基于相关集合的数据挖掘理论基础研究*)

王晓峰[1,2] 王天然[1]

(中国科学院沈阳自动化所 沈阳110030)[1] (沈阳化工学院 沈阳110142)[2]

A Study on Data Mining Theoretical Frameworks Based on Correlativity Sets

WANG Xiao-Feng[1,2] WANG Tian-Ran[1]

(Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110030)[1]

(Shenyang Institute of Chemical Technology, Shenyang 110142)[2]

Abstract The plausibility relation which is generalization of fuzzy relation and probabilistic relation is proposed in the paper. We think data mining to be a process of finding the plausibility relation in database and correlativity measure to be a particular plausibility relation based on correlativity sets. The critical calculates such as the accuracy of the rough sets, the confidence and the bayesian form in data mining can be united using the correlativity measure. The GPDM (General Process of Data Mining)that represents the nature of data mining is also proposed. The data mining theoretical foundation and frameworks based on correlativity sets are also given and discussed in the paper.

Keywords Correlativity set, Data mining, Plausibility relation, Correlativity measure

1 引言

到目前为止,人们已经发现了很多数据挖掘方法,如Apriori算法[1]、决策树方法、贝叶斯(Bayes)方法、人工神经网络方法,等等;各式各样的数据挖掘理论被提出与采用,如模糊集合、粗糙集(Rough sets)理论、数理统计、机器学习、人工神经网络、决策树、模式识别、高性能计算等[3];各种各样的数据被挖掘,从传统的关系数据库到各种文本数据、空间数据、图像与视频数据,以及Web信息、生物信息,等等;各种各样的数据挖掘平台被不断地开发,如美国SAS公司的SAS Enterprise Miner、IBM公司的Intelligent Miner、加拿大Simon Fraser大学的DB Miner、中国科学院计算技术研究所的MS Miner[4]等。但是,随着数据挖掘理论和应用的不断深入研究与发展,人们发现"有关数据挖掘的基础研究还没有成熟"。如果能建立一个统一的理论框架,把各种挖掘方法统一起来,对比较与评价各种数据挖掘方法,建立一种标准的数据挖掘语言,推动数据挖掘的进步等具有十分重要的意义。

什么是数据挖掘的理论基础?有各种各样的回答。例如,数据归约理论、数据压缩的观点、模式发现观点,概率论,归纳数据库,微观经济学等等。数据库领域国际知名学者加拿大Simon Fraser大学的Han Jiawei(韩家炜)和M. Kamber[3]在2001年提出,一个理想的理论框架应该满足如下要求:

(1)能够对典型的数据挖掘任务(如关联、分类和聚类)进行建模;

(2)有一个概率特性;

(3)能够处理不同形式的数据,并且对数据挖掘的反复和交互的本质加以考虑。

就现有的理论来讲,还没有一种理论框架能满足这3条要求。是否存在满足这些要求的理论?还需要进一步的研究,也说明对于数据挖掘的本质仍需要深入的理解。

作者认为,由于数据挖掘对象的复杂性、用户需求的多样性,不太可能有一种万能的、通用的数据挖掘方法适用各种数据挖掘任务。但是,现有的研究结果表明,把不同数据挖掘方法中的一些基本概念、基本计算和基本步骤适当地融合在一起,形成一种统一的抽象的理论框架是可行的。例如,粗糙集中的粗糙度计算和Apriori算法中的信任度和支持度计算可统一为相关测度[6];简单贝叶斯分类中逻辑回归模型与人工神经网络中Sigmoid型激励的感知器函数是一致的;信任度的概率表示与贝叶斯方法的一致性等等[4]。这些不同数据挖掘方法中的核心概念、基本计算在数学上的一致性绝非偶然,其中包含某种必然的联系。这种隐藏在表面上看起来不相关的各种理论与方法中的内在联系,是构成一个统一的理论框架的基础。

