绝对值小于最小兴趣度阈值,说明 Y 的支持度与规则 $X \Rightarrow Y$ 的置信度差异较小,则规则 $X \Rightarrow Y$ 不新奇,不能引起用户对该规则感兴趣。

1.2 含正负项目的关联规则兴趣度

前面几种模型,从不同角度展现了兴趣度的计算评价方法。

(1) 基于模板的兴趣度,是一种主观兴趣度,需要用户参与模板的定义与兴趣度阈值的确定,通过交互式方法挖掘有趣的关联规则。但是它没有给出主观兴趣度是什么、怎么用,且模板是由用户自己具体明确的,所以比较适合于有自己的概念系统的用户。

(2) 基于概率相关性的兴趣度,主要从概率的角度展现规则的前件和后件的独立性、相关性,并定义规则的兴趣度。当前后件的相关性较小时,说明用户对该规则没有兴趣或兴趣不大。它的定义比较简洁,但没有考虑规则后件的概率,对前后件互换的一对关联规则,在兴趣度上未能区分。

(3) 基于信息量的兴趣度,从规则的简洁性和包含的信息量进行综合度量,规则越简单、信息量越大,该规则越容易引起用户的兴趣。它与基于概率相关性的兴趣度类似,也未能反映规则 $X \Rightarrow Y$ 和 $Y \Rightarrow X$ 在兴趣度上的差异。

(4) 基于差异思想的兴趣度,从规则的信任度与后件的支持度差异定义兴趣度,主要是为了消除后件的高支持率带来的对规则的高信任度的影响,从而删除所产生的用户不感兴趣、甚至错误的规则。它的一个优点是对规则 $X \Rightarrow Y$ 和 $Y \Rightarrow X$ 给予了不同的兴趣度。

本文引入基于差异思想的兴趣度模型^[6]。如果对一条规则的兴趣度越接近1,说明对该规则越感兴

趣; 如果对一条规则的兴趣度越接近-1, 说明对该规则的反面规则越感兴趣。而在含正负项目的关联规则中, 负项目是融合到频繁项集中的, 并且与正频繁项目一样被视为普通项目, 对反面规则没有意义, 所以有兴趣的含正负项目的关联规则的兴趣度应该是大于0的, 且兴趣度越接近1的规则越有价值。

2 生成有趣的含正负项目的关联规则

2.1 相关性性质

含正负项目的关联规则挖掘方法同传统的关联规则挖掘方法类似, 由两大步骤组成: (1) 挖掘一般化频繁项集; (2) 从频繁项集中挖掘含正负项目的关联规则。

Apriori性质不仅可以在频繁项集生成阶段优化频繁项集挖掘, 而且还可以应用于关联规则提取阶段。首先提出一个类似于Apriori性质的定理。

定理 1 一个给定的频繁项集 $l = X \cup Y$, 如果规则 $X \Rightarrow Y$ 是不可信的, 那么 l 中任何后件包含项集 Y 的规则都是不可信的。

证明 如果规则 $X \Rightarrow Y$ 不可信, 即 $\text{Conf}(X \Rightarrow Y) < \text{Min_Conf}$ 。

设任意项集 $\beta \subset X$, 那么 $\text{Conf}\{(X - \beta) \Rightarrow (Y \cup \beta)\} = P(XY) / P(X - \beta)$ 。显然有 $P(X - \beta) \geq P(X)$, 则 $P(XY) / P(X - \beta) \leq P(XY) / P(X)$, 即 $\text{Conf}\{(X - \beta) \Rightarrow Y \cup \beta\} \leq \text{Conf}(X \Rightarrow Y) < \text{Min_Conf}$, 故得证 l 中任何后件包含项集 Y 的规则都是不可信的。

根据该定理, 采取后件项目个数从少到多的顺序考察规则的置信度, 当短后件规则的置信度不满足要求时, 包含它的长后件规则的置信度也必定不满足要求, 因此可以减少对规则置信度的计算量, 提高效率。

定义 设 l 是一个频繁项集, $Y \subset l$, 如果规则 $(l - Y) \Rightarrow Y$ 的置信度 $\text{Conf}(l - Y \Rightarrow Y) \geq \text{Min_Conf}$, 则称后件 Y 是置信项集。

对于一个频繁 k -项集 l_k , 在考虑关联规则后件的长度由短到长的过程中, 采用类似于Apriori的算法由频繁 $(n-1)$ -项集的集合与自身连接产生候选 n -项集的集合方法, 由置信 $(n-1)$ -后件项集集合 C_{n-1} 与自身连接产生候选置信 n -后件项集的集合 Cand_n , 并同时考虑关联规则兴趣度, 生成置信的且有趣的关联规则。

