# 集值决策表基于限制相容关系的分配约简

乔全喜[1,2] 秦克云[1]

(西南交通大学数学系 成都 610031)[1] (河南理工大学数学与信息科学学院 焦作 454000)[2]

**摘 要** 讨论集值决策表基于限制相容关系的分配约简方法;分配约简是保持所有决策类的粗糙上近似不变的极小属性子集;定义了分配协调集并给出了分配协调集的3个充要条件;通过实例说明该算法能够得到集值决策表的分配约简。

**关键词** 粗糙集,集值决策表,对称限制,相容关系,分配约简

**中图法分类号** TP18,TP301 **文献标识码** A

## Distributive Reduction of Set-valued Decision Table Based on Restriction Tolerance Relation

QIAO Quan-xi[1,2] QIN Ke-yun[1]

(Department of Mathematics, Southwest Jiaotong University, Chengdu 610031, China)[1]

(School of Mathematics & Information Science, Henan Polytechnic University, Jiaozuo 454000, China)[2]

**Abstract** This paper discussed distributive reduction of set-valued decision table based on symmetry restriction tolerance relation. It is the minimal attribute subset to remain all rough upper approximation of decision class unchanged. The authors defined a distributive compatible set and gave three necessary and sufficient conditions. The example shows that this algorithm can obtain the distributive reduction of set-valued decision table.

**Keywords** Rough set, Set-valued decision table, Symmetry restriction, Tolerance relation, Distributive reduction

粗糙集理论[1,2]是由波兰数学家Pawlak提出的一种处理不精确、模糊和不确定知识的数学工具。它依据对象之间的不可分辨性将论域中的对象聚类成基本知识,利用基本知识,通过上、下近似运算来描述数据对象的不确定性,从而导出概念的分类或决策规则。目前,粗糙集理论已广泛应用于属性约简[3-9]、规则提取[10,11]、模式识别与分类[12,13]等人工智能领域。粗糙集理论的基本思想是通过已有数据构成的信息系统中的属性特征将数据进行分类,然后引入上近似集和下近似集来分别刻画知识的可能性和确定性,并由决策类的下近似集归纳出确定决策规则。然而,经典粗糙集模型对知识的表达是建立在完备信息系统中的属性所诱导的等价关系基础上的,但在实际应用中,由于问题的复杂性,通常人们得到的数据是不精确和不完善的。如果信息系统中某个对象的属性是未知的,则称这种信息系统为不完备信息系统;如果一些对象的某个属性值不是取一个值,也不是取空集,而是取几个值,这种信息系统我们称为集值信息系统。近年来,一些学者对集值信息系统的知识发现与属性约简作了许多研究工作。文献[3]基于相容关系讨论了集值信息系统的最优选取与最大分布约简;文献[4]基于相容关系形成的极大相容类进一步讨论了集值信息系统的规则获取与极大相容类的相对约简;文献[8]在集值信息系统上定义了一种变精度相容关系,讨论了这种关系下的属性约简方法;文献[5]中比较详细地讨论了集值有序信息系统的优势粗糙集方法,对集值信息系统给出了更细致的刻画。文献[16]提出了描述子的概念,构建了一种新的粗糙集模型定义方法,并由协调描述子导出最优确定性决策规则,由非协调描述子导出最优结合决策规则,讨论了描述子的相对约简以及一些不确定度量方法。文献[17]基于描述子概念讨论了不完备信息系统上可信决策规则的提取,并利用区分矩阵法给出了向上与向下描述子的相对约简,得到了不完备决策信息系统的最优可信决策规则。文献[18,19]基于限制相容关系和对称限制相容关系分析比较了相应的粗糙集模型与已有粗糙集扩展模型之间的关系。文献[20]讨论了集值信息系统基于限制相容关系的属性约简方法,给出了相似水平核心属性的特征。本文针对对称限制相容关系,研究了集值决策表的分配约简的理论与方法,并通过实例说明了该算法能够得到集值决策表的分配约简。

