DOI:10.16356/j.1005-2615.2019.05.004

基于证据理论的覆盖决策信息系统约简的数值刻画

张燕兰[1,3,4] 李长清[2] 许晴媛[1,3,4]

(1.闽南师范大学计算机学院,漳州,363000;2.闽南师范大学数学与统计学院,漳州,363000;3.数据科学与智能应用福建省高等学校重点实验室,漳州,363000;4.福建省粒计算及其应用重点实验室,漳州,363000)

摘要: 决策为划分的覆盖决策信息系统的特征选择理论和方法日趋成熟。但在数据采集的过程中,有些数据集存在对象的决策缺失或决策不能完全确定情况,那么将决策刻画为覆盖更加合理。而决策为覆盖的覆盖决策信息系统特征选择的研究却很少见。本文讨论决策为覆盖的覆盖决策信息系统的特征选择,利用证据理论中的信任函数和似然函数给出覆盖决策信息系统约简的等价刻画,从而给出求约简的算法,并以实例说明该方法的有效性。

关键词: 属性约简;覆盖信息系统;覆盖粗糙集;证据理论

中图分类号: TP18 **文献标志码:** A **文章编号:** 1005-2615(2019)05-0599-10

Evidence-Theory-Based Numerical Characterization of Attribute Reduction in Covering Decision Information System

ZHANG Yanlan[1,3,4], *LI Changqing*[2], *XU qingyuan*[1,3,4]

(1. School of Computer Science, Minnan Normal University, Zhangzhou, 363000, China; 2. School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, 363000, China; 3. Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou, 363000, China; 4. Key Laboratory of Granular Computing, Minnan Normal University, Zhangzhou, 363000, China)

Abstract: The theory and methods of feature selection of the covering decision information system, whose decision is characterized by partition, are becoming mature. However, the decision attribute values of some objects in some data sets are missing or fuzzy. Then, it is reasonable to characterize the decision as a covering, but the study on feature selection of covering decision information system whose decision is characterized by covering is still very rare. In this paper, we discuss feature selection of covering decision information system whose decision is characterized by covering. We present a definition of attribute reduction in covering decision information system, and employ belief and plausibility functions from evidence theory to characterize the attribute reductions in covering decision information systems. Then, an attribute reduction algorithm based on evidence theory is proposed in covering decision information systems, and an example is used to illustrate the validity of the method.

Key words: attribute reduction; covering information systems; covering rough set; evidence theory

粗糙集理论[1-6]作为粒计算的一种重要的理论模型,能够有效地处理不精确、不一致和不完备的

基金项目: 国家自然科学基金(11701258,11871259,11526109)资助项目;福建省自然科学基金(2019J01749,2019J01748,2016J01671,2015J05011)资助项目;福建省省属高校(JK2014028)资助项目;福建省高校杰出青年科研人才培养计划资助项目。

收稿日期: 2019-09-01;**修订日期:** 2019-09-21

通信作者: 张燕兰,女,教授,E-mail: ZYL_1983_2004@163.com。

引用格式: 张燕兰,李长清,许晴媛. 基于证据理论的覆盖决策信息系统约简的数值刻画[J]. 南京航空航天大学学报,2019,51(5):599-608. ZHANG Yanlan, LI Changqing, XU qingyuan. Evidence-Theory-Based Numerical Characterization of Attribute Reduction in Covering Decision Information System[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5):599-608.

信息与知识，是一种有效的特征选择工具。粗糙集理论所处理的数据集通常以信息系统来表示，因而基于粗糙集理论的数据集的特征选择通常也被称为信息系统的属性约简。

由于收集到的数据集通常是不确定的、不完整的、不精确的、高维的或多源的，信息系统的形式具有多样性，有完备信息系统、不完备信息系统、连续值信息系统、模糊信息系统、混合信息系统和多源信息系统等。而集值属性、缺失值属性和实值属性用覆盖来刻画更加恰当，所以完备信息系统、不完备信息系统、连续值信息系统、模糊信息系统和混合信息系统等信息系统都可以用覆盖信息系统来表示，并且多源信息系统在信息融合后也可转化为覆盖信息系统。讨论覆盖信息系统的特征选择以给出不同信息系统约简理论一个统一的框架具有理论意义和实际应用价值。

覆盖粗糙集理论是覆盖信息系统特征选择的基础理论之一。由覆盖信息粒定义的覆盖上、下近似算子是覆盖粗糙集的两个基本概念。由覆盖定义不同信息粒，从而可得不同的覆盖近似算子，则引出不同的约简理论[7-13]，产生不同的约简方法：通过构造辨识矩阵求约简[7-11]，定义信息熵求约简[12]，利用组合优化方法来给出约简[13]，利用证据理论给出覆盖信息系统属性约简的等价刻画[14]，借助矩阵计算[15-16]或图论[17]来约简。这些方法在分类性能、找出约简集的极小性上体现出了优势，但针对的覆盖信息系统没有决策或者决策为划分。事实上，在数据采集的过程中条件不足等因素可能导致有些对象的决策缺失或决策不能完全确定，如连续值决策信息系统、模糊目标信息系统的决策属性值为连续值。这时直接将此类对象删去而将决策刻画为划分的做法必然导致数据部分失真，因此将决策属性刻画为覆盖更加合理，且考虑决策为覆盖的决策信息系统的特征选择尤为重要。文献[18-19]给出了决策为覆盖的决策信息系统的两类约简，但迄今，这方面的成果还很少。

Dempster-Shafer证据理论[20-21]以基本概率分配为基础，利用信任函数和似然函数构成的不确定区间来刻画证据的不确定性和未知性。集合的信任测度和似然测度可以看成是对该集合的不确定性的定量刻画，而同一集合的上近似和下近似可以看成是对该集合所表示信息的定性描述，那么证据理论与粗糙集理论之间存在必然的联系[22-29]。根据证据理论与覆盖粗糙集理论之间的联系，利用证据理论研究覆盖信息系统的特征选择方面，有一些初步的结果。比如，通过建立一对覆盖近似算子与信任函数、似然函数的联系来讨论覆盖决策信息系统的约简[14]。文献[30]定义了随机覆盖目标信息系统，并利用信任函数和似然函数给出协调和不协调随机覆盖目标信息系统约简的等价刻画。但是这些成果的覆盖决策信息系统的决策都由划分来刻画。文献[31]给出了决策为覆盖的覆盖决策信息系统的一类约简，并利用证据理论中的信任函数和似然函数给出约简的等价刻画。

本文讨论决策为覆盖的覆盖决策信息系统的一类约简。首先，由一对覆盖上、下近似算子产生一对对偶的信任函数和似然函数。然后利用信任函数和似然函数给出覆盖决策信息系统约简的等价刻画，从而给出求约简的算法，并以实例说明算法的有效性。

1 基本定义与性质

本文中，论域U是非空有限集。对$X\subseteq U$，$-X$表示X的余集，$|X|$表示X所包含的元素的个数。下面介绍覆盖粗糙集理论和证据理论的一些相关概念和性质。

