基于频繁模式图的多维关联规则挖掘算法研究

刘 波, 潘久辉
(暨南大学计算机科学系, 广东广州 510632)

摘 要: 关联规则挖掘是数据挖掘领域中重要的研究分支, 频繁项集或频繁谓词集的计算是其中的关键问题. 本文针对包括多值属性的关系数据库, 以多维关联规则挖掘为目标, 研究频繁谓词集的计算方法, 提出了 MPG 算法及 IMPG 增量算法. MPG 算法通过构建频繁模式图 MP-graph, 按照深度优先搜索方法, 动态挖掘频繁谓词集, 只需扫描数据库一次. 此外, 该方法至多增加一次数据库扫描, 就能扩展为 IMPG 算法, 进行增量关联规则挖掘. 文章分析了算法时间和空间性能, 用实验说明了算法的有效性.

关键词: 多维关联规则挖掘; 频繁谓词集; 频繁模式图; 增量式挖掘
中图分类号: TP301 文献标识码: A 文章编号: 0372-2112 (2007) 08-1612-05

Research of Algorithms Based on a Frequent Pattern Graph for Mining Multidimensional Association Rules

LIU Bo, PAN Jiu-hui
(Department of Computer Science, Jinan University, Guangzhou, Guangdong 510632, China)

Abstract: Association rule mining is an important research branch of data mining, and computing frequent itemsets or frequent predicate sets is the main problem. The paper aims at mining multidimensional association rules on a relational database which includes multi-value attributes, and studies a computing method for frequent predicate sets. It presents MPG algorithm and IMPG incremental algorithm. By constructing a frequent pattern graph and applying the depth-first-search method, MPG can find all frequent predicate sets and only scans database once. In addition, the method can be expanded into IMPG algorithm which is used for incremental association rules mining by increasing once database scan at most. The paper analyzes temporal and space performance of the algorithms, and proves their effectiveness by experiments.

Key words: multidimensional association rule; frequent predicate set; frequent pattern graph; incremental mining

1 引言

关联规则挖掘问题首先由 Agrawal 等人提出[1], 找出所有频繁项集是关联规则挖掘的关键. 著名的挖掘频繁集算法 Apriori 及产生关联规则的方法在文献[2, 3]中提出, 其后涌现了大量基于 Apriori 的改进算法. 此外, 基于频繁模式树的方法[4, 5], 超图截取的方法[6, 7] 等众多算法相继出现. 在文献[7]中证明了: 判断一个数据库中是否存在至少包括 t 个属性、支持度为 σ 的频繁项集是NP完全问题. 因此, 研究挖掘频繁集算法的主要途径是减少对数据库的扫描次数, 从而提高挖掘效率.

目前, 大部分关联规则挖掘算法是针对挖掘 2-值属性(即布尔值属性)的数据库或挖掘单维关联规则研究的, 而多值属性或多维规则更具普遍性. 单维关联规则中仅包含多次出现的单个谓词, 而多维规则中涉及两个或多个谓词. 虽然, 一些单维关联规则的方法可扩展为多维关联规则挖掘, 但应用于大规模数据库存在效率低的问题. 例如: Apriori 算法稍加修改可以找出所有的频繁谓词, 但找出所有的频繁 k-谓词集需要 k 或 $k+1$ 次数据库扫描. 文献[8, 9]提出了元规则导制的关联规则挖掘算法, 需要用规则模板描述所期望发现的规则形式, 为了发现所有多维关联规则将定义太多的规则模板. 文献[10]提出了基于遗传算法和蚂蚁算法相结合的多维规则挖掘算法, 需要设置多个参数, 结果不稳定, 不能保证得到所有频繁谓词集.