## 1 集值信息系统中的对称限制相容关系

设$U$是对象构成的非空有限集合,以下称为论域。按照Pawlak粗糙集理论,知识是对对象进行分类的能力。因此,$U$上的知识可以形式化地表示为$U$的划分,$U$上的知识库可以通过信息系统进行描述。

**定义1**[21] 一个信息系统是一个三元组$S=(U,A,F)$,其中

(1)$U=\{x_1,\cdots,x_n\}$为对象集,每个$x_i(1\leqslant i\leqslant n)$称为一个对象;

到稿日期:2012-09-15 返修日期:2012-12-24 本文受国家自然科学基金项目(60875034)资助。

乔全喜 博士生,主要研究方向为智能控制及应用,E-mail:qiaoqx@hpu.edu.cn;秦克云 博士,教授,主要研究方向为代数逻辑与智能信息处理。

(2)$A=\{a_1,\cdots,a_m\}$为属性集，每个$a_j(1\leqslant j\leqslant m)$称为一个属性；

(3)$F=\{f_l\mid 1\leqslant l\leqslant m\}$为对象属性值映射，其中$f_l:U\rightarrow V_l$，$V_l$是属性$a_l$的值域。

设$S=(U,A,F)$是信息系统。对于任意$a\in A$，由$a$可以确定论域$U$上的一个等价关系，称其为由$a$确定不可区分关系，记为$ind(a)$，定义如下：对于任意$x,y\in U$，$(x,y)\in ind(a)$当且仅当：$f_a(x)=f_a(y)$。

因此，信息系统本质上就是知识库，每一个属性决定一个知识。

集值信息系统是信息不确定与不完整的一种反映形式，具体可以定义如下：

**定义2**[21] 称三元组$S=(U,A,F)$为集值信息系统，其中$U=\{x_1,\cdots,x_n\}$为对象集，$A=\{a_1,\cdots,a_m\}$为属性集，$F=\{f_l\mid 1\leqslant l\leqslant m\}$为对象属性值映射，其中$f_l:U\rightarrow P_0(V_l)$，$V_l$是属性$a_l$的值域，$P_0(V_l)$表示$V_l$的所有非空子集构成的集合。若$S=(U,A,F)$是一个集值信息系统，$d$是论域$U$上一个决策属性，则称$S=(U,A\cup\{d\},F)$是一个集值决策表。

**定义3**[3] 设$(U,A,F)$是集值信息系统，任意属性子集$B\subseteq A$，定义二元关系：

$$R_B^{\cap}=\{(x,y)\in U^2\mid f_a(x)\cap f_a(y)\neq\emptyset(\forall a\in B)\}$$

$R_B^{\cap}$仅满足自反性与对称性，称为相容关系。

**定义4**[18] 设$(U,A,F)$是集值信息系统，任意属性子集$B\subseteq A$，定义二元关系：

$$S_a^{\alpha}=\left\{(x,y)\in U^2\,\middle|\,\frac{|f_a(x)\cap f_a(y)|}{|f_a(x)|}\geqslant\alpha\right\}$$

我们称$S_a^{\alpha}$为$\alpha$度限制相容关系。

任意$B\subseteq A$，有$S_B^{\alpha}=\bigcap\limits_{a\in B}S_a^{\alpha}$。

设$S=(U,A,F)$是集值信息系统，$B\subseteq A$，$\alpha\in(0,1]$，令$[x]_B^{\alpha}=\{y\in U\mid(x,y)\in R_B^{\alpha}\}$，称$[x]_B^{\alpha}$为$x$相对于属性集$B$的$\alpha$相容类。特别地，当$B=\{a\}$只含有一个属性时，分别将$R_B^{\alpha}$和$[x]_B^{\alpha}$记为$R_a^{\alpha}$和$[x]_a^{\alpha}$。

