DOI:10.11992/tis.201606011
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20160808.0830.006.html

基于证据理论刻画多粒度覆盖粗糙集的数值属性

车晓雅[1],李磊军[1,2],米据生[1,2]
(1.河北师范大学 数学与信息科学学院,河北 石家庄 050024; 2.河北省计算数学与应用重点实验室,河北 石家庄 050024)

摘　要:在经典多粒度粗糙集模型的基础上,基于论域中对象的极大描述和极小描述,定义了 4 种应用更为广泛的悲观多粒度覆盖粗糙集模型。然后通过集合的交、并运算与关系划分函数,构造了对象关于覆盖族的单粒度的多元覆盖及单粒度划分。在此基础上,基于证据理论,探讨了 4 种悲观多粒度覆盖粗糙集的上、下近似与信任函数和似然函数之间关系,并描述了该模型所具备的相关数值属性。对比分析表明悲观多粒度覆盖粗糙集模型既具备经典多粒度粗糙集模型能够融合多源信息的优势,又克服了其应用范围狭窄的缺点。实例分析验证了所提模型的有效性。

关键词:粗糙集理论;覆盖;粒度;证据理论;近似;特性描述
中图分类号: TP18　　**文献标志码**:A　　**文章编号**:1673-4785(2016)04-0481-06

中文引用格式:车晓雅,李磊军,米据生. 基于证据理论刻画多粒度覆盖粗糙集的数值属性[J]. 智能系统学报, 2016, 11(4): 481-486.
英文引用格式:CHE Xiaoya, LI Leijun, MI Jusheng. Evidence-theory-based numerical characterization of multi-granulation covering rough sets[J]. CAAI Transactions on Intelligent Systems, 2016, 11(4): 481-486.

Evidence-theory-based numerical characterization of multi-granulation covering rough sets

CHE Xiaoya[1], LI Leijun[1,2], MI Jusheng[1,2]
(1.College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China; 2. Hebei Key Laboratory of Computational Mathematics and Applications, Shijiazhuang 050024, China)

Abstract:Considering classical multi-granulation rough sets and using the maximal and minimal descriptors of objects in a given universe, this paper proposes four pessimistic multi-granulation covering rough set models, suitable for extensive application. Based on set union and portion functions, the notion of multi-granularity covering connected to a number of coverings and a single granularity partition in the domain are defined. On this basis, belief and plausibility functions from evidence theory are employed to define the relationship between the upper and lower approximations, the belief function, and the likelihood function, and to characterize the set approximations in the four models. Compared with classical multi-granulation rough sets, the pessimistic multi-granulation covering rough set models not only have distinct advantages and combine multi-source information, but also avoid the shortcomings of a narrow application range. Finally, a real example is used to demonstrate the effectiveness of the presented models.

Keywords: rough sets theory; covering; granulation; evidence theory; approximation; characterization

粗糙集理论由 Pawlak[1] 于 1982 年提出,是一种有效处理模糊和不确定性知识的数学工具,其在机器学习、模式识别、决策分析和数据挖掘等领域得到广泛应用[2-5]。经典粗糙集理论基于等价关系定义集合的上、下近似,然而随着现实世界中的数据在结构和形式上日益复杂化和多样化,经典粗糙集有时不再能满足实际问题的处理需求。为此众多学者从不同角度对经典粗糙集模型进行了扩展[5-8],提出了覆盖粗糙集、多粒度粗糙集、变精度粗糙集、概率粗糙集、模糊粗糙集等。其中,覆盖粗糙集是将经

收稿日期:2016-06-03.　　**网络出版日期**:2016-08-08.
基金项目:国家自然科学基金项目(61573127,61502144,61300121,61472463);河北省自然科学基金项目(A2014205157);河北省高校创新团队领军人才培育计划项目(LJRC022);河北省高校自然科学基金项目(QN2016133);河北师范大学博士科学基金项目(L2015B01);河北省教育厅研究生创新项目(sj2015001).
通信作者:车晓雅.E-mail:chexiaoya@163.com.

典粗糙集中的划分推广成更一般的覆盖,增强了其处理数据的能力[7,9]。

从粒计算的角度来看,Pawlak 粗糙集及推广形式都是基于单一二元关系,均可被称做单粒度粗糙集。然而,在许多实际应用中,需要由多个二元关系诱导出的多粒度结构对目标概念进行刻画。为此,钱宇华等[8,10]提出了基于全域中多个等价关系的经典多粒度粗糙集模型。苗夺谦等[11]在覆盖近似空间中提出 4 种乐观多粒度覆盖粗糙集模型,其中集合的第一、二型近似分别基于论域中对象极小描述的交和并,集合的第三、四型近似分别基于论域中对象极大描述的交和并。

