二型模糊粗糙属性约简模型

路 娟[1,2] 李德玉[1]
(山西大学计算机与信息技术学院 太原 030006)[1] (中北大学理学院 太原 030051)[2]

摘 要 属性约简是粗糙集理论的重要应用之一,其目的是在保持分类能力不变的前提下去掉冗余的属性,从而简化信息系统。由于经典粗糙集等价关系的要求过于严格,为了更好地解决实际问题,将粗糙集与二型模糊集结合,得到二型模糊粗糙集。利用论域和特征空间的积空间上的两个一型模糊集来构造论域的一个二型模糊划分,将模糊粗糙集属性约简的模型推广到二型模糊粗糙集框架中,得到了一个二型模糊粗糙属性约简的模型,并举例说明了用此模型进行属性约简的方法。

关键词 二型模糊集,粗糙集,二型模糊粗糙集,表示定理,属性约简

中图法分类号 TP182 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.005

Model for Type-2 Fuzzy Rough Attribute Reduction

LU Juan[1,2] LI De-yu[1]
(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)[1]
(School of Science, North University of China, Taiyuan 030051, China)[2]

Abstract Attribute reduction is an important application of rough set theory, which aims to delete redundant attributes and simplify the information system while maintaining the classifying ability of the system. But traditional rough set theory is based on an equivalent relation, which seems to be a very restrictive condition that may limit the application of the rough set model. To overcome this shortcoming, type-2 fuzzy rough set is obtained by combining rough set with type-2 fuzzy set. Two type-1 fuzzy sets, defined on the universe and the product space of feature spaces respectively, are used to construct a type-2 fuzzy partition of the universe and a model for fuzzy rough attribute reduction is expanded to the framework of type-2 fuzzy rough sets, and then a model for type-2 fuzzy rough attribute reduction is obtained. An example is given to show the application of this model.

Keywords Type-2 fuzzy set, Rough set, Type-2 fuzzy rough set, Representation theorem, Attribute reduction

1 引言

自从20世纪80年代初波兰数学家Pawlak提出粗糙集理论以来[1],由于该理论能够处理不精确和不完备的信息,它作为一种有效的知识获取工具已经被成功地应用于机器学习和知识发现、数据挖掘、模式识别和决策分析等领域。属性约简是模式识别、机器学习、数据挖掘及生物信息学中所遇到的一个重要问题,也是粗糙集理论应用最成功的核心问题之一。一般来说,在数据分析任务中,并不是所有的属性都是必要的,有些属性是冗余的。所谓属性约简,就是在保持知识库分类能力不变的条件下,删除其中不相关或者不重要的属性,使得剩余的属性构成的属性子集与原始的属性集确定的知识是相同的。通过属性约简,可以使信息系统的知识表示更加简洁,使隐藏的知识更加清晰,从而简化决策规则。

对经典粗糙集背景下不同信息系统的属性约简的研究已经取得了丰硕的成果。Skowron[2]提出了辨识矩阵的概念;张文修等[3-4]讨论了不协调目标信息系统和不完备信息系统中的知识约简问题;Wang等[5-6]分别讨论了属性动态增加的数据集的属性约简和大规模混合数据集的属性约简方法;吉晨莉等[7]研究了一致覆盖决策系统的属性约简;张颖淳等[8]讨论了基于粗糙集的属性约简在数据挖掘中的应用;唐孝和舒兰[9]提出了基于粒计算的属性约简改进算法;刘静和米据生[10]利用条件属性集的幂集上的闭算子提出了属性约简的一种新方法;刘芳和李天瑞[11]提出了一种基于边界域的不完备信息系统属性约简方法。

经典的粗糙集是基于等价关系的,而等价关系在很多情况下过于严格,从而限制了粗糙集理论的应用。例如,传统的粗糙集理论只能处理符号型的数据,但我们在实际问题中所遇到的问题往往涉及到符号型数据和实值数据,这就需要对经典粗糙集理论进行推广。Dubois和Prade[12]将粗糙集与模

到稿日期:2016-08-11 返修日期:2016-09-27 本文受国家自然科学基金项目(61672331,61573231,61432011,U1435212),山西省科技基础条件平台项目(2015091001-0102)资助。

路 娟(1981-),女,博士生,讲师,主要研究方向为模糊集及粗糙集理论、粒计算,E-mail:nancylu0312@163.com;**李德玉**(1965-),男,博士,教授,主要研究方向为粗糙集理论、粒计算、数据挖掘与知识发现,E-mail:lidysxu@163.com。

糊集结合，提出了粗糙模糊集和模糊粗糙集的概念；Jensen 和 Shen[13] 提出了模糊粗糙属性约简的一种模型；Tsang 等[14] 讨论了模糊粗糙集的属性约简的形式概念与结构，介绍了一种利用区分矩阵计算所有约简的算法，建立了模糊粗糙属性约简的数学基础；Maji 等[15] 提出一种最大化特征的相关性和重要性来进行模糊粗糙集的特征选择的方法；张照星等[16] 研究了 k-近邻模糊粗糙集的快速约简算法；Zhang 等[17] 讨论了基于模糊粗糙集的信息熵进行混合数据特征选择的方法。

二型模糊集是 Zadeh 对模糊集概念[19] 所作出的一种推广[18]。由于每个对象的隶属度本身就是一个模糊集，因此与一型模糊集合即经典的模糊集相比，二型模糊集合能够描述更多的不确定性。区间二型模糊集是二型模糊集的一个特例，指的是次级隶属度全为 1 的那些二型模糊集，由于其形式较一般的二型模糊集更为简单，因此得到了比较广泛的关注。将区间二型模糊集与粗糙集结合起来就得到了区间二型模糊粗糙集。李冬梅等[20] 指出，区间二型模糊集可以增强系统处理不确定性的能力，因此他们基于区间二型模糊粗糙集模型研究了连续域决策信息系统的属性约简，通过紧计算域给出了新的约简算法。Maji 等[21] 提出了用最大化特征的相关性和重要性来进行区间二型模糊粗糙集的特征选择的方法。

