

时态数据挖掘研究进展^{*})

Progress of Temporal Data Mining Research

张保稳 何华灿

(西北工业大学计算机系 西安710072)

Abstract Temporal data mining is one of the important branches of data mining. In this paper with the present documents first we systematically classify the present research on temporal data mining. Next, we give our generalizations and analyses to the main branches. Finally problems of the current research of temporal data mining are pointed out and solutions are proposed.

Keywords Data mining, Temporal data mining, Time series, Events sequences, Temporal patterns, Similarity search, Sequential patterns

在现实生活中,大量数据集之中的数据都带有时间特征。时态数据随处可见,遍及经济、气象、通信、医疗等等多个领域。股市每日(或月)指数、交换机的每小时的业务量、某一患者的脑电波和 Web 页的日访问量,这些都是比较常见的例子。对这些时态数据进行分析,从中获取蕴含的系统演化规律,从而完成对系统的未来行为的预测,具有重要的价值和意义。

数据挖掘(也称为数据库中的知识发现),是指从数据中提取模式的过程,这些模式是有效的、新颖的、潜在可用的和易于理解的^[1]。时态数据挖掘是数据挖掘研究的一个重要的组成部分,在时态数据挖掘的过程中必须考虑数据集之中数据间存在着的时间关系。

相对于数据挖掘较成熟的部分而言(比如关联规则的挖掘),时态数据挖掘的研究是数据挖掘的较新的一个方向。目前,国际上对于时态数据挖掘的研究逐渐成为一个新的热点,但国内在这方面的研究文献尚不多见,比较重要的工作有1998年欧阳为民等曾经从理论框架的角度对时态数据挖掘做过的介绍和分析^[2]。本文拟结合国内外研究文献中时态数据挖掘方面的进展情况,对其研究内容进行归纳总结和分析,对其主要分支加以讨论,指出其中存在的若干问题并给出解决问题的一些看法。

1 时态数据的类型

在对时态数据挖掘问题进行探讨之前,我们首先将待研究问题中的研究对象——时态数据加以分类。一般说来,这些不同类型的时态数据将形成不同类型的挖掘问题。后面我们在具体介绍时态数据挖掘分支时将分别按不同的时态数据类型进行讨论。在时态数据挖掘研究中,时态数据主要可以分为以下几类¹:

(1)数值型序列。即传统意义上狭义的时间序列,构成序列的元素是数值型的。比如股票价格的历史数据,网站的点击次数等。如果不加特殊说明,本文以后的时间序列专指这种狭义的时间序列^[3]。对时间序列而言,单个或多个时间序列都可

以作为挖掘对象。

(2)事务型序列。构成序列的元素是事务型(Transaction-al)的,我们称这种序列为事务序列。比如,某顾客在超市中某段时间内购买的商品的记录序列。对事务序列而言,一般都是这一组这类序列的集合构成挖掘的对象。

(3)事件性序列。构成序列的元素是事件,这种序列称为事件序列。比如无线通信网中的故障序列,用户的界面交互行为序列。大多数事件序列的挖掘研究是对单个事件序列进行的。

2 时态数据挖掘的分类

目前,国内外研究资料中时态数据挖掘的研究内容还比较零散,并没有系统的理论框架。尝试对时态数据挖掘进行系统性分类的文献笔者尚未见到。在对当前时态数据挖掘的主要工作进行整理之后,我们得到以下分类情况:

(1)按研究方向可分为时态模式挖掘和相似性问题研究。其中时态模式挖掘包括了各种序列的模式挖掘,时态因果和关联规则挖掘等等内容;相似性问题研究则主要是面向查询的需要,包括各种相似性搜索算法的设计和相似性查询语言的设计等。