另一方面,Dempster-Shafer(DS)证据理论产生自 20 世纪 60 年代。Dempster[12]提出了集值映射的概念,并定义了上、下概率。随后, Shafer[13]用信度函数对上、下概率重新进行诠释, 创立"证据的数学理论"。Dempster 还定义了著名的 Dempster 证据组合规则,该理论中的基本概念是信度函数,包括信任函数和似然函数,并以此来度量知识的不确定性。与粗糙集理论相类似,证据理论也是一种处理不确定性的有力工具[14-16]。许多专家对粗糙集和证据理论之间的关系进行了研究和推广。姚一豫[17]指出可以用信任函数和似然函数对粗糙集中的上、下近似算子进行解读;吴伟志等[16]将信任结构与近似空间相结合,从证据理论的角度研究 Pawlak 粗糙集的知识约简;陈德刚等[18]在统一框架下对若干覆盖近似算子进行分类,基于粒和证据理论对这些覆盖粗糙近似算子进行度量,并且用信任函数和似然函数对邻域覆盖粗糙集中上、下近似算子进行了度量,进而建立了上述函数与邻域信息系统属性约简之间的关系[19]。

将证据理论与多粒度粗糙集模型相结合是目前的研究热点之一[20-21],谭安辉[22]基于证据理论刻画了不完备信息系统中多粒度粗糙集的数值属性,指出只有悲观多粒度粗糙集的数值属性可以由信任结构刻画,并构建了一种多粒度粗糙集的属性约简算法;林国平[14]结合证据理论和多粒度粗糙集,提出一种新的融合多源信息的方法。然而,上述研究都没有考虑过如何构建多粒度覆盖粗糙集的信任结构以及如何用证据理论刻画多粒度覆盖粗糙集的数值属性。基于上述启发,本文首先在苗夺谦等[11]提出的 4 种乐观多粒度覆盖粗糙集模型的基础上定义 4 种悲观多粒度覆盖粗糙集模型,然后基于证据理论给出多粒度覆盖粗糙集的信任结构。通过集合的交运算和关系划分函数建立多粒度覆盖与单粒度划分之间的关系,进而建立了多粒度覆盖粗糙集和证据理论之间联系。

1 相关概念

1.1 Pawlak 粗糙集相关概念

定义 1 令 $S=(U,C\cup D,V_a,f)$ 是信息系统[1],其中 $U=\{x_1,x_2,x_3,\cdots,x_n\}$ 是非空有限对象集,称为论域; C 是条件属性集, D 是决策属性集, $V=\bigcup_{a\in A}V_a$, V_a 是属性 a 的属性值, $f:U\times C\cup D\to V$ 是一个信息函数,它指定 U 中每一个对象的属性值,即 $\forall a\in A,\ x\in U, f(x,a)\in V_a$。通常用 $S=(U,A)$ 代替 $S=(U,C\cup D,V_a,f)$ 。

$\forall B\subseteq C$ 决定一个二元不可辨识关系[1] R_B ,定义为

$$R_B=\{(x,y)\in U\times U\mid \forall a\in B,f(x,a)=f(y,a)\}$$

显然, R_B 是集合 U 上的等价关系,对于 $B\subseteq C$,关系 R_B 产生 U 的一个划分 $U/R_B=\{[x]_B\mid x\in U\}$,即 $[x]_B=\{y\in U\mid (x,y)\in R_B\}$ 。

定义 2 令 $S=(U,R)$ 为近似空间, R 是 U 上的等价关系, $\forall X\subseteq U$,称 $\underline{R}(X)$ 和 $\bar{R}(X)$ 为 X 关于 R 的上、下近似[1],如果

$$\underline{R}(X)=\{x\in U\mid [x]_R\subseteq U\}$$

$$\bar{R}(X)=\{x\in U\mid [x]_R\cap U\neq\varnothing\}$$

若 $\underline{R}(X)\neq\bar{R}(X)$,称 X 为粗糙集[1],否则称 X 为可定义集[1]。

1.2 覆盖粗糙集相关概念

定义 3 令 U 为论域, C 是 U 的一族子集。如果 $\varnothing\notin C$ 且 $\cup C=U$,则称 C 是 U 的一个覆盖[23];称序对 (U,C) 是覆盖近似空间[23]。

定义 4 令 (U,C) 为覆盖近似空间,其中 $C=\{C_1,C_2,\cdots,C_p\}$ 。$\forall X\subseteq U$,称 $\underline{C}(X)$ 和 $\bar{C}(X)$ 为 X 关于 C 的上、下近似[23],如果

$$\underline{C}(X)=\cup\{C_i\subseteq X,i\in\{1,2,\cdots,p\}\}$$

$$\bar{C}(X)=\cup\{C_i\cap X\neq\varnothing,i\in\{1,2,\cdots,p\}\}$$

定义 5 令 (U,C) 为覆盖近似空间, $\forall x\in U$, 集合 $\mathrm{md}_C(x)$ 和 $\mathrm{MD}_C(x)$ 分别称 x 关于 C 的极小描述和极大描述[11],如果

$$\mathrm{md}_C(x)=\{K\in C\mid x\in K\wedge(\forall S\in C\wedge x\in S\wedge S\subseteq K\Rightarrow K=S)\}$$

$$\mathrm{MD}_C(x)=\{K\in C\mid x\in K\wedge(\forall S\in C\wedge x\in S\wedge K\subseteq S\Rightarrow K=S)\}$$