下面的定理给出了对称限制相容关系的基本性质。

**定理1**[19] 设$(U,A,F)$是集值信息系统，$R_B^{\alpha}$是如上定义的二元关系，则有

(1)$R_B^{\alpha}$是自反、对称二元关系；

(2)当$B_1\subseteq B_2\subseteq A$时，$\forall\alpha\in(0,1]$，$R_A^{\alpha}\subseteq R_{B_2}^{\alpha}\subseteq R_{B_1}^{\alpha}$；

(3)当$B_1\subseteq B_2\subseteq A$时，$\forall\alpha\in(0,1]$，$[x_i]_A^{\alpha}\subseteq[x_i]_{B_2}^{\alpha}\subseteq[x_i]_{B_1}^{\alpha}$；

(4)$\{[x_i]_B^{\alpha}:x_i\in U\}$是$U$的一个覆盖。

## 2 集值决策表基于对称限制相容关系的分配约简

设$S=(U,A\cup\{d\},F)$为集值决策表，其中$d$为单值决策属性，即$U/d=\{[x]_d;x\in U\}=\{D_1,\cdots,D_r\}$形成论域$U$的划分，其中等价类称为决策类。对于$\alpha\in(0,1]$，$B\subseteq A$，称$d_B^{\alpha}:U\rightarrow P(V_d)$为$S$的由$B$确定的$\alpha$水平决策函数，定义为：

$$d_B^{\alpha}(x)=d([x]_B^{\alpha})=\{d(y)\mid y\in[x]_B^{\alpha}\}$$

即$d_B^{\alpha}(x)$是$[x]_B^{\alpha}$中对象的属性值构成的集合。

**定义5** 如果$B\subseteq A$满足$d_B^{\alpha}=d_A^{\alpha}$，则称$B$是$S$的$\alpha$分配协调集，$\alpha$分配协调集族中的极小集称为$S$的$\alpha$分配约简。

**定理2** $B$是$\alpha$分配协调集的充分必要条件是：任意$j\leqslant r$，有$\overline{R_B^{\alpha}}(D_j)=\overline{R_A^{\alpha}}(D_j)$。

证明(必要性)：设$B$是$\alpha$分配协调集。对于$\forall j\leqslant r$，由$[x]_A^{\alpha}\subseteq[x]_B^{\alpha}$可得

$$\overline{R_A^{\alpha}}(D_j)\subseteq\overline{R_B^{\alpha}}(D_j)$$

若$x\in\overline{R_B^{\alpha}}(D_j)$，不妨设$D_j=[y]d$，则$[x]_B^{\alpha}\cap D_j\neq\emptyset$，从而$d(y)\in d([x]_B^{\alpha})=d_B^{\alpha}(x)=d_A^{\alpha}(x)$

即$[x]_A^{\alpha}\cap D_j\neq\emptyset$，$x\in\overline{R_A^{\alpha}}(D_j)$，故$\overline{R_B^{\alpha}}(D_j)\subseteq\overline{R_A^{\alpha}}(D_j)$。于是有$\overline{R_B^{\alpha}}(D_j)=\overline{R_A^{\alpha}}(D_j)$。

(充分性)：对于任意$x\in U$，由$[x]_A^{\alpha}\subseteq[x]_B^{\alpha}$可得$d_A^{\alpha}(x)\subseteq d_B^{\alpha}(x)$。若$d$的属性值$i\in d_B^{\alpha}(x)$，则存在$y\in[x]_B^{\alpha}$使得$d(y)=i$，于是$[x]_B^{\alpha}\cap[y]_d\neq\emptyset$，从而$x\in\overline{R_B^{\alpha}}([y]_d)=\overline{R_A^{\alpha}}([y]_d)$，即$[x]_A^{\alpha}\cap[y]_d\neq\emptyset$。故

