粗糙集属性约简的图论方法

卢 鹏 肖健梅 王锡淮

(上海海事大学物流工程学院电气工程系 上海 200135)

摘 要 通过研究粗糙集与图论的关系,提出了以集合为权的加权多重完全多部图的概念,定义了加权多重完全多部图的邻接矩阵,得到了加权完全多部图与决策表的映射关系;给出了粗糙集决策表信息系统的图论形式和决策表信息系统属性约简的图论方法,并根据图论理论对算法进行了优化;得到了在决策表信息系统中,属性的集合不可以约简的充分必要条件;并进一步提出了基于属性置信度的计算方法和多决策属性的处理方法。编程实验结果证明该方法能有效地降低时间和空间复杂度。

关键词 加权多重完全多部图,决策表信息系统,属性约简,属性置信度

中图法分类号 TP18 **文献标识码** A

Graph Method of Rough Set Attribute Reduction

LU Peng XIAO Jian-mei WANG Xi-huai

(Department of Electrical Engineering, Logistic Engineering College, Shanghai Maritime University, Shanghai 200135, China)

Abstract Through study of rough set and graph theory, this paper put forward the concept of weighted complete multipartite multigraph which used set as weitht, defined the adjacency matrix of weighted complete multipartite multigraph, obtained the mapping relations between weighted complete multipartite multigraph and decision table, gave a gragh model of the rough set decision table information system and a method of attribute reduction in decision table Information systems based on gragh theory, optimized the algorithm, obtained the sufficient and necessary conditions of attribute reduction in decision table information system, further proposed calculation method which is based on attribute reliability and the processing method of multiple decision attributes. Programming experimental results show that this method can effectively reduce the complexity of time and space.

Keywords Weighted complete multipartite multigraph, Decision table information systems, Attribute reduction, Attribute reliability

1 引言

属性约简是粗糙集应用研究的重要内容之一。目前,常用的属性约简方法有基于正区域的属性约简[1,2]、基于差别矩阵的属性约简[3,4]、基于信息熵的属性约简[5-7]、基于分布的属性约简以及基于近似的属性约简[8]。此外,腾书华等提出了一种基于不可区分度的启发式约简算法[9]。代劲、何中市提出了基于云模型的约简方法[10]。徐久成等提出了基于相对粒度的约简算法[11]。文献[12]证明了基于信息熵的属性约简与基于分布的属性约简等价。文献[13]比较了几类不同的属性约简算法之间的差异。文献[14,15]指出基于正区域的核与基于 Skowron 差别矩阵的核不等价,给出了一个修正的差别矩阵,从而使得基于正区域的核与基于修正差别矩阵的核等价。自此之后,有许多学者对基于正区域的求核算法做了大量的研究[16-18],而对基于 Skowron 差别矩阵的求核算法的研究还比较少。本文从图论研究着手,通过对多重完全多部图的定义及研究,推导出有权多重完全多部图的邻接矩阵,并将该矩阵与粗糙集中的差别矩阵相结合,给出了粗糙集差别矩阵的图论定义和计算方法,并进一步提出了基于属性置信度的计算方法和多决策属性的处理方法。

2 粗糙集和粗糙集属性约简的基本知识

粗糙集的主要研究对象是决策表,一个决策表就是一个决策信息系统。为了从决策表中得到适应度大的规则,就需要对决策表进行约简,使得经过约简处理的决策表中的一个记录就代表具有相同规律特性的一类样本。

定义 1[1,2] 五元组 $S=(U,C,D,V,f)$ 是一个决策表,其中 $U=\{x_1,x_2,x_3,\cdots,x_n\}$ 表示研究对象的非空有限集,称为论域;$C=\{c_1,c_2,\cdots,c_r\}$ 表示条件属性的非空有限集;D 表示决策属性的非空有限集,$C\cap D=\phi$,$V=\bigcup_{a\in C\cup D}V_a$,$V_a$ 是属性 a 的值域,$f:U\times C\cup D\rightarrow V$ 是一个信息函数,它给 U 中每一个对象的所有属性赋予信息值,即对 $\forall x\in U,a\in C\cup D$,有 $f(x,a)\in V_a$;每一个属性子集 $P\subseteq(C\cup D)$ 决定了一个二元不可区分关系:

到稿日期:2011-04-24 返修日期:2011-07-22 本文受上海市教委重点学科建设项目(J50602)资助。

卢 鹏(1981-),男,博士生,主要研究方向为粗糙集理论、智能信息处理等,E-mail:lplplp20051@163.com;肖健梅(1962-),女,教授,主要研究方向为智能控制、智能信息处理等;王锡淮(1961-),男,教授,博士生导师,主要研究方向为粗糙集理论、复杂系统建模与控制等。