定义1[32] 设C是U上的一个覆盖，即$\bigcup\{K|K\in C\}=U$。任意$x\in U$，令$(x)_C=\bigcap\{K\in C|x\in K\}$。定义如下一对对偶的覆盖近似算子：对任意$X\subseteq U$，$\underline{C}(X)=\{x\in U|\forall u(x\in(u)_C\Rightarrow(u)_C\subseteq X)\}$，$\bar{C}(X)=\bigcup\{(x)_C|(x)_C\cap X\neq\varnothing\}$。

该对覆盖近似算子是重要的覆盖近似算子之一，也得到一些学者的关注。比如，Chen等[8]给出了上近似算子的定义，Qin等[32]给出了该对近似算子的定义，Zhang等给出了该对近似算子的公理组[33]，Yao等将该对近似算子归类为基于粒的覆盖近似算子[34]，Chen等[35]讨论了该对覆盖近似算子的不确定性。关于该对近似算子的性质，有如下命题。

命题1[32] 设C是U上的一个覆盖，对任意$X,Y\subseteq U$有

(1)$\underline{C}(\varnothing)=\varnothing$，$\underline{C}(U)=U$；

(2)$\underline{C}(X)\subseteq X\subseteq\bar{C}(X)$；

(3)$\underline{C}(X\cap Y)=\underline{C}(X)\cap\underline{C}(Y)$；

(4)$\bar{C}(X\cup Y)=\bar{C}(X)\cup\bar{C}(Y)$。

下面给出证据理论中两个最重要的概念：信任函数和似然函数。

定义2[20-21] 设U是非空有限集，称集函数B: $P(U)\rightarrow[0,1]$为信任函数，若它满足性质

(1)$B(\varnothing)=0$，$B(U)=1$；

(2)对于U的任意子集$X_1,X_2,\cdots,X_n(n\geqslant1)$，有$B(\bigcup_{i=1}^{n}X_i)\geqslant\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}B(\bigcap_{i\in I}X_i)$，

其中$|I|$表示I中所含元素的个数。称函数L:

$P(U)\to[0,1]$为似然函数，若L是B的对偶补，即对于任意$X\subseteq U$有：$L(X)=1-B(-X)$。

定理 1[20-21]　似然函数满足下列性质

(1) $L(\varnothing)=0$，$L(U)=1$；

(2)对于U的任意子集$X_1,X_2,\cdots,X_n(n\geqslant 1)$有

$$L(\bigcap_{i=1}^{n}X_i)\leqslant\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}L(\bigcup_{i\in I}X_i)。$$

定义 3[21]　称集函数$m:P(U)\to[0,1]$为概率分配函数，简称为mass函数，如果它满足

$$(1)\,m(\varnothing)=0;\ (2)\sum_{X\subseteq U}m(X)=1。$$

定理 2[21]　若m是U上的一个mass函数，则对任意$X\subseteq U$，由

$$B(X)=\sum_{A\subseteq X}m(A),\ L(X)=\sum_{A\cap X\neq\varnothing}m(A)$$

定义的函数B和L分别为信任函数和似然函数。

由覆盖近似算子$\underline{C}(X)$和$\bar{C}(X)$可导出信任函数和似然函数。

定理 3　设$(U,P(U),P)$是一个概率空间，C是U上的一个覆盖。对任意$X\subseteq U$，令$\mathrm{Bel}_C(X)=P(\underline{C}(X))$，$Pl_C(X)=P(\bar{C}(X))$，则$\mathrm{Bel}_C$和$Pl_C$分别为信任函数和似然函数。

证明　先证明Bel_C是信任函数。

(1) 由$\underline{C}(\varnothing)=\varnothing$，$\underline{C}(U)=U$，可得

$$\mathrm{Bel}_C(\varnothing)=P(\underline{C}(\varnothing))=P(\varnothing)=0,$$
$$\mathrm{Bel}_C(U)=P(\underline{C}(U))=P(U)=1。$$

(2) 由命题1(3)可得，$\underline{C}(\bigcup_{i=1}^{n}X_i)\supseteq\bigcup_{i=1}^{n}\underline{C}(X_i)$和$\bigcap_{i=1}^{n}\underline{C}(X_i)=\underline{C}(\bigcap_{i=1}^{n}X_i)$。则

$$\begin{aligned}\mathrm{Bel}_C(\bigcup_{i=1}^{n}X_i)&=P(\underline{C}(\bigcup_{i=1}^{n}X_i))\geqslant P(\bigcup_{i=1}^{n}\underline{C}(X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}P(\bigcap_{i\in I}\underline{C}(X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}P(\underline{C}(\bigcap_{i\in I}X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}\mathrm{Bel}_C(\bigcap_{i\in I}X_i)。\end{aligned}$$

再证明Pl_C是似然函数。

(1) 由$\bar{C}(\varnothing)=\varnothing$，$\bar{C}(U)=U$，可得$Pl_C(\varnothing)=P(\bar{C}(\varnothing))=P(\varnothing)=0$　，$Pl_C(U)=P(\bar{C}(U))=P(U)=1$。

(2) 由命题1(4)可得，$\bar{C}(\bigcap_{i=1}^{n}X_i)\subseteq\bigcap_{i=1}^{n}\bar{C}(X_i)$和$\bigcup_{i=1}^{n}\bar{C}(X_i)=\bar{C}(\bigcup_{i=1}^{n}X_i)$。于是

$$\begin{aligned}Pl_C(\bigcap_{i=1}^{n}X_i)&=P(\bar{C}(\bigcap_{i=1}^{n}X_i))\leqslant P(\bigcap_{i=1}^{n}\bar{C}(X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}P(\bigcup_{i\in I}\bar{C}(X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}P(\bar{C}(\bigcup_{i\in I}X_i))=\\&\sum_{\varnothing\neq I\subseteq\{1,2,\cdots,n\}}(-1)^{|I|+1}Pl_C(\bigcup_{i\in I}X_i)。\end{aligned}$$

推论 1　设C是有限论域U上的一个覆盖。对任意$X\subseteq U$，令$\mathrm{Bel}_C(X)=\dfrac{|\underline{C}(X)|}{|U|}$，$Pl_C(X)=\dfrac{|\bar{C}(X)|}{|U|}$，则$\mathrm{Bel}_C$和$Pl_C$分别为信任函数和似然函数。相应的mass函数为$m_C(X)=\dfrac{|j(X)|}{|U|}$，其中$n(x)=\bigcup\{(u)_C|x\in(u)_C,u\in U\}$，$j(X)=\{x\in U|n(x)=X\}$。

证明　定义函数$P:P(U)\to[0,1]$如下：对任意$X\subseteq U$，$P(X)=\dfrac{|X|}{|U|}$。由定理3可知，Bel_C和Pl_C分别为信任函数和似然函数。

