

多粒度决策粗糙集中的粒度约简方法

桑妍丽¹ 钱宇华^{1,2}

(山西大学计算机与信息技术学院 太原 030006)¹

(计算机智能与中文信息处理教育部重点实验室 太原 030006)²

摘 要 多粒度决策粗糙集模型是一种泛化的多粒度粗糙集模型,该模型结合决策粗糙集数据分析理论和多粒度思想,实现了在多个粒空间进行决策粗糙集理论的建模。在此基础上,利用贝叶斯决策理论具体分析了在多粒度粗糙集模型中乐观和悲观的融合策略下多个粒空间中的概率融合关系,推导出基于最大条件概率和最小条件概率的粗糙集近似表示,进而构建了乐观多粒度决策粗糙集模型和悲观多粒度决策粗糙集模型。在该模型中引入近似分布约简的概念,分析了多个粒空间中的粒度选择问题。基于多粒度近似分布质量定义了多粒度决策粗糙集的粒度重要度,并且基于此给出了悲观和乐观融合策略 α -下近似分布约简的粒度约简算法。通过实例验证了该算法的有效性。

关键词 多粒度决策粗糙集,贝叶斯决策理论, α -下近似分布约简,粒度约简,近似分布质量

中图法分类号 TP182 文献标识码 A DOI 10.11896/j.issn.1002-137X.2017.05.036

Granular Structure Reduction Approach to Multigranulation Decision-theoretic Rough Sets

SANG Yan-li¹ QIAN Yu-hua^{1,2}

(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)¹

(Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China)²

Abstract Multigranulation decision-theoretic rough set method (MG-DTRS) is a generalization of multigranulation rough set model through combining the decision-theoretic rough sets theory and the multigranulation idea, which is a data modeling method on decision-theoretic rough sets in the context of multiple granular spaces. Further, based on Bayesian decision theory, we made a concrete analysis about probability fusion relations used optimistic or pessimistic fusion strategies on multiple granular spaces, also, the approximate representation of the maximum conditional probability rough sets and the minimum conditional probability rough sets were proposed respectively. And then the optimistic MG-DTRS model and the pessimistic MG-DTRS model were constructed. Furthermore, a concept of the approximate distribution reduction was introduced to MG-DTRS model, and the granular structure selection problem under multiple granular spaces was investigated. Based on the multiple granular approximate distribution quality proposed in this model, the important measure of a granular structure was defined, and an α -lower approximate distribution reduction algorithm to obtain a granular structure reduction was designed under optimistic or pessimistic fusion strategies respectively. Finally, an example was employed for verifying the validity of the proposed algorithm.

Keywords Multigranulation decision-theoretic rough sets, Bayesian decision theory, α -lower approximate distribution reduction, Granular structure reduction, Approximate distribution quality

1 引言

粗糙集理论是由波兰学者 Pawlak Z 于 20 世纪 80 年代初提出的一个用于数据分析的有力的数学工具^[1-3]。经典粗糙集理论是基于二元等价关系,用已有的知识对目标概念(决策)进行近似刻画。但 Pawlak 的粗糙集并不考虑决策规则的容错性,只考虑完全确定的规则。因此,许多学者在概念上、下近似集中引入概率包含关系,进而提出了具有噪声容忍机

制的概率粗糙集模型,如决策粗糙集模型、变精度粗糙集模型、贝叶斯粗糙集模型等^[4-6],其中,Yao 等于 20 世纪 90 年代初提出的决策粗糙集理论具有强大的理论基础和合理的语义解释^[6-7]。该理论基于贝叶斯决策过程,利用损失函数来描述决策风险,讨论了“贝叶斯期望风险最小”决策语义下决策规则的获取过程,给出了概率阈值的确定方法,进而赋予了概率阈值合理的语义解释。该理论充分考虑了决策风险对决策结果的影响,为解决不确定性复杂决策问题提供了一种新的思路。

到稿日期:2016-09-03 返修日期:2016-12-30 本文受国家自然科学基金项目(61672332),山西省煤基重点科技攻关项目(MQ2014-09)资助。

桑妍丽(1978—),女,博士生,讲师,主要研究方向为数据挖掘、智能决策,E-mail:sangyl@sxu.edu.cn;钱宇华(1976—),男,博士,教授,主要研究方向为粒计算、智能决策。

