

基于模糊日历约束的时序关联规则挖掘^{*})

崔晓军^{1,2} 薛永生²

(襄樊职业技术学院信息技术系 襄樊 441050)¹ (厦门大学计算机科学系 厦门 361005)²

摘要 基于日历约束的时序关联规则挖掘由于其实用性,越来越受到研究者的关注。由于现实中用户很难对时间模式进行精确描述,因此基于模糊日历的时序关联规则挖掘更有现实意义。借助模糊概念和模糊运算,对时间区间的描述很容易实现。对于用户指定的日历模式,不同的时间区间可根据它们的隶属度具有不同的权重。在模糊日历代数的基础上,结合增量挖掘和累进计数的思想,本文提出了一种基于模糊日历约束的关联规则挖掘方法,理论分析和实验结果均表明,该算法是高效可行的。

关键词 关联规则,模糊数据挖掘,模糊集,模糊日历约束

Discover Fuzzy Calendar-based Temporal Association Rules

CUI Xiao-Jun^{1,2} XUE Yong-Sheng²

(Department of IT, Xiangfan Vocational and Technical College, Xiangfan, 441050)¹

(Department of Computer Science, Xiamen University, Xiamen 361005)²

Abstract Research of mining calendar-based temporal association rules is attracting more and more attention because of its practicability. But in real life, it is impossible for user to describe the time accurately, so fuzzy temporal association rules are more useful. Based on fuzzy calendar algebra and fuzzy operators, it is easy to describe desired temporal requirements. Time intervals may have different weights according to their membership functions. Integrated with the ideas of progression and increment, this paper presents a algorithm called BFCTAR to mine fuzzy temporal association rules. Theory analysis and experiment results indicate that this algorithm is efficient and feasible.

Keywords Association rule, Fuzzy data mining, Fuzzy set, Fuzzy calendar constraint

1 引言

近来,时序关联规则挖掘已经越来越引起研究者的关注。关于时序模式的关联规则挖掘的研究,文[1]提出周期性关联规则的挖掘,文[2]提出循环关联规则挖掘,文[3,4]提出日历关联规则挖掘。由于周期性和循环模式建立在单一的时间粒度上,而日历模式建立在多时间粒度上,这与实际生活中的年、月、日、时、分、秒等多粒度时间表示更吻合,因此基于日历的时序关联规则挖掘更有实用价值。

文[4]的方法允许用户指定感兴趣的时间区间,文[3,5]中的算法可以发现在指定日历模式下所有的时序关联规则,然而上述方法的前提均是基于精确的时间区间描述,但现实中用户很难进行精确描述,因此模糊时序关联规则挖掘更有现实意义。

在文[6]提出的模糊日历代数的基础上,综合文[7,8]提出的累进和文[9]提出的增量处理方法,本文提出了基于模糊日历约束的时序关联规则挖掘方法。

2 相关概念

2.1 模糊日历代数

自从美国科学家扎德(L. A. Zadeh)1965年提出模糊(fuzzy)集合的概念以来,模糊集理论已广泛应用在各个领域,借助模糊集理论,文[10]提出了模糊日历代数的概念,利用模糊代数,用户可以定义多时间粒度的复杂的日历,对于指定的日历模式,不同的时间区间可根据它们的隶属度具有不同的权重。

日历是一个时间区间的结构化的集合。然而,对用户来讲,精确地描述他们期望的日历模式是不太可能的,借助模糊概念和模糊运算,日历的描述就容易实现了。

日历的构造离不开层次的时间粒度,对于每个时间粒度,模糊集可以用于描述该时间粒度内所有的时间区间的贡献。对于某个时间粒度的模糊描述(例如一年的中间,一个月的最初,周末)就形成一个基本的模糊日历。