$IND(P)=\{(x,y)|(x,y)\in U\times U,\forall a\in P\wedge f(x,a)=f(y,a)\}$

关系 $IND(P)$ 构成了 U 的一个划分，用 $U/IND(P)$ 表示，简记为 U/P，U/P 中的元素 $[x]_p=\{y|\forall a\in P,f(x,a)=f(y,a)\}$ 称为 U 关于 P 的等价类。

定义 2[1,2] 在决策表 $S=(U,C,D,V,f)$ 中，$\forall R\subseteq C\cup D$，$X\subseteq U$，记 $U/R=\{R_1,R_2,\cdots,R_L\}$，则称 $\underline{R}(X)=\cup\{R_i|R_i\in U/R,R_i\subseteq X\}$ 为 X 关于 R 的下近似集。

定义 3[1,2] 在决策表 $S=(U,C,D,V,f)$ 中，设 $U/D=\{D_1,D_2,\cdots,D_k\}$ 表示由决策属性集 D 对论域 U 的划分，$U/C=\{C_1,C_2,\cdots,C_r\}$ 表示由条件属性集 C 对论域 U 的划分，其中 $C_i(i=1,2,\cdots,r)$ 称为基本块，称 $POS_C(D)=\bigcup_{D_i\in U/D}\underline{C}(D_i)$ 为 C 关于 D 的正区域。若 $POS_C(D)=U$，则称该决策表是协调的，否则称该决策表是不协调的。

定义 4[6] 在决策表 $S=(U,C,D,V,f)$ 中，若 $\forall B\subseteq C$，$POS_B(D)=POS_C(D)$；$\forall b\in B$ 均有 $POS_{B-\{b\}}(D)\neq POS_C(D)$，则称 B 是 C 相对于 D 的基于正区域的属性约简，记所有的属性约简为 $PRED_D(C)$。

定义 5[4] 在决策表 $S=(U,C,D,V,f)$ 中，设 Hu 的差别矩阵为 $M=(m_{ij})$，其元素定义如下：

$$m_{ij}=\begin{cases}\{a|a\in C,f(x_i,a)\neq f(x_j,a)\wedge f(x_i,D)\neq f(x_j,D)\}\\ \varnothing,\text{否则}\end{cases}$$

$\forall B\subseteq C$，若 B 满足：

(1) $\forall\varnothing\neq m_{ij}\in M$，有 $B\cap m_{ij}\neq\varnothing$；

(2) $\forall b\in B$，$B-\{b\}$ 不满足(1)，则称 B 是 C 相对于 D 的基于 Hu 差别矩阵的属性约简，记其所有的属性约简为 $HRED_D(C)$。

粗糙集属性约简的目标就是求出最优的属性约简。由于求出所有属性的最优约简为 NP-HARD 问题，因此现在的约简方法要么求出所有的约简，要么求局部最优的约简。

3 图论的基本知识

图 $G=(V,E)$，V 表示顶点集合，E 表示边集合。图中顶点的个数叫做图的阶，若连接同一对顶点的边数大于 1，则称这样的边为多重边，其边数称为边的重数。端点重合为一点的边叫做环，没有环及多重边的图叫做简单图，具有多重边的图叫做多重图。

定义 6[20] 每一对不同的顶点均有一条边相连的简单图，称为完全图。n 阶完全图，记作 K_n

性质 1 n 阶完全图 K_n 的边数为 C_n^2；n 阶完全图中顶点的度数和为 $n(n-1)$。

定义 7 设 $V_1,V_2,\cdots,V_m$ 是图 G 的 m 个顶点子集，使得 $V_1\cup V_2\cup\cdots\cup V_m=V(G)$，$V_i\cap V_j=\varphi\cdots(i\neq j)$ 且 G 的每一条边的两个端点一个在 V_i 中另一个在 V_j 中$(i\neq j)$，则称 G 为 m 部图(m-Partite Graph)，记作 $G=(V_1,V_2,\cdots V_m;E)$[20]；如果 V_i 中的每一个点与 V_j 中的每一个点都有一条边相连，$(i\neq j)(i=1,2,\cdots,m;j=1,2,\cdots,m)$，则称图 G 为完全 m 部图，记作 $K_{|V_1|,|V_2|,\cdots|V_m|}$，当 $|V_i|=n_i(i=1,2,\cdots,m)$ 时，将完全 m 部图 G 记作 $K_{n_1,n_2,\cdots,n_m}$。

性质 2 m 部图 G 中，$NUM(G)\leqslant\frac{1}{2}\sum_{i=1}^{m}n_i(n-n_i)$，$NUM(G)$ 表示图 G 的总边数，n 为各部分点的个数和；完全 m 部图中，$NUM(G)=\frac{1}{2}\sum_{i=1}^{m}n_i(n-n_i)$。

