第29卷 第18期 计 算 机 工 程 2003年10月
Vol.29 No.18 Computer Engineering October 2003
· 软件技术与数据库 · 文章编号：1000—3428(2003)18—0092—03 文献标识码：A 中图分类号：TP274

移动计算中基于Rough Set的时间序列研究与应用

刘 刚，李德敏，赵丽娜
(东华大学信息科学与技术学院，上海 200051)

摘 要：提出了基于Rough Set在移动计算环境中进行数据挖掘的系统模型，给出了进行数据挖掘的具体算法，介绍了由实时时态信息系统转换为时态信息系统的平均时间间隔方法，以及由时态信息系统转换为传统信息系统的增量分析方法，最后给出了实例和应用结果。

关键词：Rough Set；移动计算；时间序列；数据挖掘

Research and Application of Time Series Based on Rough Set in Mobile Computing

LIU Gang, LI Demin, ZHAO Lina
(College of Information Science and Technology, Donghua University,Shanghai 200051)

【Abstract】 The paper presents a system model based on rough set for data mining in mobile computing and provides the algorithm for data mining.It provides a method of the average time interval for transforming the real-time temporal information system to the temporal information system. And a method of increment analysis for the temporal information system to the traditional information system is also proposed. At last it supplies an example and the application results.

【Key words】 Rough set; Mobile computing; Time series; Data mining

随着网络技术和移动通信技术的迅速发展，人们越来越需要能在任何时候、任何地点访问任何数据，实现无约束自由通信和共享资源的目标。于是，出现了一种更加灵活、复杂的分布式计算环境，人们称之为移动计算(Mobile Computing)[1]。

1 移动计算系统的典型模型

在移动计算环境中系统的典型模型可用图1描述。其中主要包括3类结点：

(1)服务器(Server,即SVR)：用于存储大量信息，每个服务器维护一个本地数据库(Local Database，LDB)，服务器之间由可靠的高速互联网络连接在一起。

(2)移动支持结点(Mobile Support Station，MSS)：MSS也位于高速网络中，并且有无线联网能力。每一个MSS用于支持一个无线网络单元(Cell)，该单元内的移动客户机既可以通过无线链路与一个MSS通信，从而与整个固定网络(Fixed Network)通信，也可以接收由MSS发送的广播信息。