将粗糙集与二型模糊集结合就得到了二型模糊粗糙集的概念[22]，本文所讨论的就是在二型模糊粗糙环境下进行的属性约简。Jensen 和 Shen[13] 提出了模糊粗糙属性约简的一种模型，这种模型以模糊等价类作为其核心。本文首先找到了构造论域的二型模糊划分的一种方法，然后在此基础上将 Jensen 等的模型推广到二型模糊粗糙情形中。

2　预备知识

下面对本文中用到的基本概念和性质进行介绍，相关的符号和表示主要参考文献[22-23]。

考虑一个非空有限论域 X，一个二型模糊粗糙集 $\widetilde{A}$ 可以表示为：

$$\widetilde{A}=\{((x,u),\mu_{\widetilde{A}}(x,u))\mid x\in X,u\in J_x\subseteq[0,1]\}$$

其中，$\mu_{\widetilde{A}}(x,u)$ 是 $\widetilde{A}$ 的二型模糊隶属函数，$0\leqslant\mu_{\widetilde{A}}(x,u)\leqslant 1$。对任意取定的 $x'\in X$，

$$\mu_{\widetilde{A}}(x',u)=\int_{u\in J_{x'}} f_{x'}(u)/u,J_{x'}\subseteq[0,1]$$

称为 $\mu_{\widetilde{A}}(x,u)$ 的一个垂直切片，$f_{x'}$ 称为一个次级隶属函数。

Mendel 和 John[23] 于 2002 年提出了二型模糊集的一种新的表示方法，称为 Representation Theorem。考虑离散的 X 和 J_x，假设 X 可以离散化为 $x_1,\cdots,x_N$ 这 N 个值，在每个 x_i 上，对应的 J_{x_i} 被离散化为 M_i 个值。后文讨论的二型模糊集都是指这种 X 和 J_x 都是离散情形的集合，将其称为离散二型模糊集。

对于一个离散的二型模糊集 $\widetilde{A}$，它的嵌入二型集合 $\widetilde{A}_e$ 是一个特殊的二型模糊集合：

$$\widetilde{A}_e=\sum_{i=1}^{N}(f_{x_i}(u_i)/u_i)/x_i,u_i\in J_{x_i}\subseteq U\subseteq[0,1]$$

这个二型模糊集包含了 N 个元素，这 N 个元素即 $u_1,\cdots,u_N$ 是从每个 $J_{x_1},\cdots,J_{x_N}$ 中抽取一个元素形成的，并且带上它们所对应的次级隶属度 $f_{x_1}(u_1),\cdots,f_{x_N}(u_N)$，$\widetilde{A}$ 共有 $\prod_{i=1}^{N}M_i$ 个这样的嵌入二型集 $\widetilde{A}_e$。

对于一个离散的二型模糊集 $\widetilde{A}$，它的嵌入一型集合 A_e 是一个一型模糊集合：

$$A_e=u_i/x_i,u_i\in J_{x_i}\subseteq U\subseteq[0,1]$$

这个集合包含 N 个元素，这 N 个元素即 $u_1,\cdots,u_N$ 是从每个 $J_{x_1},\cdots,J_{x_N}$ 中抽取一个元素形成的，$\widetilde{A}$ 共有 $\prod_{i=1}^{N}M_i$ 个这样的嵌入一型集 A_e。

定理 1(Representation Theorem)[23]　对于一个离散的二型模糊集 $\widetilde{A}$，$\widetilde{A}_e^j$ 表示 $\widetilde{A}$ 的第 j 个嵌入二型集合，即

$$\widetilde{A}_e^j=\{(u_i^j,f_{x_i}(u_i^j))\mid i=1,\cdots,N\}$$

其中 $u_i^j\in\{u_{ik}\mid k=1,\cdots,M_i\}$。$\widetilde{A}$ 可以表示为它的所有嵌入二型集合的并，即

$$\widetilde{A}=\sum_{j=1}^{n}\widetilde{A}_e^j$$

其中，$n=\prod_{i=1}^{N}M_i$。

Representation Theorem 给出了二型模糊集合的一种新的表示，称为波浪形切片表示。

设 X 和 Y 是两个非空有限论域，从 X 到 Y 的一个二型模糊关系是定义在 $X\times Y$ 上的一个二型模糊集合 $\widetilde{R}\in\widetilde{F}(X\times Y)$。如果 $X=Y$，则 $\widetilde{R}$ 称为 X 上的一个二型模糊关系。

设 $\widetilde{R}$ 是从 $X=\{x_1,\cdots,x_n\}$ 到 $Y=\{y_1,\cdots,y_m\}$ 的一个离散的二型模糊关系，第 l 个嵌入二型关系 $\widetilde{R}_e^l$ 可以表示为：

$$\widetilde{R}_e^l=\{(u_{ij}^l,f_{x_iy_j}(u_{ij}^l))\mid i=1,\cdots,n;j=1,\cdots,m\}$$

其中，$u_{ij}^l\in\{u_{ijk}\mid k=1,\cdots,M_{ij}\}$。由于二型模糊关系也是二型模糊集合，同样可以利用 Representation Theorem 得到：

$$\widetilde{R}=\sum_{l=1}^{n_R}\widetilde{R}_e^l$$

其中，$n_R=\prod_{i=1}^{n}\prod_{j=1}^{m}M_{ij}$。

将二型模糊集与粗糙集结合后可以得到二型模糊粗糙集[22]。设 $X=\{x_1,\cdots,x_n\}$ 是一个非空有限论域，$\widetilde{A}$ 是 X 上的一个离散二型模糊集，$\widetilde{R}$ 是 X 上的一个离散二型模糊关系。如果 $\widetilde{A}=\sum_{j=1}^{n_A}\widetilde{A}_e^j$，$\widetilde{R}=\sum_{l=1}^{n_R}\widetilde{R}_e^l$，其中

$$\widetilde{A}_e^j=\sum_{i=1}^{N}(f_{x_i}(u_i^j)/u_i^j)/x_i$$

$$\widetilde{R}_e^l=\sum_{p,q=1}^{N}(\frac{g_{(x_p,y_q)}(v_{pq}^l)}{v_{pq}^l})/(x_p,y_q)$$