$$i=d(y)\in d([x]_A^{\alpha})=d_A^{\alpha}(x)$$

于是有

$$d_B^{\alpha}(x)\subseteq d_A^{\alpha}(x)$$

故$d_A^{\alpha}(x)=d_B^{\alpha}(x)$。

由$x$的任意性可知，$B$是$S$的$\alpha$分配协调集。

根据定理2知，$\alpha$分配约简是保持所有决策类的$\alpha$粗糙上近似不变的极小属性子集。

对于集值决策表$S=(U,A\cup\{d\},F)$，$B\subseteq A$，$x\in U$，令：

$$\delta_B(x)=\{[y]_d\mid[x]_B^{\alpha}\cap[y]_d\neq\emptyset\}$$

则有

**定理3** $B$是$\alpha$分配协调集的充分必要条件是：对于任意$x\in U$，有$\delta_B(x)=\delta_A(x)$。

证明：设$B$是$\alpha$分配协调集。对于任意$x,y\in U$，若$[y]_d\in\delta_B(x)$，则$[x]_B^{\alpha}\cap[y]_d\neq\emptyset$，从而$x\in\overline{R_B^{\alpha}}([y]_d)$，于是有$x\in\overline{R_A^{\alpha}}([y]_d)$，$[x]_A^{\alpha}\cap[y]_d\neq\emptyset$，即$[y]_d\in\delta_A(x)$。故$\delta_B(x)\subseteq\delta_A(x)$。同理可证$\delta_A(x)\subseteq\delta_B(x)$。

反之，设对于任意$x\in U$，有$\delta_B(x)=\delta_A(x)$，则当$x\in\overline{R_B^{\alpha}}([y]_d)$时有$[x]_B^{\alpha}\cap[y]_d\neq\emptyset$，故$[y]_d\in\delta_B(x)$。于是$[y]_d\in\delta_A(x)$，$[x]_A^{\alpha}\cap[y]_d\neq\emptyset$，即$x\in\overline{R_A^{\alpha}}([y]_d)$，即$B$是$\alpha$分配协调集。

注：有些文献将定理3作为分配约简的定义，其与本文的定义是等价的。

**定理4** $B$是$\alpha$分配协调集的充分必要条件是：任意$(x,y)\in D^*$，$d^{\alpha}(x,y)\neq\emptyset$时，有$B\cap d^{\alpha}(x,y)\neq\emptyset$。其中$d^{\alpha}(x,y)=\{a\in A\mid(x,y)\notin R_a^{\alpha}\}$是能区分对象$x$与$y$的条件属性构成的集合，$D^*=\{(x,y)\mid x,y\in U,d(y)\notin d_A^{\alpha}(x)\}$。

证明：设$B$是$\alpha$分配协调集，于是有$d_B^{\alpha}=d_A^{\alpha}$。若$(x,y)\in D^*$使得$d^{\alpha}(x,y)\neq\emptyset$，由$d(y)\notin d_A^{\alpha}(x)$可得$d(y)\notin d_B^{\alpha}(x)$，故$[y]_d\cap[x]_B^{\alpha}=\emptyset$，$y\notin[x]_B^{\alpha}$，故存在$b\in B$使得$(x,y)\notin R_b$，即$B\cap d^{\alpha}(x,y)\neq\emptyset$。

反之，对于任意$x\in U$，由$[x]_A^{\alpha}\subseteq[x]_B^{\alpha}$可得$d_A^{\alpha}(x)\subseteq d_B^{\alpha}(x)$。另一方面，对于$d$的任意属性值$i$，设$i\notin d_A^{\alpha}(x)$且$d(y)=i$。对于任意$u\in[y]_d$，$d(u)=d(y)=i\notin d_A^{\alpha}(x)$，故$(x,u)\in D^*$，从而由假设可得$B\cap d^{\alpha}(x,u)\neq\emptyset$，即存在$b\in B$使得$(x,u)\notin R_b$，从而$u\notin[x]_B^{\alpha}$。由$u$的任意性可得$[y]_d\cap[x]_B^{\alpha}=\emptyset$，即$i=d(y)\notin d([x]_B^{\alpha})=d_B^{\alpha}(x)$，故$d_B^{\alpha}(x)=d_A^{\alpha}(x)$。再由$x$的任意性可得$d_B^{\alpha}=d_A^{\alpha}$，故$B$是$\alpha$分配协调集。

