文章编号:1671 -6833(2014)04 -0109 -03

覆盖粗糙集模型的比较研究

孔素真[1], 金建刚[2], 孙士保[3]

(1. 河南牧业经济学院 信息工程系,河南 郑州 450011;2. 华北水利水电大学 软件学院,河南 郑州 450045;3. 河南科技大学 信息工程学院,河南 洛阳 471023)

摘　要: 在数据元素间不存在等价关系而存在覆盖关系的情况下,应用覆盖的交定义了两种覆盖粗糙集模型.采用比较的方法讨论了两种覆盖粗糙集模型中上、下近似算子的性质,并利用数学理论证明了相应观点的正确性.通过分析明确了两种覆盖粗糙集模型中上、下近似算子之间以及它们与其它粗糙集模型上、下近似算子之间的关系.通过这些研究,扩展了粗糙集理论,补充了知识发现理论与方法,为大型数据库的数据挖掘提供了一个简单的途径.

关键词: 覆盖;下近似算子;上近似算子;知识发现

中图分类号: TP18　　**文献标志码:** A　　doi:10.3969/j. issn. 1671 -6833. 2014. 04. 026

0 引言

Pawlak 提出的粗糙集[1](Rough Sets)理论是集合论的扩展,它成功地应用于不精确性问题的求解,如机器学习和模式识别. Pawlak 粗糙集理论在实际的数据处理过程中,数据元素间的等价关系很难成立.因此,研究工作主要集中在放松等价关系到相似关系[2]、一般二元关系[3]或覆盖关系[4-6]等,以及对应的约简方法[4-6].

在粗糙集模型研究中,文献[4]从对象的最小描述方面把论域上的等价关系推广到论域上的覆盖从而得到变精度覆盖粗糙集模型.文献[5]通过定义覆盖的交,在覆盖决策系统中提出一个删除多余覆盖的方法,并借助不可区分矩阵设计了一种算法计算覆盖决策系统中的所有约简.文献[6]改变了文献[5]中算法复杂度高,时间消耗大的缺点,基于不可区分矩阵设计了一种新型属性约简方法,开发了一种查找最小约简子集的启发式算法,大大降低了计算复杂度.文献[5]与文献[6]都定义了覆盖粗糙集模型,给出相应的约简算法.详细分析得知,它们只是定义了两种不同的模型,而两种模型的关系以及性质如何都没有涉及.因此,笔者在文献[5]与文献[6]定义的基础上讨论两种覆盖粗糙集模型的数学性质,分析两种模型之间的关系,并从数学角度对相应的观点进行证明.

1 基本概念

定义 1[5]　令 U 是论域,$\mathbb{C}=\{C_1,C_2,\cdots,C_n\}$ 是 U 的子集簇.称$\mathbb{C}$ 是 U 的覆盖,如果$\mathbb{C}$ 的子集都不空且 $\cup\mathbb{C}=U$,则$(U,\mathbb{C})$称为覆盖近似空间.

定义 2[5]　令$\mathbb{C}=\{C_1,C_2,\cdots,C_n\}$是 U 上的一个覆盖.对于任意的 $x\in U$,令 $C_x=\cap\{C_j:C_j\in\mathbb{C},x\in C_j\}$.则 $\mathrm{Cov}(\mathbb{C})=\{C_x:x\in U\}$ 也是 U 上的一个覆盖,称它是$\mathbb{C}$ 诱导的覆盖.

定义 3[5]　令 $\Delta=\{\mathbb{C}_i:i=1,2,\cdots,m\}$ 是 U 上的一簇覆盖.对于任意的 $x\in U$,令 $\Delta_x=\cap\{C_{ix}:C_{ix}\in\mathrm{Cov}(\mathbb{C}_i),x\in C_{ix}\}$,则 $\mathrm{Cov}(\Delta)=\{\Delta_x:x\in U\}$ 也是 U 上的一个覆盖,称它是 Δ 诱导的覆盖.