是嵌入二型集合，那么二型模糊粗糙集合是一个二元组 $(\underline{\widetilde{R}}(\widetilde{A}),\overline{\widetilde{R}}(\widetilde{A}))$，其中

$$\underline{\widetilde{R}}(\widetilde{A})=\sum_{l=1}^{n_R}\sum_{j=1}^{n_A}\underline{\widetilde{R}_e^l}(\widetilde{A}_e^j),\overline{\widetilde{R}}(\widetilde{A})=\sum_{l=1}^{n_R}\sum_{j=1}^{n_A}\overline{\widetilde{R}_e^l}(\widetilde{A}_e^j)$$

$$\underline{\widetilde{R}_e^l}(\widetilde{A}_e^j)=\sum_{i=1}^{N}\{\bigwedge_{m=1}^{N}[g_{(x_i,y_m)}(v_{im}^l)\wedge f_{x_m}(u_m^j)]/\bigwedge_{m=1}^{N}[(1-v_{im}^l)\vee u_m^j]\}/x_i$$

$$\overline{\widetilde{R}_e^l}(\widetilde{A}_e^j)=\sum_{i=1}^{N}\{\bigwedge_{m=1}^{N}[g_{(x_i,y_m)}(v_{im}^l)\wedge f_{x_m}(u_m^j)]/\bigvee_{m=1}^{N}(v_{im}^l\wedge u_m^j)\}/x_i$$

有序对 $(X,\widetilde{R})$ 称为一个二型模糊近似空间，映射 $\underline{\widetilde{R}}$:

$\widetilde{F}(X)\to\widetilde{F}(X)$和$\overline{R}:\widetilde{F}(X)\to\widetilde{F}(X)$分别称为下二型模糊粗糙近似算子和上二型模糊粗糙近似算子。

Jensen 和 Shen[13]指出，正如等价类是粗糙集中的核心概念，模糊等价类也是模糊粗糙集方法中的核心。论域的一个模糊划分所构成的一族正规模糊集恰好可以作为模糊等价类[24]。

设 P 为属性集 A 的子集，$X/P=\{F_1,\cdots,F_k\}$是 X 的模糊划分，如果 Y 是要近似的概念，则 Y 的 P-下近似和 P-上近似定义为：

$$\mu_{\underline{P}Y}(F_i)=\inf_x \max\{1-\mu_{F_i}(x),\mu_Y(x)\}$$

$$\mu_{\overline{P}Y}(F_i)=\sup_x \min\{\mu_{F_i}(x),\mu_Y(x)\}$$

其中，$i=1,\cdots,k$。对于属性约简的情形，模糊下、上近似定义为：

$$\mu_{\underline{P}Y}(x)=\sup_{F\in X/P}\min\{\mu_F(x),\inf_{y\in X}\max\{1-\mu_F(y),\mu_Y(y)\}$$

$$\mu_{\overline{P}Y}(x))=\sup_{F\in X/P}\min\{\mu_F(x),\sup_{y\in X}\min\{\mu_F(y),\mu_Y(y)\}$$

基于模糊粗糙集的属性约简建立在模糊下近似的基础上，可以用来约简包含实值属性的数据。

3 二型模糊粗糙属性约简

3.1 二型模糊划分

类似于模糊粗糙属性约简，二型模糊等价类也是二型模糊粗糙属性约简中的重要概念。由论域的一个二型模糊划分构成的一族二型模糊集正好可以作为二型模糊等价类。下面根据 Dubois 和 Prade[12]提出的模糊划分概念给出二型模糊划分的定义。

定义 1[12] 论域 X 上的一族正规模糊集 $\phi=\{F_1,\cdots,F_n\}(n<|X|)$称为 X 的一个模糊划分，如果

(1)ϕ 能够充分覆盖 X，即$\inf_x \max_{i=1,\cdots,n} F_i(x)>0$；

(2)两两互不相交，即$\forall i,j,\sup_x \min(F_i(x),F_j(x))<1$。

定义 2 论域 X 上的一族正规二型模糊集 $\phi=\{\widetilde{F}_1,\cdots,\widetilde{F}_n\}(n<|X|)$称为 X 的一个二型模糊划分，如果

(1)存在一个 $y\in(0,1]$，使得$(\sqcap_x\sqcup_{i=1}^n\widetilde{F}_i(x))(y)>0$；

(2)对任意 i,j，存在 $y\in[0,1)$使得$(\sqcup_x(\widetilde{F}_i(x)\sqcap\widetilde{F}_j(x)))(y)>0$。

其中，运算$\sqcup,\sqcap:[0,1]^{[0,1]}\times[0,1]^{[0,1]}\to[0,1]^{[0,1]}$定义为[25]：对于任意 $A=\int_u \frac{f(u)}{u}$，$B=\int_w \frac{g(w)}{w}$，$A,B\in[0,1]^{[0,1]}$：

$$A\sqcup B=\int_u\int_w \frac{f(u)\wedge g(w)}{u\vee w}$$

$$A\sqcap B=\int_u\int_w \frac{f(u)\wedge g(w)}{u\wedge w}$$

在求并的运算中，如果有多对 u 和 w 能计算得到相同的 $u\vee w$，那么选择隶属度最大的那一组；同样，在交运算中也选择隶属度最大的。

下面构造能够作为论域的划分的那一族二型模糊集。Janet Aisbett，John T. Rickard 和 David Morgenthaler[25]给出了如何从一个定义在论域 X 和特征空间 V 的积空间上的模糊集和一个定义在特征空间上的模糊集来得到空间 X 上的二型模糊集的方法。其中第一个模糊集作为 X 中元素的模糊观测值；另一个模糊集定义一个模糊概念或者模糊类，得到的二型模糊集可以解释为定义在 X 上的一个二型模糊概念。