(2)按研究内容的独立性可分为两种情况。一是将时态数据作为一种特殊的挖掘对象,定义相应的数据挖掘的框架和算法,进行专门的研究,这是当前研究的主要方向,后面的分类和分析就主要是针对这一类情况进行的;二是仅仅在数据预处理阶段,从时态数据中组建目标特征属性或者条件特征属性^[20],然后将其当作普通的非时态属性,在进行数据挖掘时仍然用原来的数据挖掘框架和算法。

(3)按研究对象可以分为时间序列的数据挖掘、事件序列的数据挖掘和事务序列的数据挖掘。

(4)按应用情形可分为单纯时态数据挖掘和混合时态数据挖掘。前者算法处理的对象主要是时态数据,后者则根据应用的情形将时态数据和非时态数据混合在一起进行数据挖掘。

^{*}本文得到国家教委博士点基金(98069923)资助。张保稳 博士生,主要从事人工智能、时间序列、数据挖掘的研究。何华灿 博士生导师,主要从事人工智能、泛逻辑理论的研究。

¹在欧阳为民的分类中涵盖了时间数据库中的时态数据,这里我们的分类主要是针对普通的数据库和数据集中的时态数据进行的,对时间数据库中的时态数据没有考虑,我们所讨论的时态数据挖掘也只限于普通的数据库和数据集中的时态数据。

(5)按算法并行处理序列个数可以分为单一序列的时态数据挖掘和多序列的时态数据挖掘。

在上面的几种分类中,第一种分类可以很好地将当前的研究工作划分开来。下面我们就按时态模式挖掘和相似性问题研究两部分对时态数据挖掘进行讨论。

3 时态模式挖掘

在时态模式挖掘方面,Chen Xiaodong 等人^[4]提出的一种时态模式挖掘的定义具有比较好的包容性,这里我们借用过来作为时态模式挖掘的定义。

定义1(时态模式) 一个时态模式是一个二元组, $(\text{Patt}, \text{TimeExp})$ 。其中 Patt 是一个一般意义上的模式,可以是趋势、偏差、分类规则、关联规则等等, TimeExp 就是一个时间表达式。它表示模式 Patt 在 $\phi(\text{TimeExp})$ 中的每个时间区间内成立²。

定义2(时态模式挖掘) 给定一个在时间域 T 上带时间戳的数据集 D 和一个时间表达式 TimeExp , 时态模式挖掘的问题就是在数据集 D 中发现具有 $(\text{Patt}, \text{TimeExp})$ 形式的所有模式,这些模式要满足与用户指定的最小频度 $\text{min-}f\%$ 相关的所有阈值约束。

目前在时态模式挖掘方面的研究工作多而杂乱,我们将按不同的时态数据类型对其中比较突出和新颖的部分进行总结和分析³。

事务数据库中的序列模式(Sequential Patterns)挖掘是时态挖掘研究中相对比较成熟的部分。它处理的数据类型就是事务序列。在这方面研究比较突出的是 IBM 公司 Almaden 研究中心 Ramakrish Strikant 和 Rakesh Agrawl 提出的 Apriori 算法^[5]。该算法的研究背景和关联规则挖掘一样,是超市的销售数据分析。Apriori 算法处理的数据对象是大型零售系统的销售数据库。算法的目标是从中发现客户购买行为中的序列模式,从而实现对客户未来行为的预期。Apriori 算法的本质是按顾客 ID 将数据库中的数据重构,每个顾客的事务数据按购买时间组成一个事务序列。这样由整个数据库中的数据生成了一个含有多个长短不一的序列的集合。然后从多个序列中间发现相同的子序列,满足一定支持度的子序列便认为是要发现的序列模式。后来, Ramakrish Strikant 和 Rakesh Agrawl 在上述研究的基础上作了进一步的扩展,引入了序列模式中相邻元素之间的时间约束,放松了序列模式元素的所有项目必须来自同一事务的限制,在项目之间引入了分类的机制,提出了 GSP(Generalized Sequential Patterns) 算法^[6]。