既然$m_C(\varnothing)=\dfrac{|j(\varnothing)|}{|U|}=0$，$\sum_{X\subseteq U}m_C(X)=\sum_{X\subseteq U}\dfrac{|j(X)|}{|U|}=\dfrac{|\bigcup_{X\subseteq U}j(X)|}{|U|}=\dfrac{|U|}{|U|}=1$，$m_C$为mass函数。且对任意$X\subseteq U$，有

$$\begin{aligned}&\mathrm{Bel}_C(X)=\frac{|\underline{C}(X)|}{|U|}=\\&\frac{1}{|U|}|\{x\in U|\forall u(x\in(u)_C\Rightarrow(u)_C\subseteq X)\}|=\\&\frac{1}{|U|}|\{x\in U|n(x)\subseteq X\}|=\frac{1}{|U|}|\bigcup_{A\subseteq X}j(A)|=\\&\frac{1}{|U|}\sum_{A\subseteq X}|j(A)|=\sum_{A\subseteq X}m_C(A)。\end{aligned}$$

可见，m_C是Bel_C的mass函数。

2　覆盖决策信息系统约简的数值刻画

文献[8]给出了决策为划分的覆盖决策信息系统的定义，并讨论其属性约简。而很多信息系统的决策并不能用划分来刻画，用覆盖来刻画更加合理。引用文献[36]中的两个例子来说明。

例 1　假设某个企业需要判断各项技术的成熟度，以便进行技术创新决策，原决策信息系统为(U,A,F,d)，其中U表示各项技术，有10项待评估的技术；A表示影响技术成熟度的属性特征，$A=\{a_1,a_2\}$，a_1表示技术的市场需求状况，a_2表示需求的变化状况；$F=\{f_l:U\to V_l(a_l\in A)\}$表示不同技术与各种属性之间的关系，即不同技术项目在不同属性下的取值，V_l是属性a_l的取值域；d表示技

术成熟度，V_d是属性d的取值域。给出决策信息系统如表1所示。这是一个条件属性值和决策属性值在[0,4]之间的连续值决策信息系统。

对任意$B\subseteq A$和d，定义关系R_B和R_d如下：

$$R_B=\{(x,y)\in U\times U|\forall a\in B,|f_a(x)-f_a(y)|\leqslant\delta_a\}$$

$$R_d=\{(x,y)\in U\times U||f_d(x)-f_d(y)|\leqslant\delta_d\}$$

式中$\delta_a>0,\delta_d>0$。则R_B和R_d是自反且对称关系，而非传递的，所以不是等价关系。于是$C_i=\{(R_{a_i})_s(x)|x\in U\}(i=1,2)$，$C_d=\{(R_d)_s(x)|x\in U\}$为覆盖。

如令$\delta_{a_1}=0.6,\delta_{a_2}=0.3,\delta_d=1.5$，则可得3个覆盖

$$C_1=\{\{x_1,x_2,x_6\},\{x_1,x_2,x_7\},\{x_3,x_7,x_8\},\{x_4,x_9,x_{10}\},\{x_4,x_8,x_9\},\{x_2,x_3,x_7\},\{x_5,x_{10}\},\{x_1,x_6\},\{x_3,x_8,x_9\},\{x_4,x_5,x_{10}\}\}$$

$$C_2=\{\{x_1,x_2,x_3,x_4,x_5\},\{x_6,x_8,x_{10}\},\{x_7,x_8,x_9\}\}$$

$$C_d=\{\{x_1,x_2,x_3,x_6,x_7\},\{x_1,x_2,x_6,x_7\},\{x_1,x_3,x_6,x_8\},\{x_4,x_5,x_9\},\{x_3,x_8,x_9\},\{x_4,x_5,x_8,x_9\},\{x_{10}\}\}$$

表1 例1中的决策信息系统

Tab.1 The decision information system of Example 1

U	a_1	a_2	d
x_1	0.2	0.0	3.8
x_2	0.8	0.2	4.0
x_3	1.8	0.0	3.2
x_4	3.2	0.2	1.8
x_5	4.0	0.0	2.0
x_6	0.0	0.8	3.8
x_7	1.2	1.2	4.0
x_8	2.2	1.0	2.8
x_9	2.8	1.2	2.2
x_{10}	3.8	0.8	0.8

例2 称(U,A,F,D)为模糊决策信息系统，其中U是对象集；A是属性集；$F=\{f_l:U\rightarrow V_l(a_l\in A)\}$是对象集在属性集上的映射或关系函数，$V_l$是属性$a_l$的取值域；$D=\{\tilde{D}_j:U\rightarrow[0,1](j\leqslant r)\}$。给出模糊决策信息系统，如表2所示。

对决策属性集D，定义关系R_D为

$$R_D=\{(x,y)\in U\times U|\forall\widetilde{D}_i\in D,|\widetilde{D}_i(x)-\widetilde{D}_i(y)|\leqslant\delta_i\}$$

式中$\delta_i>0(i=1,2)$。则R_D是自反且对称关系，而非传递的，所以不是等价关系。于是$C_D=\{(R_D)_s(x)|x\in U\}$为覆盖。

如令$\delta_1=0.2,\delta_2=0.3$，则可得覆盖$C_D=\{\{x_1,x_2,x_3\},\{x_1,x_2,x_3,x_4\},\{x_3,x_4,x_5\},\{x_4,x_5\},\{x_6\}\}$。

表2 例2中的一个模糊决策信息系统

Tab.2 Fuzzy decision information system of Example 2

U	a_1	a_2	$\widetilde{D}_1$	$\widetilde{D}_2$
x_1	1	1	0.9	0.2
x_2	1	1	1.0	0.3
x_3	1	2	0.8	0.5
x_4	1	2	0.7	0.6
x_5	1	2	0.6	0.9
x_6	2	2	0.2	1.0

例1和2表明：一些决策信息系统的条件属性和决策属性用覆盖来刻画更加合理。可见，讨论决策为覆盖的覆盖决策信息系统的属性约简具有实际意义。下面给出决策为覆盖的覆盖决策信息系统的约简的定义。

定义4[8] 设A是U上的一族覆盖，$B\subseteq A$且$B=\{C_j|j=1,\cdots,n\}$。对任意$x\in U$，称$(x)_B=\bigcap\{K|x\in K,K\in C_j,j=1,\cdots,n\}$为$x$在覆盖族$B$下的邻域，称覆盖$\mathrm{Cov}(B)=\{(x)_B|x\in U\}$为$B$诱导的$U$上的一个覆盖。

显然，对任意$B\subseteq A,x,y\in U$，有

(1) $y\in(x)_B\Leftrightarrow(y)_B\subseteq(x)_B$；

(2) $(x)_B=\bigcap_{C\in B}(x)_C$；

(3) $(x)_A\subseteq(x)_B$；

(4) $(x)_B=(x)_{\mathrm{Cov}(B)}$。

根据定义1和定义4，对任意$B\subseteq A$，可得一对关于覆盖$\mathrm{Cov}(B)$的近似算子：对任意$X\subseteq U$，

$\underline{C}_B(X)=\{x\in U|\forall u(x\in(u)_B\Rightarrow(u)_B\subseteq X)\}$，

$\bar{C}_B(X)=\bigcup\{(x)_B|(x)_B\cap X\neq\varnothing\}$。

定义5[18] 设A是U上的一族覆盖，D是U上的一个覆盖，称(U,A,D)为一个覆盖决策信息系统。设$B\subseteq A$，若对任意$D_i\in D$有$\underline{C}_A(D_i)=\underline{C}_B(D_i)$，称$B$是下近似协调集。若$B$是下近似协调集，$B$的任意真子集不是下近似协调集，称$B$是下近似约简集。$(U,A,D)$的所有下近似约简集的交称为下近似核心集，记为$\mathrm{Core}_L(A,D)$。设$B\subseteq A$，若对任意$D_i\in D$有$\bar{C}_A(D_i)=\bar{C}_B(D_i)$，称$B$是上近似协调集。若$B$是上近似协调集，$B$的任意真子集不是上近似协调集，称$B$是上近似约简集。$(U,A,D)$的所有上近似约简集的交称为上近似核心集，记为$\mathrm{Core}_U(A,D)$。