假设 F 的隶属函数为 $f:X\times V\to[0,1]$，G 的隶属函数为 $g:V\to[0,1]$。扩展原理(Extension Principle)[18]可以将定义在 V 上的模糊集扩展为定义在$[0,1]$上的模糊集，对每个 $x\in X$ 应用扩展原理，可以生成 X 上的一个二型模糊集，其隶属函数 $\widetilde{F}:X\times[0,1]\to[0,1]$定义为：

$$\widetilde{F}(x,y)=\sup_{v\in V:g(v)=y}\{f(x,v)\},x\in X,y\in[0,1]$$

下面的定理告诉我们，利用上述方法可以由特征空间 V 的一个一型模糊划分生成 X 的一个二型模糊划分。

定理 2 设 $g_i:V\to[0,1](i=1,\cdots,n)$是 V 的一个一型模糊划分。如果按照上述方法生成的一族二型模糊集 $\widetilde{F}_i:X\times[0,1]\to[0,1](i=1,\cdots,n)$是正规的，那么$\{\widetilde{F}_i\}$构成了 X 的一个二型模糊划分。

证明：只需证明二型模糊划分定义中的两个条件是成立的。

由于$\{g_i\}$是 V 的一个一型模糊划分，因此有

$$\inf_v \max_{i=1,\cdots,n} g_i(v)>0$$

也就是说，

$$\forall v\in V,\max_{i=1,\cdots,n} g_i(v)>0$$

从而$\forall v\in V,\exists i$, s. t. $g_i(v)>0$。因为 f 表示的是 X 中元素的模糊观测值，所以给定 $x\in X$，必然存在 $v_0\in V$ 使得：

$$f(x,v_0)>0$$

一方面，对于这个 v_0，存在 i_0 使得 $g_{i_0}(v_0)>0$，因此

$$\widetilde{F}_{i_0}(x,g_{i_0}(v_0))=\sup_{v\in V:g_{i_0}(v)=g_{i_0}(v_0)}\{f(x,v)\}\geqslant f(x,v_0)>0$$

另一方面，对于任意 $i\neq i_0$，

$$\widetilde{F}_i(x,g_i(v_0))=\sup_{v\in V:g_i(v)=g_i(v_0)}\{f(x,v)\}\geqslant f(x,v_0)>0$$

所以有：

$$(\sqcup_{i=1}^n\widetilde{F}_i(x))(\bigvee_{i=1}^n g_i(v_0))\geqslant\bigwedge_{i\neq i_0}\widetilde{F}_i(x,g_i(v_0))\wedge\widetilde{F}_{i_0}(x,g_{i_0}(v_0))>0$$

其中，$\bigvee_{i=1}^n g_i(v_0)\geqslant g_{i_0}(v_0)>0$。换言之，对于任意的 $x\in X$，存在 $y_x\in(0,1]$使得$(\sqcup_{i=1}^n\widetilde{F}_i(x))(y_x)>0$。这样便可以得到：

$$\sqcap_x(\sqcup_{i=1}^n\widetilde{F}_i(x))(\wedge_x y_x)\geqslant\wedge_x(\sqcup_{i=1}^n\widetilde{F}_i(x))(y_x)>0$$

其中，$\wedge_x y_x>0$。这就证明了第一个条件是成立的。

下面证明第二个条件。对任意 i,j，由$\sup_v \min(g_i(v),g_j(v))<1$，可以得到$\forall v,\min(g_i(v),g_j(v))<1$。给定 $x\in X$，存在 v_x 使得 $f(x,v_x)>0$，并且对于这个 v_x，有：

$$\widetilde{F}_i(x,g_i(v_x))=\sup_{v\in V:g_i(v)=g_i(v_x)}\{f(x,v)\}\geqslant f(x,v_x)>0$$

$$\widetilde{F}_j(x,g_j(v_x))=\sup_{v\in V:g_j(v)=g_j(v_x)}\{f(x,v)\}\geqslant f(x,v_x)>0$$

所以有：

$$(\widetilde{F}_i(x)\sqcap\widetilde{F}_j(x))(g_i(v_x)\wedge g_j(v_x))>0$$

$$g_i(v)\wedge g_j(v)<1$$

也就是说，对于任意 x，存在 $y_x<1$ 使得$(\widetilde{F}_i(x)\sqcap\widetilde{F}_j(x))(y_x)>0$。因此可以得到：

$$(\sqcup_x(\widetilde{F}_i(x)\sqcap\widetilde{F}_j(x)))(\vee_x y_x)>0$$

其中，$\vee_x y_x<1$。这样就证明了第二个条件也是成立的。

3.2　二型模糊粗糙属性约简

定义3　设 P 是属性集 A 的一个子集,$X/P=\{\widetilde{F}_1,\cdots,\widetilde{F}_m\}$是 X 的一个二型模糊划分,$\widetilde{Y}\in\widetilde{F}(X)$是要近似的二型模糊概念。如果

$$\widetilde{Y}=\sum_{l=1}^{n_Y}\widetilde{Y}_e^l=\sum_{l=1}^{n_Y}\{\sum_{i=1}^{N}[\frac{g_{x_i}(u_i^l)}{u_i^l}]/x_i\}$$

$$\widetilde{F}_k=\sum_{j=1}^{n_k}\widetilde{F}_{ke}^j=\sum_{j=1}^{n_k}\{\sum_{i=1}^{N}[\frac{f_{kx_i}(v_{ki}^j)}{v_{ki}^j}]/x_i\}$$