1.3 多粒度粗糙集相关概念

下面简要给出多粒度粗糙集的两种模型,即乐

观多粒度粗糙集和悲观多粒度粗糙集。

定义 6　令 $S=(U,A)$ 是信息系统，A 是属性集合，$A_1,A_2,\cdots,A_m\subseteq A$，其中 m 是自然数。$\forall X\subseteq U$，X 关于 $A_1,A_2,\cdots,A_m$ 的乐观多粒度上、下近似[8,10]为

$$\underline{\sum_{i=1}^{m}A_i^{O}(X)}=\{x\in U\mid [x]_{A_1}\subseteq X\vee [x]_{A_2}\subseteq X\vee\cdots\vee [x]_{A_m}\subseteq X\}$$

$$\overline{\sum_{i=1}^{m}A_i^{O}(X)}=\sim\underline{\sum_{i=1}^{m}A_i^{O}(\sim X)}$$

式中 $\sim X=U-X$。

定义 7　令 $S=(U,A)$ 是信息系统，A 是属性集合，$A_1,A_2,\cdots,A_m\subseteq A$，其中 m 是自然数。$\forall X\subseteq U$，X 关于 $A_1,A_2,\cdots,A_m$ 的悲观多粒度上、下近似[8,10]为

$$\underline{\sum_{i=1}^{m}A_i^{P}(X)}=\{x\in U\mid [x]_{A_1}\subseteq X\wedge [x]_{A_2}\subseteq X\wedge\cdots\wedge [x]_{A_m}\subseteq X\}$$

$$\overline{\sum_{i=1}^{m}A_i^{P}(X)}=\sim\underline{\sum_{i=1}^{m}A_i^{P}(\sim X)}$$

式中 $\sim X=U-X$。

2.4　证据理论相关概念

定义 8　令 U 为论域，2^U 是 U 的全体子集，集函数 $m:2^U\to[0,1]$ 称为概率指派函数[22]，即 mass 函数，如果

1）$m(\varnothing)=0$；

2）$\sum_{X\subseteq U}m(X)=1$。

若 $m(X)\neq 0$，则称 X 为 m 的焦元。

定义 9　令 U 为论域，$m:2^U\to[0,1]$ 是一个基本概率指派函数。集函数 $\mathrm{Bel}:2^U\to[0,1]$ 称为 U 上的信任函数[19]，如果 $\mathrm{Bel}(X)=\sum_{X'\subseteq X}m(X')$，$\forall X\subseteq 2^U$。集函数 $P_l:2^U\to[0,1]$ 称为 U 上的似然函数[19]，如果 $P_l(X)=\sum\limits_{X'\cap X\neq\varnothing}m(X')$，$\forall X\subseteq 2^U$。

基于相同的概率指派函数，信任函数和似然函数是对偶的，即 $\mathrm{Bel}(X)=\sim P_l(\sim X)$，其中 $\sim X=U-X$。

信任函数满足下列性质：

1）$\mathrm{Bel}(\varnothing)=0$；

2）$\mathrm{Bel}(U)=1$；

3）$\mathrm{Bel}(\cup_{i=1}^{m}X_i)\geqslant\sum_{J\subset\{1,2,\cdots,M\}}(-1)^{|J|+1\times}\mathrm{Bel}(\cap_{i\in J}X_i)$　$\forall X_1,X_2,\cdots,X_m\subseteq U$。

2　多粒度覆盖粗糙集（MGCRS）

本节选取苗夺谦等[11]提出的 4 种乐观多粒度覆盖粗糙集模型。上述模型基于论域中极小描述或极大描述的交或并定义。

谭安辉等[22]指出，在信息系统中，集合的悲观多粒度近似可以由信度函数刻画，但是集合的乐观多粒度近似一般不具备这种特性。因而本文首先基于苗夺谦等[11]提出的 4 种乐观多粒度覆盖粗糙集模型定义悲观多粒度覆盖粗糙集模型。

对于给定的近似空间 $\langle U,C\rangle$，$\forall x\in U$，x 的极小描述包含了近似空间中与 x 相关的核心对象，当讨论近似空间 $\langle U,C\rangle$ 中集合近似的问题时，极小描述可以提供关于 x 简单且关键的概括。

定义 10　令 $\langle U,C\rangle$ 为多粒度覆盖近似空间，$C_1,C_2,\cdots,C_m\in C$，其中 m 是自然数。$\forall X\subseteq U$，其关于 $C_1,C_2,\cdots,C_m$ 的第一型上、下近似定义如下：

$$\underline{\mathrm{FR}}^{P}_{\sum_{i=1}^{m}C_i}(X)=\{x\in U\mid\cap\,\mathrm{md}_{C_1}(x)\subseteq X\wedge\cap\,\mathrm{md}_{C_2}(x)\subseteq X\wedge\cdots\wedge\cap\,\mathrm{md}_{C_m}(x)\subseteq X\}$$

$$\overline{\mathrm{FR}^{P}_{\sum_{i=1}^{m}C_i}(X)}=\sim\underline{\mathrm{FR}}^{P}_{\sum_{i=1}^{m}C_i}(\sim X)$$