定义 4[5]　设$(U,\mathbb{C})$为覆盖近似空间,U 是有限非空的论域,$\mathbb{C}$ 是 U 的一个子集簇.对于 $\forall X\subseteq U$,X 关于覆盖 $\mathrm{Cov}(\Delta)$的下、上近似定义为

$$\underline{\Delta}(X)=\cup\{\Delta_x:\Delta_x\subseteq X\},\tag{1}$$

$$\overline{\Delta}(X)=\cup\{\Delta_x:\Delta_x\cap X\neq\varnothing\}.\tag{2}$$

覆盖粗糙集模型有两种表现形式,在上一种

收稿日期: 2014 -03 -07;**修订日期:** 2014 -05 -10

基金项目: 河南省重点科技攻关计划资助项目(122102210487);河南省国际科技合作计划资助项目(124300510040);洛阳市科技发展计划项目(1301008A、1301034A).

通信作者: 孙士保(1970 -),男,河南固始人,河南科技大学教授,博士,研究方向:信息处理,E-mail: sunshibao@126.com.

形式之外有另一种描述方式,具体见定义5,笔者主要分析这两种描述形式的区别与联系,为它们在不同领域应用指明方向.

定义5[6] 设$(U,\mathbb{C})$为覆盖近似空间,U是有限非空的论域,$\mathbb{C}$是U的一个子集簇.对于$\forall X\subseteq U$,X关于覆盖Cov(Δ)的下、上近似定义为

$$\underline{\Delta}'(X)=\{x\in U:\Delta_x\subseteq X\},\tag{3}$$

$$\overline{\Delta}'(X)=\{x\in U:\Delta_x\cap X\neq\varnothing\}.\tag{4}$$

从上面定义4、定义5看出,当$\mathbb{C}$是U的一簇覆盖时,它们又具有不同的性质.

2 覆盖粗糙集模型的基本性质

定理1 设$(U,\mathbb{C})$为覆盖近似空间,其中U是有限非空的论域,$\mathbb{C}$是U的一个子集簇. $\sim X=U-X$,对于$\forall X\subseteq U$,有

(1)$\underline{\Delta}(X)\subseteq X, X\subseteq\overline{\Delta}(X)$;

(2)$\underline{\Delta}(\sim X)=\sim\overline{\Delta}(X)$, $\sim\underline{\Delta}(X)=\overline{\Delta}(\sim X)$;

(3)$\underline{\Delta}(\varnothing)=\varnothing, \overline{\Delta}(\varnothing)=\varnothing$;

(4)$\underline{\Delta}(U)=U, \overline{\Delta}(U)=U$;

(5)$X\subseteq Y\Rightarrow\underline{\Delta}(X)\subseteq\underline{\Delta}(Y), \overline{\Delta}(X)\subseteq\overline{\Delta}(Y)$;

(6)$\underline{\Delta}(X\cap Y)\subseteq\underline{\Delta}(X)\cap\underline{\Delta}(Y)$,
$\overline{\Delta}(X\cup Y)\supseteq\overline{\Delta}(X)\cup\overline{\Delta}(Y)$,
$\underline{\Delta}(X\cup Y)\supseteq\underline{\Delta}(X)\cup\underline{\Delta}(Y)$,
$\overline{\Delta}(X\cap Y)\subseteq\overline{\Delta}(X)\cap\overline{\Delta}(Y)$.

(7)$\underline{\Delta}(\underline{\Delta}(X))=\underline{\Delta}(X), \overline{\Delta}(\overline{\Delta}(X))=\overline{\Delta}(X)$.

证明:(1)对于$\forall x\in U, x\in C_x$,且$C_x\in\mathrm{Cov}(\mathbb{C})$,则$x\in\Delta_x$.而对于$\forall x\in\underline{\Delta}(X)$,有$x\in\Delta_x$且$\Delta_x\subseteq X$,从而$x\in X$,则$\underline{\Delta}(X)\subseteq X$.