关于增量式关联规则挖掘的研究也较多, Cheung D 等人提出的 FUP 和 FUP2 算法[11, 12] 以及冯玉才等人提出的 IUA 算法[13], 其算法框架与 Apriori 一致, 需要多次扫描数据库, 产生大量的候选项目集. 杨明等提出了一种基于前缀广义表(PG List)的增量式更新算法[14], 仅

收稿日期: 2006-11-13; 修回日期: 2007-05-21
基金项目: 广东省科技攻关项目(No. 2003c101011)

须扫描原数据库一次, 扫描新增数据库两遍, 提高了挖掘和维护的效率, 但该方法是针对频繁模式树(FP-tree)结构[4]进行改进的, 与 FP-tree 一样具有较大的空间复杂度, 若用于多维关联规则挖掘, 前缀广义表和频繁模式树占用的空间将更大.

本文针对包括多值属性的关系数据库, 以多维关联规则挖掘为目标, 利用频繁模式图, 研究新的计算频繁谓词集算法及增量式更新算法.

2　相关概念

2.1　多维关联规则的描述

给定关系数据表 D, 每个记录用 m 个属性描述, 并且每一属性值均离散化, 有固定的取值个数. 下面给出多维关联规则相关定义及性质[1,15].

定义 **1**　多维关联规则是形如 $A \Rightarrow B$ 的蕴涵式, 其中 $A = A_1 \wedge A_2 \wedge \cdots\cdots A_m$, $B = B_1 \wedge B_2 \wedge \cdots\cdots B_n$, A_i 和 B_j 均是属性-值对(下面称为谓词项或者数据项), 并且 $A \cap B = \approx$(空集).

定义 **2**　k-谓词集是包含 k 个合取谓词项的集合.

定义 **3**　谓词集的频率是 D 中包含谓词集的记录数, 如果其除以记录总数的商大于或等于最小支持度, 则称它为频繁谓词集.

定义 **4**　同时满足最小支持度和最小置信度的规则称为强规则.

性质 **1**　频繁谓词集的所有非空子集都是频繁的.

性质 **2**　如果一个谓词集不是频繁的, 则它的所有超集也都不是频繁的.

多维关联规则的挖掘是一个两步过程: (1) 找出所有频繁谓词集; (2) 由频繁谓词集产生强关联规则, 其总体性能由第一步决定. 本文仅研究频繁谓词集的挖掘方法, 由频繁谓词集生成关联规则采用文献[2]的方法.

2.2　频繁模式图的描述及构造方法

假设数据表 D 的模式用属性集 $A = \{A_1, A_2, \cdots, A_m\}$ 描述, A_i 属性(离散型)的值域记为 $Dom(A_i)$, 每一记录 T 记为$(t_1, t_2, \cdots, t_m)$, 其中 t_i 用$A_i = u$ 表示(即为谓词项), 且 $u \in Dom(A_i)$.

定义 **5**　一个数据表图 $Data\text{-}G = (V, E)$, V 是 D 的所有数据项的集合, 按属性名分为 m 个子集, 即 $\bigcup_{i=1}^{m} V_i = V$, 其中 V_i 仅包括属性名为 A_i 的谓词项; E 是两个相邻数据项有向弧的集合, 即 $E = \{\langle v_i, v_j\rangle \mid v_i \in V_i, v_j \in V_j, i < j$, 且 v_i 与 v_j 为同一记录中相邻谓词项$\}$.

定义 **6**　一个数据表的频繁模式图 MP-graph = (V', E'), 是其 Data-G 的简化图, V' 是频繁 1-谓词集的集合, E' 是 V' 中频繁 1-谓词集连接弧的集合.

例如, 表 1 给出了 D 的一个示例, t_{ij} 代表A_i 属性的第 j 个取值, 其 Data-G 如图 1 所示(顶点旁边的数据是数据项 $A_i = t_{ij}$ 出现的频率). 若最小支持度为 40%, 其 MP-graph 如图 2, 顶点代表频繁 1-谓词集, 有向弧连接两个频繁 1-谓词集.