本文将从知识与知识的表示问题入手,探讨数据挖掘的本质,提出相关集合作为数据挖掘基础的新观点,分析现有的主要数据挖掘方法在基本概念、基本理论、基本计算上的共性、彼此之间的内在的联系以及相互融合的问题,论证相关集合理论与方法作为数据挖掘理论框架基础的可能性。

2 数据挖掘的本质

2.1 数据挖掘中的知识表示

集合是最基本的数学概念之一,是把具体事物抽象为数学符号的第一步。关系是建立在集合上的另一个最基本的概念。如果一个集合代表一个概念,或代表一个事物,那么关系就是概念或事物之间的联系(包括函数、运算)。知识是关于事物的状态、运动规律以及一个事物同周围其它事物之间的联系。在数据挖掘研究领域,人们通常把知识分成4种:关联规则、分类规则、顺序规则、聚类等。其中,关联规则描述了数据(事物)之间的内在联系;聚类是自动辨别类别;分类是已知类别求分类函数;顺序规则是关联或分类的时序化,是一种特殊的关联规则或分类规则。这4种规则可概括为两类:一类是"辨别"事物的知识,如聚类和分类则根据数据特征"辨别"事物,使数据形成不同的集合;另一类是描述事物之间的"内在联系",如关联规则。"辨别"事物和描述"联系"这两类知识在数学上都可以用"关系"来表示。"辨别"事物是建立特征属性集合和概念集合的关系,事物之间的联系则是概念集间的"关系"。所以用集合、关系等来表述数据挖掘是最合适不过的,比其它理论更贴近数据、贴近知识。

*)本文获国家自然科学基金重点课题资助(79931000)和863高科技研究发展计划(2001AA413410)资助。

对于数据挖掘来说，事物是以数据形式存在数据库中的，挖掘事物的数据特征或事物之间的联系(知识)就是从数据库中发现这些“关系”。由于代表事物的数据往往是不完全的、非完备的，因此挖掘出来的知识是一种不完备的、非精确的关系，是一种推测结果。为能方便地描述这种推测结果，我们引入一种似然关系。

2.2 似然关系

设 X、Y 为两个有限集合，R 是 $X\times Y$ 上的二元关系，如果对任意 $x\ R\ y$，存在有 η 与之对应，其中，$\eta\in[0,1]$，$x\in X$，$y\in Y$，那么称 R 为 $X\times Y$ 上的二元似然关系，η 是关系 R 的似然度。

很明显，似然关系是模糊关系和概率关系的推广。当 η 是隶属值时，R 是模糊关系；如果 η 是概率值，那么 R 是概率关系。一般情况下 η 是相关强度，连同 R 一起表示挖掘的知识。“似然”表示似乎是对的，似然关系意味着关系 R 似乎成立。究竟成立的程度如何？由 η 决定。当 $\eta=1$ 时，R 为普通关系；$\eta=0$ 时，关系为假(无关系)；一般情况下 $0<\eta<1$，表示关系 R 似乎成立；η 接近1时，成立的可信度大些，反之小些。

注意：美国 Stanford 大学的 Daphne Koller 教授等人1999年在 IJCAI99会议(第16届国际人工智能大会)上提出过一种 Probabilistic Relational Models(PRM)[2]概率关系模型，并采用贝叶斯网络方法直接从数据中学习这种概率关系模型，具有参数学习和结构学习两种学习方式。

Daphne Koller 等人认为数据挖掘的结果是一种概率关系，但概率关系不能表达关系的所有情况，关系的不确定性也不完全都是随机性事件，也可能是因为数据模糊，边界不清，也可能是数据不完备，不足以建立准确的关系。对模糊数据来说，模糊数学中关于模糊关系的定义讨论了这一方面的影响，这些结果已被人们普遍接受。另外，我国学者何华灿教授根据模糊集合理论的缺陷提出了“关系柔性”的概念，并指出模糊命题之间的关系可能是连续可变的[10]。数据挖掘的结果是否也是一种“关系柔性”？似然关系为我们提供了一种新思路。