**定理5**[9] $\Delta^*$的极小析取范式中的所有合取范式恰为$S$

的所有 $\alpha$ 分配约简，其中 $\Delta^* = \bigwedge_{(x,y)\in D^*} \vee d^\alpha(x,y)$。

需要注意的是，尽管 $R_B^\alpha$ 是对称的二元关系，但 $d(y)\notin d_A^\alpha(x)$ 与 $d(x)\notin d_A^\alpha(y)$ 一般说来不是等价的，见例 1。因此，$\alpha$ 分配约简的区分矩阵未必是对称矩阵。

例 1　考虑表 1 所列的集值决策表。

表 1　集值信息系统

| | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | d |
|---|---|---|---|---|---|---|
| $x_1$ | {2} | {1,2} | {1} | {2,3} | {1} | 1 |
| $x_2$ | {1,2} | {3} | {1,2} | {2} | {1,2} | 3 |
| $x_3$ | {1} | {2,3} | {2} | {1,2} | {1,2} | 2 |
| $x_4$ | {1} | {2} | {2} | {2} | {1} | 2 |
| $x_5$ | {3} | {2} | {1,2} | {1} | {3} | 3 |
| $x_6$ | {1,3} | {2} | {1,2} | {1,2} | {2,3} | 3 |
| $x_7$ | {2} | {1,2,3} | {1,2} | {2} | {1,3} | 1 |
| $x_8$ | {1} | {3} | {2} | {2} | {2} | 3 |

取 $\alpha=0.5$，经计算可得：

$d_A^{0.5}(x_1)=d([x_1]_A^{0.5})=d(\{x_1,x_7\})=\{1\}$

$d_A^{0.5}(x_2)=d([x_2]_A^{0.5})=d(\{x_2,x_3,x_8\})=\{2,3\}$

$d_A^{0.5}(x_3)=d([x_3]_A^{0.5})=d(\{x_2,x_3,x_4,x_6,x_8\})=\{2,3\}$

$d_A^{0.5}(x_4)=d([x_4]_A^{0.5})=d(\{x_3,x_4\})=\{2\}$

$d_A^{0.5}(x_5)=d([x_5]_A^{0.5})=d(\{x_5,x_6\})=\{3\}$

$d_A^{0.5}(x_6)=d([x_6]_A^{0.5})=d(\{x_3,x_5,x_6\})=\{2,3\}$

$d_A^{0.5}(x_7)=d([x_7]_A^{0.5})=d(\{x_1,x_7\})=\{1\}$

$d_A^{0.5}(x_8)=d([x_8]_A^{0.5})=d(\{x_2,x_3,x_8\})=\{2,3\}$

注意到 $d(x_2)=3\notin\{2\}=d([x_4]_A^{0.5})$，因此 $(x_4,x_2)\in D^*$，即 $x_4$ 与 $x_2$ 需要区分。经计算可得 $S_{x_4x_2}^{a_1}=0.5$，$S_{x_4x_2}^{a_2}=0$，$S_{x_4x_2}^{a_3}=0.5$，$S_{x_4x_2}^{a_4}=1$，$S_{x_4x_2}^{a_5}=0.5$，故有 $d^{0.5}(x_4,x_2)=\{a_2\}$。由于 $d(x_4)=2\in\{2,3\}=d([x_2]_A^{0.5})$，因此 $(x_2,x_4)\notin D^*$，故 $x_2$ 与 $x_4$ 不需要区分，此时令 $d^{0.5}(x_2,x_4)=\varnothing$。对于任意 $x,y\in U$，可同理计算 $d^{0.5}(x,y)$，得到下面的区分矩阵：