定义 8[20] 设图 G 的顶点集 $V(G)=\{v_1,v_2,\cdots,v_n\}$，图 G 的邻接矩阵(adjacency matrix)$A(G)=(a_{ij})_{n\times n}$，$a_{ij}$ 表示点 v_i 和点 v_j 之间边的条数。若存在 $a_{ij}>1$，则称图 G 为多重图。

4 粗糙集属性约简的图论方法

4.1 基于决策表的多重完全图

定义 9 若图 G 的任意两点之间均有重边，且所有边的重数均相等，则称图 G 为经典多重完全图。

方法 1 依照决策表构造经典多重完全图

设在定义 1 给出的系统 S 中不存在不确定信息和冲突信息，相同的样本按照一个样本来处理，经过处理后形成新的样本空间，为方便起见，将经过处理后的样本空间仍然记作 U(以下所提到的 U 都是经过处理后的样本空间)。以 U 中的样本作为图中的点，称为样本点。各样本之间的属性关系为边，称为属性边。两个样本点之间的属性边的集合称为属性边集。由于决策表里有 $|C\cup D|$ 个属性，并且对每一个属性，任意两个样本间都有这个属性关系，因此，任意两个样本点之间都要连 $|C\cup D|$ 条边，这样就构成了一个经典多重完全图。

定理 1 忽略具体属性值的完备无冲突决策表，系统依照方法 1 构造经典多重完全图，存在唯一的多重完全图与其对应。

证明：设经过样本处理后的样本空间为 U，由于忽略了具体属性值，因此决策系统可以简化为 $S=(U,C\cup D)$，即仅表示样本点与属性个数。按照方法 1 构造，由于 U 与 $C\cup D$ 恒定，所以顶点与边都是被确定的，因此构造的图是唯一确定的。我们把这个多重完全图称为此决策表的图，简称为决策表图。

例 1 构造由决策表 1 确定的经典多重完全图。

表 1 示例决策表

样本	属性 a	属性 b	决策属性 d
X_1	1	0	1
X_2	2	2	2
X_3	1	1	0
X_4	2	0	1
X_5	0	2	0
X_6	0	1	1

构造决策表 1 的多重完全图，如图 1 所示。

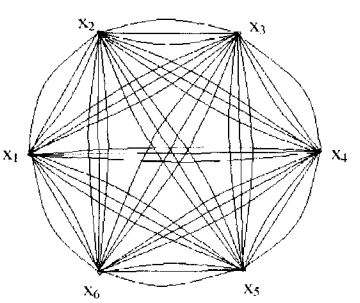

图 1 基于表 1 的经典三重六点完全图

4.2 基于决策表的经典多重完全多部图

定义 10 如果 m 部图 G 中存在两点之间有重边，则称 G 为多重 m 部图；如果多重 m 部图中边的重数都不超过 s，则称为 s 重 m 部图；如果 s 重 m 部图中边的重数都等于 s，则称为

经典 s 重 m 部图；相应地把有重边的完全 m 部图称为多重完全 m 部图；把边的重数都不超过 s 的多重完全 m 部图称为 s 重完全 m 部图；把边的重数都等于 s 的多重完全 m 部图称为经典 s 重完全 m 部图，记为 $K_{(n_1,n_2,\cdots,n_m;s)}$ 或者 $K(n_1,n_2,\cdots,n_m;s)$。

方法 2　依照决策属性作为分部条件构造经典多重完全多部图

D 为多决策属性时，可以将其转化为单决策属性：设 $D=\{d_1,d_2,\cdots,d_s\}$，每个决策属性分别有 $t_1,t_2,\cdots,t_s$ 个值，则 D 共有 $\prod_{i=1}^{s}t_i$ 种情况，令 $k=\prod_{i=1}^{s}t_i$，这样就将多决策属性转化为单决策属性，仍用 D 表示。

当 D 为单决策属性时，将决策属性值代入由方法 1 得到的经典多重完全图中，依据决策属性值对样本点进行分类，可以分为 $U/D=\{D_1,D_2,\cdots,D_k\}$ k 个部分，各部分别包含 n_1，$n_2,\cdots,n_k$ 个样本点，$n_1+n_2+\cdots+n_k=n$。同部的点之间不连边，并且将代表决策属性的边都去掉。这样在方法 1 构造的经典多重完全图的基础上得到了经典多重完全 k 部图。新得到的图代表了决策属性值不同的样本点之间的关系，忽略了决策属性值相同的样本点之间的关系，在一定程度上对决策表所包含的关系进行了约简。根据文献[14]中的研究，对属性求核和约简来说，该操作是合适的。