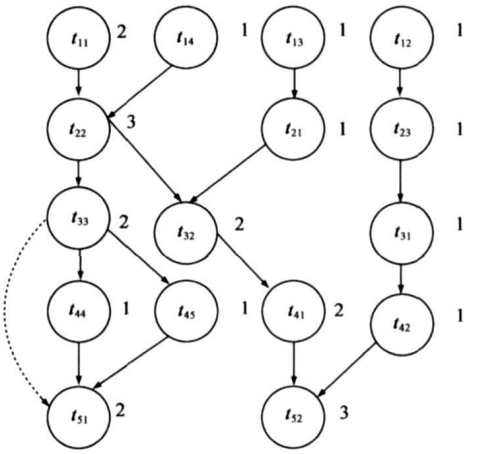

图 1　数据表图示例

构造一个数据表图仅需要扫描一次数据库, 根据扫描时统计的每一数据项频率, 可将其简化为频繁模式图, 并得到频繁 1-谓词集. 接着, 通过搜索频繁模式图, 可以求出其他频繁 k-谓词集($k \geqslant 2$).

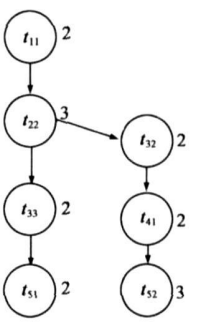

图 2　频繁模式图示例

将数据表图简化为频繁模式图的步骤如下:

(1) 找出数据表图中所有频率小于最小支持度乘以记录总数的顶点集 V_1.

例如: 图 1 中, $V_1 = \{t_{14}, t_{13}, t_{12}, t_{21}, t_{23}, t_{31}, t_{44}, t_{45}, t_{42}\}$;

(2) 在数据表图中, 增加弧$\langle t_{ij}, t_{kl}\rangle$, 其中 t_{ij} 和 t_{kl} 均为频繁 1-谓词集, 且 t_{ij} 到 t_{kl} 路径中的顶点(t_{ij} 和 t_{kl} 除外)都属于 V_1.

例如: 在图 1 中, 增加弧$\langle t_{33}, t_{51}\rangle$

(3) 去掉数据表图中的所有包含在 V_1 中的顶点及以它们为弧头或弧尾的弧, 则得到频繁模式图.

表 **1**　数据表示例

记录标识	$A1$	$A2$	$A3$	$A4$	$A5$
1	t_{11}	t_{22}	t_{33}	t_{44}	t_{51}
2	t_{14}	t_{22}	t_{33}	t_{45}	t_{51}
3	t_{11}	t_{22}	t_{32}	t_{41}	t_{52}
4	t_{13}	t_{21}	t_{32}	t_{41}	t_{52}
5	t_{12}	t_{23}	t_{31}	t_{42}	t_{52}

3　频繁谓词集挖掘算法

3.1　频繁模式图辅助信息

为了避免在频繁模式图构建时再次扫描数据库,

当构建数据表图时, 累计顶点数据项对应的记录标识(假设每一记录的标识唯一), 作为频繁模式图的辅助信息, t_{ij}的辅助信息记为 info-t_{ij}.

例如: 在图 2 中, 各顶点的辅助信息如下: info-t_{11}= {1, 3}, info-t_{22}= {1, 2, 3}, info-t_{33}= {1, 2}, info-t_{32}= {3, 4}, info-t_{41}= {3, 4}, info-t_{51}= {1, 2}, t_{52}= {3, 4, 5}.

3.2 频繁谓词集产生算法

3.2.1 算法基本思想

定义 **7** 在频繁模式图中, 从入度为零的顶点开始按深度优先方法搜索到出度为零的顶点称为一条搜索路径.

定理 **1** 在一条搜索路径上, 若一对顶点辅助信息的交集中包含的记录标识数除以总的记录数大于或等于最小支持度, 则这对顶点构成一个频繁 2-谓词集; 否则, 这对顶点不能构成一个频繁 2-谓词集.