那么,二型模糊 P-下近似和 P-上近似可以定义为:

$$\mu_{\underline{P\widetilde{Y}}}(\widetilde{F}_k)=\sum_{l=1}^{n_Y}\sum_{j=1}^{n_k}\{\bigwedge_{i=1}^{N}f_{kx_i}(v_{ki}^j)\wedge g_{x_i}(u_i^l)\}/\bigwedge_{i=1}^{N}[(1-v_{ki}^j)\vee u_i^l]\}$$

$$\mu_{\overline{P\widetilde{Y}}}(\widetilde{F}_k)=\sum_{l=1}^{n_Y}\sum_{j=1}^{n_k}\{\bigwedge_{i=1}^{N}[f_{kx_i}(v_{ki}^j)\wedge g_{x_i}(u_i^l)]/\bigvee_{i=1}^{N}[v_{ki}^j\wedge u_i^l]\}$$

其中,$k=1,\cdots,m$。

对于属性约简的情形,二型模糊 P-下近似和 P-上近似可以被重新定义为:

$$\mu_{\underline{P\widetilde{Y}}}(x)=\bigsqcup_{\widetilde{F}\in X/P}\{\mu_{\widetilde{F}}(x)\sqcap\mu_{\underline{P\widetilde{Y}}}(\widetilde{F})\}$$

$$\mu_{\overline{P\widetilde{Y}}}(x)=\bigsqcup_{\widetilde{F}\in X/P}\{\mu_{\widetilde{F}}(x)\sqcap\mu_{\overline{P\widetilde{Y}}}(\widetilde{F})\}$$

正如模糊粗糙属性约简是建立在模糊下近似的概念基础上的,二型模糊粗糙属性约简也是通过二型模糊下近似的概念来进行的。

定义4　设 P 和 Q 是属性集 A 的两个子集,一个对象 $x\in X$ 属于二型模糊正域的隶属度定义为:

$$\mu_{POS_P(Q)}(x)=\bigsqcup_{\widetilde{Y}\in X/Q}\mu_{\underline{P\widetilde{Y}}}(x),x\in X$$

二型模糊粗糙依赖函数定义为:

$$\gamma_P(Q)=\frac{|\mu_{POS_P(Q)}(x)|}{|X|}=\frac{\sum_{x\in X}C_{POS_P(Q)}(x)}{|X|}$$

其中,$C_{POS_P(Q)}(x)$是 $\mu_{POS_P(Q)}(x)$的形心。

类似于模糊粗糙属性约简,我们需要考虑多属性的情形。一般地,

$$X/P=\otimes\{X/IND(\{a\})|a\in P\}$$

X/P 中的每个集合都表示一个等价类,因此一个对象属于这样一个等价类的程度可以通过计算每个成分模糊等价类 F_i $(i=1,\cdots,n)$的交集得到:

$$\mu_{F_1\cap\cdots\cap F_n}(x)=\sqcap_{i=1}^{n}\mu_{F_i}(x)$$

3.3　二型模糊粗糙属性约简的例子

表1列给出了一个简单的例子。$X=\{e_1,e_2,\cdots,e_6\}$是对象集(患者),$A=C\cup Q$ 是属性集,其中 $C=\{$头疼,肌肉疼,体温$\}$是条件属性集,$Q=\{$流感$\}$是决策属性集。不同于其他的条件属性“头疼”、“肌肉疼”和决策属性“流感”,对于每个患者,属性“体温”的属性值是实数集的一个子集。这是因为患者通常会多次测量自己的体温,而这些测量结果可能是不同的。

表1　二型模糊粗糙属性约简的例子

患者	头疼	肌肉疼	体温/℃	流感
e_1	是	是	36.3,36.4	否
e_2	是	是	37.9,38.1	是
e_3	是	是	38.8,39.0	是
e_4	否	是	36.5	否
e_5	否	否	37.8,38.0	否
e_6	否	是	39.4,39.5	是

设 $c_1=$头疼,$c_2=$肌肉疼,$c_3=$体温,则

$$X/Q=\{Y_1,Y_2\}=\{\{e_2,e_3,e_6\},\{e_1,e_4,e_5\}\}$$

$$X/\{c_1\}=\{E_1,E_2\}=\{\{e_1,e_2,e_3\},\{e_4,e_5,e_6\}\}$$

$$X/\{c_2\}=\{F_1,F_2\}=\{\{e_1,e_2,e_3,e_4,e_6\},\{e_5\}\}$$

于是计算可得:

$$\gamma_{\{c_1\}}(Q)=\frac{|POS_{\{c_1\}}(Q)|}{|X|}=\frac{|\sum_{Y\in X/Q}\underline{\{c_1\}Y}|}{|X|}=0$$

$$\gamma_{\{c_2\}}(Q)=\frac{|POS_{\{c_2\}}(Q)|}{|X|}=\frac{|\sum_{Y\in X/Q}\underline{\{c_2\}Y}|}{|X|}=1/6$$

利用模糊集 $g_N:V\to[0,1]$,$g_H:V\to[0,1]$来定义概念“体温正常”和“体温高”,其中 $V=[36,42]$。

$$g_N(x)=\begin{cases}1, & x\in[36,37]\\ -x+38, & x\in(37,38)\\ 0, & x\in[38,42]\end{cases}$$

$$g_H(x)=\begin{cases}0, & x\in[36,37]\\ x-37, & x\in(37,38)\\ 0, & x\in[38,42]\end{cases}$$