$$\begin{pmatrix}
\varnothing & \{a_2\} & \{a_1,a_3\} & \{a_1,a_3\} & \{a_1,a_4,a_5\} & \{a_1,a_5\} & \varnothing & \{a_1,a_2,a_3,a_5\} \\
\{a_2\} & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_2\} & \varnothing \\
\{a_1,a_3\} & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_1\} & \varnothing \\
\{a_1,a_3\} & \{a_2\} & \varnothing & \varnothing & \{a_1,a_4,a_5\} & \{a_5\} & \{a_1,a_2\} & \{a_2,a_5\} \\
\{a_1,a_4,a_5\} & \varnothing & \{a_1,a_5\} & \{a_1,a_4,a_5\} & \varnothing & \varnothing & \{a_1,a_2,a_4\} & \varnothing \\
\{a_1,a_5\} & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_1,a_2\} & \varnothing \\
\varnothing & \{a_2\} & \{a_1\} & \{a_1,a_2\} & \{a_1,a_2,a_4\} & \{a_1,a_2\} & \varnothing & \{a_1,a_2,a_5\} \\
\{a_1,a_2,a_3,a_5\} & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \{a_1,a_2,a_5\} & \varnothing
\end{pmatrix}$$

故区分函数为：

$$\begin{aligned}\Delta^* &= a_2\wedge(a_1\vee a_3)\wedge(a_1\vee a_4\vee a_5)\wedge(a_1\vee a_5)\wedge(a_1\vee a_2\vee a_3\vee a_5)\wedge(a_1\vee a_2\vee a_4\vee a_5)\wedge a_1\wedge a_5\wedge(a_1\vee a_2)\wedge(a_2\vee a_5)\wedge(a_1\vee a_2\vee a_4)\wedge(a_1\vee a_2\vee a_5)\\ &= a_1\wedge a_2\wedge a_5\end{aligned}$$

从而 $S$ 有唯一的 0.5 分配约简 $\{a_1,a_2,a_5\}$。

若取 $\alpha=0.7$，经计算可得：

$d_A^{0.7}(x_1)=d([x_1]_A^{0.7})=d(\{x_1\})=\{1\}$

$d_A^{0.7}(x_2)=d([x_2]_A^{0.7})=d(\{x_2\})=\{3\}$

$d_A^{0.7}(x_3)=d([x_3]_A^{0.7})=d(\{x_3\})=\{2\}$

$d_A^{0.7}(x_4)=d([x_4]_A^{0.7})=d(\{x_4\})=\{2\}$

$d_A^{0.7}(x_5)=d([x_5]_A^{0.7})=d(\{x_5\})=\{3\}$

$d_A^{0.7}(x_6)=d([x_6]_A^{0.7})=d(\{x_6\})=\{3\}$

$d_A^{0.7}(x_7)=d([x_7]_A^{0.7})=d(\{x_7\})=\{1\}$

$d_A^{0.7}(x_8)=d([x_8]_A^{0.7})=d(\{x_8\})=\{3\}$

于是，相应的区分矩阵为：

$$\begin{pmatrix}
\varnothing & A & A & A-\{a_5\} & A & A & \varnothing & A \\
A & \varnothing & A-\{a_5\} & A-\{a_4\} & \varnothing & \varnothing & A-\{a_3,a_4\} & \varnothing \\
A & A-\{a_5\} & \varnothing & \varnothing & A & A-\{a_4\} & A & A-\{a_1,a_3\} \\
A-\{a_5\} & A-\{a_4\} & \varnothing & \varnothing & A-\{a_2\} & A-\{a_2\} & A-\{a_4\} & \{a_2,a_5\} \\
A & \varnothing & A & A-\{a_2\} & \varnothing & \varnothing & A-\{a_3\} & \varnothing \\
A & \varnothing & A-\{a_4\} & A-\{a_2\} & \varnothing & \varnothing & A-\{a_3\} & \varnothing \\
\varnothing & A-\{a_3,a_4\} & A & A-\{a_4\} & A-\{a_3\} & A-\{a_3\} & \varnothing & A-\{a_4\} \\
A & \varnothing & A-\{a_1,a_3\} & \{a_2,a_5\} & \varnothing & \varnothing & A-\{a_4\} & \varnothing
\end{pmatrix}$$