(3)移动客户机(Mobile Client，MC)：MC用于客户的日常操作处理和移动通信，具有移动性，可出现在任意一个无线网络单元中。

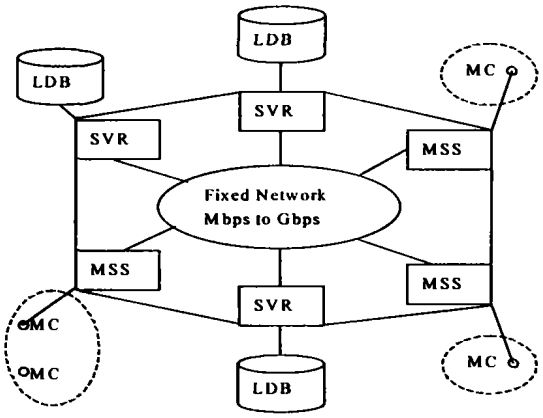

图1 移动计算系统的典型模型

2 基于Rough Set挖掘的系统模型

在移动计算环境中,运用Rough Set方法进行数据挖掘的系统模型可用图2表示。

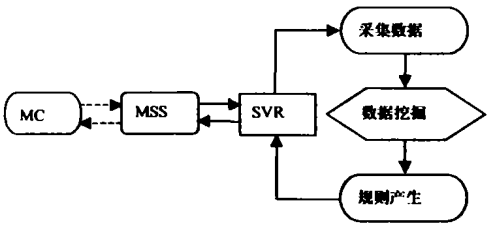

图2 基于Rough Set数据挖掘的系统模型

(1)MC(移动客户机)与MSS(移动支持结点)建立通信连接，将客户对决策的具体要求通过移动通信网络发送给MSS，再由MSS经高速网络传送给SVR(服务器)。

(2)SVR根据客户的具体要求进行数据采集，选取相应的功能子库和数据区间，确定操作数据的范围。

(3)对选定的数据进行挖掘。首先将RTTIS转换为TIS，再由TIS转换为IS，最后运用Rough Set方法对IS进行数据挖掘，提取出内在的规则。

(4)对产生的规则进行处理，表达成用户容易理解的形式。

(5)SVR对于产生的规则数据，通过MSS返回给MC。

3 信息系统

定义1 在Rough Set理论中信息系统IS是一个二元组：

I=<U，A>

其中U为对象的非空有限集合，A为属性的非空有限集合。A=C∪D，且C∩D=Φ，C是条件属性集合，D是决策属性。

3.1 时态信息系统

Rough Set方法挖掘的数据通常是时间轴上某一点的快

基金项目：上海市局管基金资助项目(00JG05047)
作者简介：刘 刚(1977－)，男，硕士生，研究方向为移动计算中的数据挖掘及应用；李德敏，博士、副教授；赵丽娜，助教
收稿日期：2002-09-30

照，由这些数据构成的信息系统可以被认为是三维空间，它包括对象、属性和属性值。其中对象和属性构成二维，而对象在属性上的值构成第三维。在分析时间序列时，人们往往对属性值随时间的变化感兴趣。因此，为了能应用Rough Set方法来分析时间序列，必须将这种变化反映在信息系统中，使传统的信息系统能表示与时间相关的数据，即将时间包括于传统的信息系统的3个维的某一个维中。属性值维中的数据值用时间函数来代替。时态信息系统用来表示在相同时间间隔上的变化的数据。

定义2 时态信息系统TIS:

I_t=<U，A∪t，λ >

其中U为对象集合，A为属性集合，t为序列属性，$t \notin A$，λ为序列属性t上的排序关系，λ ={(x,y): x,y ∈ N and x<y}。

3.2 实时时态信息系统

在移动计算环境中存在着大量与时间相关的数据，这些数据大多为实时数据，由于数据间的时间间隔长度是不均等的，因此，引入实时时态信息系统。

定义3 实时时态信息系统RTTIS:

I_n=<U，A∪t∪ δ ，λ >

其中U为对象集合；A为属性集合;t为序列属性，$t \notin A$;λ为序列属性t上的排序关系,λ ={(x,y): x,y ∈ N and x<y};δ 为时间属性,$\delta \notin A$, $\delta(x_2)$(简记为δ_2)表示自对象$x_1 \in U$发生以来到对象$x_2 \in U$发生的时间,这里$t(x_1)<t(x_2)$,并且不存在$y \in U$，使其满足$(y>x_1) \wedge (y<x_2)$。

4 TIS转换为IS的增量分析方法

将TIS转化为IS，是将Rough Set方法应用于时间序列挖掘的基础。在文献[4]中提出了一种将TIS转换为IS的方法。它依赖于向后跟踪的时间长度，在预先确定了向后跟踪时间长度δ 后，构成转换后新的信息系统的属性集A'，通常|A'|=|A|*δ 。这种挖掘方法依赖于向后跟踪的时间长度，而且当δ 较大时属性集将成倍增大。在文献[5]中提出的转换方法，它对每一个属性$a \in A$，增加两个新的属性a_1和a_2，a_1表示状态的起始点，设置$a_1(t_{i+1})=a(t_i)$，a_2表示在该区间内的增量$a_2(t_{i+1})=|a(t_{i+1})-a(t_i)|$。显然，转化后信息系统的属性个数为|A|*2。以上两种方法的共同缺点是转换后属性集都将增大。

我们提出一种新的转换方法：增量分析法。该方法的主要思想是将传统信息系统中在时间轴上某一点的快照变为在一段时间上的增量变化。设某一属性$a \in A$在状态t_i的值为$a(t_i)$,在t_{i+1}上的值为$a(t_{i+1})$,则在状态t_{i+1}的属性值设置为: $a(t_{i+1}) = a(t_{i+1}) - a(t_i)$。

该方法的优点是不需要预先确定跟踪的时间长度，分析时间序列也不受跟踪时间长度的限制；同时其转化后属性集的个数保持不变，因此降低了数据挖掘的复杂性，提高了挖掘效率。而且在初始值给定的情况下，可以由初始值与增量进行累加，求出任意时间段的数据值，从而使信息系统的信息量保持不变。

5 RTTIS转换为TIS的平均间隔方法

由于RTTIS描述的是时间区间间隔不等的数据集合，而TIS描述的是在时间区间间隔相等的数据集合，因此，在进行RTTIS向TIS转换的时候，主要考虑将不等的时间区间间隔转化为相等的时间区间间隔。