定义 1(基本模糊日历 Basic Fuzzy Calendars) 给定一个时间粒度 U ,基本模糊日历 A ,通过隶属函数 μ_A 描述了该时间粒度下所有的时间区间的模糊表示。形式地, $\forall T_i \in U$, $\mu_A: U \rightarrow [0, 1]$, 函数值 $\mu_A(T_i)$ 代表时间区间 T_i 对于基本模糊日历 A 的隶属度。

图 1 给出了基本模糊日历的实例。通常选用梯形和三角形分布为模糊日历的隶属函数,但用户也可根据自己任意地指定隶属函数的图形,还可以使用“很”、“多”、“少”等形容词来进行模糊描述。对于模糊日历来说,直观上容易看出,重要的时间区间隶属度也较大,对日历的贡献也较大,如对图 1 中的模糊日历“周初”,周 1 和周 2 最重要,隶属度为 1,周 3 次之,隶属度为 0.5,周 4 以后则没有关系,隶属度为 0。模糊日历同样可以用于精确的时间区间描述,如图 2 所示。对于时序关系(如“之前”、“之后”等),模糊日历也可通过阶梯状的隶属函数进行描述,如图 3 所示。

在现实世界中,复杂的时间表达(如年末月中,但不在周初)是很常见的、有用的描述。在基本模糊日历的基础上,借助与(and)、或(or)、非(not)、异或(xor)和减(sub)五种运算可以很方便地构造任意复杂的模糊日历。

^{*})基金项目:福建省自然科学基金项目(A0310008),福建省高新技术研究开放计划重点项目(2003H043)。崔晓军 硕士研究生,副教授,主要研究方向为数据仓库、数据挖掘等。薛永生 教授,主要研究方向为数据库理论与应用、分布式数据库、数据仓库、数据挖掘、网络技术等。

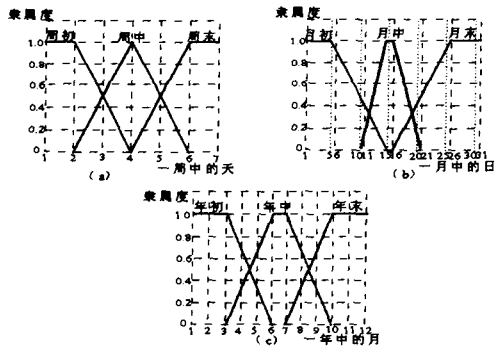


图1 不同时间粒度下的基本模糊日历(a)周(b)月(c)年

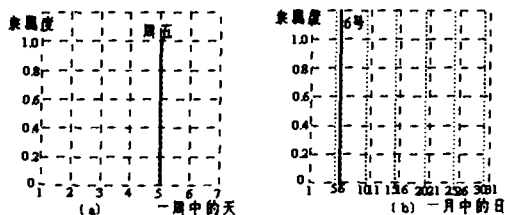


图2 精确的时间区间描述(a)周五(b)6号

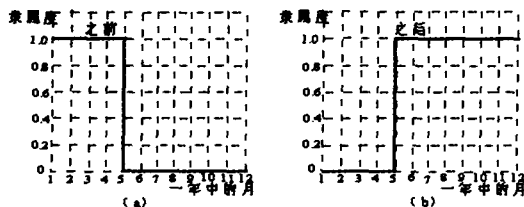


图3 时序关系的描述:(a)5月前(b)5月后

定义2(模糊日历 Fuzzy Calendars) 可以递归地定义如下:(1)一个基本模糊日历是一个模糊日历;(2)令 A 和 B 是两个模糊日历,则 A and B, A or B, not A, A xor B 和 A sub B 运算的结果也是模糊日历。

这些运算符的语义描述如下:

定义3 令 A 和 B 是两个模糊日历,分别具有隶属度 μ_A 和 μ_B , 则

(1) A and B 表示为 $A \wedge B$, 运算结果的隶属函数 $\mu_{A \wedge B}$ 定义为:

$$\mu_{A \wedge B} \equiv \mu_A \times \mu_B \quad (2.1)$$

(2) A or B 表示为 $A \vee B$, 运算结果的隶属函数 $\mu_{A \vee B}$ 定义为:

$$\mu_{A \vee B} \equiv \mu_A + \mu_B - \mu_A \times \mu_B \quad (2.2)$$

(3) not A 表示为 $\neg A$, 运算结果的隶属函数 $\mu_{\neg A}$ 定义为:

$$\mu_{\neg A} \equiv 1 - \mu_A \quad (2.3)$$

(4) A xor B 表示为 $A \oplus B$, 运算结果的隶属函数 $\mu_{A \oplus B}$ 定义为:

$$\mu_{A \oplus B} \equiv \mu_A \times (1 - \mu_B)^2 + \mu_B \times (1 - \mu_A)^2 + \mu_A \times \mu_B \times (1 - \mu_A \times \mu_B) \quad (2.4)$$

(5) A sub B 表示为 $A - B$, 运算结果的隶属函数 $\mu_{A - B}$ 定义为:

$$\mu_{A - B} \equiv \mu_A - \mu_A \times \mu_B \quad (2.5)$$

在这里用代数的和、积来定义 or、and 运算,而没有用 max、min,主要原因在于 max 或 min 在某些场合不适用,例如考虑两个模糊集的 and 运算时,希望大的模糊集对结果也有大的影响,如果采用 min 运算就会导致结果没有影响了。同理, max 运算在处理 or 运算时也不适合。

对于基本模糊日历“周初,周中,周末,月初,月中,月末,年初,年中,年末”,可以简记为 wb, wm, we, mb, mm, me, yb, ym, ye, 其隶属函数分别表示为 $\mu_{wb}, \mu_{wm}, \mu_{we}, \mu_{mb}, \mu_{mm}, \mu_{me}, \mu_{yb}, \mu_{ym}, \mu_{ye}$, 则一个复杂的模糊日历可以分解为以上基本模糊日历的组合。

例1 一个复杂模糊日历 C1: 月中且年末,或者,周末且年初。这个日历的隶属函数 μ_{C1} 可以表示为:

$$\begin{aligned} \mu_{C1} &= \mu_{(mm \wedge ye)} \vee (\mu_{we} \wedge yb) \\ &= \mu_{(mm \wedge ye)} + \mu_{(we \wedge yb)} - \mu_{(mm \wedge ye)} \times \mu_{(we \wedge yb)} \\ &= \mu_{mm} \times \mu_{ye} + \mu_{we} \times \mu_{yb} - \mu_{mm} \times \mu_{ye} \times \mu_{we} \times \mu_{yb} \end{aligned}$$

例2 一个复杂模糊日历 C2: 不在年末,并且,在年中但不在月末。这个日历的隶属函数 μ_{C2} 可以表示为:

$$\begin{aligned} \mu_{C2} &= \mu_{(\neg ye) \wedge (ym \wedge \neg me)} \\ &= \mu_{(\neg ye)} \times \mu_{(ym \wedge \neg me)} \\ &= (1 - \mu_{ye}) \times \mu_{ym} \times (1 - \mu_{me}) \end{aligned}$$

例3 一个复杂模糊日历 C3: 在月初或月末,但不是同时。这个日历的隶属函数 μ_{C3} 可以表示为:

$$\begin{aligned} \mu_{C3} &= \mu_{(mb \oplus me)} \\ &= \mu_{mb} \times (1 - \mu_{me})^2 + \mu_{me} \times (1 - \mu_{mb})^2 \\ &\quad + \mu_{mb} \times \mu_{me} \times (1 - \mu_{mb} \times \mu_{me}) \end{aligned}$$