似然关系描述的是一种不确定的、非精确的或非确切的关系。它可能是一种概率关系，或模糊关系，或因数据不完备形成的一种部分关系。在似然关系定义中对此并没有提出具体要求，所以可代表各种具体的挖掘结果，如关联规则、分类规则、聚类规则等。根据似然关系和前面对知识的讨论结果，有以下数据挖掘公式：

数据挖掘＝集合＋似然关系＋挖掘算法

三者的联系关系如图1所示。

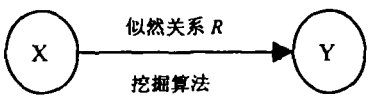

图1 数据挖掘的本质

其中，X、Y 表示数据集，为挖掘对象；R 是似然关系，为挖掘结果；挖掘算法是一种由一系列关键计算步骤组成的具体挖掘过程。这样，数据挖掘被抽象为一个简单的基本模型，从这一基本模型出发，进一步深入研究、寻找各种挖掘方法的关键步骤和共同点(一致性)，就有可能建立起数据挖掘的理论框架。

2.3 相关测度——一种似然度的计算方法

似然关系的定义比较容易理解，但问题的关键是似然度如何计算？特别是数据挖掘中的似然度如何计算？作者在文[6]中提出了相关测度的概念，并证明了关联规则的信任度、Rough sets 理论的粗糙度(一种模糊度)等都是一种相关测度，形式上是一致的。这里将进一步讨论似然关系、似然度与相关测度之间的联系，从而给出相似测度的计算方法。

为讨论方便，下面先介绍相关测度的定义(详细内容见文[6])，再进行比较分析。

相关测度定义 设 U 是非空有限集合，$P(U)$ 是 U 的幂集，R 是定义在 $P(U)$ 上的二元关系，若 $A\in P(U)$，给定集合 $X\subseteq U$，A 与 X 有关系 R，则称 $F_A(X):P(U)\to[0,1]$ 为集合 A 与集合 X 相对于关系 R 的相关测度。其中，$F_A(X)$ 是由关系 R 诱导产生的 $P(U)$ 到 $[0,1]$ 上的映射，满足

(1) $F_A(\Phi)=0$；

(2) $F_A(A)=1$；

(3) $0\leqslant F_A(X)\leqslant 1$；

(4) $F_A(X_i\cup X_j)\leqslant F_A(X_i)+F_A(X_j)$

其中，X_i、$X_j\subseteq U$，A 与 X_i、X_j 有关系 R。

注意1 R 是定义在幂集上的或集合间的二元关系。比如，A 与 X 有共同的属性关系、A 包含 X(包含关系)等等。通过关系两个集合元素建立了一种联系(或关联)，而相关测度是这种关联程度的一个量化。

注意2 相关测度描述了给定关系和相关集合的数字特征，但实际的 $F_A(X)$ 映射要根据关系的具体定义来考虑。如果关系是用某种集合运算定义的，那么只要它满足测度定义的条件就可作为 $F_A(X)$ 映射直接运用；如果是用语言或特征说明定义的，则必须把它用集合运算精确地描述出来，把关系化为集合间的运算形式。比如，设 $U=\{t1,t2,t3,t4,t5\}$，$P(U)$ 是 U 的幂集，R 是 $P(U)$ 上的包含关系。R 可以精确地表示为：$R=\{\langle A,B\rangle \mid A,B\in P(U),A\supseteq B\}$，$A$ 的相关集合是 $[A]_R=\{B\mid A\supseteq B,B\in P(U)\}$，这时 $F_A(X)=|[X]_R\cap[A]_R|/|[A]_R|$，满足测度定义的条件要求，是 R 的一种相关测度。