对于$\forall x\in X$,且$x\in\Delta_x$,则$\Delta_x\cap X\neq\varnothing$,即$x\in\overline{\Delta}(X)$,故$X\subseteq\overline{\Delta}(X)$;

(2)对于$\forall x\in U$,如果$x\in\underline{\Delta}(\sim X)$,则$x\in\cup\{\Delta_x:\Delta_x\subseteq U-X\}$,进而$x\notin\cup\{\Delta_x:\Delta_x\cap X\neq\varnothing\}$,即$x\notin\overline{\Delta}(X)$,$x\in\sim\overline{\Delta}(X)$,有$\underline{\Delta}(\sim X)\subseteq\sim\overline{\Delta}(X)$.

对于$\forall x\in U$,如果$x\in\sim\overline{\Delta}(X)$,则$x\notin\overline{\Delta}(X)$,即$x\notin\cup\{\Delta_x:\Delta_x\cap X\neq\varnothing\}$,可知$x\in\cup\{\Delta_x:\Delta_x\subseteq U-X\}$,即$x\in\underline{\Delta}(\sim X)$,有$\sim\overline{\Delta}(X)\subseteq\underline{\Delta}(\sim X)$.

因此,$\underline{\Delta}(\sim X)=\sim\overline{\Delta}(X)$,另式证明类似.

(3)对于$\forall x\in U$,如果$x\in\underline{\Delta}(\varnothing)$,有$x\in\Delta_x$且$\Delta_x\subseteq\varnothing$,从而$x\in\varnothing$,则$\underline{\Delta}(X)\subseteq\varnothing$.由于$\varnothing\subseteq\underline{\Delta}(X)$,则$\underline{\Delta}(\varnothing)=\varnothing$.另式证明类似.

(4)对于$\forall x\in U, x\in C_x$,且$C_x\in Cov(\mathbb{C})$,则$x\in\Delta_x$.由于$\Delta_x\subseteq U$,则$x\in\underline{\Delta}(U)$,因此$U\subseteq\underline{\Delta}(U)$.又由于$\underline{\Delta}(U)\subseteq U$,则$\underline{\Delta}(U)=U$.类似$\overline{\Delta}(U)=U$.

(5)对于$\forall x\in U$,如果$x\in\underline{\Delta}(X)$,即$x\in\cup\{\Delta_x:\Delta_x\subseteq X\}$,当$x$确定好后,$\Delta_x$值是固定的,由$X\subseteq Y$可知$x\in\cup\{\Delta_x:\Delta_x\subseteq Y\}$,即$x\in\underline{\Delta}(Y)$,故$X\subseteq Y\Rightarrow\underline{\Delta}(X)\subseteq\underline{\Delta}(Y)$.

类似可证$\overline{\Delta}(X)\subseteq\overline{\Delta}(Y)$.

(6)对于$\forall x\in U$,如果$x\in\underline{\Delta}(X\cap Y)$,则$x\in\cup\{\Delta_x:\Delta_x\subseteq X\cap Y\}$,当$x$确定好后,$\Delta_x$值是固定的,可知$x\in\cup\{\Delta_x:\Delta_x\subseteq X\}$且$x\in\cup\{\Delta_x:\Delta_x\subseteq Y\}$,所以可以得到$x\in(\cup\{\Delta_x:\Delta_x\subseteq X\})\cap(\cup\{\Delta_x:\Delta_x\subseteq Y\})$,故$\underline{\Delta}(X\cap Y)\subseteq\underline{\Delta}(X)\cap\underline{\Delta}(Y)$.

同理可证:$\overline{\Delta}(X\cup Y)\supseteq\overline{\Delta}(X)\cup\overline{\Delta}(Y)$,$\underline{\Delta}(X\cup Y)\supseteq\underline{\Delta}(X)\cup\underline{\Delta}(Y)$,$\overline{\Delta}(X\cap Y)\subseteq\overline{\Delta}(X)\cap\overline{\Delta}(Y)$.