2.2 含正负项目的关联规则挖掘算法

MARPN(mining association rules with positive

and negative items)算法

输入: 事务数据库 D , 最小支持度 Min_Supp , 最小置信度 Min_Conf , 最小兴趣度 Min_Intr ;

输出: 有趣的含正负项目的关联规则 R 。

算法如下:

构造频繁模式FPN树^[7], 从FPN_tree中挖掘频繁项集^[7],

For each $i \in l_k$ //对 l_k 中的每一个频繁//项集 $l_k(k \geq 2)$

if $\text{Conf}(l_k - i \Rightarrow i) \geq \text{Min_Conf}$ then
{

$C_1 \leftarrow i$; // C_1 存放置信1-后件项集

$$\text{Intr}(l_k - i \Rightarrow i) = \frac{\text{Conf}(l_k - i \Rightarrow i) - \text{Supp}(i)}{\max\{\text{Conf}(l_k - i \Rightarrow i), \text{Supp}(i)\}}$$

if $\text{Intr}(l_k - i \Rightarrow i) \geq \text{Min_Intr}$

then $R = \cup\{l_k - i \Rightarrow i\}$;

}

for $(n = 2; C_{n-1} \neq \Phi \ \&\ \& \ n < k; n++)$

{

//由置信 $(n-1)$ 后件项集构造候选 n 后件项集

$\text{Cand}_n = C \text{ Apriori_Gen}(C_{n-1}, \text{Min_Supp})$;

for each $c \in \text{Cand}_n$

if $\text{Conf}(l_k - c \Rightarrow c) \geq \text{Min_Conf}$

{

$C_n \leftarrow c$; // C_n 存放置信 n -项集

$$\text{Intr}(l_k - c \Rightarrow c) = \frac{\text{Conf}(l_k - c \Rightarrow c) - \text{Supp}(c)}{\max\{\text{Conf}(l_k - c \Rightarrow c), \text{Supp}(c)\}}$$

If $\text{Intr}(l_k - c \Rightarrow c) \geq \text{Min_Intr}$

then $R = \cup\{l_k - c \Rightarrow c\}$;

}

}

procedure $C \text{ Apriori_Gen}(C_{n-1}$:置信 $(n-1)$ -后件项集, Min_Conf :最小置信度阈值)

for each itemset $c_1 \in C_{n-1}$

for each itemset $c_2 \in C_{n-1}$

if $(c_1[1] = c_2[1]) \wedge (c_1[2] = c_2[2]) \wedge \dots \wedge (c_1[n-2] = c_2[n-2]) \wedge (|c_1[n-1]| < |c_2[n-1]|)$

then{

$h = c_1 \bowtie c_2$

//置信 $(n-1)$ 项集连接生成候选置信 n 后件项集

add h to Cand_n

}

其中 C_{n-1} 为置信 $(n-1)$ -后件项集; Min_Supp 为最小支持度阈值。

2.3 含正负项目的关联规则的矛盾性讨论

考虑了负项目以后,生成的含正负项目的关联规则可能额外增多,并且可能同时出现类似于 $X \Rightarrow Y$ 和 $X \Rightarrow \bar{Y}$ (或者 $\bar{X} \Rightarrow Y$)的矛盾规则。文献[9]通过事先指定兴趣度值及需要挖掘的规则数量,从一定程度上减少了矛盾规则存在的数量。而本文利用引入的兴趣度量对矛盾规则进行剪枝。

定理 2 如果规则 $X \Rightarrow Y$ 是感兴趣的,则 $X \Rightarrow \bar{Y}$ 和 $\bar{X} \Rightarrow Y$ 都是不感兴趣的。

证明 因为:

$$\begin{aligned} \text{Intr}(X \Rightarrow Y) &= \frac{\text{Conf}(X \Rightarrow Y) - \text{Supp}(Y)}{\max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(XY)/P(X) - P(Y)}{\max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(XY) - P(X)P(Y)}{P(X) \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} \\ \text{Intr}(X \Rightarrow \bar{Y}) &= \frac{\text{Conf}(X \Rightarrow \bar{Y}) - \text{Supp}(\bar{Y})}{\max\{\text{Conf}(X \Rightarrow \bar{Y}), \text{Supp}(\bar{Y})\}} = \\ &= \frac{P(X\bar{Y})/P(X) - P(\bar{Y})}{\max\{\text{Conf}(X \Rightarrow \bar{Y}), \text{Supp}(\bar{Y})\}} = \\ &= \frac{[P(X) - P(XY)]/P(X) - (1 - P(Y))}{\max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{[P(X) - P(XY)] - (1 - P(Y))P(X)}{P(X) \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(X) - P(XY) - P(X) + P(X)P(Y)}{P(X) \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(X)P(Y) - P(XY)}{P(X) \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} \\ \text{Intr}(\bar{X} \Rightarrow Y) &= \frac{\text{Conf}(\bar{X} \Rightarrow Y) - \text{Supp}(Y)}{\max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(\bar{X}Y)/P(\bar{X}) - P(Y)}{\max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{[P(Y) - P(XY)]/[1 - P(X)] - P(Y)}{\max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{[P(Y) - P(XY)] - P(Y)[1 - P(X)]}{[1 - P(X)] \max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(Y) - P(XY) - P(Y) + P(X)P(Y)}{[1 - P(X)] \max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} = \\ &= \frac{P(X)P(Y) - P(XY)}{[1 - P(X)] \max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} \end{aligned}$$

如果规则 $X \Rightarrow Y$ 是感兴趣的,即 $\text{Intr}(X \Rightarrow Y) \geq \text{Min_Intr} > 0$,则:

$$\frac{P(XY) - P(X)P(Y)}{P(X) \cdot \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} > 0$$

得到 $P(XY) > P(X)P(Y)$, 所以有:

$$\text{Intr}(X \Rightarrow \bar{Y}) = \frac{P(X)P(Y) - P(XY)}{P(X) \max\{\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)\}} < 0$$

$$\text{Intr}(\bar{X} \Rightarrow Y) = \frac{P(X)P(Y) - P(XY)}{[1 - P(X)] \max\{\text{Conf}(\bar{X} \Rightarrow Y), \text{Supp}(Y)\}} < 0$$

在含正负项目的关联规则中,负项目融合到频繁项集中,将其与正频繁项目一样看成是普通项目,对反面规则没有意义,不考虑绝对值。所以显然有:

$$\text{Intr}(X \Rightarrow \bar{Y}) < 0 < \text{Min_Intr}$$

$$\text{Intr}(\bar{X} \Rightarrow Y) < 0 < \text{Min_Intr}$$

故如果规则 $X \Rightarrow Y$ 是感兴趣的,则 $X \Rightarrow \bar{Y}$ 和 $\bar{X} \Rightarrow Y$ 都是不感兴趣的。

矛盾规则分为两种情况:(1)矛盾规则 $X \Rightarrow Y$ 和 $X \Rightarrow \bar{Y}$ 中后件 Y 只含有一个项目(或规则 $X \Rightarrow Y$ 和 $\bar{X} \Rightarrow Y$ 中前件 X 只含有一个项目);(2)前件和后件都含有多个项目。定理2只适用于对情况(1)的剪枝,对情况(2)由于规则的前、后件都含有多个项目,表示的是多个项目的取舍,例如 $Y = \{i_1, i_2, \dots, i_k\} = \{i_1 \wedge i_2 \wedge \dots \wedge i_k\} (k \geq 2)$,它的取补项集是各项目对应的负项目的析取 $\bar{Y} = \{\bar{i}_1 \vee \bar{i}_2 \vee \dots \vee \bar{i}_k\}$,它可以视为所有对应的项集的超集。对于任何前件完全相同,后件含有互补项目(即正项目和它对应的负项目,或者后件完全相同,前件含有互补项目)的一对关联规则,比如 $\{a, \bar{b}\} \Rightarrow \{c, d\}$ 和 $\{a, \bar{b}\} \Rightarrow \{c, \bar{d}\}$,当它们的支持度、置信度和兴趣度都满足最小阈值要求时,对它们的取舍由用户自行决定。

3 实验及结果分析

本文实验在一台CPU PIII800 MHz、内存256 M、硬盘40 G的兼容机上完成。操作系统为Windows 2000,全部程序用Visual C++6.0编写。实验采用chess数据集,该数据集包括3 196个事务和75个项目,是一个高度关联的数据集,数据集原始占用244 kb。

实验分别在几种不同的参数阈值设置下进行,实现了基于兴趣度的含正负项目的关联规则挖掘,并且与不使用兴趣度的规则挖掘结果进行了比较。由于兴趣度度量的范围是 $[-1, 1]$,所以实验直接将最小兴趣度设置为小于-1的数,作为不使用兴趣度的挖掘条件。