某个具体的时间 T 相对于模糊日历 A 的隶属度(时间 T 在 A 内)记为 $\sigma_A(T)$, 利用以上运算符可以方便地计算出来。

例4 时间 $T=2006/08/20$ (周日)对于模糊日历 C1 的隶属度 $\sigma_{C1}(T)$, 根据图 1 的隶属度函数和式 6, 可以计算如下:

$$\begin{aligned} \mu_{C1} &= \mu_{mm} \times \mu_{ye} + \mu_{we} \times \mu_{yb} - \mu_{mm} \times \mu_{ye} \times \mu_{we} \times \mu_{yb} \\ &= \mu_{mm}(20) \times \mu_{ye}(8) + \mu_{we}(7) \times \mu_{yb}(8) \\ &\quad - \mu_{mm}(20) \times \mu_{ye}(8) \times \mu_{we}(7) \times \mu_{yb}(8) \\ &= 0.2 \times 0.33 + 1.0 \times 0 - 0.2 \times 0.33 \times 1.0 \times 0 \\ &= 0.066 \end{aligned}$$

2.2 模糊时序关联规则

设 $I = \{i_1, i_2, \dots, i_m\}$ 是数据项的集合, 任务相关的数据 D 是数据库事务集合, D 中的每个事务 T 都具有一个时间戳 TID, 且 $T \subseteq I$, 时间信息指示事务发生的时间。假定事务数据库 D 被划分为 n 个集合 D_1, D_2, \dots, D_n , 每个 D_i 包含所有在最小时间粒度决定的时间段 T_i 内发生的事务(如最小粒度为天, 则每个 D_i 包含一天内发生的事务)。模糊时序关联规则挖掘的任务就是发现在模糊日历模式给定的时间区间内有趣的关联规则。这些有趣的规则的加权支持度和加权置信度应该大于用户指定的最小支持度和最小置信度阈值。

设 FC 为用户指定的模糊日历, w_i 为相应于分区 D_i 的时间段 T_i 的隶属度, w_i 的计算过程如例 4。对于给定的项集 $A \subseteq I$, 事务 t 包含 A 当且仅当 $A \subseteq t$, 设 $|D_i(A)|$ 表示分区 D_i 中包含项集 A 的事务数, 则项集 A 在 D 中的加权计数记为 $\sigma_D(A)$ 。

$$\sigma_D(A) = \sum_{k=1}^n (|D_k(A)| \times W_k) \quad (2.6)$$

项集 A 是频繁的, 当且仅当其加权计数 $\sigma_D(A)$ 满足支持度阈值 minsupp, 即:

$$\sigma_D(A) \geq \left[\sum_{k=1}^n (|D_k| \times W_k) \right] \times \text{minsupp}\% \quad (2.7)$$

定义4(模糊关联规则 Fuzzy Association Rule) 是给定模糊日历 FC 下的形如 $X \Rightarrow^{FC} Y$ 的蕴含式, 其中 $X \subseteq I, Y \subseteq I$ 且 $X \cap Y = \emptyset$ 。模糊关联规则 $X \Rightarrow^{FC} Y$ 具有加权支持度 s%,

当且仅当 $\sigma_D(XUY) = [\sum_{k=1}^n (|P_k| \times W_k)] \times s\%$, $X \Rightarrow^{FC} Y$ 具有加权置信度 $c\%$, 当且仅当 $\sigma_D(XUY) / \sigma_D(X) = c\%$ 。

3 理论依据及基本思想

综合文[7,8]提出的累进和文[9]提出的增量处理方法, 本文提出了基于模糊日历约束的时序关联规则挖掘方法, 其基本思想如下: 首先, 把事务数据库划分为 n 个分区 (D_1, D_2, \dots, D_n), 每个分区只包括由模糊日历中的最小时间粒度所确定的基本时间段 (如最小时间粒度为天, 则基本时间段为一天) 的事务, 对每个分区 D_i 计算其权重 (隶属度) w_i , 对于分区 D_i , 定义其加权的计数阈值 m_i 为:

$$m_i = W_i \times |D_i| \times \text{minsupp}\% \quad (3.1)$$

其中, $|D_i|$ 是分区 D_i 内的事务数, minsupp 是用户指定的支持度阈值。

分区的累积加权的计数阈值 M_i 定义为:

$$\begin{cases} M_i = m_i \\ M_{ij} = M_{i(j-1)} + m_j \end{cases} \quad (1 \leq i \leq j \leq n) \quad (3.2)$$