定义 11　令$\langle U,C\rangle$为多粒度覆盖近似空间，$C_1,C_2,\cdots,C_m\in C$，其中 m 是自然数。$\forall X\subseteq U$，其关于 $C_1,C_2,\cdots,C_m$ 的第二型上、下近似定义如下：

$$\underline{\mathrm{SR}}^{P}_{\sum_{i=1}^{m}C_i}(X)=\{x\in U\mid\cup\,\mathrm{md}_{C_1}(x)\subseteq X\wedge\cup\,\mathrm{md}_{C_2}(x)\subseteq X\wedge\cdots\wedge\cup\,\mathrm{md}_{C_m}(x)\subseteq X\}$$

$$\overline{\mathrm{SR}^{P}_{\sum_{i=1}^{m}C_i}(X)}=\sim\underline{\mathrm{SR}}^{P}_{\sum_{i=1}^{m}C_i}(\sim X)$$

x 的极大描述包含近似空间中所有与 x 相关的对象，当讨论近似空间 $\langle U,C\rangle$ 中集合近似的问题时，极大描述可以提供一个详细且综合的对于 x 的概括。

定义 12　令 $\langle U,C\rangle$ 为多粒度覆盖近似空间，$C_1,C_2,\cdots,C_m\in C$，其中 m 是自然数。$\forall X\subseteq U$，其关于 $C_1,C_2,\cdots,C_m$ 的第三型上、下近似定义如下：

$$\underline{\mathrm{TR}}^{P}_{\sum_{i=1}^{m}C_i}(X)=\{x\in U\mid\cap\,\mathrm{MD}_{C_1}(x)\subseteq X\wedge\cap\,\mathrm{MD}_{C_2}(x)\subseteq X\wedge\cdots\wedge\cap\,\mathrm{MD}_{C_m}(x)\subseteq X\}$$

$$\overline{\mathrm{TR}^{P}_{\sum_{i=1}^{m}C_i}(X)}=\sim\underline{\mathrm{TR}}^{P}_{\sum_{i=1}^{m}C_i}(\sim X)$$

定义 13　令 $\langle U,C\rangle$ 为多粒度覆盖近似空间，$C_1,C_2,\cdots,C_m\in C$，其中 m 是自然数。$\forall X\subseteq U$，，其中 m 是自然数。$\forall X\subseteq U$，其关于 $C_1,C_2,\cdots,C_m$ 的第 4 型上、下近似定义如下：

$$\underline{\mathrm{LR}}^{P}_{\sum_{i=1}^{m}C_i}(X)=\{x\in U\mid\cup\,\mathrm{MD}_{C_1}(x)\subseteq X\wedge\cup\,\mathrm{MD}_{C_2}(x)\subseteq X\wedge\cdots\wedge\cup\,\mathrm{MD}_{C_m}(x)\subseteq X\}$$

$$\overline{\mathrm{LR}^{P}_{\sum_{i=1}^{m}C_i}}(X) = \sim \underline{\mathrm{LR}}^{P}_{\sum_{i=1}^{m}C_i}(\sim X)$$

显然,如果 C 是 U 上的一族划分,上述 4 种多粒度覆盖粗糙集模型将退化为经典悲观多粒度粗糙集模型。因此,上述 4 种模型是对经典多粒度粗糙集模型的推广,并且也是粗糙集模型和覆盖粗糙集模型的推广。

3 MGCRS 与证据理论之间联系

本节讨论证据理论和多粒度覆盖粗糙集之间的联系。由于基于信任结构所导出的信任函数和似然函数是度量多粒度覆盖粗糙集中上、下近似的基础,因此首先给出上述 4 种类型悲观多粒度覆盖粗糙集模型相应的信任函数和似然函数来度量集合的上、下近似(记作 $\underline{\nabla}_j,\overline{\nabla_j}(j=1,2,3,4)$)。

下面假设 P 是一个平均概率分布,即 $\forall X\subseteq U$, $P(x)=|X|\times|U|^{-1}$, $|\cdot|$ 是集合的势。

定义 14 令 $S=(U,A)$ 为信息系统, $C=\{C_1,C_2,\cdots,C_m\}$ 是 U 上一族覆盖, $\forall x\in U$,用 $\nabla_{ij}(x)$ $(i=1,2,\cdots,m;j=1,2,3,4)$ 分别表示 $\cap\,\mathrm{md}_{c_i}(x)$ 、$\cup\,\mathrm{md}_{c_i}(x)$ 、$\cap\,\mathrm{MD}_{c_i}(x)$ 、$\cup\,\mathrm{MD}_{c_i}(x)$,称 $\nabla_j(x)=\cup\{\nabla_{ij}(x)\mid i=1,2,\cdots,m;j=1,2,3,4\}$ 为 x 关于覆盖族 C 的多元覆盖。