考虑 $f:X\times V\to[0,1]$,具体定义为:

$$f(e_1)=\frac{1}{36.3}+\frac{1}{36.4},f(e_2)=\frac{1}{37.9}+\frac{1}{38.1}$$

$$f(e_3)=\frac{1}{38.8}+\frac{1}{39.0},f(e_4)=\frac{1}{36.5}$$

$$f(e_5)=\frac{1}{37.8}+\frac{1}{38.0},f(e_6)=\frac{1}{39.4}+\frac{1}{39.5}$$

据此可以得到二型模糊等价类 $X/\{c_3\}=\{\widetilde{N},\widetilde{H}\}$,其中

$$\widetilde{N}=\frac{1/1}{e_1}+\frac{1/0.1+1/0}{e_2}+\frac{1/0}{e_3}+\frac{1/1}{e_4}+\frac{1/0.2+1/0}{e_5}+\frac{1/0}{e_6}$$

$$\widetilde{H}=\frac{1/0}{e_1}+\frac{1/0.9+1/1}{e_2}+\frac{1/1}{e_3}+\frac{1/0}{e_4}+\frac{1/0.8+1/1}{e_5}+\frac{1/1}{e_6}$$

利用定义3,可以得到

$$\mu_{\underline{\{c_3\}}Y_1}(\widetilde{N})=\sum_{j=1}^{4}\{1/\bigwedge_{i=1,4,5}[1-\widetilde{N}(e_i,v_i^j)]\}=1/0$$

$$\mu_{\underline{\{c_3\}}Y_1}(\widetilde{H})=1/0.2+1/0$$

$$\mu_{\underline{\{c_3\}}Y_2}(\widetilde{N})=1/0.9+1/1$$

$$\mu_{\underline{\{c_3\}}Y_2}(\widetilde{H})=1/0$$

因此

$$\begin{aligned}\mu_{\underline{\{c_3\}}Y_1}(e_1)&=[\mu_{\widetilde{N}}(e_1)\sqcap\mu_{\underline{\{c_3\}}Y_1}(\widetilde{N})]\sqcup[\mu_{\widetilde{H}}(e_1)\sqcap\mu_{\underline{\{c_3\}}Y_1}(\widetilde{H})]\\&=(1/1\sqcap 1/0)\sqcup(1/0\sqcap(1/0.2+1/0)]=1/0\end{aligned}$$

$$\begin{aligned}\mu_{\underline{\{c_3\}}Y_2}(e_1)&=[\mu_{\widetilde{N}}(e_1)\sqcap\mu_{\underline{\{c_3\}}Y_2}(\widetilde{N})]\sqcup[\mu_{\widetilde{H}}(e_1)\sqcap\mu_{\underline{\{c_3\}}Y_2}(\widetilde{H})]\\&=(1/1\sqcap(1/0.9+1/1))\sqcup(1/0\sqcap 1/0)\\&=1/0.9+1/1\end{aligned}$$

$$\mu_{\underline{\{c_3\}}Y_1}(e_2)=1/0.2+1/0,\mu_{\underline{\{c_3\}}Y_2}(e_2)=1/0.1+1/0$$

$$\mu_{\underline{\{c_3\}}Y_1}(e_3)=1/0.2+1/0,\mu_{\underline{\{c_3\}}Y_2}(e_3)=1/0$$

$$\mu_{\underline{\{c_3\}}Y_1}(e_4)=1/0,\mu_{\underline{\{c_3\}}Y_2}(e_4)=1/0.9+1/1$$

$$\mu_{\underline{\{c_3\}}Y_1}(e_5)=1/0.2+1/0,\mu_{\underline{\{c_3\}}Y_2}(e_5)=1/0.2+1/0$$

$$\mu_{\underline{\{c_3\}}Y_1}(e_6)=1/0.2+1/0,\mu_{\underline{\{c_3\}}Y_2}(e_6)=1/0$$

由上面这些数值可以得到一个对象 $e_i\in X$ 属于二型模糊正域的隶属度:

$$\begin{aligned}\mu_{POS_{\{c_3\}}(Q)}(e_1)&=\mu_{\underline{\{c_3\}}Y_1}(e_1)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_1)\\&=1/0\sqcup(1/0.9+1/1)=1/0.9+1/1\end{aligned}$$

$$\mu_{POS_{\{c_3\}}(Q)}(e_2)=\mu_{\underline{\{c_3\}}Y_1}(e_2)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_2)$$
$$=1/0.2+1/0.1+1/0$$
$$\mu_{POS_{\{c_3\}}(Q)}(e_3)=\mu_{\underline{\{c_3\}}Y_1}(e_3)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_3)$$
$$=1/0.2+1/0$$
$$\mu_{POS_{\{c_3\}}(Q)}(e_4)=\mu_{\underline{\{c_3\}}Y_1}(e_4)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_4)$$
$$=1/0.9+1/1$$
$$\mu_{POS_{\{c_3\}}(Q)}(e_5)=\mu_{\underline{\{c_3\}}Y_1}(e_5)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_5)$$
$$=1/0.2+1/0$$
$$\mu_{POS_{\{c_3\}}(Q)}(e_6)=\mu_{\underline{\{c_3\}}Y_1}(e_6)\sqcup\mu_{\underline{\{c_3\}}Y_2}(e_6)$$
$$=1/0.2+1/0$$

因此,

$$\gamma_{\{c_3\}}(Q)=\frac{|\mu_{POS_{\{c_3\}}(Q)}(e)|}{|X|}=\frac{\sum_{i=1}^{6}C_{POS_{\{c_3\}}(Q)}(e_i)}{|X|}=\frac{2.3}{6}$$