例3 设(U,A,D)为一个覆盖决策信息系统，其中$U=\{x_1,x_2,x_3,x_4,x_5\}$，$A=\{C_1,C_2,C_3\}$是$U$的3个覆盖，$D$是$U$上的一个覆盖，且

$$C_1=\{\{x_1,x_2\},\{x_2,x_3,x_4\},\{x_5\}\}$$
$$C_2=\{\{x_1,x_2\},\{x_3,x_4\},\{x_2,x_5\}\}$$
$$C_3=\{\{x_1,x_2\},\{x_3,x_4\},\{x_4,x_5\}\}$$
$$D=\{\{x_1,x_2,x_3\},\{x_1,x_4\},\{x_5\}\}$$

可得点在不同覆盖族下的邻域,如表3所示。

表3　例3中的点的邻域

Tab.3　Neighborhoods of elements in Example 3

x	x_1	x_2	x_3	x_4	x_5
$(x)_{C_1}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_2,x_3,x_4\}$	$\{x_2,x_3,x_4\}$	$\{x_5\}$
$(x)_{C_2}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_3,x_4\}$	$\{x_2,x_5\}$
$(x)_{C_3}$	$\{x_1,x_2\}$	$\{x_1,x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_4,x_5\}$
$(x)_{\{C_1,C_2\}}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_3,x_4\}$	$\{x_5\}$
$(x)_{\{C_1,C_3\}}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_5\}$
$(x)_{\{C_2,C_3\}}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_5\}$
$(x)_A$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_5\}$

由定义1,可得 $\bar{C}_A(\{x_1,x_2,x_3\})=\{x_1,x_2,x_3,x_4\}$, $\bar{C}_A(\{x_1,x_4\})=\{x_1,x_2,x_3,x_4\}$, $\bar{C}_A(\{x_5\})=\{x_5\}$。根据定义5可知 $\{C_1\}$ 和 $\{C_2,C_3\}$ 为覆盖上近似约简集, $\{C_1,C_2\}$, $\{C_1,C_3\}$ 和 $\{C_2,C_3\}$ 是覆盖下近似约简集。

注1　由例3可见,一个覆盖决策信息系统的上近似约简集和下近似约简集未必相同。接下来,利用信任函数和似然函数给出下近似约简集和上近似约简集的刻画。

在覆盖决策信息系统 (U,A,D) 中,对任意 $B\subseteq A$,将覆盖 $\mathrm{Cov}(B)$ 诱导的信任函数 $\mathrm{Bel}_{\mathrm{Cov}(B)}$ 和似然函数 $Pl_{\mathrm{Cov}(B)}$ 分别记为 Bel_B 和 Pl_B。

定理4　设 (U,A,D) 为覆盖决策信息系统,其中 $D=\{D_1,D_2,\cdots,D_m\}$。B 是 (U,A,D) 的下近似协调集的充要条件是 $\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{m}\mathrm{Bel}_A(D_i)$。$B$ 是 (U,A,D) 的下近似约简集的充要条件是 $\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{m}\mathrm{Bel}_A(D_i)$, 且对任意 $C\subseteq B$, $\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)>\sum_{i=1}^{m}\mathrm{Bel}_C(D_i)$。

证明　设 B 是下近似协调集,则对任意 $D_i\in D$ 有 $\underline{C}_A(D_i)=\underline{C}_B(D_i)$。于是

$$\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{m}\frac{|\underline{C}_B(D_i)|}{n}=\sum_{i=1}^{m}\frac{|\underline{C}_A(D_i)|}{n}=\sum_{i=1}^{m}\mathrm{Bel}_A(D_i) \tag{1}$$

反之,对任意 $D_i\in D$ 有 $\underline{C}_A(D_i)\supseteq\underline{C}_B(D_i)$,所以 $|\underline{C}_A(D_i)|\geqslant|\underline{C}_B(D_i)|$。而由 $\sum_{i=1}^{n}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{n}\mathrm{Bel}_A(D_i)$ 可得:对任意 $D_i\in D$ 有 $|\underline{C}_A(D_i)|=|\underline{C}_B(D_i)|$,所以 $\underline{C}_A(D_i)=\underline{C}_B(D_i)$。可见, B 是下近似协调集。

若 B 是下近似约简集,则 B 是下近似协调集,于是 $\sum_{i=1}^{n}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{n}\mathrm{Bel}_A(D_i)$。对任意 $C\subseteq B$, C 不是下近似协调集,所以 $\sum_{i=1}^{n}\mathrm{Bel}_C(D_i)\neq\sum_{i=1}^{n}\mathrm{Bel}_A(D_i)=\sum_{i=1}^{n}\mathrm{Bel}_B(D_i)$。又由 $\sum_{i=1}^{n}\mathrm{Bel}_C(D_i)\leqslant\sum_{i=1}^{n}\mathrm{Bel}_B(D_i)$ 可知 $\sum_{i=1}^{n}\mathrm{Bel}_C(D_i)<\sum_{i=1}^{n}\mathrm{Bel}_B(D_i)$。反之,根据 $\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)=\sum_{i=1}^{m}\mathrm{Bel}_A(D_i)$ 可知 B 是下近似协调集。对任意 $C\subseteq B$, $\sum_{i=1}^{m}\mathrm{Bel}_B(D_i)>\sum_{i=1}^{m}\mathrm{Bel}_C(D_i)$,所以 C 不是下近似协调集。可见 B 是下近似约简集。

定理5　设 (U,A,D) 为一覆盖决策信息系统,其中 $D=\{D_1,D_2,\cdots,D_m\}$。B 是 (U,A,D) 的上近似协调集的充要条件是 $\sum_{i=1}^{m}Pl_B(D_i)=\sum_{i=1}^{m}Pl_A(D_i)$。$B$ 是 (U,A,D) 的上近似约简集的充要条件是 $\sum_{i=1}^{m}Pl_B(D_i)=\sum_{i=1}^{m}Pl_A(D_i)$, 且对任意 $C\subseteq B$ 有 $\sum_{i=1}^{m}Pl_B(D_i)<\sum_{i=1}^{m}Pl_C(D_i)$。

证明与定理4的证明类似。

注2　根据定理5,可设计上近似约简集的算法。由定义1可知计算一个集合的下近似比计算其上近似复杂,所以没有直接利用定理4来设计下近似约简集的算法,而是通过保持决策类的余集的上近似不变的上近似约简集得到下近似约简集。