例 2　构造由决策表 1 中的决策属性作为分部条件所确定的经典多重完全多部图，如图 2 所示。

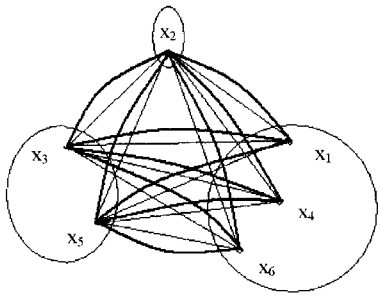

图 2　基于表 1 的六点经典二重完全三部图 $K(1,2,3;2)$

4.3　基于决策表的加权多重完全多部图

定义 11　如果图 G 的每一条边 e 都相应的有一个数 l_e 称为边 e 的权，则 G 连同它边上的权称为加权图，也称为有权图；当图 G 为 m 部图时称为加(有)权 m 部图；当图 G 为多重 m 部图时称为加(有)权多重 m 部图；当图 G 为多重完全 m 部图时称为加(有)权多重完全 m 部图；相应地把加权的 s 重完全 m 部图称为加(有)权 s 重完全 m 部图；加权的经典 s 重完全 m 部图称为加(有)权经典 s 重完全 m 部图。

方法 3　依照条件属性作为加权条件构造加权多重完全多部图

我们的目标是进行决策表的属性约简，为了达到这一目标，将定义 11 中的权值进行扩展，权值由数扩展为集合，具体操作如下。

(1)将经典多重完全多部图进行加权处理，每条边的权值为一个集合 $\{i,h_{i1},h_{i2}\}(i=1,2,\cdots,r)$，这里将图论中权的概念由数字推广到集合。该集合中 i 表示此边连接的两个样本点的第 i 个条件属性，h_{i1},h_{i2} 表示两个样本点在 C_i 属性上的取值，这样就得到了加权经典多重完全多部图。

(2)由于粗糙集研究的依赖关系主要为相异关系，因此对上述加权经典多重完全多部图进行进一步处理，去掉权值数组中属性值相同的边，就得到了非经典加权多重完全多部图。

(3)经过上步处理后的权值数组中各属性数值均不同，且粗糙集属性约简对条件属性的具体取值不做要求，所以可以进一步简化图的权，边的权变为 $\{i\}$，去掉两个样本点在该属性的具体取值。

(4)再进一步将任意两点间的边都用一条边表示，边上的权用原来两点间的所有边上的权集合来表示。这样得到的图就能表示决策表中决策属性值不同的样本点之间的相异性。

例 3　构造由决策表 1 中的决策属性作为分部条件，条件属性作为加权条件所确定的加权多重完全多部图，如图 3 所示。

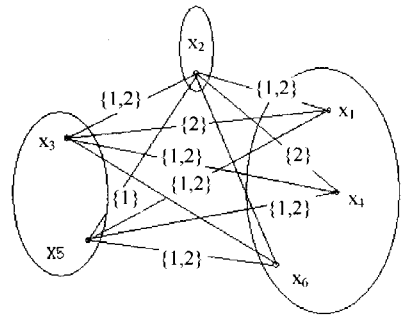

图 3　基于表 1 的加权六点非经典二重完全三部图 $K(1,2,3)$

4.4　基于决策表的加权多重完全多部图的性质

定义 12　如果在多重多部图中，不同部的任意两点间均有边相连，则称该多重多部图是完全的。反之是不完全的。

定义 13　在多重完全多部图中，如果某两点之间的重边个数为 1，即两点间只有一条边，则此边称为多重多部完全图的单边。

定义 14　在多重多部图中，如果去掉某条边后不影响图的完全性，则称该边为可去边。反之为不可去边。

相应地可以得到可去边类的定义。

定义 15　在多重完全多部图中，如果去掉某类边(具有某种性质的所有的边的集合)后不影响图的完全性，则该类边为可去边类，反之为不可去边类。

定理 2　多重完全多部图中，一条边为不可去边的充要条件为：该边为多重图的单边。

证明：充分性显然。

(必要性)由定义 14 可知，如果多重完全多部图 G 中的边 e 为不可去边，则去掉边 e 后的图 $G-e$ 不能保持完全性，因此 $G-e$ 中一定存在着两点 V_i,V_j 属于不同的部且 V_i,V_j 间无任何边相连。由于去掉边 e 之前 G 是一个完全图，因此边 e 所连接的两点就是 V_i,V_j，这两点间只有 e 这一条边，由定义 13 可知，该边为单边。

将定理 2 推广，可得定理 3。

定理 3　多重完全多部图 G 中，某一类边 E_s 为不可去边类的充要条件为：G 中存在两点 V_i,V_j，这两点间所连的边集合是 E_s 的子集合。

证明：(充分性)如果 G 中存在两点 V_i,V_j，这两点间所连的边集合是某一类边 E_s 的子集合，则将 E_s 从 G 中都去掉后，连接 V_i,V_j 两点的所有边必然都被去掉，因此 V_i,V_j 间将无边相连，此时 $G-E_s$ 不是完全图，所以 E_s 为不可去边类。