首先生成候选频繁 2 项集 (C_2), 然后通过 C_2 生成所有的候选频繁项集 (C), 最后通过一次扫描数据库 D 生成频繁项集 (L), 关联规则即可产生。

候选频繁 2 项集 (C_2) 的生成采用累进的方法^[8], 一个 2 项集 I 对于分区 D_j 是部分频繁的, 当且仅当存在一个分区 D_i ($1 \leq i \leq j \leq n$) 使得 I 在分区 D_i, D_{i+1}, \dots, D_j 中的累进加权计数大于等于加权计数阈值 M_{ij} 。 $WC_{ij}(I)$ 的定义如下:

$$\begin{cases} WC_{ii}(I) = \sigma_{D_i}(I) \\ WC_{ij}(I) = WC_{i(j-1)}(I) + \sigma_{D_j}(I) \end{cases} \quad (3.3)$$

其中 $\sigma_{D_i}(I)$ 是项集 I 在分区 D_i 中的加权计数, 定义如下:

$$\sigma_{D_i}(I) = W_i \times |D_i(I)|$$

从第一个分区 D_1 开始依次处理每个分区, 最初候选频繁 2 项集 (C_2) 为空。对于分区 D_j 中的每个 2 项集, 存在两种情况: (1) 如果 $I \notin C_2$ 且 $\sigma_{D_j}(I) \geq m_j$, 则将 I 加入到 C_2 中, 并且将 D_j 和 $\sigma_{D_j}(I)$ 分别记录为 I 的起始分区和累进加权计数; (2) 如果 $I \in C_2$, 假设已记录的 I 的起始分区和累进加权计数分别为 D_i 和 V , 则若 $V + \sigma_{D_j}(I) \geq M_{ij}$ 则仍将 I 保留在 C_2 中, 只将其累进加权计数改为 $V + \sigma_{D_j}(I)$, 若 $V + \sigma_{D_j}(I) < M_{ij}$ 则将 I 从 C_2 中删除, 同时删除关于项集 I 的所有记录信息。很明显, 当处理完最后一个分区 D_n 后, 保留在 C_2 中的 2 项集一定是关于分区 D_n 部分频繁的, 证明见文[8], 同时, 由于在处理过程中不断删除不满足条件的项集, C_2 也很接近于 L_2 , 这使得 C_2 成为一个好的候选集。

然后利用 C_2 来连接生成其它的候选 k 项集 ($k \geq 3$): $C_{k+1} = C_k \circ C_k$, \circ 代表文[27]中的连接操作。对于任意的 $R \in C_{k+1}$, R 的所有 k 项子集一定包含在 C_k 中 (Apriori 性质), 因此, 候选项集的集合 C 是所有候选 k 项集 ($k \geq 2$) 的并集, 即 $C = \bigcup_{k \geq 2} C_k$ 。

最后通过扫描数据库 D 一次, 由 C 生成频繁项集 L 。对于 C 中的每个项集 I , 已经记录了 $\sigma_D(I)$ 。如果 $\sigma_D(I) \geq M_{1n}$ (数据库 D 的加权计数阈值), 则 I 即是 D 的频繁项集, 否则 I 即是非频繁的。数据库 D 的频繁项集 $L = \bigcup_{k \leq 2} L_k$, 通过频繁项集 L 可方便地得到模糊时序关联规则, 挖掘任务完成。

4 算法描述

算法 1 BFCTAR (基于模糊日历约束的时序关联规则挖掘算法)