在事件序列方面, Heikki Mannila 等人在对无线通讯网络故障管理数据库进行处理时提出了从事件序列进行模式发现的问题⁴。给定一个输入的事件序列、一类偏序事件集合模板,事件模式发现就是从序列中发现满足频度阈值的符合偏序模板的模式^[7]。在算法的实现上,该文有几点可以借鉴之处,对于我们处理序列挖掘问题很有帮助。其一是使用了从短频繁模式集中生成长的频繁模式的剪枝技术,其二是在算法设计时充分利用了相邻滑窗间的相似性。其三是将复杂模式

分解为简单模式再进行挖掘的方法。

从时间序列中抽取模式是时态模式挖掘的一个比较新颖的方向。1999年 Richard J. Povinelli 在他的博士论文中提出了一种时间序列数据挖掘(Time Series Data Mining)的框架^[9]。在他的研究中,他并没有试图对整个时间序列做预测和分析,而仅对时序中的事件(Event)的出现加以模式发现和预测。首先,他定义了一种事件标志函数(Event Characterization Function),然后在其基础之上定义数据挖掘的目标函数,再进行数据挖掘。事件标志函数标志着对应时间子序列的可划为事件的程度。从本质上讲, Richard J. Povinelli 是将时序分析预测目标的范围由原来的系统未来行为的全集缩减到满足事件阈值的子集、而相应的时序建模也由原来的数学表达式改为由规则表示的模式集表达。我们认为,这一尝试对如何用数据挖掘的技术辅助进行时间序列分析具有重要意义。

4 相似性研究

对象之间相似性的定义和度量研究在统计理论、机器学习以及数据挖掘等方面具有重要的意义。在时态数据挖掘中,相似性问题的研究更是和数据库查询紧密结合在一起。比如,需要在数据库中发现具有相似价格浮动曲线的股票,发现具有相同销售模式的产品等。当前在时态数据挖掘方面相似性的研究主要是面向查询需要的。

相似性查询可以分为两类。一类是整体匹配,查询序列和数据库中的记录序列具有相同的长度。一类是子序列匹配,查询序列比数据库中的记录序列要短,需要在记录序列中寻找和查询序列相似的子序列。

相似性问题的研究主要包括相似性搜索算法的设计和相似性查询语言设计两部分。对于相似性查询语言部分,我们这里暂不讨论。下面我们主要介绍相似性算法的设计,它是相似性问题的核心。

相似性搜索算法的设计包括以下三个部分:相似性的定义、相似性度量模型的建立和算法的实现。

一般相似性的定义根据应用需求而定,而相似性度量模型则是依据所定义的相似性进行数学抽象而成。在相似性定义方面,有的比较简单、粗糙,例如, Agrawal 等人就提出了一种相似性,它是根据直观意义上时序数据的上升、下降的趋势定义的^[10]。通过这种相似性可以比较粗糙地从数据库中发现具有相似形状的时序。有的相似性的定义则比较复杂,例如 Gautam Das 等提出了一种称为 F-相似的相似性模型^[11];设 F 是一个函数集,对于两个待比较时序而言,如果它们有满足一定长度要求的子序列,且存在 F 中的一个函数 f 使得其中的一个子序列可以近似地映射到另一个子序列,则称两时序具有 F-相似性。这种相似性对时间序列中的异常点(outliers)、基线以及比例因子不敏感。后来 Agrawal 等又提出了一种 ϵ -相似性度量模型,这种相似性可以容忍时序中由于噪声的存在而引起的局部不匹配,并对时间轴上的偏移以及幅度的缩放比例不敏感^[12]。

在算法实现上,一般都要对原始时序进行处理,使其表示形式适合相似性模型的要求和数据库查询的需要。Keogh, E

² $\phi(\text{TimeExp}) = \{p_1, p_2, \dots, p_n\}$, 是对时间表达式的解释。其中 p_i 为时间域上的一个区间。