在文献[4]中提出了一种将RTTIS转化为TIS的思想。该思想是找出各时间区间间隔的最大公约数，然后将RTTIS的区间间隔限定为所求得的最大公约数，以便得到相等的时间区间间隔。通常情况下，转化后要增加大量的新对象。文献[5]在文献[4]的基础上进行了改进，提出了最小时间间隔方法，它把RTTIS中各时间间隔的最小值作为统一的时间间隔，使转化后的TIS中只增加较少的新对象。

5.1 平均时间间隔方法

上述两种方法的最大缺点是转换后的TIS中都将增加新对象。当一个信息系统的对象或属性数增加时，在其上进行挖掘时所进行的工作量也就越大。因此，在进行RTTIS转化时，应尽量避免对象的增加，以免增加计算的复杂性。为此，提出平均时间间隔方法。该方法采用RTTIS中各时间间隔总和的平均值作为统一的时间间隔，从而使转化后的对象数保持不变，同时由于所取的时间间隔比较适中，因此基本不影响转换的精确度。在下述的形式化表示中，对各对象的属性值采用线性插值法[5]进行计算。

Input: I_n=<U，A∪t∪ δ ，λ >
Output: I_t=<U，A∪t，λ >
Begin

$$\delta_{avg}=\frac{\delta(x_2)+\delta(x_3)+\cdots+\delta(x_n)}{n-1}$$; //取平均值

U={x_1};
Z= x_1;
for i=1 to n-1
Begin
Z=Z+ δ_{avg};
find k ∈ {1,2,⋯,n},$x_k \le Z \le x_{k+1}$;
for j=1 to |A|

$$a_j(Z)= a_j(x_k)+\frac{a_j(x_{k+1})-a_j(x_k)}{x_{k+1}-x_k}\times(Z-x_k);$$

//计算对象Z的属性值
U'= U'∪ {Z};
End
End

5.2 计算复杂性比较

在RTTIS中，设δ_2，δ_3，⋯，δ_n为其对象在δ 属性所对应的属性值，δ_{min}为其中的最小值，δ_{avg}为总和的平均值。在实际应用中，由于RTTIS中对象往往较多，因此，用最大公约数法获得的时间间隔通常为1，从而使转化后TIS中的对象成倍增加，转化后TIS中的对象个数为

$O_{max}=\delta_2+\delta_3+\ldots+\delta_n+1$

采用最小时间间隔方法转化后TIS中的对象个数为:

$O_{min}=(\delta_2+\delta_3+\ldots+\delta_n)/\delta_{min}+1$

我们提出的平均时间间隔方法，其转化后TIS中的对象个数为

$O_{avg}=(\delta_2+\delta_3+\ldots+\delta_n)/\delta_{avg}+1=n$

在通常情况下，由于$\delta_{avg} \ge \delta_{min} \ge 1$，因此有

$O_{max} \ge O_{min} \ge O_{avg}$

可见，在转换后TIS中的对象数方面，平均时间间隔方法要优于最大公约数方法和最小时间间隔方法，它使转化后的对象个数保持不变，从而避免了计算复杂性的增加，提高了数据挖掘效率。

6 移动计算中数据挖掘的具体算法

基于以上描述给出如下算法步骤:

step1 MC将客户对决策的具体要求经MSS发送给SVR。

step2 SVR根据客户的具体要求进行数据采集，构造信息系统。

step3 运用平均时间间隔方法将RTTIS转化为TIS。

step4 运用增量分析方法将TIS转化为IS。

step5 应用Rough Set方法对IS进行规则提取。首先根据等价关系[3]对IS进行分类，然后构造区分矩阵[3]，最后由区分函数[3]推导出决策规则。

step6 SVR对于产生的决策规则进行处理，并通过MSS返回给MC。

7 实例和应用结果

实例为一个实际应用对象随时间变化的数据。表1是由该对象的数据经过加工整理而成的实时时态信息系统。

表1 加工后的实时时态信息系统

t	a	b	δ	d
1	4	5	4	3
2	5.25	4.25	3	1.75
3	4	3	5	3
4	5.5	1.5	2	4.5
5	4	3	6	3
6	5	2	4	2

由表1可知δ_{avg}为4，t_1的δ 属性值恰好为4，故各属性值不需要改变，而t_2的δ 属性值为$3<\delta_{avg}$，因此要采用平均时间间隔方法对它的各个属性值进行线性插值，同时向t_3借一个时间间隔，由于t_3被t_2借去一个时间间隔后剩下的δ 属性值恰好为4，因此各属性值也不需要改变，依此类推。采用平均时间间隔方法转化后的时态信息系统如表2所示。