输入: 事务数据库 D ; 最小支持度阈值 minsupp ; 最小置信度阈值 minconf ; 模糊日历 c

输出: 模糊日历下的频繁项集 L

算法过程:

```

1) 将  $D$  划分为  $n$  个不相交的分区  $D_1, D_2, \dots, D_n$ , 每个分区包含最小时间粒度所确定的一个时间段内的记录;
2) 计算每个分区的权重 (隶属度)  $w_i$ ;
3) 计算每个分区的加权计数阈值  $m_i (m_i = W_i \times |D_i| \times \text{minsupp})$ ;
4)  $M_{11} = m_1$ ;
5) FOR  $i = 1$  to  $n$ 
6)   For  $j = 2$  to  $n$ 
7)      $M_{ij} = M_{i(j-1)} + m_j$ ;
8)   ENDFOR
9) ENDFOR
10)  $C_2 = \emptyset$ ;
11) FOR each 2-项集  $I \in D_1$ 
12)   IF  $\sigma_{D_1}(I) \geq m_1$ 
13)      $C_2 = C_2 \cup \{I\}$ ;
14)      $D_s(I) = D_1$ ; //  $D_s(I)$  为项集  $I$  起始分区
15)      $V(I) = \sigma_{D_1}(I)$  //  $V(I)$  为  $I$  的累积加权计数
16)   ENDIF
17) ENDFOR
18) FOR  $j = 2$  to  $n$ 
19)   For each 2-项集  $I, D_j$ 
20)     { IF  $I \notin C_2$  and  $\sigma_{D_j}(I) \geq m_j$ 
21)        $C_2 = C_2 \cup \{I\}$ ;
22)        $D_s(I) = D_j$ ;
23)        $V(I) = \sigma_{D_j}(I)$ 
24)     } ENDFOR
25)   IF  $I \in C_2$ 
26)     IF  $V(I) + \sigma_{D_j}(I) \geq M_{D_s(I)j}$ 
27)        $V(I) = V(I) + \sigma_{D_j}(I)$ 
28)     ELSE
29)        $C_2 = C_2 - \{I\}$ 
30)     ENDFOR
31)   ENDFOR
32) ENDFOR
33) ENDFOR
34)  $C = \emptyset$ 
35) WHILE ( $C_k \neq \emptyset$  and  $k \geq 2$ )
36)    $C_{k+1} = C_k \circ C_k$ 
37)    $C = C_k \cup C_{k+1}$ 
38) ENDFOR
39)  $L = \emptyset$ 
40) FOR each  $I \in C$ 
41)    $\sigma_D(I) = \sum_{D_i} \sigma_{D_i}(I)$ 
42)   IF  $\sigma_D(I) \geq M_{1n}$ 
43)      $L = L \cup \{I\}$ 
44)   ENDFOR
45) ENDFOR
46) RETURN  $L$ ;

```

5 实验设计与分析

5.1 实验设计

在支持度阈值不同的情况下, 对 BCTAR 算法和文[10]提出的 Apriori+ 算法从执行时间、I/O 代价、生成的平均候选项集数这几个指标进行比较。实验数据采用文[11]中介绍的方法生成的人造数据: T10. I4. D100K, 其中 T 代表事务平均大小, I 代表可能的最大频繁项集的项数, D 代表事务数据库的事务数, 以 K 为单位。实验采用的模糊日历如例 1 (月中且年末, 或者, 周末且年初)。事务数据库划分为 100 个分区, 每个分区中的 1000 条记录对应于该模糊日历的最小的时间粒度 (天)。实验环境中的硬件平台为联想开天 M6200 (P4 3.0G, 1G RAM), 操作系统 Windows 2003 Sever, 数据库系统 SQL Server2000, 编程环境为 Jbuilder2005。