根据相关测度的定义可知，给定一个二元关系 R，可建立一个从论域集合幂集到闭区间 $[0,1]$ 之间的映射，当映射满足可加性等基本性质时，就成为关系 R 的相关测度。很明显，这种相关测度值是 $[0,1]$ 间的一个数值结果。比较似然关系的定义，不难发现，二者其实是统一的。在似然关系的定义中只是要求集合中任意两个元素有关系 R 时有一个 $[0,1]$ 之间的数值(似然度)与之对应，并没有给出具体的计算方法和约束条件，而相关测度却对 $F_A(X)$ 的计算规则提出了具体的约束条件。似然度是从元素角度描述关系的成立情况，相关测度则是从集合角度描述关系的可信赖程度。二者如何统一呢？注意，在数据挖掘中往往一个问题有两种表述。例如，一个概念可能包含有多个属性，这多个属性构成了一个集合。而一个概念又可能从属于某个概念集合，是一个集合的元素。这样，可以从元素角度讨论概念，也可从集合角度讨论概念。一个概念既是一个(概念集合的)元素，又是一个(属性)集合。同理，属性是属性值的集合，也具有二重含义。如果两个概念之间有关系 R，那么 R 的成立程度也有两种度量方式：一种是基于元素的，一种是基于集合的。似然度是基于元素的关系度量，相关测度是基于集合的关系度量，二者度量的是一个关系，所以可令似然度等于相关测度。其关系如图2所示。

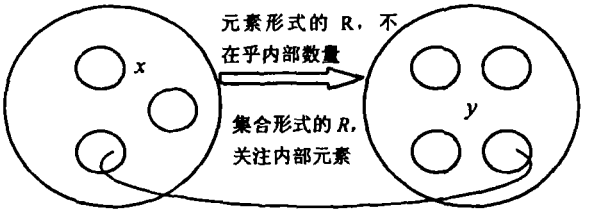

图2 同一关系的两种表达

数据挖掘是从数据库中挖掘知识，是一种从“低级”的属性集中发现“高级”的概念关系，从具体数据到抽象概念的过程。由属性集之间的相关测度获得概念之间的似然度，正是这种从“低级”数据中挖掘“高级”概念思想的具体体现。

3 相关集合——数据挖掘的基础

如上所述，数据挖掘的结果可以表示为一种似然关系，其似然度可用相关测度“挖掘”出来。这种“挖掘”是否具有普遍意义？能否满足现有的各种挖掘方法，构成数据挖掘理论框架的基础？下面对这些问题做进一步的深入探讨。

3.1 相关测度——经典挖掘方法的统一基础

在文[6]中我们论证了关联规则的信任度是一种相关测度，一种基于集合交集运算的信任度。其主要思路为：

设 X、Y 为两个相关联的项目集，关联规则 $X\Rightarrow Y$ 在事务数据库中的可信度是指包含 X 项目和 Y 项目的事务数与包含 X 的事务数之比，记为 confidence($X\Rightarrow Y$)。注意到 X、Y 均有二重含义：从符号本身来讲，是一些项目名称；另外它们分别又是含有对应项目的事务集合。所以有：

$$\text{confidence}(X\Rightarrow Y)= |X\cap Y|/|X|$$

但，根据相关测度的定义有

$$|X\cap Y|/|X|=F_A(X)$$

其中，$X\cap Y$ 表示 X 项目和 Y 项目在相同的事务中是含有 X 项目和含有 Y 项目的两个事务集合的交集，X 表示包含 X 项目的事务集。

所以，$|X\cap Y|/|X| =(|X\cap Y|/|D|)/(|X|/|D|)=P(X\cap Y)/P(X)= P(X|Y) * P(Y)/P(X)=P(Y|X)$。

这里，$|D|$为事务数据库中的事务总数；$P(X)=|X|/|D|$，表示项目 X 出现在事务数据库中的概率；$P(X\cap Y)=|X\cap Y|/|D|$，表示 X 项目和 Y 项目同时出现在事务数据库中的概率；$P(Y)=|Y|/|D|$，表示项目 Y 出现在事务数据库中的概率。