证明: 假设一对顶点为 T_1 和 T_2, 其辅助信息的交集为 T, $N(T)$ 表示 T 中包含的记录数, 则有:

$$N(T)=N(\text{info-}T_1\cap\text{info-}T_2) \tag{1}$$

按照支持度的计算公式[1], 得公式(2), 其中: N 是总的记录数.

$$\text{Support}(T)=\frac{N(T)}{N} \tag{2}$$

按照定义 3, 对于最小支持度 Minsup, 如果下式满足, 则 T 是频繁谓词集.

$$\text{Support}(T)\geqslant\text{Minsup} \tag{3}$$

证毕.

定理 **2** 在一条搜索路径上, 若存在 $k(k>2)$ 个顶点, 它们之间的所有顶点对辅助信息的交集中包含的相同记录标识数除以总的记录数大于或等于最小支持度, 则这 k 个顶点构成一个频繁 k-谓词集.

该定理的证明与定理 1 的证明类似.

例如: 在图 2 中, 存在两条搜索路径, 有下列顶点对的辅助信息交集满足定理 1 的条件:

(1) 搜索路径 $t_{11}-t_{22}-t_{33}-t_{51}$

info-$t_{11}\cap$ info-t_{22}= {1, 3}

info-$t_{22}\cap$ info-t_{33}= {1, 2}

info-$t_{33}\cap$ info-t_{51}= {1, 2}

info-$t_{22}\cap$ info-t_{51}= {1, 2}

(2) 搜索路径 $t_{11}-t_{22}-t_{32}-t_{42}-t_{52}$

info-$t_{11}\cap$ info-t_{22}= {1, 3}

info-$t_{32}\cap$ info-t_{41}= {3, 4}

info-$t_{41}\cap$ info-t_{52}= {3, 4}

info-$t_{32}\cap$ info-t_{52}= {3, 4}

根据定理 1, {t_{11}, t_{22}}, {t_{22}, t_{33}}, {t_{33}, t_{51}}, {t_{22}, t_{51}}, {t_{32}, t_{41}}, {t_{41}, t_{52}}, {t_{32}, t_{52}} 是频繁 2-谓词集. 根据定理 2, {t_{22}, t_{33}, t_{51}}, {t_{32}, t_{41}, t_{52}} 是频繁 3-谓词集.

定理 **3** 频繁模式图中, 不同搜索路径上的两个不同顶点一定不能构成频繁谓词集.

证明: 假设 T_1 和 T_2 是频繁模式图中不同搜索路径上任意两个不同顶点, 根据数据表图的定义(定义 5)和频繁模式图的定义(定义 6), T_1 和 T_2 不可能在数据表的同一条记录中存在, 所以不可能构成频繁谓词集.

推论 **1** 一个数据库的任一频繁谓词集中的所有顶点, 一定存在于其频繁模式图中的某一条搜索路径上.

因此, 对于一个频繁模式图, 计算搜索路径上所有顶点对辅助信息的交集, 可以求出所有频繁 k-谓词集 ($k>1$).

3.2.2 MPG 算法描述

图 3 中的 MPG 算法描述了求频繁项集的过程, 其说明如下:

(1) G 为频繁模式图;

(2) V 为频繁模式图的顶点集;

(3) Path 为当前搜索路径;

(4) J 的初始状态为所有搜索路径的起点集合;

(5) Mset 为频繁谓词集集合.