³ 在时态模式挖掘方面,和时间数据库相关的时态关联规划和因果关系的挖掘也是重要的部分,以后我们将专文讨论。

⁴ 原文中为 episodes, 定义为具有某种偏序关系的 events 的集合。因为英文中 episodes 和 events 都有“事件”的意思,翻译过来区别不明显,笔者在下面用它的内涵作为代替。

等人在相似性的研究中使用 STB 的索引方法^[13]。他们使用分段线性表示的形式表示时序,并将子时序分桶(bins),再将桶用位矢量加以表示,然后建立索引,进行查询。当前更为常见的一类方法是基于特征概念的相似性研究,这类方法首先从原始时间序列中提取相关目标特征,然后通过比较特征向量的相似度来比较时序之间的相似度。并且为了适应数据库查询的需要,一般都对特征参数进行索引或者按某种方式加以组织。例如,Agrawl 和 Faloutsos 等人利用 DFT 对时序进行处理,提取频域特征参数,在这些参数之间建立相似性的度量,然后对特征向量的参数按 R-树的方式加以索引,对时序的比较转化为特征参数之间的比较^[14]。Faloutsos 等人将上面的方法扩展到了可以在数据库中寻找和查询序列(或者其子序列)匹配的子序列^[15]。Yka Huhtala 等人则是通过提取时序的小波特征进行相似性的比较^[16]。

上面讨论的相似性的研究工作主要是针对时间序列的时态数据类型。同时间序列的相似性研究相比,事件序列的相似性问题方面所做的工作则很少,Heikki Mannila 等人定义了一种面向事件序列的相似性及其测度模型^[17]。这里的相似性是用将事件序列进行转换时所需的操作代价来度量的。至于事务序列类型的时态数据,一般不专门考虑其相似性问题。

5 存在的问题和解决策略

数据挖掘是一个面向应用的研究领域,当前很多的研究都是用一种数据挖掘技术解决特定的应用问题。时态数据挖掘也不例外。总而言之,与关联规则挖掘等数据挖掘的成熟部分相比,当前的时态数据挖掘研究的内容比较杂乱、零散而不系统。我们认为,时态数据挖掘的研究还处于一个比较初级的阶段,目前的研究中还存在着下列问题:

(1)在时态模式挖掘方面,针对数值型时态数据的研究较少,绝大多数是针对事件型和事务型的。在相似性研究方面,情况恰巧相反,针对数值型时态数据的研究占绝大多数,针对事件型时态数据的则很少。这使得在对数值型时态数据的数据挖掘中,模式挖掘和相似性研究之间形成了一种割裂的现象。解决这个问题的一种方法是将数值型时态数据转化为事件型时态数据的形式,然后按已有的事件序列模式挖掘算法对数值型时态数据进行处理。目前已经有 Guralnikden 等人在这方面进行了初步研究^[18]。

(2)在时间序列分析领域内,当前的研究现状是:由于实际应用中时间序列具有不规则、混沌等非线性特征,使得预测系统未来的全部行为几乎不可能,对系统行为的精确预测效果也难以令人满意,传统的时间序列分析方法不再适用^[9]。这使得人们不得不转向对系统的关键行为和带有粒度的预测以及建模进行研究。在解决时间序列问题的思路,由原来的应用概率论、随机过程等纯数学的方法,逐渐转变为引入模式识别、机器学习等人工智能技术和数学手段相结合的方法。数据挖掘本身就是将人工智能技术和各种数学方法相结合从数据中发现知识的过程,因此将数据挖掘的思想引入到时间序列分析中去是大有作为的。而在当前的时态数据挖掘研究中,对时间序列的处理主要是从数据库角度出发,研究相似性搜索。如果可以从时间序列问题的角度出发,从时间序列中提取知识,并将之和传统的时间序列分析技术相结合,这必将有利于时间序列问题的解决。Richard J. Povinelli 等人已经在这方面作出了一些有益的尝试。