$\forall X\subseteq U,x\in U$, X 在上述 4 种悲观多粒度覆盖粗糙集模型中上、下近似的定义分别基于论域中 x 极小描述或极大描述的交或并所得 x 的相关元。因为 X 定义于多粒度环境中,所以无论 x 的极小描述还是 x 的极大描述均同时与覆盖 $C_1,C_2,\cdots,C_m$ 相关。即 $\forall j\in\{1,2,3,4\}$ 如果集合 $\{(\nabla_{ij}(x))i=1,2,\cdots,m\}$ 中所有元均为 X 子集,则 x 属于 $\underline{\nabla}_j(X)$,如果至少存在一个 $\nabla_{ij}(x)$ 与 X 相交不为空,则 x 属于 $\overline{\nabla_j}(X)$。鉴于此, $\forall x\in U$,本文对集族 $\{\nabla_{ij}(x)\mid i=1,2,\cdots,m\}$ 取并集后得 $\nabla_j(x)$, $\nabla_j(x)$ 是论域 U 上一个单粒度覆盖,从而悲观多粒度覆盖粗糙集转化为单粒度覆盖粗糙集。进一步,定理 1 借助关系划分函数,将覆盖与划分建立联系,将覆盖粗糙集转化为经典粗糙集,进而在定理 2 中得出证据理论与多粒度悲观覆盖粗糙集之间联系。

定理 1 令 U 为论域, C 为 U 上覆盖, $\forall x\in U$, $j\in\{1,2,3,4\}$,定义关系划分函数 $f_j:C\to U$, $f_j(X)=\{x\in U\mid X=\nabla_j(x),j=1,2,3,4\}$, $\forall x\in U$,则 $f_j(X)$ 是 U 上的一个划分。

证明 $\forall j\in\{1,2,3,4\}$,首先须证, $\forall X,X'\subseteq U,X\neq X'$,有 $f_j(X)\cap f_j(X')\neq\varnothing$ 成立。

假设, $\exists x\in U$,使 $x\in f_j(X)\cap f_j(X')$,则有 $\nabla_j(x)=X=X'$ 成立,与 $X\neq X'$ 矛盾。$\therefore\ \forall X,X'\subseteq U,X\neq X',f_j(X)\cap f_j(X')=\varnothing$ 成立。

其次须证,有 $U=\cup_{X\subseteq U}f_j(X)$ 成立。

$\forall x\in U$,有 $x\in f_j(\nabla_j(x))$, $\nabla_j(x)\neq\varnothing$, $\nabla_j(x)\subseteq U$,则 $\cup_{X\subseteq U}f_j(X)=U$ 成立。

由划分定义知, $f_j(X)$ 为 U 上划分。

定理 2 令 $S=(U,A)$ 为信息系统, $C=\{C_1,C_2,\cdots,C_m\}$ 是 U 上一族覆盖。$\forall X\subseteq U,x\in U$,概率指派函数 $m:2^U\to[0,1]$ 定义如下:

$$m_{\nabla_j}(X)=\begin{cases}P(f_j(X)), & X=\nabla_j(x)\in\nabla_j^*,\\ \quad j=1,2,3,4 & \\ 0, & 其他\end{cases}$$

式中: $\nabla_j^*=\{\nabla_j(x)\mid x\in U,j=1,2,3,4\}$,则 U 上相应的信任函数 $\mathrm{Bel}_{\nabla_j}(X)$ 和似然函数 $P_{l\nabla_j}(X)$ 为

$$\mathrm{Bel}_{\nabla_j}(X)=P(\underline{\nabla}_j(X)),j=1,2,3,4$$

$$P_{l\nabla_j}(X)=P(\overline{\nabla_j}(X)),j=1,2,3,4$$

证明 $\forall j=1,2,3,4$

1) $\forall X\subseteq U$,显然有 $U=\cup\sum_{X\subseteq U}f(X)$。则

$$\sum_{X\subseteq U}m_{\nabla_j}(X)=\sum_{X\subseteq U}P(f_j(X))=|U|^{-1}\times\left(\sum_{X\subseteq U}|f_j(X)|\right)=\sum_{X\subseteq U}|f_j(X)|\times|U|^{-1}=|U|^{-1}\times\left|\sum_{X\subseteq U}f_j(X)\right|=|U|\times|U|^{-1}=1$$

2) 下证 $\mathrm{Bel}_{\nabla_j}(X)=P(\underline{\nabla}_j(X))$

$$\mathrm{Bel}_{\nabla_j}(X)=\sum_{X'\subseteq X}m_{\nabla_j}(X')=\sum_{X'\subseteq X}P(f_{\nabla_j}(X'))=\sum_{X'\subseteq X}|f_j(X')|\times|U|^{-1}=\sum_{X'\subseteq X}|f_j(X')|\times|U|^{-1}=|\cup_{X'\subseteq X}f_j(X')|\times|U|^{-1}$$

$$\because f_j(X)=\{x\in U\mid\nabla_j(x)=X\}$$

$$\therefore\ \cup_{X'\subseteq X}f_j(X')=\cup_{X'\subseteq X}\{x\in U\mid\nabla_j(x)=X'\}=\{x\in U\mid\nabla_j(x)\in X\}$$

$$\therefore\ \mathrm{Bel}_{\nabla_j}(X)=|\cup_{X'\subseteq X}f_j(X')|\times|U|^{-1}=|U|^{-1}\times|\{x\in U\mid\nabla_j(x)\subseteq X\}|$$