5.2 实验结果与分析

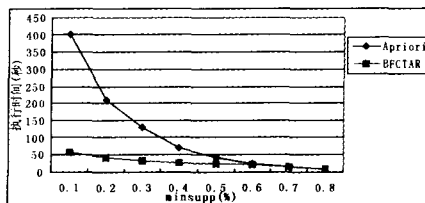


图 4 BFCTAR 算法和 Apriori+ 算法执行时间比较

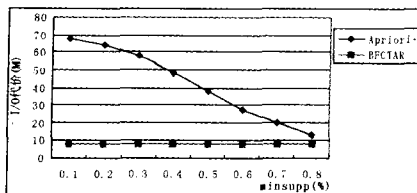


图5 BFCTAR 算法和 Apriori+ 算法 I/O 代价比较

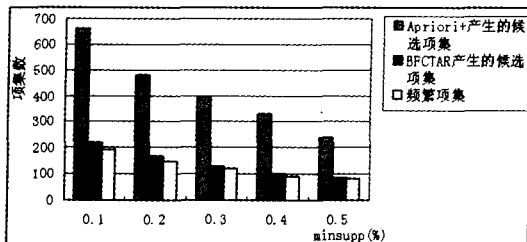


图6 BFCTAR 算法和 Apriori+ 算法生成的候选项集数比较

BFCTAR 算法与 Apriori+ 算法的比较结果如图 4, 5, 6 所示。可看出, 基于模糊日历约束的 BFCTAR 算法的执行效率优于 Apriori+ 算法, 特别是在支持度阈值较小时, 其 I/O 代价也明显小于 Apriori+ 算法; 其产生的候选项集数接近于频繁项集数, 明显小于 Apriori+ 算法所产生的候选项集数。其原因主要在于 BFCTAR 算法在生成候选项集的过程中采用累进的方法, 每扫描一个分区即记录候选项集的相应的信息, 减少了扫描数据库的次数, 并且在处理过程中不断删除不满足条件的项集, 所以生成的候选项集数也很接近于频繁项集。