(3)数据挖掘中存在着不确定性问题,粗糙集和模糊集的

理论已经应用于数据挖掘问题中^[19]。如何结合不确定性理论进行时态数据挖掘的问题尚待我们进行研究和探讨。

(4)从更深层次上讲,当前的时态数据挖掘的研究比较零散,缺少统一的理论框架。在对时态数据挖掘各个分支深入研究的基础上,提出系统的理论框架结构,对时态数据挖掘的发展将有重要的指导意义。

结束语 本文结合了当前时态数据挖掘研究进展情况,首先将时态数据分为时间序列、事务序列和事件序列三类。然后我们对时态数据挖掘的研究内容按照不同的标准进行了系统的分类。在时态数据模式挖掘和相似性研究两大分支中,我们按不同时态数据类型分别对当前的研究内容进行了归纳、总结和分析。最后我们指出了当前研究中存在的若干问题并给出了关于问题解决的一些看法。

参考文献

- 1 Agrawal R, Mannila H, Srikant R, et al. Fast Discovery of Association Rules. In: Fayyad M, Piatetsky-Shapiro G, Smyth P, eds. *Advances in Knowledge Discovery and Data Mining*, Menlo Park, California: AAAI/MIT Press, 1996. 307~328
- 2 欧阳为民,蔡庆生. 数据库中的时态数据挖掘研究. *计算机科学*, 1998, 25(4): 60~63
- 3 杨叔子,吴雅. *时间序列分析的工程应用*. 华中理工大学出版社, 1992
- 4 Chen X, Petrounias I. A framework for temporal data mining. In: Quirchmayr G, et al. eds. *Proc. Ninth Intl. Conf. on Database and Expert Systems Applications, DEXA'98*, Vienna, Austria, *Lecture Notes in Computer Science*, 1460. Springer-Verlag, 1998. 796~805
- 5 Agrawl R, Strikant R. Mining Sequential Patterns. In: *Proc. of the 11th Int'l Conf. on Data Engineering*, Taipei, Taiwan, IEEE Computer Society Press, March 1995. 3~14
- 6 Agrawl R, Strikant R. Mining Sequential Patterns: Generalizations and Performance Improvements. In: *Proc. Intl. Conf. on Extending Database Technology. EDBT'96*. Avignon France
- 7 Mannila H, Toivonen H, Verkamo A I. Discovering frequent episodes in sequences. In: *Proc. First Intl. Conf. on Knowledge Discovery and Data Mining (KDD-95)*, Montreal, Quebec, Canada. AAAI Press, Menlo Park, California. 1995. 210~215
- 8 Weigend A S, Gershenfeld N A. *Time Series Prediction: Forecasting the Future and Understanding the Past*, eds. Reading, MA: Addison-Wesley, 1993
- 9 Povinelli R. Identifying Temporal Patterns for Characterization and Prediction of Financial Time Series Events. In *Proc. International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining, TSDM2000*, Lyon, France. *Lecture Notes in Artificial Intelligence*, 2007. Roddick, J. F. and Hornsby, K., Eds., Springer. 2000
- 10 Agrawal R, Psaila G, Wimmers E L, Zaot M. Querying shapes of histories. In: Dayal U, et al. eds. *Proc. Twenty-first Intl. Conf. on Very Large Databases (VLDB '95)*, Zurich, Switzerland. Morgan Kaufmann Publishers, Inc. San Francisco, USA. 1995. 502~514
- 11 Das G, Gunopulos D, Mannila H. Finding similar time series. In: J. Komorowski, et al., eds. *Proc. of the 1st European Symposium on Principle of Data Mining and Knowledge Discovery (PKDD'97)*, vol. 1263 of LNAI, Springer, 1997. 88~100
- 12 Agrawal R, et al. Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. In: U. Dayal, P. M. D. Gray, S. Nishio, eds. *Proc. of the 21st VLDB Conference Zurich, Switzerland 1995*, Morgan Kaufmann, 1995. 490~501