进一步,可证 $\nabla_j(x)\subseteq X\Leftrightarrow x\in\underline{\nabla}_j(X)$

$\therefore\ \mathrm{Bel}_{\nabla_j}(X)=|\{x\in U\mid\nabla_j(x)\subseteq X\}|\times|U|-1=|U|^{-1}|\underline{\nabla}_j(X)|=P(\underline{\nabla}_j(X))$

同理可证, $P_{l\nabla_j}(X)=P(\overline{\nabla_j}(X))$。

其中, $j=1,2,3,4$ 代表用相应信度函数分别刻画 4 种多粒度覆盖粗糙集的近似。

下面用例子进一步解释其具体含义。

例 1 考虑一个房子的评价问题。设 $U=\{x_1,x_2,\cdots,x_6\}$ 是 6 所房子的集合,令 $A=\{$公摊面积,颜色,价格,环境$\}$ 是属性集合, $B=\{$购买意见$\}$ 是决策集合。"公摊面积"的属性值是{较大,普通,

较小},"颜色"的属性值是{优,良,差},"价格"的属性值是{高,中,低},"环境"的属性值是{安静,较吵,很吵},"购买意见"的决策值是{支持,中立,反对}。有 3 个专家{甲,乙,丙}对 6 所房子进行评价,他们的评价结果互相独立。评价结果列于表 1。

表 1　一个关于房屋评价的信息系统

Table 1　An information system of a house evaluation problem

U	公摊面积	颜色	价格	环境	购买意见
x_1	{较大,普通}	{优}	{高}	{安静}	{支持,中立}
x_2	{较大,普通}	{良}	{高,中,低}	{安静,较吵}	{支持,中立,反对}
x_3	{较大,普通}	{优,良}	{中,低}	{较吵,很吵}	{中立,反对}
x_4	{普通}	{差}	{低}	{很吵}	{中立}
x_5	{普通}	{差}	{中}	{很吵}	{反对}
x_6	{普通,较小}	{优,良}	{高,低}	{很吵}	{支持,反对}

由属性集 A 诱导出的一族覆盖 $C=\{C_i, i=1,2,\cdots,4\}$ 和等价关系 $R=\{R_i, i=1,2,\cdots,4\}$ 及由决策集 B 诱导出覆盖的 C_5,如下:$C_1=\{\{x_1,x_2,x_3\},\{x_1,x_2,x_3,x_4,x_5,x_6\},\{x_6\}\}$, $C_2=\{\{x_1,x_3,x_6\},\{x_2,x_3,x_6\},\{x_4,x_5\}\}$, $C_3=\{\{x_1,x_2,x_6\},\{x_2,x_3,x_5\},\{x_2,x_3,x_4,x_6\}\}$, $C_4=\{\{x_1,x_2\},\{x_2,x_3\},\{x_3,x_4,x_5,x_6\}\}$, $C_5=\{\{x_1,x_2,x_6\},\{x_1,x_2,x_3,x_4\},\{x_2,x_3,x_5,x_6\}\}$, $R_1=\{\{x_1,x_2,x_3\},\{x_4,x_5\},\{x_6\}\}$, $R_2=\{\{x_1\},\{x_2\},\{x_3,x_6\},\{x_4,x_5\}\}$, $R_3=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6\}\}$, $R_4=\{\{x_1\},\{x_2\},\{x_3\},\{x_4,x_5,x_6\}\}$

选定专家决策 $D_1=\{x_1,x_2,x_6\}$,用信任函数和似然函数刻画决策 D_1 在经典悲观多粒度粗糙集模型和四型悲观多粒度覆盖粗糙集模型中的信任区间及不确定性,为用户做购买决定提供不同参考意见。

1) 对 $\forall x_t \in U, t\in\{1,2,\cdots,6\}$。

① $\nabla_{i1}(x_t)=\cap\, \mathrm{md}_{C_i}(x_t)$, $i\in\{1,2,3,4\}$。

C_1 中, $\nabla_{11}(x_1)=\{x_1,x_2,x_3\}$, $\nabla_{11}(x_2)=\{x_1,x_2,x_3\}$, $\nabla_{11}(x_3)=\{x_1,x_2,x_3\}$, $\nabla_{11}(x_4)=U$, $\nabla_{11}(x_5)=\{x_1,x_2,x_3,x_4,x_5,x_6\}$, $\nabla_{11}(x_6)=\{x_6\}$;