在粒计算视角下^[8],包括决策粗糙集模型在内的这些粗糙集模型,其目标概念的上/下近似都是由单一关系诱导的粒空间中的信息粒来近似表示。在面对如多源信息系统\分布式信息系统以及高维数据分析等知识背景时,目标概念和目标决策往往需要在多个粒空间中进行刻画,而多粒度粗糙集方法就是一种基于多个粒空间上的数据建模理论^[9-10]。在多粒度决策过程中,根据“求同存异”和“求同排异”的融合策略分别发展了乐观多粒度粗糙集和悲观多粒度粗糙集^[10-11]。文献^[12]进一步将多粒度数据分析的理念引入决策粗糙集中,建立了多粒度决策粗糙集模型,在这个框架下给出了3类多粒度决策粗糙集模型,分别为均值多粒度粗糙集模型、乐观多粒度粗糙集模型和悲观多粒度粗糙集模型。在多粒度粗糙集的研究中,如何在多粒度意义下进行粒度选择(粒度约简)是该模型特有的问题^[13]。粒度选择是指在不影响目标概念或目标决策的前提下选择一个无冗余的粒空间子集。近年来,国内的一些学者针对悲观多粒度粗糙集、变精度多粒度粗糙集等模型上的粒度选择问题做了一些探讨性的研究^[14-17],而多粒度决策粗糙集模型是基于决策粗糙集理论的一种泛化的多粒度粗糙集模型,因此研究该模型的粒度选择问题为解决实际风险决策问题提供了一定的理论基础。

本文针对多粒度意义下决策粗糙集贝叶斯决策过程的讨论,根据乐观和悲观两种粒度融合策略,具体给出了多粒度决策粗糙集的数据分析模型,并在此基础上重点发展了适合于多粒度决策粗糙集模型的 α -下近似分布粒度约简方法。为此引入近似分布约简的概念,定义了近似分布质量作为多粒度决策粗糙集模型中的不确定性度量方式,提出了粒度重要度的度量方式,分别在悲观、乐观融合概率意义下探讨在多个粒空间中如何进行粒度选取的问题。

2 相关概念

本节首先介绍多粒度粗糙集的基本概念。

2.1 乐观多粒度粗糙近似

在乐观多粒度粗糙集中,目标概念通过多个等价粒空间按照“求同存异”策略来近似刻画。“求同存异”策略是指每个决策者根据自己的粒空间进行决策,而不反对其他决策者所在的粒空间的决策,是一种乐观的决策策略^[10]。在多个独立的粒空间中,某个对象所在的信息粒只要存在一个信息粒包含在目标概念中,就将其放入下近似。其上近似则定义为目标概念补集的下近似的补集。

定义 1^[10] 设 $S=(U,AT)$ 是一个完备信息系统,其中 $R_1,R_2,\cdots,R_m\subseteq AT, \forall X\subseteq U$ 。X 关于 $R_1,R_2,\cdots,R_m$ 的乐观多粒度下/上近似分别定义为:

$$\begin{aligned} \sum_{i=1}^m R_i^O(X) &= \{x: [x]_{R_1} \subseteq X \vee [x]_{R_2} \subseteq X \vee \cdots \vee [x]_{R_m} \subseteq X, x \in U\} \\ \overline{\sum_{i=1}^m R_i^O(X)} &= \sim(\sum_{i=1}^m R_i^O(\sim X)) \end{aligned}$$

根据定义 1,X 的乐观多粒度粗糙集的边界域定义为:

$$Bn_{\sum_{i=1}^m R_i}^O(X) = \overline{\sum_{i=1}^m R_i^O(X)} - \sum_{i=1}^m R_i^O(X)$$

2.2 悲观多粒度粗糙近似

在悲观多粒度粗糙集中,目标概念仍然通过多个等价粒空间采用“求同排异”策略来近似刻画。即在决策过程中,所有决策者使用共同满意的方案进行决策,而存在分歧的方案则不能用于决策,是一种悲观或者保守的决策策略^[10]。在多个独立的粒空间中,某个对象所在的信息粒只有都包含在目标概念中才能将其放入下近似。其上近似则定义为目标概念补集的下近似的补集。

定义 2^[10] 设 $S=(U,AT)$ 是一个完备信息系统,其中 $R_1,R_2,\cdots,R_m\subseteq AT, \forall X\subseteq U$ 。X 关于 $R_1,R_2,\cdots,R_m$ 的悲观多粒度下/上近似分别定义为:

$$\begin{aligned} \sum_{i=1}^m R_i^P(X) &= \{x: [x]_{R_1} \subseteq X \wedge [x]_{R_2} \subseteq X \wedge \cdots \wedge [x]_{R_m} \subseteq X, x \in U\} \\ \overline{\sum_{i=1}^m R_i^P(X)} &= \sim(\sum_{i=1}^m R_i^P(\sim X)) \end{aligned}$$

根据定义 2,X 的悲观多粒度的边界域定义为:

$$Bn_{\sum_{i=1}^m R_i}^P(X) = \overline{\sum_{i=1}^m R_i^P(X)} - \sum_{i=1}^m R_i^P(X)$$