3　覆盖决策信息系统属性约简的方法

定义6　设 (U,A,D) 为一覆盖决策信息系统,定义 $C\in A$ 在 A 中的重要性为

$$\mathrm{Sig}_{A\setminus\{C\}}(C)=\sum_{i=1}^{m}Pl_{A\setminus\{C\}}(D_i)-\sum_{i=1}^{m}Pl_A(D_i) \tag{2}$$

命题2　$\mathrm{Core}_U(A,D)=\{C\in A|\mathrm{Sig}_{A\setminus\{C\}}(C)>0\}$。

证明　任意 $C\in\mathrm{Core}_U(A,D)$, $A\setminus\{C\}$ 不是上近似协调集。否则, $A\setminus\{C\}$ 至少包含一个上近似约简集,那么该约简集不包含 C,与 $C\in\mathrm{Core}_U(A,D)$ 矛盾。于是根据定理5有 $\sum_{i=1}^{m}Pl_{A\setminus\{C\}}(D_i)>$

$\sum_{i=1}^{m} Pl_A(D_i)$，故 $\mathrm{Sig}_{A\setminus\{C\}}(C)>0$。反之，如果 $\mathrm{Sig}_{A\setminus\{C\}}(C)>0$，则 $\sum_{i=1}^{m} Pl_{A\setminus\{C\}}(D_i)>\sum_{i=1}^{m} Pl_A(D_i)$。可见，$A\setminus\{C\}$不是上近似协调集，那么$A$的所有上近似约简集必包含$C$，所以$C\in \mathrm{Core}_U(A,D)$。

命题2表明上近似核心集中的覆盖都是重要的。

定义7 设(U,A,D)为一覆盖决策信息系统，其中$D=\{D_1,D_2,\cdots,D_m\}$。设$B\subseteq A$，定义覆盖$C\in A\setminus B$对B的重要性为

$$\mathrm{Sig}_B(C)=\sum_{i=1}^{m} Pl_B(D_i)-\sum_{i=1}^{m} Pl_{B\cup\{C\}}(D_i) \quad (3)$$

定义7将覆盖$C\in A\setminus B$对B的重要性通过似然函数值变化来度量，$\mathrm{Sig}_B(C)$的值越大，说明覆盖$C\in A\setminus B$对B越重要。根据定义6和定义7中重要性的定义与性质来寻找覆盖决策信息系统$(U,A,D)(D=\{D_1,D_2,\cdots,D_m\})$最小约简的算法。

算法1

第1步 令$\mathrm{Core}_U(A,D)=\varnothing$，计算$\sum_{i=1}^{m} Pl_A(D_i)=M$。规定$\sum_{i=1}^{m} Pl_{\varnothing}(D_i)=M$。

第2步 计算$C\in A$在A中的重要性$\mathrm{Sig}_{A\setminus\{C\}}(C)$，作$\mathrm{Core}_U(A,D)=\{C\in A|Sig_{A\setminus\{C\}}(C)>0\}$。若$\sum_{i=1}^{m} Pl_{\mathrm{Core}(A,D)}(D_i)=M$，则$\mathrm{Core}_U(A,D)$为$A$的最小约简，输出$\mathrm{Core}_U(A,D)$。否则，转第3步。

第3步 令$B=\mathrm{Core}_U(A,D)$，对覆盖集$A\setminus B$，重复作：

(1) 对每个$C\in A\setminus B$，计算$\mathrm{Sig}_B(C)$。

(2) 选取$C_0\in A\setminus B$，使$\mathrm{Sig}_B(C_0)=\mathop{\mathrm{Max}}\limits_{C\in A\setminus B}\mathrm{Sig}_B(C)$，作$B=B\cup\{C_0\}$。

(3) 判断是否$\sum_{i=1}^{m} Pl_B(D_i)=M$，若成立输出$B$，否则转(1)。

例4 设(U,A,D)为一个覆盖决策信息系统，其中$U=\{x_1,x_2,x_3,x_4,x_5\}$，$A=\{C_1,C_2,C_3\}$是$U$的3个覆盖，$D$是$U$的一个覆盖，且

$$C_1=\{\{x_1,x_2\},\{x_2,x_3,x_4\},\{x_4,x_5\}\}$$
$$C_2=\{\{x_1,x_2,x_3\},\{x_3,x_4\},\{x_5\}\}$$
$$C_3=\{\{x_1,x_2\},\{x_3,x_4\},\{x_4,x_5\}\}$$
$$D=\{D_1=\{x_1,x_2,x_3\},D_2=\{x_2,x_5,$$
$$D_3=\{x_3,x_4\},D_4=\{x_5\}\}$$

可得U中点在不同覆盖集下的邻域如表4所示。

由定义1，可知$\bar{C}_A(D_1)=\{x_1,x_2,x_3\}$，$\bar{C}_A(D_2)=\{x_1,x_2,x_5\}$，$\bar{C}_A(D_3)=\{x_3,x_4\}$，$\bar{C}_A(D_4)=\{x_5\}$。根据定义5可知$\{C_1,C_2\}$和$\{C_2,C_3\}$为覆盖上近似约简集，也是覆盖下近似约简集。

表4 例4中的点的邻域

Tab.4 Neighborhoods of elements in Example 4

x	x_1	x_2	x_3	x_4	x_5
$(x)_{C_1}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_2,x_3,x_4\}$	$\{x_4\}$	$\{x_4,x_5\}$
$(x)_{C_2}$	$\{x_1,x_2,x_3\}$	$\{x_1,x_2,x_3\}$	$\{x_3\}$	$\{x_3,x_4\}$	$\{x_5\}$
$(x)_{C_3}$	$\{x_1,x_2\}$	$\{x_1,x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_4,x_5\}$
$(x)_{\{C_1,C_2\}}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$
$(x)_{\{C_1,C_3\}}$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3,x_4\}$	$\{x_4\}$	$\{x_4,x_5\}$
$(x)_{\{C_2,C_3\}}$	$\{x_1,x_2\}$	$\{x_1,x_2\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$
$(x)_A$	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3\}$	$\{x_4\}$	$\{x_5\}$

按照算法1给出上近似约简集的求解。

第1步 令$\mathrm{Core}_U(A,D)=\varnothing$，计算$M=\sum_{i=1}^{4} Pl_A(D_i)=\frac{3}{5}+\frac{3}{5}+\frac{2}{5}+\frac{1}{5}=\frac{9}{5}$。

第2步 由$\bar{C}_{A\setminus\{C_1\}}(D_1)=\{x_1,x_2,x_3\}$，$\bar{C}_{A\setminus\{C_1\}}(D_2)=\{x_1,x_2,x_5\}$，$\bar{C}_{A\setminus\{C_1\}}(D_3)=\{x_3,x_4\}$，$\bar{C}_{A\setminus\{C_1\}}(D_4)=\{x_5\}$可得$\sum_{i=1}^{4} P\,l_{A\setminus\{C_1\}}(D_i)=\frac{3}{5}+\frac{3}{5}+\frac{2}{5}+\frac{1}{5}=\frac{9}{5}$，则$\mathrm{Sig}_{A\setminus\{C_1\}}(C_1)=\sum_{i=1}^{4} Pl_{A\setminus\{C_1\}}(D_i)-\sum_{i=1}^{4} Pl_A(D_i)=0$。