算法名称: MPG

```
输入: 数据集 D, 最小支持度 Minsup
输出: 频繁谓词集集合 Mset
1. 创建频繁模式图 G;
2. Mset = V;
3. J = { V 中入度为 0 及出度为非 0 的结点}
4. while ( ! J)
5.    Path = ≅;
6.       从 J 中选择一起始点 T, 加入到 Path 中, 并从 J 中删除;
7.    Node = T;
8.    Repeat
9.          while(Node 的出度不为 0)
10.              选 Node 的下一邻接点 T-next 加入到 Path 中;
11.              Node = T-next;
12.         end while
13.         计算并比较 Path 中所有顶点对的辅助信息交集;
14.         根据 Minsup, 将得到的 Path 路径上的频繁谓词集并入 Mset
中;
15.         Path = {T};
16.         Node = T;
17.   Until (从 T 出发的所有顶点均遍历过)
18. end while
19. 输出 Mset
```

图 3 MPG 算法

3.3 频繁谓词集增量式挖掘方法

频繁谓词集的更新分为两种情况: (1) 最小支持度的阈值变化; (2) 数据表记录的增加或减少. 第(2)种情况需要首先对增量数据集进行扫描处理, 再计算所有

谓词集的变化频率;而第(1)种情况则不需要上述的处理.下面仅对第(2)种情况提出了增量更新算法 IMPG,其主要步骤如下:

(1)扫描增量数据表中的每一记录;

(2)若增加标识为 id 的记录 R,其中包括数据项 t_{ij},则 info-t_{ij} = info-t_{ij} $\cup\{id\}$,数据表记录总数增加 1;若删除标识为 id 的记录 R,其中包括数据项 t_{ij},则 info-t_{ij} = info-t_{ij} - $\{id\}$,数据表记录总数减 1.

(3)根据 info-t_{ij} 信息,t_{ij} 频率可能发生下面两种情况的变化,需按下述步骤调整 MP-graph 和频繁谓词集.

(a) t_{ij} 频率< minsup* 记录总数,若 t_{ij} 在原 MP-graph 中为弧头或弧尾,则在 MP-graph 中去掉 t_{ij} 及以其为弧头和弧尾的弧;若 t_{ij} 在原 MP-graph 中位于 $\cdots t_{mn} \rightarrow t_{ij} \rightarrow t_{kl} \cdots$ 路径上,则在 MP-graph 中去掉 t_{ij} 及与其邻接弧,并增加弧 $t_{mn} \rightarrow t_{kl}$. 此外,在原来包括 t_{ij} 的频繁谓词集中删除 t_{ij}.

(b) t_{ij} 频率 ≥minsup* 记录总数,若 t_{ij} 已在 MP-graph 中,则 MP-graph 不变.若 t_{ij} 不在 MP-graph 中,则 t_{ij} 需加入 MP-graph 的顶点集,并重新扫描一次包括在 info-t_{ij} 中的标识符所对应的记录,在 MP-graph 中增加与 t_{ij} 连接的弧.接着,对 MP-graph 中包括 t_{ij} 顶点的路径进行遍历,计算并比较这些路径上顶点辅助信息集的交集,进而求出新增频繁谓词集.

3.4 算法性能分析与比较

给定关系数据表 D, m 是 D 的属性数目, n 为记录总数, $n_i(1 \leqslant i \leqslant m)$ 是属性 $A_i((1 \leqslant i \leqslant m)$ 的离散值数目,令: $k = \mathrm{Max}\{n_i, 1 \leqslant i \leqslant m\}$. 根据数据表图和频繁模式图的定义,一条搜索路径最多包括的顶点数为 m, MP-graph 在初始状态下,总的顶点数目 $\leqslant m * k$,总路径数 $\leqslant n_1 * n_2 * n_3 * n_4 \cdots\cdots n_{m-1} * n_m \leqslant k^m$.

(1)从空间复杂度方面比较

MP-graph 有效地将数据表信息压缩到相应的路径中,与 FP-tree[4] 和 PG-List[14] 相比,虽然增加了顶点的辅助信息集,但其为线性结构,占用空间较小. FP-tree 和 PG-List 占用的空间可能很大,设数据表包括 n 条记录,每条记录包括 m 个属性,除头表节点, FP-tree 节点数目可高达 $m * n$(当 $n \gg k$ 时,远大于 MP-graph 总的顶点数目), PG-List 与 FP-tree 空间复杂度相同,对于大规模数据库,不能在内存中一次处理.因此,总体上讲, MP-graph 结构比 FP-tree 和 PG-List 简单,空间复杂度低,易于动态调整.