3 多粒度决策粗糙集

本节在多粒度决策粗糙集理论的基础上,利用贝叶斯决策过程,通过计算各种分类决策的风险损失值,找出具有最小期望风险的决策,依据乐观融合策略和悲观融合策略推导出在多粒度意义下的决策规则,进而得到了乐观多粒度决策粗糙集和悲观多粒度决策粗糙集的数据分析模型,并给出其相关性质。

3.1 多粒度粗糙集意义下的贝叶斯风险决策理论

区别于经典的粗糙集模型,多粒度粗糙集是通过多个等价关系导出的粒结构来进行多粒度数据建模。多个粒结构之间保持独立,但又遵循一定的融合策略^[12]。

给定 m 个粒度结构 $R_1,R_2,\cdots,R_m\subseteq AT, \forall X\subseteq U$,多粒度粗糙集的下、上近似可以分别表示为^[12]:

$$\begin{aligned} \sum_{i=1}^m R_i(X) &= f_l(R_1,R_2,\cdots,R_m) \\ \overline{\sum_{i=1}^m R_i(X)} &= f_u(R_1,R_2,\cdots,R_m) \end{aligned}$$

其中, f_l 为下近似融合函数, f_u 为上近似融合函数。给定 m 个粒结构,利用融合函数计算多粒度粗糙集的下、上近似。在实际应用中,根据不同的语义和需求,融合函数具有很多形式。

根据 Bayes 风险决策,可以得到给定对象 x ,在 m 个粒度结构下采取行动 a_i 的期望风险为:

$$R(a_i|x_1,x_2,\cdots,x_m) = \sum_{k=1}^m \sum_{j=1}^l \lambda_k(a_i|\omega_j)P(\omega_j|x_k)$$

其中, $\lambda_k(a_i|\omega_j)$ 表示在第 k 个粒结构的状 态 ω_j 下采取行动 a_i 的风险代价; $P(\omega_j|x_k)$ 为第 k 个粒结构下对象 x 在状态 ω_j 下的条件概率。这里假设损失值 $\lambda_k(a_i|\omega_j)=\lambda_l(a_i|\omega_j)(k,l\leq m)$,即在各 个粒结构下的损失函数值都相等。此时,在多粒度粗糙集中的信息融合可以简化为在相同论域中的概率集合。即期望风险可表示为:

$$R(a_i|x_1,x_2,\cdots,x_m) = \sum_{j=1}^l \lambda(a_i|\omega_j)P_{\sum_{i=1}^m R_i}(\omega_j|x_1,x_2,\cdots,x_m)$$

其中, $P_{\sum_{i=1}^m R_i}(\omega_j | x_1, x_2, \dots, x_m)$ 是在不同的多粒度融合策略下的概率统一表示形式。

考虑具有两种状态的状态集合 $\Omega = \{X, X^c\}$, 给定 m 个粒结构, 根据贝叶斯决策程序, 选择一种序列决策行动使所有对象的决策总风险尽可能小, 其中决策行动包括正域决策、负域决策和边界域决策, 记为决策集 $A = \{a_1, a_2, a_3\}$, 可以进一步得到其对应的期望损失为:

$$R_1 = R(a_1 | x_1, x_2, \dots, x_m) \\ = \lambda_{11} P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) + \lambda_{12} P_{\sum_{i=1}^m R_i}(X^c | x_1, x_2, \dots, x_m)$$

$$R_2 = R(a_2 | x_1, x_2, \dots, x_m) \\ = \lambda_{21} P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) + \lambda_{22} P_{\sum_{i=1}^m R_i}(X^c | x_1, x_2, \dots, x_m)$$

$$R_3 = R(a_3 | x_1, x_2, \dots, x_m) \\ = \lambda_{31} P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) + \lambda_{32} P_{\sum_{i=1}^m R_i}(X^c | x_1, x_2, \dots, x_m)$$

记 $\lambda_{i1} = \lambda(a_i | X)$, $\lambda_{i2} = \lambda(a_i | X^c)$, $i = 1, 2, 3$ 。

这样, 考虑误分类带来的损失, 依据贝叶斯决策过程对每个对象进行决策时, 为了获得尽可能小的风险决策, 可以得到如下关系。

(P) 若 $R_1 \leq R_2$ 且 $R_1 \leq R_3$, 则 $x \in POS(X)$;

(N) 若 $R_2 \leq R_1$ 且 $R_2 \leq R_3$, 则 $x \in NEG(X)$;

(B) 若 $R_3 \leq R_1$ 且 $R_3 \leq R_2$, 则 $x \in BND(X)$ 。

决策代价函数值的大小满足: $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$, $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, 进而可得到如下含阈值形式的决策规则, 即:

(P1) 若 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \geq \alpha$ 且 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \geq \gamma$, 则 $x \in POS(X)$;

(N1) 若 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \leq \beta$ 且 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \leq \gamma$, 则 $x \in NEG(X)$;