因此，$F_A(X)=$ confidence($X\Rightarrow Y$)$= |X\cap Y|/|X|$

$=P(Y|X)=P(X|Y) * P(Y)/P(X)$，

即 $X\Rightarrow Y$ 规则的信任度与条件概率 $P(Y|X)$及相关测度是等价的。也可以说，相关测度概括了信任度和条件概率。

注意到简单贝叶斯公式与条件概率的关系，并且贝叶斯公式满足相关测度的定义要求，所以用相关测度能够表达贝叶斯公式、信任度、条件概率。

另外，在文[6]中我们还证明了粗糙集理论中的粗糙度同信任度一样，也是一种相关测度，所以分类挖掘的重要工具——粗糙集理论同以贝叶斯公式为核心的贝叶斯网络以及以信任度计算为关键的关联规则挖掘方法(注意：支持度是信任度的特例)在相关测度上统一了，这些数据挖掘方法具有相同的相关测度。

这样，相关测度为这些经典的数据挖掘方法建立了一个共同的理论框架基础。这基本满足 Jiawei Han(韩家炜)和 M. Kamber 提出的理想理论框架的第1项基本要求。

另外从相关集合出发，我们还发展了一批有效的数据挖掘方法，其中有基于相关集合映射的双空间搜索方法、利用粗糙集约简性质自顶向下挖掘频繁项的方法，也有专门用于知识库化简的相关集合方法[5~9]，这些方法包括分类规则和关联规则的挖掘，各种广义粗糙集合、信任度、概率计算、模糊集合等都可以放到一个相关集合的框架之中，这从另一方面反映了相关集合和相关测度的普适性。

3.2 GPDM(一个一般性的数据挖掘过程)

相关测度为各种数据挖掘方法提供了一个抽象的、形式化的计算公式。根据这一公式和相关集合的知识表示方法，我们有下面的一般数据挖掘过程。

设论域 U 为给定的有限集合，其中每个元素都是一个属性，属性的不同组合构成了不同的概念。假设有两个概念集合 C、D，$c_i\in C$，$d_j\in D$ 是两个具体的概念，则概念 $c_i\subseteq U$，$d_j\subseteq U$ 都是属性的集合。假定关系 $c_i\rightarrow d_j$ 成立，然后验证是否成立。如果 $c_i\rightarrow d_j$ 的相关测度大于事先给定的信任阈值，则 $c_i\rightarrow d_j$ 成立，否则不成立。表示成算法有，一般的数据挖掘过程 GPDM：

(1)依据某种规则，如概念(集合)的支持度，从给定的论域中产生候选概念(集合)，例如 C、D。如果没有候选概念(集合)，则结束数据挖掘过程。

(2)从概念(集合)中选取候选概念属性(或概念)，例如 $c_i\in C$，$d_j\in D$。

(3)试产生一个蕴涵式(或因果关系)如 $c_i\rightarrow d_j$。

(4)如果概念属性(或概念)间的似然度(通过相关测度获得)满足事先定义的指标，则蕴涵式(因果关系)$c_i\rightarrow d_j$ 连同似然度一起构成一条规则(知识)。

(5)还有没处理的候选概念属性(或概念)吗？是，转到(2)。否则转到(1)计算下一组概念(集合)。

注意，不同的候选概念(集合)产生方法及不同的支持度算法构成了各种不同的数据挖掘算法。

例如，在决策树方法中，根据信息熵的大小选择概念属性。在前面的相关集合数据挖掘方法中，把条件属性值集的组合作为候选概念集合；依据有相同对象(交集)关系按定义6来计算两类属性值之间的相关强度，确认蕴涵式是否成立等等。

在经典的关联规则挖掘方法——Apriori 算法中，根据支持度大小选择项目集(概念属性)。根据规则 $X\Rightarrow Y$ 在事务数据库 D 中的可信度(confidence)来确定规则是否成立。