(7)对于$\forall x\in U$,如果$x\in\underline{\Delta}(\underline{\Delta}(X))$,即$x\in\cup\{\Delta_x:\Delta_x\subseteq\underline{\Delta}(X)\}$,对于$\Delta_x\subseteq\underline{\Delta}(X)$,所以$\cup\{\Delta_x:\Delta_x\subseteq\underline{\Delta}(X)\}\subseteq\underline{\Delta}(X)$,有$x\in\underline{\Delta}(X)$,从而$\underline{\Delta}(\underline{\Delta}(X))\subseteq\underline{\Delta}(X)$.

对于$\forall x\in U$,如果$x\in\underline{\Delta}(X)$,由于$\underline{\Delta}(X)=\cup\{\Delta_x:\Delta_x\subseteq X\}$,有$x\in\cup\{\Delta_x:\Delta_x\subseteq\underline{\Delta}(X)\}$,从而$x\in\underline{\Delta}(\underline{\Delta}(X))$,$\underline{\Delta}(X)\subseteq\underline{\Delta}(\underline{\Delta}(X))$.

因此$\underline{\Delta}(\underline{\Delta}(X))=\underline{\Delta}(X)$.另式证明类似.

定理2 设$(U,\mathbb{C})$为覆盖近似空间,其中U是有限非空的论域,$\mathbb{C}$是U的一个子集簇. $\sim X=U-X$,对于$\forall X\subseteq U$,有

(1)$\underline{\Delta}'(X)\subseteq X, X\subseteq\overline{\Delta}'(X)$;

(2)$\underline{\Delta}'(\sim X)=\sim\overline{\Delta}'(X)$, $\sim\underline{\Delta}'(X)=\overline{\Delta}'(-X)$;

(3)$\underline{\Delta}'(\varnothing)=\varnothing, \overline{\Delta}'(\varnothing)=\varnothing$;

(4)$\underline{\Delta}'(U)=U, \overline{\Delta}'(U)=U$;

(5)$X\subseteq Y\Rightarrow\underline{\Delta}'(X)\subseteq\underline{\Delta}'(Y), \overline{\Delta}'(X)\subseteq\overline{\Delta}'(Y)$;

(6)$\underline{\Delta}'(X\cap Y)\subseteq\underline{\Delta}'(X)\cap\underline{\Delta}'(Y)$,
$\overline{\Delta}'(X\cup Y)\supseteq\overline{\Delta}'(X)\cup\overline{\Delta}'(Y)$;
$\underline{\Delta}'(X\cup Y)\supseteq\underline{\Delta}'(X)\cup\underline{\Delta}'(Y)$,
$\overline{\Delta}'(X\cap Y)\subseteq\overline{\Delta}'(X)\cap\overline{\Delta}'(Y)$,

(7)$\underline{\Delta}'(\underline{\Delta}'(X))=\overline{\Delta}'(X), \overline{\Delta}'(\overline{\Delta}'(X))=\overline{\Delta}'(X)$.

证明:类似定理1.

3 两种覆盖粗糙集模型区别与联系

(1)$\underline{\Delta}(X)$与$\underline{\Delta}'(X)$、$\overline{\Delta}(X)=\overline{\Delta}'(X)$是覆盖粗糙集模型的两种表示形式,其中下近似确定了肯定包含于被近似对象的元素,上近似确定了可能包含于被近似对象的元素,它们为数据关系不满足等价性而满足覆盖的不确定性知识发现问题提供了一种新的思路.

(2)设$(U,\mathbb{C})$为覆盖近似空间,其中U是有限非空的论域,当$\mathbb{C}$是U的一个划分时,C_x和Δ_x

是包含 x 的等价类,即 $C_x = \Delta_x = [x]_R$. 这样 $\underline{\Delta}(X)$ 与 $\underline{\Delta}'(X)$、$\overline{\Delta}(X) = \overline{\Delta}'(X)$ 等价,并且是经典粗糙集模型的两种不同表示形式[1]. 因此,涉及的两种覆盖粗糙集模型是经典粗糙集模型的扩展,是它的一般形式.