(B1) 若 $\beta \leq P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \leq \alpha$, 则 $x \in BND(X)$ 。

$$\text{此时, } \alpha = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})}$$

$$\gamma = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}$$

$$\beta = \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}$$

决策损失值进一步满足条件 $(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22})$, 即 $\alpha > \gamma > \beta$ 。

可以得到如下简化的决策规则:

(P2) 若 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \geq \alpha$, 则 $x \in POS(X)$;

(N2) 若 $P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) \leq \beta$, 则 $x \in NEG(X)$;

(B2) 若 $\beta < P_{\sum_{i=1}^m R_i}(X | x_1, x_2, \dots, x_m) < \alpha$, 则 $x \in BND(X)$ 。

此时, 在 m 个粒结构中可以通过不同的融合策略得到的条件概率来判断任意一个对象 x 是否包含在目标概念中, 这里考虑两种策略, 即乐观融合策略和悲观融合策略。

3.2 乐观多粒度决策粗糙集模型

在乐观的决策策略下, 在多粒度空间中只要存在一个粒结构对象 x 能够划分到目标概念 X 的条件概率 $P(X | x_k)$ 满足阈值条件, 对象 x 即可被划分到目标概念的正域。基于此, 可以得到如下规则:

(OP) 若 $\exists j \in \{1, 2, \dots, m\}$, 使得 $P(X | [x]_{R_j}) \geq \alpha$ 成立, 则 $x \in POS(X)$;

(ON) 若 $\forall j \in \{1, 2, \dots, m\}$, 使得 $P(X | [x]_{R_j}) \leq \beta$ 成立, 则 $x \in NEG(X)$;

(OB) 否则 $x \in BND(X)$ 。

定义 3 设 $S = (U, AT)$ 是一个完备信息系统, 给定 m 个粒结构 $R_1, R_2, \dots, R_m \subseteq AT$, $\forall X \subseteq U$, 则乐观多粒度决策粗糙集的下、上近似分别定义为:

$$\underline{\sum_{i=1}^m R_i}^{O, \alpha}(X) = \{x; \max_{j=1}^m P(X | [x]_{R_j}) \geq \alpha, x \in U\}$$

$$\overline{\sum_{i=1}^m R_i}^{O, \beta}(X) = U - \{x; \max_{j=1}^m P(X | [x]_{R_j}) \leq \beta, x \in U\}$$

根据定义 3, X 的乐观多粒度决策粗糙集边界域定义为:

$$Bn_{\sum_{i=1}^m R_i}^O(X) = \overline{\sum_{i=1}^m R_i}^{O, \beta}(X) - \underline{\sum_{i=1}^m R_i}^{O, \alpha}(X)$$

定理 1 给定 $\{R_1, R_2, \dots, R_m\}$ 为 m 个粒结构, $\forall X \subseteq U$, 下列性质成立。

$$(1) \underline{\sum_{i=1}^m R_i}^{O, \alpha}(X) \supseteq \underline{R_i}^{\alpha}(X), i \leq m.$$

$$(2) \overline{\sum_{i=1}^m R_i}^{O, \beta}(X) \subseteq \overline{R_i}^{\beta}(X), i \leq m.$$

$$(3) \underline{\sum_{i=1}^m R_i}^{O, \alpha}(X) = \bigcup_{i=1}^m \underline{R_i}^{\alpha}(X).$$

$$(4) \overline{\sum_{i=1}^m R_i}^{O, \beta}(X) = \bigcap_{i=1}^m \overline{R_i}^{\beta}(X).$$

其中, $\underline{R_i}^{\alpha}(X) = \{x \in U | P(X | [x]_{R_i}) \geq \alpha\}$, $\overline{R_i}^{\beta}(X) = \{x \in U | P(X | [x]_{R_i}) \geq \beta\}$ 。

$$(5) \forall X_1 \subseteq X_2 \subseteq U \Rightarrow \underline{\sum_{i=1}^m R_i}^{O, \alpha}(X_1) \subseteq \underline{\sum_{i=1}^m R_i}^{O, \alpha}(X_2).$$

$$(6) \alpha_1 \leq \alpha_2 \Rightarrow \underline{\sum_{i=1}^m R_i}^{O, \alpha_1}(X) \supseteq \underline{\sum_{i=1}^m R_i}^{O, \alpha_2}(X).$$

3.3 悲观多粒度决策粗糙集模型

在悲观的决策策略下, 在多粒度空间中所有的粒结构都满足对象 x 能够划分到目标概念 X 的条件概率 $P(X | x_k)$ 满足阈值条件, 则对象 x 即可划分到目标概念的正域。可以得到如下规则:

(PP) 若 $\forall j \in \{1, 2, \dots, m\}$, 使得 $P(X | [x]_{R_j}) \geq \alpha$ 成立, 则 $x \in POS(X)$;