(下转第103页)

- theorem proving to problem solving. *Artificial Intelligence*, 1971, 2:189~208
- 3 McDermott D. et al. Planning: What it is, What it could be. An introduction to the Special Issue on Planning and Scheduling. *Artificial Intelligence*, 1995, 76:1~16
 - 4 Weld D S. Recent Advances in AI Planning. *Artificial Intelligence (to appear)*. 1999. <ftp://ftp.cs.washington.edu/pub/ai/pi2.ps>
 - 5 Yang Qiang. *Intelligent Planning A Decomposition and Abstraction Based Approach*. Berlin: Springer, 1997
 - 6 Kautz H, McAllester D, Selman B. Encoding plans in propositional logic. In: Proc. 5th Int. Conf. Principles of Knowledge Representation and Reasoning, 1996
 - 7 Kautz H, Selman B. Pushing the envelope: Planning, propositional logic, and stochastic search. In: Proc. 13th Nat. Conf. AI. 1996. 1194~1201
 - 8 Kautz H, Selman B. Blackbox: A new approach to the application of theorem proving to problem solving. In: AIPS98 Workshop on Planning as Combinatorial Search. 1998. 58~60
 - 9 Kautz H, Selman B. The role of domain specific knowledge in the planning as satisfiability framework. In: Proc. 4th Intl. Conf. AI Planning Systems. 1998
 - 10 Tate A, Drabble B, Dalton J. O-Plan: a Knowledge-Based Planner and its Application to Logistics. AAAI Press, Menlo Park, California. May 1996. <ftp://ftp.aiia.ed.ac.uk/pub/documents/1996/96-arpi-oplan-and-logistics.ps>
 - 11 Blum A, Furst M. Fast planning through planning graph analysis. In: Proc. 14th Int. Joint Conf. AI. 1995. 1636~1642
 - 12 Canny J. Unpublished Observation. 1985
 - 13 Gupta N, Nau D S. Complexity results for blocks-world planning. In: Proc. Ninth National Conf. on Artificial Intelligence. 1991. 629~633
 - 14 Chenoweth S V. On the NP-hardness of blocks world. In: Proc. of the Ninth National Conference on Artificial Intelligence (AAAI-91). Anaheim, CA. 1991. 623~628
 - 15 Gupta N, Nau D S. On the Complexity of Blocks-World Planning. *Artificial Intelligence*, 1992, 56(2-3): 223~254
 - 16 Chapman D. Planning for conjunctive goals. *Artificial Intelligence*, 1987, 32(3): 333~377
 - 17 Bylander T. Complexity results for planning. In: Proc. of IJCAI91. 1991. 274~279
 - 18 Nebel B. On the computational complexity of temporal projection, planning, and plan validation. *Artificial Intelligence*, 1994, 66(1): 125~160
 - 19 Selman B. Near-Optimal Plans, Tractability, and Reactivity. In: Proc. KR-94. 1994. 521~529
 - 20 Soderland S, Weld D. Evaluating nonlinear planning: [Technical Report TR 91-02-03]. University of Washington CSE. 1991
 - 21 McAllester D, Rosenblitt D. Systematic Nonlinear Planning. AAAI-1991. <ftp://ai.mit.edu/pub/dam/aaai91c.ps>
 - 22 <http://arpi.isx.com/arpi.html>
 - 23 <http://planet.dfki.de>
 - 24 <http://www-aig.jpl.nasa.gov/public/planning/index.html>
 - 25 <http://www.eumetsat.de/>
 - 26 Aarup M, et al. Optimum-AIV, a knowledge-based planning and scheduling system for spacecraft AI. *Intelligent Scheduling*. Morgan Kaufmann. 1992
 - 27 Chien S A, et al. Automated generation of tracking plans for a network of communication antennas. In: Proc. 1997 IEEE AeroSpace Conf. 1997
 - 28 Chien S. Using AI techniques to automatically generate image processing procedures: a preliminary report. Proc. AIPS 94
 - 29 Etzioni O, Weld D S. A softbot-based interface to the Internet. *Communications of the ACM*, 1994, 37(7)
 - 30 Weld D S. Planning-based control of software Agents. AIPS-1996. ftp://june.cs.washington.edu/pub/ai/weld_aips96.ps.Z
 - 31 Etzioni O. A Softbot-Based Interface to the Internet. CACM. July 1994 <ftp://ftp.cs.washington.edu/pub/etzioni/softbots/cacm.ps>
 - 32 Weld D S. Planning to Gather Information. AAAI. Aug. 1996. <ftp://ftp.cs.washington.edu/pub/ai/occam-aaai96.ps>
 - 33 Channou D, et al. Very Fast Motion Planning for Dexterous Robots. IEEE International Symposium on Assembly and Task Planning Aug. 1995
 - 34 Olawsky D, Gini M. Deferred Planning and Sensor Use, in Innovative Approaches to Planning, Scheduling and Control. In: Proc. 1990 DARPA Workshop. M. Kaufmann. 1990
 - 35 Etzioni O. Efficient Information Gathering on the Internet. FOCS 1996. <ftp://ftp.cs.washington.edu/pub/etzioni/softbots/focs96.ps.gz>
 - 36 Erdem E, Lifschitz V, Wong M. Wire Routing and Satisfiability Planning. In: Proc. of the First Intl. Conf. of Computational Logic, 2000
 - 37 Nau D, Gupta S, Regli W. AI planning vs manufacturing-operation planning: a case study. IJCAI 1995. <http://elara.salford.ac.uk/plansig/papers/nau-ijcai95.pdf>