在粗糙集理论中根据知识依赖值(一种测度)决定约简项。在贝叶斯网络方法中根据概率密度分布函数决定候选参数，由概率值推断参数或结构是否成立等等。

这些方法基本都符合上面的一般挖掘过程 GPDM，所以 GPDM 概括了一般的数据挖掘，包括分类规则、关联规则等在内的各种挖掘方法，是一个数据挖掘的基本框架。

4 面向关系数据库的 GPDM 系统框架结构

对于数据挖掘来说，关系数据库的一个字段就是一种属性，若干个属性的集合表示一个概念，一个广义元组(记录)就是这种概念的一个具体实例或者对象。用相关集合方法从数据库中发现知识，就是分析属性集合(概念)之间的相关性，确定似然关系。我们已经讨论了 GPDM 的处理过程，下面将进一步给出 GPDM 系统的框架结构。

设 $S=\{U,A,H,V,f\}$是一个给定的信息系统。其中：U 是对象的有限集合，$U=\{x_1,x_2,\cdots,x_n\}$；A 是属性的有限集合，$A=C\cup D$，C 是条件属性集合，D 是决策属性集合；H 是概念的有限集合，$H=\{A_1,A_2,\cdots,A_m\}$，$A_i\subseteq A$ 是部分属性集合；$V=\bigcup_{p\in A}V_p$，V_P是属性 P 的值域；$f:U\times A\rightarrow V$ 是对象属性到值域的映射，若 $p\in A$，$x\in U$，则 $f(x,p)\in V_P$。

一个具体的关系数据库就是一个这样的信息系统，其中字段是属性，行是对象，第 x 行处字段 p 的值就是对象 x 对应属性 p 的属性值。若干个字段的组合代表一个概念。

根据上面的 GPDM 过程，有图3所示的 GPDM 系统结构。把数据挖掘分成了5个层次：概念层、属性层、属性值层、属性值-对象层、对象层。

其中，所有要挖掘的概念集合构成了概念空间；概念空间

经过候选规则把可能存在某种关系的概念(属性集合)放在一起,形成具体的概念层挖掘对象;概念是属性的集合,由属性表达概念,所有概念属性的集合构成了属性层;每个属性是若干属性值组成的集合,属性值表达属性,所有属性值在一起形成了属性值层;数据库的全体广义元组(记录)构成了对象层;把含有同样属性值的对象放在一起,构成一个集合,表达属性值。所有这样的属性值-对象集合组成了一个集合类:属性值-对象层。

概念层是一种适合人类思维的抽象层,在最上面;对象层是数据库中的数据,一种适合机器工作的数据层;中间是从上到下的过渡层。通过由上向下经过逐层转换,把概念之间可能存在的关系转化成元组对象集合之间的似然关系。属性值-对象层中集合之间的相关测度决定了似然关系是否成立。当似然关系成立时,作为数据挖掘结果由下向上逐层传递到概念层。