(PN) 若 $\exists j \in \{1, 2, \dots, m\}$, 使得 $P(X | [x]_{R_j}) \leq \beta$ 成立, 则 $x \in NEG(X)$;

(PB) 否则 $x \in BND(X)$ 。

定义 4 给定 m 个粒结构 $R_1, R_2, \dots, R_m \subseteq AT$, $\forall X \subseteq U$, 则悲观多粒度决策粗糙集的下、上近似分别定义为:

$$\underline{\sum_{i=1}^m R_i}^{P, \alpha}(X) = \{x; \min_{j=1}^m P(X | [x]_{R_j}) \geq \alpha, x \in U\}$$

$$\overline{\sum_{i=1}^m R_i}^{P, \beta}(X) = U - \{x; \min_{j=1}^m P(X | [x]_{R_j}) \leq \beta, x \in U\}$$

根据定义 4, X 的悲观多粒度决策粗糙集边界域定义为:

$$Bn_{\sum_{i=1}^m R_i}^P(X) = \overline{\sum_{i=1}^m R_i}^{P, \beta}(X) - \underline{\sum_{i=1}^m R_i}^{P, \alpha}(X)$$

定理 2 给定 $A=\{R_1, R_2, \dots, R_m\}$ 为 m 个粒结构, $\forall X \subseteq U$, 下列性质成立。

$$(1) \sum_{i=1}^m \underline{R_i}^{P,\alpha}(X) \subseteq \underline{R_i}^{\alpha}(X), i \leq m.$$

$$(2) \sum_{i=1}^m \underline{R_i}^{P,\beta}(X) \supseteq \underline{R_i}^{\beta}(X), i \leq m.$$

$$(3) \sum_{i=1}^m \underline{R_i}^{P,\alpha}(X) = \bigcap_{i=1}^m \underline{R_i}^{\alpha}(X).$$

$$(4) \sum_{i=1}^m \underline{R_i}^{P,\beta}(X) = \bigcup_{i=1}^m \underline{R_i}^{\beta}(X).$$

其中, $\underline{R_i}^{\alpha}(X) = \{x \in U | P(X | [x]_{R_i}) \geq \alpha\}$, $\underline{R_i}^{\beta}(X) = \{x \in U | P(X | [x]_{R_i}) \geq \beta\}$.

$$(5) \forall X_1 \subseteq X_2 \subseteq U \Rightarrow \sum_{i=1}^m \underline{R_i}^{P,\alpha}(X_1) \subseteq \sum_{i=1}^m \underline{R_i}^{P,\alpha}(X_2).$$

$$(6) \alpha_1 \leq \alpha_2 \Rightarrow \sum_{i=1}^m \underline{R_i}^{P,\alpha_1}(X) \supseteq \sum_{i=1}^m \underline{R_i}^{P,\alpha_2}(X).$$

4 多粒度决策粗糙集模型的粒度约简

对于多粒度下的目标决策近似,有些粒结构在目标决策中意义不大,去除后并不影响目标决策的近似刻画,这就产生了多粒度粗糙集中的粒度选择问题,选择出来的粒空间集合称为一个粒度约简。粒度约简的本质就是在给定的多个粒空间上求取一个无冗余的粒空间集合。

4.1 多粒度决策粗糙集的 α -下近似分布约简

为了不改变多粒度空间下对目标决策的刻画,引入分布约简的概念^[18],考虑悲观多粒度决策粗糙集和乐观多粒度决策粗糙集的粒度约简问题。为简化起见,假定对象对每个决策类 $Y_k \in U/d$ 的误分类代价 λ_{ij} 相同。

定义 5 设 $S=(U, AT \cup \{d\})$ 是一个完备的决策信息系统, $A \subseteq AT = \{R_1, R_2, \dots, R_n\}$, $U/d = \{Y_1, Y_2, \dots, Y_r\}$, 则多粒度决策粗糙集的 α -下近似分布定义为:

$$\mu_{AT}^{\Delta,\alpha}(d) = \{ \sum_{R_i \in AT} R_i^{\Delta,\alpha}(Y_1), \sum_{R_i \in AT} R_i^{\Delta,\alpha}(Y_2), \dots, \sum_{R_i \in AT} R_i^{\Delta,\alpha}(Y_r) \}$$

其中, $\Delta \in \{P, O\}$ 表示悲观多粒度决策粗糙集和乐观多粒度决策粗糙集两种情况。

若 $\mu_A^{\Delta,\alpha}(d) = \mu_{AT}^{\Delta,\alpha}(d)$, 则称 A 为 S 的粒度 α -下近似分布一致集。若 A 是 S 的 α -下近似分布一致集, 且对于 $\forall A' \subset A$, 都有 $\mu_{A'}^{\Delta,\alpha}(d) \neq \mu_{AT}^{\Delta,\alpha}(d)$, 则称 A 为 S 的多粒度 α -下近似分布约简。