由$\bar{C}_{A\setminus\{C_2\}}(D_1)=\{x_1,x_2,x_3,x_4\}$，$\bar{C}_{A\setminus\{C_2\}}(D_2)=\{x_1,x_2,x_4,x_5\}$，$\bar{C}_{A\setminus\{C_2\}}(D_3)=\{x_3,x_4,x_5\}$，$\bar{C}_{A\setminus\{C_2\}}(D_4)=\{x_4,x_5\}$可得$\sum_{i=1}^{4} P\,l_{A\setminus\{C_2\}}(D_i)=\frac{4}{5}+\frac{4}{5}+\frac{3}{5}+\frac{2}{5}=\frac{13}{5}$，则$\mathrm{Sig}_{A\setminus\{C_2\}}(\mathrm{C}_2)=\sum_{i=1}^{4} Pl_{A\setminus\{C_2\}}(D_i)-\sum_{i=1}^{4} Pl_A(D_i)=\frac{4}{5}$。

由$\bar{C}_{A\setminus\{C_3\}}(D_1)=\{x_1,x_2,x_3\}$，$\bar{C}_{A\setminus\{C_3\}}(D_2)=\{x_1,x_2,x_5\}$，$\bar{C}_{A\setminus\{C_3\}}(D_3)=\{x_3,x_4\}$，$\bar{C}_{A\setminus\{C_3\}}(D_4)=\{x_5\}$可得$\sum_{i=1}^{4} Pl_{A\setminus\{C_3\}}(D_i)=\frac{3}{5}+\frac{3}{5}+\frac{2}{5}+\frac{1}{5}=\frac{9}{5}$，则$\mathrm{Sig}_{A\setminus\{C_3\}}(C_3)=\sum_{i=1}^{4} Pl_{A\setminus\{C_3\}}(D_i)-\sum_{i=1}^{4} Pl_A(D_i)=0$。

综上可得，$B=\mathrm{Core}_U(A,D)=\{C_2\}$。

由$\bar{C}_B(D_1)=\{x_1,x_2,x_3,x_4\}$，$\bar{C}_B(D_2)=\{x_1,x_2,x_3,x_5\}$，$\bar{C}_B(D_3)=\{x_1,x_2,x_3,x_4\}$，$\bar{C}_B(D_4)=\{x_5\}$可得

$$\sum_{i=1}^{4} Pl_B(D_i)=\frac{4}{5}+\frac{4}{5}+\frac{4}{5}+\frac{1}{5}=\frac{13}{5}\neq M$$

可见，$B=\{C_2\}$不是上近似约简集。

第 3 步　计算

$$\mathrm{Sig}_B(C_1)=\sum_{i=1}^{4}Pl_B(D_i)-\sum_{i=1}^{4}Pl_{B\cup\{C_1\}}(D_i)=\frac{13}{5}-\frac{9}{5}=\frac{4}{5}$$

$$\mathrm{Sig}_B(C_3)=\sum_{i=1}^{4}Pl_B(D_i)-\sum_{i=1}^{4}Pl_{B\cup\{C_3\}}(D_i)=\frac{13}{5}-\frac{9}{5}=\frac{4}{5}$$

选取 $C_1\in A\setminus B$,使 $\mathrm{Sig}_B(C_1)=\max\limits_{C\in A\setminus B}\mathrm{Sig}_B(C)$,作 $B=B\cup\{C_1\}$。由 $\sum\limits_{i=1}^{4}Pl_B(D_i)=\frac{9}{5}=M$,可知 $B=\{C_1,C_2\}$是上近似约简集。

注 3　(1)覆盖决策信息系统$(U,A,D)(D=\{D_1,D_2,D_3,D_4\})$的下近似约简集就是覆盖决策信息系统$(U,A,D')$ $(D'=\{-D_1,-D_2,-D_3,-D_4\})$的上近似约简集,所以可利用求上近似约简集的算法给出下近似约简集。

(2) 划分是特殊的覆盖,所以本文的理论显然适用于决策刻画为划分的覆盖决策信息系统。

(3) 文献[14]中,Chen 等由一对覆盖近似算子得到信任函数和似然函数,从而利用信任函数给出决策为划分的覆盖决策信息系统属性约简的算法。本文所讨论的覆盖近似算子与文献[14]中的覆盖近似算子不同,而且本文是根据决策类的上近似得到的似然函数来构造算法,而文献[14]是根据决策类的下近似得到的信任函数来构造算法。当本文中的覆盖决策信息系统的决策刻画为划分时,算法 1 得到的约简与文献[14]中算法得到的约简可能不同,本文以文献[14]中的例子来进行说明。

例 5　考虑房子评估问题。设 $U=\{x_1,x_2,\cdots,x_{10}\}$是 10 套房子,$A=$\{price; structure; color; surroundings\}是属性集,“price”的属性值为\{high; middle; low\},“structure”的属性值为\{reasonable; ordinary; poor\},“color”的属性值为\{good; bad\},“surroundings”的属性值为\{quiet; slightly noisy; noisy; very noisy\}。有 4 个专家$\{A,B,C,D\}$来评估这些房子的属性值,专家之间的意见相互独立且同等重要。评估结果如表 5。所示

由决策信息系统的条件属性可得一族覆盖 $A=\{C_1,C_2,C_3,C_4\}$。

Price: $C_1=\{\{x_1,x_2,x_3,x_4,x_6,x_7,x_8,x_9,x_{10}\},\{x_3,x_4,x_6,x_7\},\{x_3,x_4,x_5,x_6,x_7\}\}$。

Structure: $C_2=\{\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\},\{x_6,x_7,x_8,x_9\},\{x_{10}\}\}$。

Color: $C_3=\{\{x_1,x_2,x_3,x_6,x_8,x_9,x_{10}\},\{x_2,x_3,x_4,x_5,x_6,x_7,x_9\}\}$。

Surroundings: $C_4=\{\{x_1,x_2,x_3,x_6\},\{x_2,x_3,x_4,x_5,x_6,x_7\},\{x_6,x_8,x_9,x_{10}\},\{x_6,x_7,x_9\}\}$。

决策属性构成的划分为 $D=\{D_1=\{x_1,x_2,x_3,x_6\},D_2=\{x_4,x_5,x_7\},D_3=\{x_8,x_9,x_{10}\}\}$。

由覆盖族 A 可得,$(x_1)_A=\{x_1,x_2,x_3,x_6\}$,$(x_2)_A=\{x_2,x_3,x_6\}$,$(x_3)_A=\{x_3,x_6\}$,$(x_4)_A=\{x_3,x_4,x_6,x_7\}$,$(x_5)_A=\{x_3,x_4,x_5,x_6,x_7\}$,$(x_6)_A=\{x_6\}$,$(x_7)_A=\{x_6,x_7\}$,$(x_8)_A=\{x_6,x_8,x_9\}$,$(x_9)_A=\{x_6,x_9\}$,$(x_{10})_A=\{x_{10}\}$。

第 1 步　令 $\mathrm{Core}_U(A,D)=\varnothing$,计算

$$M=\sum_{i=1}^{3}Pl_A(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{4}{10}=\frac{18}{10}。$$