小结 本文在模糊集理论和模糊日历代数的基础上, 提出了一种基于模糊日历约束的时序关联规则挖掘算法, 理论分析和实验结果表明该算法是有效的。虽然是基于日历约束

的模糊规则挖掘算法, 但对于其它属性的模糊约束也具有普遍意义。

参考文献

- Han J, Dong G, Yin Y. Efficient Mining of Partial Periodic Patterns in Time Series Databases. In: Proceedings of the International Conference on Data Engineering, 1999. 106~115
- Ozden B, Ramaswamy S, Silberschatz A. Cyclic Association Rules. In: Proceedings of the 15th International Conference on Data Engineering, 1998. 412~421
- Li Y, Ning P, Wang X S, Jajodia S. Discovering Calendar-based Temporal Association Rules. Data and Knowledge Engineering, 2003, 44(2): 193~218
- Ramaswamy S, Mahajan S, Silberschatz A. On the Discovery of Interesting Patterns in Association Rules. In: Proceedings of the International Very Large Database Conference, 1998. 368~379
- Lee W J, Jiang J Y, Lee S J. An Efficient Algorithm to Discover Calendar-based Temporal Association Rules. In: Proceedings of 2004 IEEE International Conference on Systems, Man, and Cybernetics, 2004. 3122~3127
- Lee W J, Lee S J. Fuzzy Calendar Algebra and Its Applications to Data Mining. In: Proceedings of 11th International Symposium on Temporal Representation and Reasoning, 2004. 71~78
- Lee C H, Lin C R, Chen M S. Sliding-window filtering: An efficient algorithm for incremental mining. In: Proc. ACM 10th Int. Conf. Information Knowledge Management, Atlanta, GA, Nov. 2001. 263~270
- Lee C H, Ou J C, Chen M S. Progressive weighted miner: An efficient method for time-constraint mining. In: Proc. 7th Pacific-Asia Conf. Knowledge Discovery Data Mining, Seoul, Korea, Apr. - May 2003. 449~460
- Cheung D W, Lee S D, Kao B. A general incremental technique for maintaining discovered association rules. In: Proc. 5th Int. Conf. Database Systems Advanced Applications, Melbourne, Australia, Apr. 1997. 185~194
- Ale J M, Rossi G H. An Approach to Discovering Temporal Association Rules. In: Proc. of the 2000 ACM Symposium on Applied Computing, 2000. 294~300
- Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proc. 1994 Int. Conf. Very Large Databases (VLDB'94)
- Grzymala-Busse J W. Data with missing attribute values; Generalization of indiscernibility relation and rule induction. Transactions on Rough Sets, Lecture Notes in Computer Science Journal Subline, Springer-Verlag, 2004, 1: 78~95
- Kryszkiewicz M. Rough set approach to incomplete information systems. In: Proceedings of the Second Annual Joint Conference on Information Sciences, September 28-October 1, Wrightsville Beach, NC, 1995. 194~197
- Stefanowski J. Algorithms of Decision Rule Induction in Data Mining. Poznan, Poland; Poznan University of Technology Press, 2001
- Stefanowski J, Tsoukias A. On the extension of rough sets under incomplete information. In: Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, RSFDGrC'1999, Yamaguchi, Japan, 1999. 73~81
- Stefanowski J, Tsoukias A. Incomplete information tables and rough classification. Computational Intelligence, 2001, 17: 545~566
- Greco S, Matarazzo B, Slowinski R. Dealing with missing data in rough set analysis of multi-attribute and multicriteria decision problems. In: Zanakias S. H, Doukidis G, Zopounidis Z, eds. Decision Making: Recent developments and Worldwide Applications. Dordrecht, Boston, London; Kluwer Academic Publishers, 2000. 295~316
- Slowinski R, Vanderpooten D. A generalized definition of rough approximations based on similarity. IEEE Transactions on Knowledge and Data Engineering, 2000, 12: 331~336
- Yao Y Y. Two views of the theory of rough sets in finite universes. International J of Approximate Reasoning, 1996, 15: 291~317
- Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation operators. Information Sciences, 1998, 111: 239~259
- Yao Y Y. On the generalizing rough set theory. In: Proc. of the 9th Int Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'2003), Chongqing, China, Oct. 2003. 44~51

(上接第 169 页)

- Pensa R G, Boulicaut J-F. Towards Fault-Tolerant Formal Concept Analysis. In: the 9th Congress of the Italian Association for Artificial Intelligence, Milan, Italy, September Springer-Verlag LNAI 3673, 2005. 212~223
- Grzymala-Busse J W. Three Approaches to Missing Attribute Values-A Rough Set Perspective. In: Workshop on Foundations of Data Mining, associated with the fourth IEEE International Conference on Data Mining, Brighton, UK, November, 2004, 1: 4
- Saquer J, Deogun J. Concept approximations based on rough sets and similarity measures Int [J]. Appl Math and Comp Sci, 2001, 11(3): 655~674
- Grzymala-Busse J W, Siddhaye S. Rough Set Approaches to Rule Induction from Incomplete Data. In: Proceedings of the IPMU'2004, the 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia, Italy, July 2004. 923~930
- Grzymala-Busse J W. Rough Set Strategies to Data with Missing Attribute Values. In: Proceedings of the Workshop on Foundations and New Directions in Data Mining, associated with the third IEEE International Conference on Data Mining, Melbourne, FL, USA, November 2003. 56~63
- Kryszkiewicz M. Rules in incomplete information systems. Information Sciences: an International Journal, 1999, 113(3-4): 271~292
- Burmeister P, Holzer R. Treating Incomplete Knowledge in Formal Concept Analysis. In: Ganter B, Stumme G, Wille R, eds. Formal Concept Analysis: State of the Art. LNAI 3626. Springer, Heidelberg, 2005
- Wang Guo-yin. Extension of rough set under incomplete information systems. Journal of Computer Research and Development, 2002, 39(10): 1238~1243
- Priss U. Formal concept analysis in information science. Annual Review of Information Science and Technology, 2006, 40: 521~543