表1 不同参数设置下含正负项目的关联规则挖掘结果比较

支持度	置信度	兴趣度	含正负项目的关联规则数目	
			考虑兴趣度	不考虑兴趣度
0.94	0.85	0.4	11 274	50 285
		0.3	16 814	
		0.4	36 414	
	0.80	0.3	68 120	393 560
		0.4	278 718	
		0.3	332 006	650 920
0.88	0.80	0.4	719 604	
		0.3	1 268 175	4 092 154

在实验中,支持度设置应较高^[10],否则规则数量将更多。在引入了“兴趣度”这个评测参数后,生成的含正负项目的规则数量至少减少一半,并且随着最小兴趣度的增加,对无趣规则的删减效果更加明显。

4 总 结

本文引入关联规则的第3个评价参数——兴趣度,对现有的几种基本的兴趣度模型进行分析,选择了其中一种适合于含正负项目的关联规则的兴趣度模型作为判断关联规则是否有趣的方法;借用Apriori关于置信度的性质和Apriori方法挖掘含正负项目的关联规则,并对矛盾关联规则进行了分析。实验结果表明,可挖掘出完备的含正负项目的关联规则。由于事务项中引入负项目后,事务中的各项存在极大的相关性,虽然挖掘出了完备的包含正负项目的频繁项集以及关联规则,但是数量较多,而很多规则可能都由负项目构成,对应的规则可能在某些领域没有太大的应用价值。因此,怎样利用有效的技术^[8]减少频繁候选项集的存储空间和限制关联规则数量,保证仅挖掘出某领域有实用价值的关联规则是今后进一步研究的方向。

参 考 文 献

[1] AGRAWAL R, IMIELINAKI T, SWAMI A. Mining association rules between sets of items in large databases [C]//Proc of the ACM SIGMOD Conference on Management of Data. Washington, D C, USA: ACM, 1993: 207-216.

[2] KLEMETTINEN M, MANNILA H, RONKAINEN P P, et al. Finding interesting rules from large sets of discovered association rules[C]//Proc of the Third Int'l Conference on Information and Knowledge Management. New York: ACM, 1994: 401-407.

[3] SRIKANT R, AGRAWAL R. Mining quantitative association rules in large relational table[C]//Proc of the ACM SIGMOD Conf on Management of Data. Montreal, Canada: ACM, 1996: 1-12.

[4] SYMTH P, GOODMAN R M. An information theoretic approach to rule induction from database[J]. IEEE Transaction on Knowledge Data Engineering, 1992, 4(4): 301-316.

[5] 周欣,沙朝锋,朱扬勇,等. 兴趣度——关联规则的又一个阈值[J]. 计算机研究与发展, 2000, 37(5): 627-633. ZHOU Xin, SHA Chao-feng, ZHU Yang-yong, et al. Measure——another threshold in association rules[J]. Journal of Computer Research and Development, 2000, 37(5): 627-633.

[6] 周皓峰,朱扬勇,施伯乐. 一个基于兴趣度的关联规则的挖掘算法[J]. 计算机研究与发展, 2002, 39(4): 450-457. ZHOU Hao-feng, ZHU Yang-yong, SHI Bai-le. A mining algorithm for association rules based on interest measure[J]. Journal of Computer Research and Development, 2002, 39(4): 450-457.

[7] 张玉芳,熊忠阳,彭燕,等. 基于FP-Tree含正负项目的频繁项集挖掘算法[J]. 模式识别与人工智能, 2008, 21(2): 246-253. ZHANG Yu-fang, XIONG Zhong-yang, PENG Yan, et al. Mining frequent itemitem sets with positive and negative items based on FP-Tree[J]. Pattern Recognition and Artificial Intelligence, 2008, 21(2): 246-253.

[8] 黄进,尹治本. 关联规则挖掘的Apriori算法的改进[J]. 电子科技大学学报, 2003, 32(1): 76-79. HUANG Jin, YIN Zhi-ben. Improvement of apriori algorithm for mining association rules[J]. Journal of University of Electronic Science and Technology of China, 2003, 32(1): 76-79.

[9] THIRUVADY D R, WEBB G I. Mining negative rules using GRD[C]//Proc of PAKDD. Sydney, Australia: Springer Verlag, 2004: 161-165.

[10] GAN M, ZHANG M, WANG S. One extended form for negative association rules and the corresponding mining algorithm[C]//Proc of the Fourth International Conference on Machine Learning and Cybernetics. Piscataway, NJ: [s.n.], 2005: 18-21

编辑 税红