(2)从扫描数据库次数分析

基于 MP-graph 的 MPG 算法仅需扫描数据库一次,而增量挖掘方法在最坏情况下也只需扫描原数据库和新增数据库中的一部分记录两遍,克服了基于 Apriori 框架的算法多次扫描数据库的缺点.这与基于 FP-tree 和 PG-List 的方法相似,它们也最多仅需扫描数据库两遍.

(3)与典型的多维关联规则挖掘方法比较

文献[9]是最为典型的多维规则挖掘算法之一,其基本思想是:采用不同的数据立方体(data cube)结构,给定包括 k 个谓词($k < m$)的元规则模板,通过对数据立方体切片,由($k-1$)-谓词集产生 k-谓词集;或者, k-谓词集直接由 k 数据立方体的聚集层计算, k-维方体的单元用于存放对应 k-谓词集的计数或支持度.这种方法必须按照用户的兴趣定义规则模板,创建多个立方体,借助于 OLAP 技术,适合于对多维数据库的频繁谓词集挖掘,但不便于对一般事务或关系型数据库进行频繁谓词集挖掘(因为需转换为多维模型).

4 实验及结果分析

为了测试算法的时间性能,我们采用文献[14]相似的环境和实验方案,用 VC6.0 实现了 MPG 算法、IMPG 算法以及基于 Apriori 的多维谓词集挖掘算法 MApriori 和其增量算法 IApriori.利用网上提供的 mushroom database[16]对上述算法进行了下面 3 组实验,该数据库共有 23 个属性, 8124 条记录.

(1)随机从数据集中抽取 5 个不同的记录集(大小分别为: 1000, 2000, 4000, 6000 和 8000)进行测试,最小支持度阈值为 0.005,结果如图 4 所示.

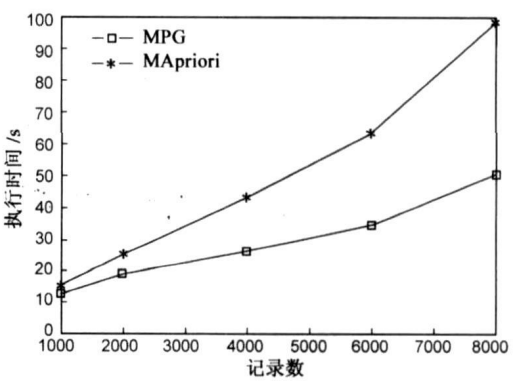

图 4　算法的执行时间

(2)随机从数据集中抽取 8000 个记录,对 6 个不同的最小支持度阈值(0.005, 0.01, 0.02, 0.03, 0.04, 0.05)进行测试,结果如图 5 所示.

(3)将随机抽取的 8000 个记录分成两部分 DB(6000 个记录)和 Ine-DB(2000 个记录),最小支持度阈值为 0.02,从 Ine-DB 中抽取 6 个不同的记录数(60, 600, 800, 1200, 1600, 2000)作为 DB 的 6 种不同增量情况,对 IMPG 和 IApriori 算法进行测试,结果如图 6 所示.