故区分函数为：

$$\begin{aligned}\Delta^* &= (a_2\vee a_5)\wedge(a_1\vee a_2\vee a_3\vee a_4)\wedge(a_1\vee a_3\vee a_4\vee a_5)\\ &= (a_1\wedge a_2)\vee(a_1\wedge a_5)\vee(a_2\wedge a_3)\vee(a_2\wedge a_4)\vee(a_2\wedge a_5)\vee(a_3\wedge a_5)\vee(a_4\wedge a_5)\end{aligned}$$

从而 $S$ 有 7 个 0.7 分配约简：$\{a_1,a_2\}$，$\{a_1,a_5\}$，$\{a_2,a_3\}$，$\{a_2,a_4\}$，$\{a_2,a_5\}$，$\{a_3,a_5\}$ 和 $\{a_4,a_5\}$。

**结束语**　集值信息系统对象关于属性取集合值是信息不确定的一种反映形式。本文讨论集值决策表基于相对限制相容关系的分配约简方法，分配约简是保持所有决策类的粗糙上近似不变的极小属性子集。通过实例说明该算法能够得到集值决策表的分配约简。以后，我们将在本文的基础上进一步讨论集值决策表基于相对限制相容关系的正域约简方法。

## 参考文献


[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982(11): 341-356

[2] Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about Data[M]. Kluwer Academic Publishers, Boston, 1991

[3] Zhang W X, Mi J S. Incomplete information system and its optimal selections[J]. Computers and Mathematics with Applications, 2004, 48: 691-698

[4] Guan Y-Y, Wang H-K. Set-valued information systems[J]. Information Sciences, 2006, 176: 2507-2525

[5] Qian Y, Dang C, Liang J, et al. Set-valued ordered information systems[J]. Information Sciences, 2009, 179: 2809-2832

[6] Li F, Yin Y Q. Approaches to knowledge reduction of covering

decision systems based on information theory[J]. Information Sciences,2009,179:1694-1704
[7] 洪晓蕾,王燕,莫执文,等.集值不完备信息系统上的一种知识约简方法[J].四川师范大学学报:自然科学版,2007,30(3):266-269
[8] 陈子春,秦克云.集值信息系统在相容关系下的属性约简[J].模糊系统与数学,2009,23(1):150-154
[9] 陈子春.集值信息系统的知识发现与属性约简研究[D].成都:西南交通大学,2011
[10] Tsumoto S. Mining diagnostic tules from clinical databases using rough sets and medical diagnostic model[J]. Information Siences,2004,162:65-80
[11] Tsai Y C,Cheng C H,Chang J R. Entropy-based fuzzy rough claasification approach for extracting rules[J]. Expert Systems with Application,2006,31(2):436-443
[12] Hu Q H,Yu D R,Xie Z X. Neighborhood classifiers[J]. Expert Systems with Application,2008,34:866-876
[13] Li Y,Shiu S C K,Pal S K. Combining feature reduction and case selection in building CBR classifiers[J]. IEEE Transactions on Knowledge and Data Engineering,2006,18:415-429
[14] 宋笑雪,解争龙,张文修.集值决策信息系统的知识约简与规则提取[J].计算机科学,2007,34(4):182-184
[15] 宋笑雪,张文修.基于集值决策属性的集值信息系统[J].计算机工程与应用,2007,43(17):8-10
[16] Leung Y,Wu W Z,Zhang W X. Knowledge acquisition in incomplete information systems: a rough set approach[J]. European Journal of Operational Research,2006,168(1):164-180
[17] Yang X,Xie J,Song X,et al. Credible rules in incomplete decision system based on descriptors[J]. Knowledge-Based Systems,2009(22):8-17
[18] 吴鹏,杨勇,张阿红.基于集值的 Rough 集扩充模型[J].计算机工程与应用,2008,44(32):134-136
[19] 鲍忠奎,杨善林.集值信息系统的粗糙集扩展模型[J].计算机工程与应用,2011,47(35):22-24
[20] 乔全喜,秦克云.集值信息系统基于对称限制相容关系的属性约简[J].计算机工程与应用,2011,47(35)
[21] 张文修,梁怡,吴伟志.信息系统与知识发现[M].北京:科学出版社,2003