C_2 中, $\nabla_{21}(x_1)=\{x_1,x_3,x_6\}$, $\nabla_{21}(x_2)=\{x_2,x_3,x_6\}$, $\nabla_{21}(x_3)=\{x_3,x_6\}$, $\nabla_{21}(x_4)=\{x_4,x_5\}$, $\nabla_{21}(x_5)=\{x_4,x_5\}$, $\nabla_{21}(x_6)=\{x_3,x_6\}$;

C_3 中, $\nabla_{31}(x_1)=\{x_1,x_2,x_6\}$, $\nabla_{31}(x_2)=\{x_2\}$, $\nabla_{31}(x_3)=\{x_2,x_3\}$, $\nabla_{31}(x_4)=\{x_2,x_3,x_4,x_6\}$, $\nabla_{31}(x_5)=\{x_2,x_3,x_5\}$, $\nabla_{31}(x_6)=\{x_2,x_6\}$;

C_4 中, $\nabla_{41}(x_1)=\{x_1,x_2\}$, $\nabla_{41}(x_2)=\{x_2\}$, $\nabla_{41}(x_3)=\{x_3\}$, $\nabla_{41}(x_4)=\{x_3,x_4,x_5,x_6\}$, $\nabla_{41}(x_5)=\{x_3,x_4,x_5,x_6\}$, $\nabla_{41}(x_6)=\{x_3,x_4,x_5,x_6\}$。

② $\nabla_1(x_t)=\cup\{\nabla_{i1}(x_t)\mid i=1,2,\cdots,m\}$,则

$$\nabla_1(x_1)=\{x_1,x_2,x_3,x_6\}$$

$$\nabla_1(x_2)=\{x_1,x_2,x_3,x_6\}$$

$$\nabla_1(x_3)=\{x_1,x_2,x_3,x_6\},\nabla_1(x_4)=U$$

$$\nabla_1(x_5)=U$$

$$\nabla_1(x_6)=\{x_2,x_3,x_4,x_5,x_6\}$$

是关于覆盖族 C 的多粒度覆盖。

③ 由关系划分函数知, $\{x_1,x_2,x_3,x_6\}=f_1(\nabla_1(x_1)=\nabla_1(x_2)=\nabla_1(x_3))$, $f_1(\nabla_1(x_4))=(\nabla_1(x_5))=U$, $f_1(\nabla_1(x_6))=U-\{x_1\}$, 所以 $\{\{x_1,x_2,x_3\},\{x_4,x_5\},\{x_6\}\}$ 是 U 上一个划分。

根据第 1 型模型知, $\underline{\nabla_1}(D_1)=\{\{x_6\}\}$, $\overline{\nabla_1}(D_1)=\{\{x_6\},\{x_1,x_2,x_3\}\}$。$U$ 上焦元为 $\{\{x_1,x_2,x_3\},\{x_4,x_5\},\{x_6\}\}$,其概率指派函数值分别为 $m(\{x_6\})=6^{-1}$, $m(\{x_1,x_2,x_3\})=2^{-1}$, $m(\{x_4,x_5\})=3^{-1}$。

相应的信任函数和似然函数值为, $\mathrm{Bel}_{\nabla_1}(D_1)=P(\underline{\nabla_1}(D_1))=6^{-1}$, $P_{l\nabla_1}(X)=P(\overline{\nabla_1}(X))=2\times 3^{-1}$。则专家决策 D_1 在第一型悲观多粒度覆盖粗糙集模型中的信任区间为 $[6^{-1},2\times 3^{-1}]$,其不确定性为 $2\times 3^{-1}-6^{-1}=2^{-1}$。$R$ 中, D_1 的近似集为 $\varnothing$、U,则相应信任区间为 $[0,1]$

显然,本文所提出的模型相较于经典多粒度粗糙集模型,更具实际应用价值。

2) $j=2,3,4$ 时,分别有 $\nabla_{i1}(x)=\cup\,\mathrm{md}_{c_i}(x)$, $\cap\,\mathrm{MD}_{c_i}(x)$, $\cup\,\mathrm{MD}_{c_i}(x)$, $i\in\{1,2,3,4\}$。计算过程与 1) 相似,本文不再一一计算。

4　结论

经过上述讨论,本文得出以下结论:

1) 通过对现有多粒度覆盖粗糙集模型进行分析,构造了 4 种悲观多粒度覆盖粗糙集模型,以使其能够与证据理论更好地结合;

2) 基于集合的交、并运算和关系划分函数,实现了悲观多粒度覆盖粗糙集到单粒度多元覆盖粗糙集再到单粒度经典粗糙集的转化,进而实现简化上述 4 种模型的目的;

3) 结合证据理论,刻画了上述模型的近似及其

不确定性。

在后续研究中,可以进一步给出基于信任函数和似然函数的多粒度覆盖粗糙集属性约简算法。

参考文献:

[1]PALAWK Z. Rough set[J]. International journal of computer & information sciences, 1982, 11(5): 341-356.

[2]CHEN Degang, KWONG S, HE Qiang, et al. Geometrical interpretation and applications of membership functions with fuzzy rough sets[J]. Fuzzy sets and systems, 2012, 193: 122-135.