定理 3 设 $S=(U, AT \cup \{d\})$ 是一个信息系统, $X \subseteq U$, 粒度集 $A=\{R_1, R_2, \dots, R_m\} \subseteq AT$, 若 $A' \subseteq A$, 则:

$$\sum_{R_i \in A} \underline{R_i}^{P,\alpha}(X) \subseteq \sum_{R_i \in A'} \underline{R_i}^{P,\alpha}(X)$$

$$\sum_{R_i \in A} \underline{R_i}^{O,\alpha}(X) \supseteq \sum_{R_i \in A'} \underline{R_i}^{O,\alpha}(X)$$

证明: 由定理 2 可知, $\sum_{R_i \in A} \underline{R_i}^{P,\alpha}(X) = \sum_{i=1}^m \underline{R_i}^{P,\alpha}(X) =$

$$\bigcap_{i=1}^m \underline{R_i}^{\alpha}(X), \sum_{R_i \in A'} \underline{R_i}^{P,\alpha}(X) = \bigcap_{R_i \in A'} \underline{R_i}^{\alpha}(X).$$

显然, $\bigcap_{i=1}^m \underline{R_i}^{\alpha}(X) \subseteq \bigcap_{R_i \in A'} \underline{R_i}^{\alpha}(X)$, 即 $\sum_{R_i \in A} \underline{R_i}^{P,\alpha}(X) \subseteq$

$$\sum_{R_i \in A'} \underline{R_i}^{P,\alpha}(X) \text{ 成立。}$$

同理可证明 $\sum_{R_i \in A} \underline{R_i}^{O,\alpha}(X) \supseteq \sum_{R_i \in A'} \underline{R_i}^{O,\alpha}(X)$ 成立。

基于多粒度粗糙集模型,用多等价关系的下近似来刻画目标决策。由上述定理可知,在乐观多粒度决策粗糙集中,随着粒空间的增加,下近似呈递增的趋势;而在悲观多粒度粗糙集中,随着粒空间的增加,放入下近似中的对象数呈递减的趋势。为了真实刻画目标决策,用下近似域的变化来衡量粒度的重要度。

4.2 粒度重要度量

在多个粒度空间中,哪些粒度对于目标决策是重要的,哪些粒度是冗余的。首先给出如下粒度重要性的定义。

定义 6 设 $S=(U, AT \cup \{d\})$ 是一个完备的决策信息系统, $A=\{R_1, R_2, \dots, R_m\} \subseteq AT$, 给定粒度 $R \in A$, $Y \in U/d$ 。若 $\sum_{i=1}^m \underline{R_i}^{\Delta,\alpha}(Y) \neq \sum_{i=1, R_i \neq R}^m \underline{R_i}^{\Delta,\alpha}(Y)$, 则称粒度 R 在 A 中关于 Y 是重要的; 若 $\sum_{i=1}^m \underline{R_i}^{\Delta,\alpha}(Y) = \sum_{i=1, R_i \neq R}^m \underline{R_i}^{\Delta,\alpha}(Y)$, 则称粒度 R 在 A 中关于 Y 是不重要的。

在多粒度决策粗糙集中,其下近似分布可能存在同一个对象被分到不同决策类的不一致情况,传统的近似质量已经不能刻画这种不一致,因此定义近似分布质量作为多粒度决策粗糙集的度量方式。

定义 7 设 $S=(U, AT \cup \{d\})$ 是一个完备的决策信息系统, 粒度集 $A=\{R_1, R_2, \dots, R_m\} \subseteq AT$ 为 m 个粒结构。用粒度集 A 表示 $Y_j \in U/d (j=1, 2, \dots, r)$ 的近似分布质量定义为:

$$\gamma_{\sum_{i=1}^m R_i}^{\Delta,\alpha}(d) = \left(\frac{\sum_{j=1}^r \sum_{i=1}^m \underline{R_i}^{\Delta,\alpha}(Y_j)}{|U|} \right) / r$$

推论 1 若 $A' \subseteq A$, 则 $\gamma_{A'}^{\Delta,\alpha}(d) \leq \gamma_A^{\Delta,\alpha}(d)$; 若 $A' \subseteq A$, 则 $\gamma_{A'}^{P,\alpha}(d) \geq \gamma_A^{P,\alpha}(d)$ 。