将实验结果与文献[14]公布的单维规则挖掘实验结果比较,我们的多维关联规则挖掘算法执行时间稍高

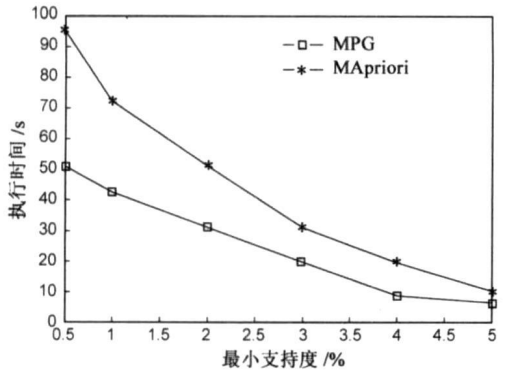

图 5 算法的执行时间

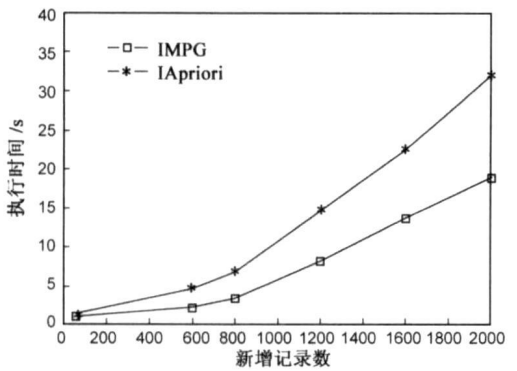

图 6 算法的执行时间

于基于 FP-tree 和基于 PG-List 的单维关联规则挖掘算法执行时间,这是合理的,因为多维关联规则挖掘需要较多时间区分不同谓词(不仅仅区分属性值). 但我们的算法执行时间比多维关联规则挖掘算法 MApriori 和增量算法 IApriori 的执行时间少.

5 结束语

本文提出了频繁模式图的概念和基于频繁模式图的快速求解频繁谓词集的算法. 仅需扫描一次或两次(对于增量挖掘)数据库,并通过计算频繁模式图的搜索路径上顶点对的辅助信息交集,就能得到所有频繁谓词集. 实验表明MPG 算法和其增量算法 IMPG 是有效可行的. 该方法比其他一些方法在时间性能或空间性能上存在优势,适合多维关联规则挖掘.

参考文献:

[1] Jiawei Han, Micheline Kamber. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2001. 150- 170.

[2] R Agrawal, T Imielinski, A Swami. Mining association rules between sets of items in large databases[A]. Proceedings of 1993 ACM-SIGMOD International Conference on Management of Data(SIGMOD' 93)[C]. New York: ACM Press, 1993. 207 - 216.

[3] R Agrawal, R Srikant. Fast algorithms for mining association rules[A]. Proceedings of the 20th International Conference on Very Large Data Bases(VLDB' 94)[C]. Morgan Kaufmann, 1994. 487- 499.

[4] J Han, J Pei, Y Yin. Mining frenquent patterns without candidate generation[A]. Proceedings of 2000 ACM-SIGMOD International Conference on Management of Data(SIGMOD' 00)[C]. New York: ACM Press, 2000. 1- 12.

[5] 宋余庆, 朱玉全, 等. 一种基于频繁模式树的约束最大频繁项目集挖掘及其更新算法[J]. 计算机研究与发展, 2005, 42(5): 777- 783.
Song Yuqing, Zhu Yuquan, et al. An Algorithm and Its Updating Algorithm Based on Frequent Pattern Tree for Mining Constrained Maximum Frequent Itemsets[J]. Journal of Computer Research and Development, 2005, 42(5): 777-783. (in Chinese)

[6] D Gunopulos, R Khardon, H Mannila, H Toivonen. Data mining, hypergraph transversals, and machine learning[A]. Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems[C]. New York: ACM Press, 1997. 209- 216.

[7] D Gunopulos, R Khardon, H Mannila, et al. Discovering all most specific sentences[J]. ACM Transactions on Database Systems, 2003, V(N): 1- 36.

[8] Yongjian Fu, Jiawei Han. Meta-rule-guided mining of association rules in relational databases[A]. Proceedings of the 1st International Workshop on Integration of Knowledge Discovery with Deductive and Object-oriented Databases[C]. 1995. 39- 46.