(必要性)如果某一类边 E_s 为不可去边类，则由定义 17

知 $G-E_s$ 不是完全图。因此在 $G-E_s$ 中必然存在两点 V_i，V_j，这两点间无任何边相连。由于 G 是一个完全图，因此 V_i，V_j 间在 G 中连的边都被去掉了，所以 V_i，V_j 间所连的边集合是 E_s 的子集合。

由定理 3 可以直接得到定理 4。

定理 4 多重完全多部图 G 中，某一类边 E_s 为可去边类的充要条件为：G 中任意两点 V_i，V_j $(i\neq j)$，两点间所连的边集合不是 E_s 的子集合。

定理 5 在决策表信息系统中，某些属性的集合可以约简的充分必要条件是：该决策表信息系统构成的加权多重完全多部图 G 的邻接矩阵中的所有非空元素都不是这个属性集合的子集合。

定理 6 在决策表信息系统中，某些属性的集合不可以约简的充分必要条件是：该决策表信息系统构成的加权多重完全多部图 G 的邻接矩阵中存在的非空元素是这个属性集合的子集合。

结合粗糙集知识，可得定理 7。

定理 7 粗糙集属性约简相当于对图进行去边操作，约简一个条件属性相当于去掉代表这个条件属性的一类边。加权多重完全多部图中如果出现单边，则该单边为不可去边，即该边代表的条件属性不可约简。如果决策表信息系统中存在可约简属性，则此单边所代表的属性为核属性之一。

证明：由前面定义可知，给定一个决策系统，按本文方法构造出来的加权多重完全多部图是唯一确定的，即对决策系统的操作相当于对图的操作。在图的构造中，某类边的集合就代表某一条件属性，所以去边类的操作就代表决策表的属性约简，反之亦然。如果图中存在单边，则该边为不可去边，该边所属的一类边为不可去边类，该边类所表示的属性为不可约简属性。若所有的属性边类均存在单边，则该决策表信息系统在绝对意义上不可约简，即没有核属性。如果决策表信息系统可约简，则该单边所表示的条件属性为核属性之一。

SKOWRON 的经典算法中，未能对差别矩阵中出现单属性（即核属性）这一性质给出合理的证明，叶存毅教授对此进行了说明[14]，本文给出了明确的解释。文献[14]中提出了决策表冲突的情况，本文的决策表中不能出现冲突样本，如果出现，则该决策表信息系统不能构成完全多部图（两不同部之间的样本点无边相连）。应用本方法，可以很直观地解决如上问题。

4.5 基于决策表的加权多重完全多部图的邻接矩阵

由邻接矩阵的定义可推导出基于决策表的加权多重完全多部图的邻接矩阵。该邻接矩阵为一分块对称阵，矩阵的元素表示决策表中各个样本点之间的差异。对决策系统 $S=(U,C,D,V,f)$ 来说，样本空间中样本点决策属性值为 1，2，…，k 的个数分别为 $n_1,n_2,\cdots,n_k$，$n_1+n_2+\cdots+n_k=n$。则该邻接矩阵表示为：

$$\overbrace{\begin{pmatrix} \phi_{n_1\times n_1} & A_{n_1\times n_2} & A_{n_1\times n_3} & \cdots & A_{n_1\times n_k} \\ A^{\mathrm{T}}_{n_1\times n_2} & \phi_{n_2\times n_2} & A_{n_2\times n_3} & \cdots & A_{n_2\times n_k} \\ A^{\mathrm{T}}_{n_1\times n_3} & A^{\mathrm{T}}_{n_2\times n_3} & \phi_{n_3\times n_3} & \cdots & A_{n_3\times n_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A^{\mathrm{T}}_{n_1\times n_k} & A^{\mathrm{T}}_{n_2\times n_k} & A^{\mathrm{T}}_{n_3\times n_k} & \cdots & \phi_{n_k\times n_k} \end{pmatrix}}^{n=n_1+n_2+\cdots+n_k\text{的分块矩阵}}$$

得到的矩阵为对称阵，对角线为空集合的方阵，矩阵 $A_{n_i\times n_j}$ 表示 n_i 部与 n_j 部的边的权集合。

4.6 加权多重完全多部图中边的合并和点的收缩

4.6.1 边的合并

如果两点间连接有 k 条边，可以将这 k 条边合并为一条 k 权边，合并后的新边的权用原来各个边的权的并的集合来表示。例如：点 X_1，X_2 为加权多重完全多部图中不同部的两点，两点间有边 e_1,e_2,e_3，权值分别为 i,j,k，则 X_1，X_2 之间的边合并后为一条多权边 e_{12}，权值为 $\{i,j,k\}$，也可以表示为 $i\cup j\cup k$。