依据推论 1 可以得到,在多粒度框架下随着粒度空间的单调变化,多粒度决策粗糙集近似质量也呈现单调变化,并且能够完全表示出多粒度决策空间中下近似分布的变化。

为了真实刻画目标决策,借助于多粒度下的近似分布质量 γ 来衡量粒度的重要度。下面分别给出粒度内部重要性和外部重要性的度量定义。

定义 8 设 $S=(U, AT \cup \{d\})$ 是一个完备的决策信息系统, $A \subseteq AT = \{R_1, R_2, \dots, R_n\}$, $\forall R \in A$, 粒度 R 在粒度集 A 的关于 d 的粒度内部重要性度量 $Sig_{inner}^{\Delta,\alpha}(R, A, d)$ 定义为:

$$Sig_{inner}^{\Delta,\alpha}(R, A, d) = |\gamma_A^{\Delta,\alpha}(d) - \gamma_{A \setminus \{R\}}^{\Delta,\alpha}(d)|; \Delta \in \{P, O\}$$

定义 9 设 $S=(U, AT \cup \{d\})$ 是一个完备的决策信息系统, $A \subseteq AT = \{R_1, R_2, \dots, R_n\}$, $\forall R \in AT - A$, 粒度 R 在粒度集 A 的关于 d 的粒度外部重要性度量 $Sig_{outer}^{\Delta,\alpha}(R, A, d)$ 定义为:

$$Sig_{outer}^{\Delta,\alpha}(R, A, d) = |\gamma_A^{\Delta,\alpha}(d) - \gamma_{A \cup \{R\}}^{\Delta,\alpha}(d)|; \Delta \in \{P, O\}$$

上述重要性度量满足以下性质:

$$(1) Sig_{outer}^{\Delta,\alpha}(R, A, d) \geq 0, Sig_{inner}^{\Delta,\alpha}(R, A, d) \geq 0;$$

(2) 粒度 R 在粒度集 A 中关于 d 不是下近似重要的, 当且仅当 $Sig_{inner}^{\Delta,\alpha}(R, A, d) = 0$;

(3) $Sig_{outer}^{\Delta,\alpha}(R, A, d)$ 的值越大, 说明在已知粒度集 A 的条件下, 粒度 R 对决策 d 就越重要。

定义 10 设 $S=(U, AT \cup \{d\})$ 是一个决策信息系统, 粒度集 $A=\{R_1, R_2, \dots, R_m\} \subseteq AT$ 的核定义为:

$$Core(A)=\{R_i \in A | Sig_{outer}^{\Delta}(R_i, A, d) > 0\}$$

定义 11 设 $S=(U, AT \cup \{d\})$ 是一个决策信息系统, 粒度集 $A \subseteq AT, U/d=\{Y_1, Y_2, \dots, Y_r\}$.

$$(1) \gamma_A^{\Delta, \alpha}(d) = \gamma_{AT}^{\Delta, \alpha}(d);$$

$$(2) \text{对于 } \forall R \in A, \text{ 都有 } \gamma_{A-(R)}^{\Delta, \alpha}(d) \neq \gamma_{AT}^{\Delta, \alpha}(d).$$

则称 A 为 S 的多粒度 α -下近似分布约简。

以上基于多粒度决策粗糙集的近似分布质量给出了下近似分布的粒度约简定义, 相应的粒度约简算法的介绍如下。

4.3 多粒度决策粗糙集中的粒度约简算法

该算法首先根据粒度内部重要度获得一个原始多粒度空间下不可去除的粒度构成的集合, 以这个粒度集作为粒度核集, 将 $Sig_{outer}^{\Delta}(R, A, d)$ 作为粒度重要性的度量, 每次都选择将当前多粒度空间下被认为是最重要的粒度添加到约简粒度集中, 将近似分布质量 γ 作为约简算法的停止条件。在此算法中, 可以根据多粒度决策粗糙集模型的乐观和悲观融合策略的特点分别对算法做适当优化, 以提高算法的效率。

定理 4 设 $S=(U, AT)$ 是一个完备的信息系统, 其中 $A=\{R_1, R_2, \dots, R_m\} \subseteq AT, \forall X \subseteq U$ 。则有

$$\sum_{i=1}^m R_i^{O, \alpha}(X) = \{x: \exists i(i=1 \dots m), P(X|[x]_{R_i}) \geq \alpha, x \in U\}$$

推论 2 $\forall x \in U$, 若 $\exists i$, 使得 $P(X|[x]_{R_i}) \geq \alpha$, 则 $[x]_{A_i} \subseteq$

$$\sum_{i=1}^m R_i^{O, \alpha}(X)。$$

在求取下近似集合的过程中, 随着粒度空间的增加, 乐观多粒度决策粗糙集下近似集合中的对象呈现递增的趋势, 因此, 在根据推论 2 求取乐观多粒度空间下近似集合的过程中, 只需要判断对象 x 在任意一个粒结构中的条件概率是否满足阈值条件, 若满足则该对象的等价类也满足目标概念的下近似集合。