[3]LIANG Jiye, CHIN K S, Dang Chuangyin, et al. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International journal of general systems, 2002, 31(4): 331-342.

[4]LIANG Jiye, WANG Feng, DANG Chaungyin, et al. A group incremental approach to feature selection applying rough set technique[J]. IEEE transactions on knowledge and data engineering, 2014, 26(2): 294-308.

[5]TAN Anhui, LI Jinjin, LIN Guoping. Extended results on the relationship between information systems[J]. Information sciences, 2015, 290: 156-173.

[6]BONIKOWSKI Z, BRYNIARSKI E, WYBRANIEC-SKARDOWSKA U. Extensions and intentions in the rough set theory [J]. Information sciences, 1998, 107(1/2/3/4): 149-167.

[7]FENG Tao, MI Jusheng, WU Weizhi. Covering-based generalized rough fuzzy sets[M]//WANG Guoying, PETERS J F, SKOWRON A, et al. Rough Sets and Knowledge Technology. Berlin Heidelberg: Springer, 2006: 208-215.

[8]QIAN Y H, LIANG J Y. Rough set method based on multi-granulations[C]//Proceedings of the 5th IEEE International Conference on Cognitive Informatics. Beijing: IEEE, 2006: 297-304.

[9]徐伟华, 刘士虎, 张文修. 一般二元关系下信息系统知识的粒度描述[J]. 计算机工程与应用, 2011, 47(18): 40-44.
XU Weihua, LIU Shihu, ZHANG Wenxiu. Granularity representation of knowledge in information system based on general binary-relation[J]. Computer engineering and applications, 2011, 47(18): 40-44.

[10]QIAN Yuhua, LIANG Jiye, YAO Yiyu, et al. MGRS: a multi-granulation rough set [J]. Information sciences, 2010, 180(6): 949-970.

[11]LIU Caihui, MIAO Duoqian, QIAN Jin. On multi-granulation covering rough sets[J]. International journal of approximate reasoning, 2014, 55(6): 1404-1418.

[12]DEMPSTER A P. Upper and lower probability inferences based on a sample from a finite univariate population[J]. Biometrika, 1967, 54(3/4): 515-528.

[13]SHAFER G. A mathematical theory of evidence[J]. Technometrics, 1978, 20(1): 242.

[14]LIN Guoping, LIANG Jiye, QIAN Yuhua. An information fusion approach by combining multigranulation rough sets and evidence theory[J]. Information sciences, 2015, 314: 184-199.

[15]林国平. 覆盖广义粗糙集与信任函数[J]. 漳州师范学院学报: 自然科学版, 2010(2): 1-4.
LIN Guoping. Connections between covering generization rough set and dempster-shafer theory of evidence[J]. Journal of Zhangzhou normal university: natural science, 2010(2): 1-4.

[16]WU Weizhi, MI Jushneg. Knowledge reduction in incomplete information systems based on dempster-shafer theory of evidence[M]//WANG Guoying, PETERS J F, SKOWRON A, et al. Rough Sets and Knowledge Technology. Berlin Heidelberg: Springer, 2006: 254-261.

[17]YAO Y Y, LINGRAS P J. Interpretations of belief functions in the theory of rough sets[J]. Information sciences, 1998, 104(1/2): 81-106.

[18]CHEN Degang, ZHANG Xiaoxia, LI Wanlu. On measurements of covering rough sets based on granules and evidence theory[J]. Information sciences, 2015, 317: 329-348.

[19]CHEN Degang, LI Wanlu, ZHANG Xiao, et al. Evidence-theory-based numerical algorithms of attribute reduction with neighborhood-covering rough sets [J]. International journal of approximate reasoning, 2014, 55(3): 908-923.

[20]WU Weizhi, LEUNG Y, ZHANG Wenxiu. Connections between rough set theory and Dempster-Shafer theory of evidence[J]. International journal of general systems, 2002, 31(4): 405-430.

[21]吴伟志, 米据生, 李同军. 无限论域中的粗糙近似空间与信任结构[J]. 计算机研究与发展, 2012, 49(2): 327-336.
WU Weizhi, MI Jusheng, LI Tongjun. Rough approximation spaces and belief structures in infinite universes of discourse[J]. Journal of computer research and development, 2012, 49(2): 327-336.

[22]TAN Anhui, WU Weizhi, LI Jinjin, et al. Evidence-theory-based numerical characterization of multi-granulation rough sets in incomplete information systems[J]. Fuzzy sets and systems, 2016, 294: 18-35.

[23]ZAKOWSKI B W. Approximations in the space (u, π) [J]. Demonstratio mathematica, 1983, 16(3): 761-769.

作者简介:

车晓雅,女, 1991年生, 硕士研究生, 主要研究方向为人工智能的数学基础。

李磊军,男,1985年生,讲师,博士,主要研究方向为粗糙集,概念格,粒计算与集成学习等,已发表学术论文10余篇,其中被SCI检索5篇。