定理 8 在加权多重完全多部图中，任意两个不同部的点 X_i 和 X_j 之间有且只有一条多权边 e_{ij}，同部的点之间没有边连接。

证明：由定义 12 可得。

4.6.2 点的收缩

加权多重完全多部图的两点收缩操作，就是合并与这两点相关联的所有多权边。

合并的方法：

(1)当两条多权边恰有一个公共顶点，且另外的两个顶点分别为要收缩的两点时，将这样的两条多权边合并成一条新的多权边，新边的权值按以下规则得到：将两条多权边的权值化为范式形式，进行 $\cap$ 操作后得到新的权值，仍用集合形式表示。“,”表示 $\cup$，“$*$”表示 $\cap$。例如，$\{a,\{b*c\}\}$ 可表示为 $a\cup(b\cap c)$，$a\cap(b\cup c)$ 可以表示为 $\{a*\{b,c\}\}$。

(2)收缩的这两点之间的边成为收缩后的新点自身的多权环。

(3)其他的边依次保留，收缩前与收缩的这两点相连接的在边收缩后都连接到收缩后的新点上。

由于合并后的多权边表示的是两点间的特定关系，单纯地合并多权边将不再有图论的意义。例如，将边 e_{12} 和边 e_{34} 进行合并，合并后的边将无法表示。为了使多权边合并后的结果仍有图论意义，将不再单独进行边的合并，而是按照上述方法进行点的收缩操作。

对决策表的加权多重完全多部图依次进行点的收缩，最后将图收缩为两个复合点，两点间的权值就表示整个图的边集合的合并结果，也就是粗糙集决策表的属性约简。

4.7 基于属性置信度的近似计算方法

当决策表信息系统构造的图中所有条件属性表示的边类均不可进行去边操作，即所有的条件属性均存在单边，且样本空间很大时，可以采用一种近似计算方法，分别统计各个属性所代表的单边数 NUM_i。

定义 16 属性置信度 $Confidence(C_i)=NUM_i/\sum_{j=1,2,\cdots k}NUM_j$，其中 NUM_i 表示第 i 个属性的单边个数。

根据属性置信度的定义，在近似计算中可以进行快速属性约简。例如，某一属性的属性置信度$\leqslant$1%时，在约简置信度 99%的条件下，可将该属性约简。

4.8 算法描述

由于图的计算可以转化为对图的邻接矩阵进行计算，因此我们对邻接矩阵进行处理。算法如下：

输入信息：决策系统 $S=(U,C,D,V,f)$，其中 $U=\{x_1,x_2,x_3,\cdots,x_n\}$ 为样本空间，$C=\{c_1,c_2,\cdots,c_r\}$ 为条件属性集合，$D=\{1,2,\cdots,k\}$ 为决策属性集合。

(1)对样本点进行遍历，按照决策属性值将样本点分为 k 类，每类分别包含 $n_1, n_2, \cdots, n_k$ 个点，形成图的点集合 $V(n_1, n_2, \cdots, n_k)$。

(2)依照本文第 4.1 节－第 4.5 节的方法对样本点进行处理，得到分块矩阵 A_k。

(3)判断其是否符合 4.7 节中节属性置信度近似计算的条件，如符合，按属性置信度近似计算。否则，依照 4.6 节的方法进行点的收缩。从 A_k 的第一个非空值表示的样本点开始按列向后依次进行点收缩操作，当收缩为一列后再按行收缩，最后收缩为两个样本点。在此过程中，将 A_k 中表示单边的元素记入核属性集合。

(4)收缩完毕，计算两个样本点之间的权值就得到粗糙集的属性约简。

5 复杂度分析

按照经典的粗糙集属性约简方法，时间和空间复杂度都为 $|C||M|^2$，按照本文的方法可以大幅降低时间和空间复杂度。设 $M=\sum_{i=1}^{k} N_i$，表示根据决策属性分成的 k 块，空间和时间复杂度降为 $|C|\|\sum_{i,j=1,2,\cdots,k,i\neq j}|N_i|*|N_j|\|$，当 N_i 中有一项特别大时，复杂度近似降为 $k|C||M|$。

在一定置信度条件约束下可近似地进行去边即约简操作的时间复杂度为 $\sum_{i,j=1,2,\cdots,k,i\neq j}|N_i|\times|N_j|$。实际应用中，如果用户有特殊要求，应根据用户的关注度在各个属性上分配用户关注度并与置信度结合，进而得出用户满意的置信度，再按照用户关注来进行属性约简。