(3)设 $(U, \mathbb{C})$ 为覆盖近似空间,其中 U 是有限非空的论域,$\mathbb{C}$ 是 U 的一个子集簇. $\sim X = U - X$,对于 $\forall X \subseteq U$,$\underline{\Delta}'(X) \subseteq \underline{\Delta}(X)$,$\overline{\Delta}'(X) \subseteq \overline{\Delta}(X)$.

证明:对于 $\forall x \in U$,$x \in C_x$,且 $C_x \in Cov(\mathbb{C})$,则 $x \in \Delta_x$. 对于 $\forall x \in \underline{\Delta}'(X)$,总有 $x \in \Delta_x$,也就有 $x \in \underline{\Delta}(X)$. 因此,$\overline{\Delta}'(X) \subseteq \overline{\Delta}(X)$. 同理可证.

(4)设 $(U, \mathbb{C})$ 为覆盖近似空间,其中 U 是有限非空论域,$\mathbb{C}$ 是 U 的一个子集簇. 对于 $\forall X \subseteq U$,$\underline{\Delta}'(X)$ 与 $\overline{\Delta}'(X)$ 用于对象元素比较分散的数据集进行处理,它以单个数据元素为基础聚类数据;而 $\underline{\Delta}(X)$ 与 $\overline{\Delta}(X)$ 可以很好地对块状数据进行处理,因为它是以 Δ_x 数据块(对象的覆盖类数据块)为基础来聚类数据,可以大大提高数据处理的效率.

4　结论

粗糙集理论在不确定性知识发现中发挥着重要的作用. 在经典的粗糙集理论中,决策表中的每一个条件属性诱导一个划分. 当这一条件不能满足时,它限制了粗糙集理论的应用. 因此,为了拓展粗糙集理论的应用范围,把数据对象之间的等价关系扩展到覆盖关系. 笔者通过研究文献[5]与文献[6]中的覆盖粗糙集模型的定义,讨论了两种覆盖粗糙集模型中上、下近似算子的性质、证明了相应观点的正确性、分析了两种覆盖粗糙集模型中上、下近似算子之间以及与其它粗糙集模型上、下近似算子之间的关系. 通过这些研究,扩展了粗糙集理论,完善了知识发现理论与方法,为大型数据库的数据挖掘提供了一个简单的途径.

参考文献:

[1] PAWLAK Z, SKOWRON A. Rough sets: Some extensions[J]. Information Sciences, 2007(177): 28-40.

[2] SLOWINSKI R, VANDERPOOTEN D. A generalized definition of rough approximations based on similarity[J]. IEEE Transactions on Data and Knowledge Engineering, 2000, 2: 331-336.

[3] HU Q, YU D, LIU J, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178: 3577-3594.

[4] ZHU W, WANG F Y. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences, 2003, 152: 217-230.

[5] CHEN De-gang, WANG Chang-zhong, HU Qing-hua. A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets[J]. Information Sciences, 2007, 177: 3500-3518.

[6] 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.

A Comparative Study of Covering Rough Set Model

KONG Su-zhen[1], JIN Jian-gang[2], SUN Shi-bao[3]

(1. Department of Information Engineering, Henan University of Animal Husbandry and Economy, Zhengzhou 450011, China; 2. Software College, North China University of Water Resources & Electric Power, Zhengzhou 450045, China; 3. Information Engineering College, Henan University of Science & Technology, Luoyang 471023, China)

Abstract: By means of intersection of covering, two kinds of covering rough set models are defined in the case of absence of equivalence relations in addition to covering relations between the data elements. The properties of lower approximation operator and upper approximation operator are discussed by using comparable methods in two covering rough set model, and the correctness of the corresponding point is proved by using mathematical theory. Through analysis of the definitions of two kinds of covering rough set models, the relationships of the lower approximation operator and upper approximation operator and with other rough set models are illustrated. Through these researches, the rough set theory is expanded, the theory and method of knowledge discovery is improved, and a simple way is provided for data mining of large databases.

Key words: cover; lower approximation operator; upper approximation operator; knowledge discovery