(上接第126页)

- 13 Keogh E J, Pazzani M. An Indexing Scheme for Fast Similarity Search in Large Time Series Databases. In: Proc. 11th Intl. Conf. on Scientific and Statistical Database Management, SSDBM'99, Cleveland, OH. IEEE Computer Society. 1999. 56~67
- 14 Agrawal R, Faloutsos C, Swami A N. Efficient similarity search in sequence database. In: D. Lomet, ed. Proc. of the 4th Intl. Conf. of Foundations of Data Organization and Algorithms (FODO'93), Vol. 730 of Lecture Notes in Computer Science, Springer, 1993. 69~84
- 15 Faloutsos C, Ranganathan M, Manolopoulos Y. Fast subsequence matching in time-series databases. In: Proc. 1994. ACM-SIGMOD Intl. Conf. Management of Data, Minneapolis, May 1994. 419~429
- 16 Huhtala Y, Karkkainen J, Toivonen H. Mining for similarities in aligned time series using wavelets. Part of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools and Technology. Orlando, Florida. 1999, 3695: 150~160
- 17 Mannila H, Ronkainen P. Similarity of Event Sequences (revised version). In: Proc. of the Fourth Intl. Workshop on Temporal Representation and Reasoning (TIME'97), 10th-11th May, 1997, Daytona Beach, Florida, USA, 36~139.
- 18 Guralnik V, Srivastava J. Event Detection from Time Series Data. In: Chaudhuri S, Madigan D, eds. In: Proc. Fifth Intl. Conf. on Knowledge Discovery and Data Mining, San Diego, CA, USA. ACM Press, 1999. 33~42
- 19 Tsumoto S. Automated Discovery of Plausible Rules Based on Rough Sets and Rough Induction. In: Proc. of the 3th Pacific-Asia Conf. PAKDD-99, Beijing, China, April, 1999. 210~219
- 20 Weiss S, Indurkha N. Predictive Data Mining, Morgan Kaufmann Publishers, Inc., San Francisco, USA, 1998