6 实例分析和结论

为了实验算法的正确性和性能，选取表 2 中所列的决策表和 UCI 机器学习数据库中 2 个具有离散属性和决策属性的典型数据库在 Windows XP 2002 版，CPU2.6Hz，内存 2G，Matlab 上编程实现。分别采用基于互信息的 MIBARK 约简算法[21](简称 M)、基于条件信息熵的 CEBARKNC 约简算法[5](简称 C)和本文的算法(简称 T)在 3 个不同的数据上进行比较。由于样本 Lung Cancer 中仅有一个丢失信息，对计算结果影响不大，因此用左右平均补全操作，使决策表完备。M 算法先计算决策表的核，然后对非核属性进行约简。C 算法直接在原有条件属性集上进行约简。本文在约简的同时计算决策表的核，并且引入属性约简置信度这一概念，在样本很大的情况下，可以在一定置信度情况下快速地计算约简。例如对样本集合 Car Evaluation 的计算就用到了这个方法，该样本集合中有 1728 个样本，6 个条件属性，经计算，6 个条件属性在构造加权多重多部图后均存在单边，即在保持决策表不冲突的条件下 6 个条件属性都不可约简。对比 M 和 C 算法可以发现，两种算法都没能对条件属性进行约简。通过属性置信度的计算，该样本集合的 6 个条件属性中有一个属性置信度为 3.3%，按照本文方法，在 95%的约简置信度下该属性可约简。

表 2 关于气象信息的决策表系统

U	条件属性				决策属性
	Outlook(a1)	Temperature(a2)	Humidity(a3)	Windy(a4)	(d)
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

从表 3 可以得出以下结论：

(1)M 算法和 T 算法均计算决策表的核属性，因此在有核属性的情况下，M 算法和 T 算法明显优于 C 算法。当无核时，M 算法没有良好的处理办法，因此在表现上不如 C 算法和 T 算法。

(2)在样本空间很大且所有属性均绝对不可约简时，M 算法和 C 算法均未提出一种可行的解决办法。T 算法在一定置信度的情况下给出了约简。

(3)在算法执行时间上，T 算法明显优于前两种算法，尤其在大样本处理上有较好的应用。

实际测试结果如下：N=(样本个数)，M=(条件属性个数)，C=(核属性个数)，r=(约简后的条件属性个数)，con=约简置信度，用%表示，$C1$=算法计算的属性核，T=算法时间，用秒表示。

表 3 样本测试结果

数据集	N	M	C	M算法				C算法				T算法			
				r	Con	C1	t	r	con	C1	t	r	con	C1	t
气象信息	14	4	2	3	100	2	0.15	3	100	0	0.21	3	100	2	0.09
Car Evaluation	1728	6	0	6	100	0	4.56	6	100	0	7.05	5	95	0	1.38
Lung Cancer	32	56	0	4	100	0	3.05	6	100	0	2.07	4	100	0	1.01

结束语 现实中的各种科学研究方法之间存在一定的联系。本文从图论角度对粗糙集决策系统进行处理，研究了两者之间的联系，给出了一个基于图论的粗糙集属性约简办法并论证了该方法的可行性。理论分析与实验结果表明，基于图论的属性约简算法定义严谨，结果直观，可以有效降低空间和时间复杂度，平均复杂度降低为 $|C|\|\sum_{i,j=1,2,\cdots,k,i\neq j}|N_i|\times|N_j|\|$。在大样本空间中一定置信度要求的情况下，可以进一步将复杂度降为 $\sum_{i,j=1,2,\cdots,k,i\neq j}|N_i|\times|N_j|$。

参 考 文 献

[1] Pawlak Z. Rough set[J]. Communication of the ACM, 1995, 38(11): 89-95

[2] Pawlak Z. Rough set theory and its application to data analysis [J]. Cybernetics and systems, 1998, 9(5): 661-668

[3] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[M]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, 1992: 331-362

(下转第 272 页)

结束语 为了充分利用共享的 Web 服务，如何将现有的 Web 服务连接与合并起来，生成复杂的复合 Web 服务，从而提供更强大、更完整的商业功能，成为现有服务集合增加价值的关键。本文在前人研究的基础上，提出了 Web 服务动态 QoS 模型。与已有的其它 Web 服务 QoS 模型相比，动态 QoS 模型考虑了服务的动态性以及服务组合中的网络特性，使得服务的 QoS 属性更加真实地反映 Web 服务当前的状态。另外，本文将用户约束引入到组合过程中，最大限度地避免了服务重计算现象的发生。但是，Web 组合服务自适应性方面和 Web 服务的动态监控方面还有许多要研究的工作，如何在组合算法中增加自适应性，以及如何更准确地监控 Web 服务，都是今后工作的重点。

参 考 文 献

[1] 王勇，代桂平，侯亚荣. 信任感知的聚合动态选择方法[J]. 计算机学报，2009，32(8)：1668-1675

[2] Aggarwal R，Verrna K，Miller J，et al. Constraint driven Web service composition in METEOR-S[C]// Proc of the IEEE International Conference on Services Computing. Shanghai，China：IEEE，2004：23-30

[3] Zeng L Z，Benatallah B，Ngu A H H，et al. QoS-aware middleware for Web services composition[J]. IEEE Trans on Software Engineering，2004，30(5)：31-327