可以看出,属性 c_3 可以带来二型模糊粗糙依赖度的最大增长,因此这个属性被加入潜在约简集合中。继续迭代此过程,可以由 $X/\{c_1,c_3\}=\{E_1\cap\widetilde{N},E_2\cap\widetilde{N},E_1\cap\widetilde{H},E_2\cap\widetilde{H}\}$ 和 $X/\{c_2,c_3\}=\{F_1\cap\widetilde{N},F_2\cap\widetilde{N},F_1\cap\widetilde{H},F_2\cap\widetilde{H}\}$ 计算得出 $\gamma_{\{c_1,c_3\}}=4.1/6$,$\gamma_{\{c_2,c_3\}}=5.75/6$。将属性 c_2 加入约简候选集可以带来二型模糊粗糙依赖度的更大增长,因此新的约简候选集就变成了 $\{c_2,c_3\}$。最后,把属性 c_1 也加入潜在约简集合,从而可以得到 $\gamma_{\{c_1,c_2,c_3\}}=5.8/6$。然而,若把例子中的模糊属性分明化,则得到表 2 中对应的粗糙属性约简的问题,可以看到 $\{c_1,c_3\}$ 和 $\{c_2,c_3\}$ 都是约简,$\{c_3\}$ 是核,$\gamma'_{\{c_1,c_3\}}(Q)=\gamma'_{\{c_2,c_3\}}(Q)=1$。

表 2 粗糙属性约简的例子

患者	头疼	肌肉疼	体温	流感
e_1	是	是	正常	否
e_2	是	是	高	是
e_3	是	是	很高	是
e_4	否	是	正常	否
e_5	否	否	高	否
e_6	否	是	很高	是

从上面的讨论可以得到,当我们把本来模糊的问题简单分明化时会损失一些信息,如果用二型模糊的方法来讨论则可以明显得到 3 个属性的重要性排名:c_3,c_2,c_1。

结束语 本文讨论了二型模糊环境下的属性约简,将一型模糊划分的概念推广到二型的情形,提出了论域的二型模糊划分的概念,并在此基础上提出了一个二型模糊粗糙属性约简模型。二型模糊粗糙的属性约简由于刻画了更多的不确定性,因所得到的约简不能像经典粗糙集的属性约简那样由依赖度确切得到,但是可以在不损失信息的前提下得到各个属性重要性的刻画。本文对二型模糊粗糙集的定义是基于 Mendel 所提出的二型模糊集合的波浪形切片表示,也即 Representation Theorem。Mendel[26] 指出,二型模糊集合的 Representation Theorem 在进行理论推导时是非常有用的,但是并不推荐将其应用于计算的目的,这是因为在波浪形切片表示中,即使对于一个简单的二型模糊集合,嵌入二型模糊集的个数 n_A 也是非常巨大的,所以本文只给出了一个较简单的例子来说明这个二型模糊粗糙属性约简模型的应用方法,而没有进行真实数据的数值实验。本文所提出的属性约简方法是对二型模糊粗糙集应用的一个有益探索。未来希望能够找到更适合二型模糊粗糙集属性约简的计算方法,以更好地用二型模糊粗糙集理论解决实际问题。

参 考 文 献

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

[2] SKOWRON A,RAUSZER C. The discernibility matrices and functions in information systems[M]// Slowinski R,ed. Intelligent Decision Support- Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers,1992:331-362.

[3] 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.

[4] ZHANG W X,MI J S,WU W Z. Approaches to knowledge reductions in inconsistent systems[J]. International Journal of Intelligent Systems,2003,18:989-1000.

[5] WANG F,LIANG J Y,QIAN Y H. Attribute reduction:A dimension incremental strategy[J]. Knowledge-based Systems,2013,39:95-108.

[6] WANG F,LIANG J Y. An efficient feature selection algorithm for hybrid data[J]. Neurocomputing,2016,193:33-41.

[7] JI C L,YANG Y. Attribute reduction of consistent covering decision system[J]. Computer Science,2012,39(6A):288-290,303. (in Chinese)
吉晨莉,杨勇. 一致覆盖决策系统的属性约简[J]. 计算机科学,2012,39(6A):288-290,303.

[8] ZHANG Y C,SU B H,CAO J. Study on application of attributive reduction based on rough sets in data mining[J]. Computer Science,2013,40(8):223-226. (in Chinese)
张颖淳,苏伯洪,曹娟. 基于粗糙集的属性约简在数据挖掘中的应用研究[J]. 计算机科学,2013,40(8):223-226.

[9] TANG X,SHU L. Improved algorithm of attribute reduction based on granular computing[J]. Computer Science,2014,41(11A):313-315,346. (in Chinese)
唐孝,舒兰. 基于粒计算的属性约简改进算法[J]. 计算机科学,2014,41(11A):313-315,346.

[10] LIU J,MI J S. Attribute reduction based on closure operators[J]. Computer Science,2014,41(10):249-251,265. (in Chinese)
刘静,米据生. 基于闭算子的属性约简[J]. 计算机科学,2014,41(10):249-251,265.

[11] LIU F,LI T R. Method for attribute reduction based on rough sets boundary regions[J]. Computer Science,2016,43(3):242-245,284. (in Chinese)
刘芳,李天瑞. 基于边界域的不完备信息系统属性约简方法[J]. 计算机科学,2016,43(3):242-245,284.

[12] DUBOIS D,PRADE H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General Systems,1990,17(2/3):191-209.

[13] JESEN R,SHEN Q. Fuzzy-rough sets assisted attribute selection[J]. IEEE Transaction on Fuzzy Systems,2007,15(1):73-89.

[14] TSANG E C C,CHEN D G,YEUNG D S,et al. Attribute reduction using fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems,2008,16(5):1130-1141.

[15] MAJI P,GARAI P. On fuzzy -rough attribute selection:Criteria of max-dependency, max-relevance, min-redundancy, and max-significance [J]. Applied Soft Computing, 2013, 13(9): 3968-3980.

[16] ZHANG Z X,FAN X Q,ZHAO S Y,et al. Fast reduction algorithm research based on k-nearest neighbor fuzzy rough set[J]. Journal of Frontiers of Computer Science and Technology,2015,9(1):14-23.(in Chinese)
张照星,范星奇,赵素云,等. k-近邻模糊粗糙集的快速约简算法研究[J]. 计算机科学与探索,2015,9(1):14-23.

[17] ZHANG X, MEI C L, CHEN D G, et al. Feature selection in mixed data:A method using a novel fuzzy rough set-based information entropy [J]. Pattern Recognition,2016,56(1):1-15.