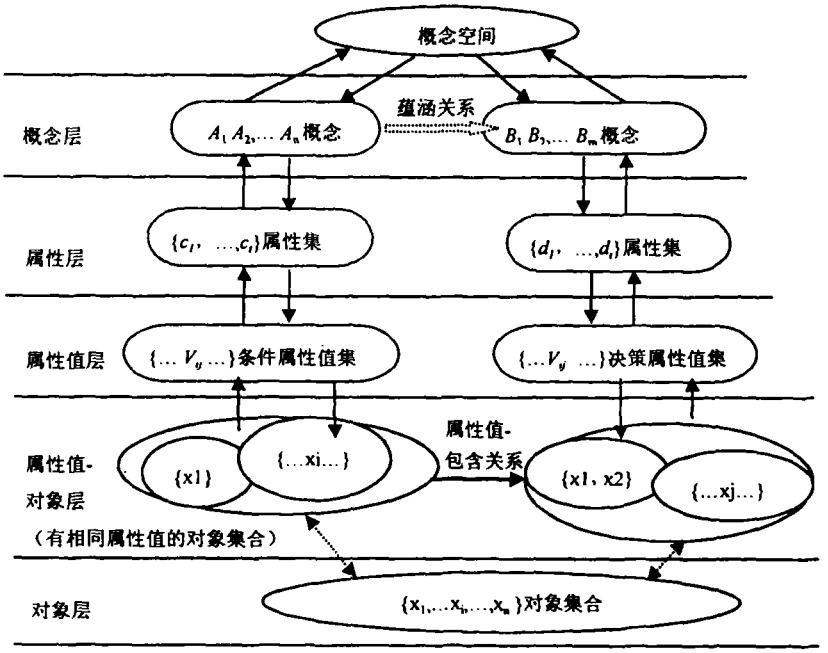

图3 GPDM 系统的结构模型

注意:概念是事物在人脑中的反映,概念空间是面向人的,而属性、属性值以及广义元组是存储在计算机数据库中的数据。概念与属性之间的转换过程是人机交互过程,可以事先由人规定好概念与概念或概念与属性的关系,也可以让计算机按某种规则(程序)把属性组织起来构成概念。前者是人工建立的"概念树",后者是机器进行结构归纳或推断的结果。李德毅院士提出的隶属云方法[11]是一种层次间的转换方法,贝叶斯网络[3,4]则是一种具体的层次结构归纳方法。在文[8,9]中,数据库知识发现的相关集合算法(Mset Algorithm for KDD)则是按照图3层次结构,把寻找概念间的蕴涵式转换为寻找相关对象集合的包含关系,用集合的相关强度来判断条件对象集合和决策对象集合间的包含关系是否成立。

结束语 本文提出的新概念主要有似然关系和似然度。我们从数据挖掘的概念出发分析了数据挖掘的本质,并把数据挖掘结果概括为一种似然关系,提出了用相关测度计算似然度的方法,最后给出了一个数据挖掘的描述方法及 GPDM 结构模型。本文提出的似然关系是模糊关系和概率关系的推广与概括,而数据挖掘就是从数据库中发现这种隐含在数据背后的似然关系。笔者认为,相关测度是一种具体的似然度,一种基于相关集合的似然度。利用相关测度能把粗糙集理论中的粗糙度、关联规则中的信任度、贝叶斯公式等(注意:支持度是信任度的特例)数据挖掘的关键计算统一起来,并据此建立一个表达数据挖掘本质的 GPDM 一般描述方法、数据挖掘的抽象模型。基于相关集合能够构建一个描述数据挖掘本质的一般数据挖掘的理论基础和框架结构,该架构可满足 Han Jiawei 和 M. Kamber 理想理论框架的基本要求,这种理论框架应容纳现有经典的挖掘方法,成为开发数据挖掘语言及统一建模的重要理论基础。本文对数据挖掘理论基础做了较为深入的探讨,仍有许多问题有待于深入研究,如似然关系的性质,不同格式、不同结构(尤其是半结构)的数据挖掘(或转换)问题等。

参 考 文 献

1 Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: Proc. of the ACM SIGMOD Conf. on Management of data, 1993. 207～216
2 Friedman N, Getoor L, Koller D, Pfeffer A. Learning Probabilistic Relational Models. In: Proc. of the 16th Intl. Joint Conf. on Artificial Intelligence (IJCAI), Stockholm, Sweden, Aug. 1999. 1300～1307
3 Han Jiawei, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2000
4 史忠植. 知识发现. 清华大学出版社, 2002. 1
5 Wang Xiaofeng, Yin Danna, Cheng Shihchuan. Mutuality Sets. [J] 沈阳化工学院学报, 1999, 3: 67～76
6 王晓峰, 王天然. 相关测度与增量式信任度和支持度的计算. [J]软件学报, 2002. 11
7 王晓峰, 王天然. 基于双空间搜索的频繁项挖掘方法. [J]计算机科学, 2002. 4
8 王晓峰, 尹丹娜, 郑诗诠. 相关集合及其在知识库化简中的应用. [J]清华大学学报, 1998, 7: S2: 6～9
9 王晓峰, 唐忠, 等. 相关集合数据库知识发现中的应用. [J]南京大学学报, 计算机专辑, 2000(36): 11, 52～57
10 何华灿, 王华, 刘永怀, 王拥军, 杜永文. 泛逻辑学原理. 科学出版社, 2001. 8
11 李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器. [J]计算机研究与发展, 1995, 32(6)