[4] 杨胜文，史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. 计算机学报，2005，28(4)：589-594

[5] 李曼，王大治，杜小勇，等. 基于领域本体的 Web 服务动态组合[J]. 计算机学报，2005，28(4)：644-650

[6] Hashemian S V. A Graph-based Approach to Web Services Composition[C]// Proc of the 2005 Symposium on Applications and the Internet Society. Trento，Italy：IEEE/IPSJ，2005：183-189

[7] 刘书雷，刘云翔，张帆，等. 一种服务聚合中 QoS 全局最优服务动态选择算法[J]. 软件学报，2007，18(3)：646-656

[8] Peer J. Bringing Together Semantic Web and Web Services[C]// Proc of the 1st International Semantic Web Conference(ISWC 2002). London，UK：Springer-Verlag，Lecture Notes in Computer Science，Vol. 2342，2002：279-291

[9] Zhang Lian-jie，Li Bing，Chao Tian，et al. On demand Web services-based business process composition[C]// Proc of the IEEE International Conference on System，Man，and Cybernetics. Washington，USA：IEEE，2003：4057-4064

[10] 代钰，杨雷，张斌，等. 支持组合服务选取的 QoS 模型及优化求解[J]. 计算机学报，2006，29(7)：1167-1178

[11] Martin D，Burstein M，Denker G，et al. OWL-S：Semantic markup for Web services[EB/OL]. http://www. ai. sri. com/daml/services/owl-s/1. 2/，2006

[12] Harmelen F，Hendler J，Horrocks I，et al. OWL Web Ontology Language Reference[EB/OL]. http://www. w3. org/tr/owl-ref，2004

[13] 张成文，苏森，陈俊亮. 基于遗传算法的 QoS 感知的 Web 服务选择[J]. 计算机学报，2009，29(7)：1029-1037

[14] 刘峰，谭庆平，杨艳萍. 基于图论的 Web 服务合成算法[J]. 华中科技大学学报：自然科学版，2005，12(3)：202-204

[15] Zhang R，Arpinar B，Aleman Meza B. Automatic composition of semantic Web services[C]// Proceedings of International Conference on Web Services. Las Vegas，USA，2003：38-41

(上接第 254 页)

[4] Hu X H，Cercne N. Learning in relational databases：A rough set approach[J]. International Journal of Computational Intelligence，1995，11(2)：323-338

[5] Wang G Y，Yu H，Yang D C. Decision table reduction based on conditional information entropy[J]. Chinese Journal of Computer，2002，25(7)：759-766

[6] Wang G Y. Calculation method for core attributions of decision table[J]. Chinese Journal of Computer，2003，26(5)：611-615

[7] Wang G Y. Rough reduction in algebra view and information view [J]. International Journal of Intelligent System，2003，18(4)：679-688

[8] Zhang W X，Mi J S，Wu W Z. Knowledge reductions in inconsistent information systems[J]. Chinese Journal of Computer，2003，26(1)：12-18

[9] 腾书华，魏荣华，孙即样，等，基于不可区分度的启发式快速完备约简算法[J]. 计算机科学，2009，36(8)：196-199

[10] 代劲，何中市. 基于云模型的决策表规则约简[J]. 计算机科学，2010，37(6)：265-267

[11] 徐久成，史进玲，孙林. 一种基于相对粒度的决策表约简算法[J]. 计算机科学，2009，36(3)：205-207

[12] Xu Z Y，Yang B R，Song W. Comparative study of different attribute reduction based on decision table[J]. Chinese Journal of Electronics，2006，15(4A)：953-956

[13] Xu Z Y. Comparative Research of Different Attribute Reduction Definitions[J]. Journal of Chinese Computer Systems，2008，5：848-853

[14] 叶东毅，陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报，2002，30(7)：1086-1088

[15] 赵军，王国胤，吴中福，等. 一种高效的属性核计算方法[J]. 小型微型计算机系统，2003，24(11)：1950-1953

[16] 聂红梅，周家庆. 一个新的差别矩阵及其求核方法[J]. 四川大学学报：自然科学版，2007，44(2)：277-283

[17] 周创德，田卫东. 基于约束函数的差别矩阵及其求核算法[J]. 计算机工程，2008，34(15)：60-63

[18] 葛浩，杨传健，李龙澍. 基于可分辨矩阵的快速求核算法[J]. 计算机工程与设计，2009，30(5)：1021-1024

[19] 汪小燕. 一种改进的差别矩阵及其求核方法[J]. 安徽工业大学学报：自然科学版，2009，26(1)：86-90

[20] Bondy J A，Murty U S R. Graph Theory[S]. ISBN：978-1-84628-969-9，Springer

[21] 苗夺谦，胡桂荣. 知识约简的一种启发算法[J]. 计算机研究与发展，1999，36(6)：681-684