[18] ZADEH L A. The concept of a linguistic variable and its application to approximate reasoning-1[J]. Information Science,1975,8(3):199-249.

[19] ZADEH L A. Fuzzy sets[J]. Information and Control,1965,8(3):338-356.

[20] LI D M,LI T,ZHAO T. Continuous attribute reduction algorithm based on interval type-2 fuzzy rough sets[J]. Application research of computers,2015,32(5):1379-1382.(in Chinese)
李冬梅,李涛,赵涛. 基于区间二型模糊粗糙集的连续属性约简算法[J]. 计算机应用研究,2015,32(5):1379-1382.

[21] MAJI P,GARAI P. IT2 fuzzy-rough sets and max relevance-max significance criterion for attribute selection [J]. IEEE Transactions on Cybernetics,2015,45(8):1657-1668.

[22] LU J,LI D Y,ZHAI Y H,et al. A model for type-2 fuzzy rough sets[J]. Information Sciences,2016,328(c):359-377.

[23] MENDEL J M,JOHN R I. Type-2 fuzzy sets made simple[J]. IEEE Transactions on Fuzzy Systems,2002,10(2):117-127.

[24] DUBOIS D,PRADE H. Putting rough sets and fuzzy sets together[J]. Intell. Decision Support,1992,11:203-232.

[25] AISBETT J,RICKARD J T,MORGENTHALER D. Multivariate modeling and type-2 fuzzy sets[J]. Fuzzy sets and systems,2011,163(1):78-95.

[26] MENDEL J M. Advances in type-2 fuzzy sets and systems[J]. Information Sciences,2007,177(1):84-110.

(上接第 22 页)

[17] MA S S,CHEN G S,FANG X Q. Research on Prognostic and Health Management System of Complex Equipment[J]. Computer Measurenment & Control,2010,18(1):1-4.(in Chinese)
马飒飒,陈国顺,方兴桥. 复杂装备故障预测与健康管理系统初探[J]. 计算机测量与控制,2010,18(1):1-4.

[18] CHEN J,PENG Y,LI Q,et al. Data-driven PHM software system for airborne equipment[J]. Journal of Electronic Measurement and Instrumentation, 2015,29(10):1536-1543. (in Chinese)
陈静,彭宇,李祺,等. 数据驱动的机载设备 PHM 软件系统[J]. 电子测量与仪器学报,2015,29(10):1536-1543.

[19] CUI J,HE J,LIU Q,et al. Research on the Architecture Design of UAV PHM System[J]. Computer Measurement & Control,2016,24(6):133-135.(in Chinese)
崔嘉,贺静,刘奇,等. 无人机 PHM 系统体系结构设计研究[J]. 计算机测量与控制,2016,24(6):133-135.

[20] GUO Y M,CAI X B,ZHANG B Z,et al. Prognostics and Health management technology for new-generation weapon[J]. Computer Engineering and Application, 2008, 44(13): 199-203. (in Chinese)
郭阳明,蔡小斌,张宝珍,等. 新一代装备的预测与健康状态管理技术[J]. 计算机工程与应用,2008,44(13):199-203.

[21] GUO Y M,CAI X B,ZHANG B Z,et al. Review of Prognostics and Health Management Technology[J]. Computer Measurement & Control,2008,16(9):1213-1218.(in Chinese)
郭阳明,蔡小斌,张宝珍,等. 故障预测与健康状态管理技术综述[J]. 计算机测量与控制,2008,16(9):1213-1218.

[22] JIA H,XU H Q,QU Q L. Maintenance support for ship equipment [J]. Instrumentation Customer,2006,13(3):8-10.(in Chinese)
贾慧,胥辉旗,屈清林. 基于 PHM 的舰船装备维修保障研究[J]. 仪器仪表用户,2006,13(3):8-10.

[23] LONG F,WANG Z,CHENG X J,et al. A PHM Blueprint Based on Information Fusion for Military Digital Devices[J]. Microelectronics & Computer,2010,27(9):86-90.(in Chinese)
龙凤,王忠,程绪建,等. 一种基于信息融合的军用电子产品 PHM 方案设计[J]. 微电子学与计算机,2010,27(9):86-90.

[24] WANG H Z,YANG J P,WANG S H. Research on the Maintenance Support System for Radar Equipment Based on the PHM [J]. Journal of the Academy of Equipment Command & Technology,2008,19(4):83-86.(in Chinese)
王晗中,杨江平,王世华. 基于 PHM 的雷达装备维修保障研究[J]. 装备指挥技术学院学报,2008,19(4):83-86.

[25] ZHANG B Z,ZENG T X. Advanced Prognostics and Health Management Technology[J]. Measurement & Control Technology,2003,22(11):4-6.(in Chinese)
张宝珍,曾天翔. 先进的故障预测与状态管理技术[J]. 测控技术,2003,22(11):4-6.

[26] DISCENZO F M,NICKERSON W,MITCHELL C E,et al. Open systems architecture enables health management for next generation system monitoring and maintenance[R]. Development Program White Paper,2001.

[27] LV Z B,SUN Q,WANG J,et al. Research on Prognostic and Health Management (PHM) Model Verification Engineering Process[J]. Computer Measurement & Control,2016,24(9):281-283.(in Chinese)
吕镇邦,孙倩,王娟,等. PHM 模型的工程化验证方法研究[J]. 计算机测量与控制,2016,24(9):281-283.

[28] WANG Y Y. The Study of Fault Diagnosis Based on Information Fusion Technology[D]. Shijiazhuang:Hebei University of Science and Technology,2010:10-12.(in Chinese)
王艳英. 基于信息融合技术的故障诊断方法研究[D]. 石家庄:河北科技大学,2010:10-12.

[29] 陈建辉. 战车原理[M]. 北京:兵器工业出版社,2006:1-2.