(上接第 172 页)

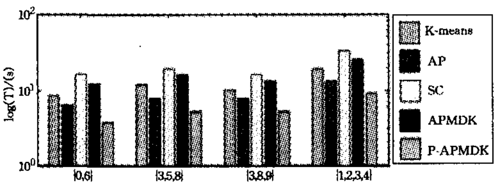


图 12 不同算法聚类时间对比图

从以上的实验结果可以得出如下结论:

1. 基于流形距离核的混合测度能够揭示数据集的整体结构信息,能够根据数据间局部密度自适应调整聚类参数的特点,能够灵活适用于各种复杂结构的数据集,拓宽了 AP 算法的应用范围。

2. 对于流形结构比较单一或者相互分离的数据集,APMDK 和 P-APMDK 算法都能获得比较理想的聚类效果,同时,对于大规模复杂的数据集,采用 P-APMDK 算法能够在改善聚类性能的同时明显提高算法的运行效率。

3. 对同一数据集,P-APMK 算法的聚类效果稍差于 APMDK 算法,在对运行速度有较高要求的场合,可以使用 P-APMDK 算法牺牲部分精度,提高处理速度。

**结束语** 本文针对仿射传播算法只能应用于超球形数据集的不足,提出了基于混合测度的仿射传播算法(APMDK),该算法能够同时满足聚类的局部一致性和全局一致性假设,克服了原有算法不能处理非凸形结构聚类的缺陷。在此基础上,为了进一步降低计算复杂度,能够在保持聚类性能的同时大大提高算法运算速度,提出了一种基于"分治"思想的并行算法——P-APMDK 算法。实验结果验证了本文算法在处理大规模复杂数据上的优越性。研究如何改进当前算法的参数设置问题,从而进一步提高算法的精度,是我们今后重要的研究方向。

## 参考文献

[1] Frey B J,Dueck D. Clustering by passing messages between data points[J]. Science,2007,315(5814):972-976
[2] Jain A K. Data clustering:50 years beyond K-means[J]. Pattern Recognition lett,2009,9(11)
[3] von Luxburg U. A tutorial on spectral clustering[R]. Technical report. Max Planck Institute for Biological Cybernetics,2006
[4] Wang K, Zhang J, Li D, et al. Adaptive Affinity Propagation Clustering[J]. Acta Automatica Sinica,2007,33(12):1242-1246
[5] 肖宇,于剑.基于近邻传播算法的半监督聚类[J].软件学报,2008,19(11):2803-2813
[6] 董俊,王锁萍,熊范纶.可变相似性度量的近邻传播聚类[J].电子与信息学报,2010,32(3):509-514
[7] Zhou D,Bousquet O. Learning with Local and Global Consistency [A]//Proceeding of advances in Neural Information Processing Systems [C]. Cambridge:MIT Press,2004:321-328
[8] Zelnik-Manor L,Perona P. Self-tuning spectral clustering[M]. Advances in Neural Information Processing Systems (NIPS), Cambridge,MA:MIT Press,2004
[9] Ertoz L,Steinbach M,Kumar V. A new shared nearest neighbor clustering algorithm and its applications[C]// Workshop on Clustering High Dimensional Data and its Applications,Second SIAM International Conference on Data Mining. Arlington,VA, USA,2002
[10] Zhang X L,Furtlehner C. Toward autonomic grids: Analyzing the job flow with affinity streaming[C]//KDD '09:Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009
[11] Song Y,Chen W-Y,Bai H,et al. Parallel spectral clustering[C]// Proceedings of Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). 2008:374-389
[12] Asuncion A,Newman D J. UCI machine learning repository. University of California,Irvine,School of Information and Computer Sciences[C/OL]. http://archive. ics. uci. edu/ml/datasets. html,2007