表2 采用平均时间间隔方法转化后的时态信息系统

t	a	b	d
1	4	5	3
2	5	4	2
3	4	3	3
4	5	2	4
5	4	3	3
6	5	2	2

表3 由表2采用增量分析方法转化后的传统信息系统

ID	a	b	d
1	1	−1	−1
2	−1	−1	1
3	1	−1	1
4	−1	1	−1
5	1	−1	−1

由表2可见，采用平均时间间隔方法转化后的对象个数保持不变，而且基本不影响转换的精确度。而采用最大公约数方法，由于时间属性δ 的最大公约数为1，转化后时态信息系统中将含有24个对象。采用最小时间间隔方法，由于δ_{min}为2，因此转化后的时态信息系统中含有12个对象。由此可见，采用平均时间间隔方法，可以有效地避免计算复杂性的增加，提高了数据挖掘的效率。

表3是由表2采用增量分析方法转化后的传统信息系统。由于第一个对象前没有其它对象无法进行增量分析，因此第一个对象被省略。运用Rough Set方法对表3进行规则提取，具体决策结果如下所示:

确定的规则:

E_1:(a, -1)∧(b, -1)→(d, 1)

E_2:(b, 1)→(d, -1)

可能的规则:

E_3:(a, 1)→(d, -1)

从规则E_1可以看出当属性a、b减小时属性d增加，从规则E_2可以看出当属性b增加时属性d减小。从规则E_3可以看出当属性a增加时属性d可能减小。因此,从中可以发现属性d的变化趋势与属性a、b的变化趋势成反比,且比率为1:1。

8 结束语

在移动计算环境中，以Rough Set方法为基础对时间序列进行数据挖掘，可以有效地预测事物发展的趋势。本文提出了在移动计算环境中数据挖掘的系统模型，给出了进行数据挖掘的具体算法，介绍了由实时时态信息系统转换为时态信息系统的平均时间间隔方法，以及由时态信息系统转换为传统信息系统的增量分析方法。这两种方法可以有效地避免计算复杂性的增加，从而大大提高了数据挖掘的效率，对于股票、医疗、企业和军事等领域有广泛的应用价值。

参考文献

1 胡虔怀,郑若忠.移动数据库及其关键技术.Applications of the Computer Systems,2001

2 Pawlak Z.Rough Sets.International Journal of Computer and Information Sciences,1982,11(15)

3 Mollestad T,Komorowski J.A Rough Set Framework for Mining Propositional Default Rules.Proceedings of 9th International Symposium of Foundations of Intelligent Systems (ISMIS'96), 1996:448-457

4 Anders T B.Mining Time Series Using Rough Sets——A Case Study. Proceeding of First European Sysposium,PKDD97,Norway:Trondheim, 1997:351-358

5 尹旭日,商 琳,何佳洲等.Rough 集挖掘时间序列的研究.南京大学学报,2001,37(2)

☆☆☆

(上接第26页)

(1) 任务组成或关系变化或未完成任务的输入/输出数据变化时，可利用本算法及时重构任务时序模型。

(2) 当任务的资源需求或资源供应情况变化时，可以用本算法及时调整任务时序模型，以保证任务的顺利执行。

(3) 任务时序模型的调整可以通过设置触发器来自动完成，触发器由任务输入输出数据、资源变化触发。

4 结论

PDM已成功地实现了产品数据的统一管理，其下一步推广目标将是协同产品开发过程管理，而过程管理模型是实现过程管理的基础。本文分析了数据和资源等对并行产品开发的影响，研究了基于数据和资源约束的并行产品开发过程建模，并具体实现了建模算法。该算法能有效地提高产品开发并行程度和简化产品开发管理。

参考文献

1 Yan Jin, Li Zhao, Raghunath A. ActiveProcess: A Process-driven and Agent -based Approach to Supporting Collaborative Engineering. ASME Design Engineering Technical Conferences, DETC99/CIE-9099, 1999

2 刘清华. 分布环境下协同产品开发过程管理的研究与实践[学位论文]. 武汉：华中科技大学，2001

3 耿素云,屈婉玲,张立昂. 离散数学. 北京：清华大学出版社, 1992