[9] Micheline Kamber, Jiawei Han, Jenny Y Chiang. Metarule-guided mining of multi-dimensional association rules using data cubes[A]. Proceedings of 1997 International Conference on Knowledge Discovery and Data Mining[C]. Menlo Park, CA: AAAI Press, 1997. 207- 210.

[10] 沈国强, 覃征. 一种新的多维关联规则挖掘算法[J]. 小型微型计算机系统, 2006, 27(2): 291- 294.
Shen Guo-qiang, Qin Zheng. New multidimensional association rule mining algorithm[J]. MINI-MICRO SYSTEMS, 2006, 27(2): 291- 294. (in Chinese)

[11] Cheung D, Han J, Ng V, Wong V. Maintenance of discovered association rules in large databases: an incremental updating technique[A]. Proceedings of the 12th International Conference on Data Engineering[C]. Washington: IEEE Computer Society Press, 1996. 106- 114.

[12] Cheung D, lEE S, Kao B. A general incremental technique for maintaining discovered association rules[A]. Proceedings of the 5th International Conference on Database Systems for Advanced Applications[C]. Singapore: World Scientific Press, 1997. 185- 194.

[13] 冯玉才, 冯剑琳. 关联规则的增量式更新算法[J]. 软件学报, 1998, 9(4): 301- 306.

Feng Yu-cai, Feng Jian-lin. Incremental updating algorithms for mining association rules[J]. Journal of Software, 1998, 9(4): 301- 306. (in Chinese)
[14] 杨明,孙志挥. 一种基于前缀广义表的关联规则增量式更新算法[J]. 计算机学报, 2003, 26(10): 1318- 1325.
Yang Ming, Sun Zhi-Hui. An incremental updating algorithm based on prefix general list for association rules[J]. Chinese Journal of Computers, 2003, 26(10): 1318- 1325. (in Chinese)
[15] 毛国君, 段立娟, 王实, 实云. 数据挖掘原理与算法[M]. 北京: 清华大学出版社, 2005. 64- 70.
[16] Hettich S, Bay S D. The UCI KDD Archive[DB/OL]. http://kdd.ics.uci.edu, 1999-12-12.

作者简介:

刘 波 女, 1965年生于长沙, 暨南大学副教授. 研究方向为数据挖掘、数据库、信息集成与数据仓库、智能化信息处理.
E-mail: ddxllb@163.com

潘久辉 男, 1956年生于湖南, 暨南大学教授. 研究方向为数据库、信息集成与数据仓库、程序语言理论、网格计算.

电子学报

2007年第8期 Acta Electronica Sinica No.8 2007

(总期288期) (Monthly) (Series No.288)

主管单位	中国科学技术协会	China Association for Science and Technology
主办单位	中国电子学会	Published by the Chinese Institute of Electronics, Beijing
编辑	《电子学报》编辑委员会	Edited by Editorial Board of Acta Electronica Sinica
主编	王守觉	Chief Editor: WANG Shou-jue
总编辑	刘力	Director: LIU Li
通信处	北京165信箱 (邮政编码100036)	Add: Editorial Office of Acta Electronica Sinica (P O Box 165, Beijing 100036, China)
电话	(010) 68279116, 68285082	Tel: 86-10-68279116, 68285082
传真	(010) 68173796	Fax: 86-10-68173796

Home page: http://www.elecjournal.org; http://dzxu.chinajournal.net.cn

Email: cje@elecjournal.org; wanghui@ejournal.org.cn

排版印刷	北京新瑞铭印刷有限公司	Printed by Xinruiming Co. Ltd., Beijing, China
国内总发行	北京市报刊发行局	Distributed by Domestic: Beijing Baokan Faxingju, China
国外总发行	中国国际图书贸易总公司	Foreign: China International Book Trading Corporation
国内订购处	全国各邮电局	Subscription Office——All Local Post Offices in China

国内统一刊号: CN11- 2087/TN 邮发代号(国内/国外): 2- 891/M436 国内定价 ' 32.00