第 2 步　由覆盖族$A\setminus\{C_1\}$可得,$\bar{C}_{A\setminus\{C_1\}}(D_1)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9\}$,$\bar{C}_{A\setminus\{C_1\}}(D_2)=\{x_2,x_3,x_4,x_5,x_6,x_7\}$,$\bar{C}_{A\setminus\{C_1\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。于是

$$\sum_{i=1}^{3}Pl_{A\setminus\{C_1\}}(D_i)=\frac{9}{10}+\frac{6}{10}+\frac{4}{10}=\frac{19}{10}$$

则 $\mathrm{Sig}_{A\setminus\{C_1\}}(C_1)=\sum\limits_{i=1}^{3}Pl_{A\setminus\{C_1\}}(D_i)-\sum\limits_{i=1}^{3}Pl_A(D_i)=\frac{1}{10}$。

表 5　房子评估问题的决策系统

Tab.5　Decision system of house evaluation problem

U	Price	Structure	Color	Surrounding	Decision
x_1	{high}	{reasonable}	{good}	{quiet}	sale
x_2	{high}	{reasonable}	{good; bad}	{quiet; slightly noisy}	sale
x_3	{high; middle; low}	{reasonable}	{good; bad}	{quiet; slightly noisy}	sale
x_4	{high; middle; low}	{reasonable}	{bad}	{slightly noisy}	further-evaluation
x_5	{low}	{reasonable}	{bad}	{slightly noisy}	further-evaluation
x_6	{high; middle; low}	{reasonable; ordinary}	{good; bad}	{quiet; slightly noisy; noisy; very noisy}	sale
x_7	{high; middle; low}	{reasonable; ordinary}	{bad}	{slightly noisy; very noisy}	further-evaluation
x_8	{high}	{ordinary}	{good}	{noisy}	reject
x_9	{high}	{ordinary}	{good; bad}	{noisy; very noisy}	reject
x_{10}	{high}	{poor}	{good}	{noisy}	reject

由覆盖族 $A\setminus\{C_2\}$ 可得 $\bar{C}_{A\setminus\{C_2\}}(D_1)=U$，$\bar{C}_{A\setminus\{C_2\}}(D_2)=\{x_3,x_4,x_5,x_6,x_7\}$，$\bar{C}_{A\setminus\{C_2\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。可见

$$\sum_{i=1}^{3}Pl_{A\setminus\{C_2\}}(D_i)=\frac{10}{10}+\frac{5}{10}+\frac{4}{10}=\frac{19}{10}$$

则 $\mathrm{Sig}_{A\setminus\{C_2\}}(C_2)=\sum_{i=1}^{3}Pl_{A\setminus\{C_2\}}(D_i)-\sum_{i=1}^{3}Pl_A(D_i)=\frac{1}{10}$。

由覆盖族 $A\setminus\{C_3\}$ 可得 $\bar{C}_{A\setminus\{C_3\}}(D_1)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_{A\setminus\{C_3\}}(D_2)=\{x_3,x_4,x_5,x_6,x_7\}$，$\bar{C}_{A\setminus\{C_3\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。于是

$$\sum_{i=1}^{3}Pl_{A\setminus\{C_3\}}(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{4}{10}=\frac{18}{10}$$

可得 $\mathrm{Sig}_{A\setminus\{C_3\}}(C_3)=\sum_{i=1}^{3}Pl_{A\setminus\{C_3\}}(D_i)-\sum_{i=1}^{3}Pl_A(D_i)=0$。

由覆盖族 $A\setminus\{C_4\}$ 可得 $\bar{C}_{A\setminus\{C_4\}}(D_1)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_{A\setminus\{C_4\}}(D_2)=\{x_3,x_4,x_5,x_6,x_7\}$，$\bar{C}_{A\setminus\{C_4\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。那么

$$\sum_{i=1}^{3}Pl_{A\setminus\{C_4\}}(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{4}{10}=\frac{18}{10}$$

可见 $\mathrm{Sig}_{A\setminus\{C_4\}}(C_4)=\sum_{i=1}^{3}Pl_{A\setminus\{C_4\}}(D_i)-\sum_{i=1}^{3}Pl_A(D_i)=0$。

综上可得，$B=\mathrm{Core}_U(A,D)=\{C_1,C_2\}$。

又由 $\bar{C}_B(D_1)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_B(D_2)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_B(D_3)=\{x_6,x_7,x_8,x_9,x_{10}\}$得

$$\sum_{i=1}^{3}Pl_B(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{9}{10}=\frac{23}{10}\neq M$$

可见，$B=\{C_1,C_2\}$不是上近似约简集。

第3步 由覆盖族 $B\cup\{C_3\}$ 可得 $\bar{C}_{B\cup\{C_3\}}(D_1)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_{B\cup\{C_3\}}(D_2)=\{x_3,x_4,x_5,x_6,x_7\}$，$\bar{C}_{B\cup\{C_3\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。可见

$$\sum_{i=1}^{3}Pl_{B\cup\{C_3\}}(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{4}{10}=\frac{18}{10}$$

于是 $\mathrm{Sig}_B(C_3)=\sum_{i=1}^{3}Pl_B(D_i)-\sum_{i=1}^{3}Pl_{B\cup\{C_3\}}(D_i)=\frac{5}{10}$。

由覆盖族 $B\cup\{C_4\}$ 可得 $\bar{C}_{B\cup\{C_4\}}(D_1)=\{x_1,x_2,x_3,\cdots,x_9\}$，$\bar{C}_{B\cup\{C_4\}}(D_2)=\{x_3,x_4,x_5,x_6,x_7\}$，$\bar{C}_{B\cup\{C_4\}}(D_3)=\{x_6,x_8,x_9,x_{10}\}$。可得

$$\sum_{i=1}^{3}Pl_{B\cup\{C_4\}}(D_i)=\frac{9}{10}+\frac{5}{10}+\frac{4}{10}=\frac{18}{10}$$

则 $\mathrm{Sig}_B(C_4)=\sum_{i=1}^{3}Pl_B(D_i)-\sum_{i=1}^{3}Pl_{B\cup\{C_4\}}(D_i)=\frac{5}{10}$。

选取 $C_3\in A\setminus B$，使 $\mathrm{Sig}_B(C_3)=\max_{C\in A\setminus B}\mathrm{Sig}_B(C)$，作$B=B\cup\{C_3\}$。由 $\sum_{i=1}^{3}Pl_B(D_i)=\frac{18}{10}=M$，可知 $B=\{C_1,C_2,C_3\}$是上近似约简集。

可见，得到的约简是{Price, Structure, Color}，与文献[14]得到的约简{Structure, Color}或{Structure, Surroundings}不同。但是，本文似然函数的计算比文献[14]中信任函数的计算简单。

4 结论

本文讨论了决策为覆盖的覆盖决策信息系统的特征选择，由一对覆盖近似算子生成了一对对偶的信任函数和似然函数，利用信任函数和似然函数给出了覆盖决策信息系统约简的等价刻画，从而给出求约简的算法，并以实例说明算法的有效性。

参考文献：

[1] PAWLAK Z. Rough sets theoretical aspects of reasoning about data[M]. Boston: Kluwer Academic Publishers, 1991.

[2] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356.

[3] 周献中，黄兵，李华雄，等. 不完备信息系统知识获取的粗糙集理论与方法[M]. 南京：南京大学出版社，2010: 1-30.
ZHOU Xianzhong, HUANG Bing, LI Huaxiong, et al. Rough sets theory & approaches for knowledge acquisition in incomplete information systems[M]. Nanjing: Nanjing University Press, 2010: 1-30.

[4] 张文修，吴伟志，梁吉业，等. 粗糙集理论与方法[M]. 北京：科学出版社，2001: 1-25.
ZHANG Wenxiu, WU Weizhi, LIANG Jiye, et al. Rough set theory and methods[M]. Beijing: Science Press, 2001: 1-25.

[5] 梁吉业，李德玉. 信息系统中的不确定性与知识获取[M]. 北京：科学出版社，2005: 1-118.
LIANG Jiye, LI Deyu. The uncertainty and knowledge acquiring in information systems[M]. Beijing: Science Press, 2005: 1-118.

[6] 苗夺谦，李德毅，姚一豫，等. 不确定性与粒计算[M]. 北京：科学出版社，2011: 57-73.
MIAO Duoqian, LI Deyi, YAO Yiyu, et al. Uncertainty and granular computing[M]. Beijing: Science Press, 2011: 57-73.

[7] TSANG E C C, CHEN D G, YEUNG D S. Approximations and reducts with covering generalized

rough sets[J]. Computers and Mathematics with Applications, 2008, 56: 279-289.

[8] CHEN D G, WANG C Z, HU H Q. A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets[J]. Information Sciences, 2007, 177: 3500-3518.

[9] WANG C Z, HE Q, CHEN D G, et al. A novel method for attribute reduction of covering decision systems[J]. Information Sciences, 2014, 254: 181-196.

[10] 杨田.覆盖粗糙集约简理论及应用[D].长沙:湖南大学, 2010.
YANG Tian. The reduction theory of covering rough sets and its application[D]. Changsha: Hunan University, 2010.

[11] WANG C Z, SHAO M W, SUN B Q, et al. An improved attribute reduction scheme with covering based rough sets[J]. Applied Soft Computing, 2015, 26: 235-243.

[12] LI F, YIN Y Q. Approaches to knowledge reduction of covering decision systems based on information theory[J]. Information Sciences, 2009, 179: 1694-1704.

[13] ZHANG X, MEI C L, CHEN D G, et al. Multi-confidence rule acquisition oriented attribute reduction of covering decision systems via combinatorial optimization[J]. Knowledge-Based Systems, 2013, 50: 187-197.

[14] CHEN D G, LI W L, ZHANG X, et al. Evidence-theory-based numerical algorithms of attribute reduction with neighborhood-covering rough sets[J]. International Journal of Approximate Reasoning, 2014, 55: 908-923.

[15] TAN A H, LI J J, LIN G P, et al. Fast approach to knowledge acquisition in covering information systems using matrix operations[J]. Knowledge-Based Systems, 2015, 79: 90-98.

[16] TAN A H, LI J J, LIN Y J, et al. Matrix-based set approximations and reductions in covering decision information systems[J].International Journal of Approximate Reasoning, 2015, 59: 68-80.

[17] CHEN J K, LIN Y J, LIN G P, et al. Attribute reduction of covering decision systems by hypergraph model[J]. Knowledge-Based Systems, 2017, 118: 93-104.

[18] ZHANG Y L, LI J J. The relative reduction of covering generalized rough sets[C]//The 2008 IEEE International Conference on Granular Computing, GrC 2008. Hangzhou, China: IEEE, 2008: 809-812.

[19] 张燕兰, 李进金.覆盖决策系统的相对约简[J].工程数学学报, 2009, 26(5):929-935.
ZHANG Yanlan, LI Jinjin. On relative reduction of knowledge in covering decision systems[J]. Chinese Journal of Engineering Mathematics, 2009, 26(5): 929-935.

[20] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.

[21] SHAFER G. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976.

[22] PAWLAK Z. Rough probability[J]. Bulletin of the Polish Academy of Science, 1984, 32: 607-615.

[23] SKOWRON A, GRZYMALA-BUSSE J. From rough set theory to evidence theory[M]. Wiley, New York: Advance in the Dempster-Shafer Theory of Evidence, 1994: 193-236.

[24] WONG S K M, LINGRAS P J. The compatibility view of Shafer-Dempster theory using the concept of rough set Methodologies for Intelligent Systems[M]. Amsterdan, Netherlands: North Holland, 1989: 33-42.

[25] YAO Y Y, LINGRAS P J. Interpretations of belief functions in the theory of rough sets[J]. Information Sciences, 1998, 104: 81-106.

[26] WU W Z, LEUNG Y, ZHANG W X. Connections between rough set theory and Dempster-Shafer theory of evidence[J]. International Journal of General Systems, 2002, 31: 405-430.

[27] WU W Z, ZHANG M, LI H Z, et al. Knowledge reduction in random information systems via Dempster-Shafer theory of evidence[J]. Information Sciences, 2005, 174: 143-164.

[28] TAN A, WU W, TAO Y. On the belief structures and reductions of multigranulation spaces with decisions[J]. International Journal of Approximate Reasoning, 2017, 88: 39-52.

[29] TAN A, WU W, TAO Y. A unified framework for characterizing rough sets with evidence theory in various approximation spaces[J].Information Sciences, 2018, 454-455: 144-160.

[30] 方连花, 李克典.随机覆盖目标信息系统的属性约简[J].计算机工程与应用, 2014, 50(2): 107-111.
FANG Lianhua, LI Kedian. Attribute reduction in random covering objective information system[J]. Com-

puter Engineering and Applications, 2014, 50(2): 107-111.

[31] 张燕兰，李长清. 基于证据理论的覆盖决策信息系统的属性约简[J]. 模式识别与人工智能，2018，31(9)：797-808.
ZHANG Yanlan, LI Changqing. Attribute reduction of covering decision information system based on evidence theory[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(9): 797-808.

[32] QIN K, GAO Y, PEI Z. On covering rough sets [C]//The Second International Conference on Rough Sets and Knowledge Technology (RSKT 2007, Lecture Notes in Computer Science. [S. l.]: Springer, 2007: 34-41.

[33] ZHANG Y L, LI J J, WU W Z. On axiomatic characterizations of three pairs of covering based approximation operators[J]. Information Sciences, 2010, 180: 274-287.

[34] YAO Y Y, YAO B X. Covering based rough set approximations[J]. Information Sciences, 2012, 200: 91-107.

[35] CHEN D G, ZHANG X X, LI W L. On measurements of covering rough sets based on granules and evidence theory[J]. Information Sciences, 2015, 317: 329-348.

[36] 张文修，仇国芳. 基于粗糙集的不确定性决策[M]. 北京：清华大学出版社，2005：12-24.
ZHANG Wenxiu, QIU Guofang. Uncertain decision making based on rough sets[M]. Beijing: Tsinghua University Press, 2005: 12-24.

（编辑：陈珺）