定理 5 设 $S=(U, AT)$ 是一个完备的信息系统, 其中 $R_1, R_2, \dots, R_m \subseteq AT, \forall X \subseteq U$, 则有:

$$\sim(\sum_{i=1}^m R_i^{P, \alpha}(X)) = \{x: \exists i(i=1 \dots m), P(X|[x]_{R_i}) < \alpha,$$

$x \in U\}$

推论 3 $\forall x \in U$, 若 $\exists i$, 使得 $P(X|[x]_{R_i}) < \alpha$, 则 $[x]_{R_i} \not\subseteq$

$$\sum_{i=1}^m R_i^{P, \alpha}(X)。$$

在求取下近似集合的过程中, 随着粒度空间的增加, 悲观多粒度粗糙集下近似集合中的对象呈现递减的趋势, 因此根据推论 3 可以在求取多粒度空间下近似集合的过程中将不满足下近似条件的对象及其等价类从集合中去除, 以加速求得下近似集合。

依据上述推论, 可以加速求得下近似集合。下面分别给出在乐观和悲观意义下在多个等价粒空间计算集合 X 的下近似的加速算法。

算法 1 乐观多粒度决策下近似求解算法

输入: 信息系统 $S=(U, AT)$, 集合 $X \subseteq U$, 损失函数 $\lambda_{ij}, i, j=1, 2, 3$;

输出: 在粒度集 AT 下, 论域 U 上由 AT 表示的 X 的乐观多粒度决策下近似集合 L^O 。

1. 设置 $L^O \leftarrow \emptyset, U' \leftarrow U$;

2. 由损失函数计算 α 的值;

3. i from 1 to $|AT|$ 在论域 U 上计算 $[x]_{R_i}$;

//求取各个粒结构下的等价类。

4. i from 1 to $|AT|$

$\forall x \in U'$, 循环执行直到 $U' = \emptyset$;

4.1 If $P(X|[x]_{R_i}) \geq \alpha$ then $L^O \leftarrow L^O \cup [x]_{R_i}$;

//根据定理 5, 在某一个粒度下满足阈值要求, 则加入下近似集合。

4.2. $U' \leftarrow U' \setminus ([x]_{R_i} \cap U')$;

$U' \leftarrow U \setminus L^O$;

5. 输出下近似集合 L^O 。

算法 2 悲观多粒度决策下近似求解算法

输入: 信息系统 $S=(U, AT)$, 集合 $X \subseteq U$, 损失函数 $\lambda_{ij}, i, j=1, 2, 3$;

输出: 在粒度集 AT 下, 论域 U 上由 AT 表示的 X 的悲观多粒度决策下近似集合 L^P 。

1. 设置 $L^P \leftarrow U, U' \leftarrow U, L' \leftarrow U$;

2. 由损失函数计算 α 的值;

3. i from 1 to $|AT|$ 在论域 U 上计算 $[x]_{R_i}$;

4. i from 1 to $|AT|$

$\forall x \in U'$, 循环执行直到 $U' = \emptyset$;

4.1 If $P(X|[x]_{R_i}) < \alpha$ then $L' \leftarrow L' \setminus ([x]_{R_i} \cap L')$;

//在某一个粒度结构下不满足下近似要求, 则剔除。

4.2 $U' \leftarrow U' \setminus ([x]_{R_i} \cap U')$;

4.3 If $L' = \emptyset$ 可结束循环;

$U' \leftarrow L'$;

5. $L^P \leftarrow L'$;

6. 输出下近似集合 L^P 。

依据上述原理, 以 $Sig_{outer}^{\Delta}(R, A, d)$ 作为粒度重要性的启发式信息, 设计出如下多粒度决策粗糙集的下近似分布约简算法。

算法 3 α -下近似分布粒度约简算法

设 $S=(U, AT \cup \{d\})$ 是一个完备的决策系统, $AT=\{R_1, R_2, \dots, R_{|AT|}\}, U/d=\{Y_1, Y_2, \dots, Y_r\}$ 。

输入: 决策表 $S=(U, AT \cup \{d\})$;

输出: 该决策表的一个粒度约简 RED。

1. 设置 $Core_d(AT) \leftarrow \emptyset$;

2. 计算 U/d , 根据算法 1 和算法 2 分别计算每个决策类 Y_k 的下近似集合。

3. 对于每一个 $R_i \in AT$, 计算 $Sig_{inner}^{\Delta}(R_i, AT, d)$;

If $Sig_{inner}^{\Delta}(R_i, AT, d) > 0$, then $Core_d(AT) = Core_d(AT) \cup R_i$;

4. $RED \leftarrow Core_d(AT)$;

5. 将其余粒度记入集合 $Q = AT - RED$;

6. While $\gamma_{\sum_{R_i \in RED} R_i}^{\Delta, \alpha}(d) \neq \gamma_{\sum_{R_i \in AT} R_i}^{\Delta, \alpha}(d)$ Do

6.1 $\forall R_i \in Q$ Do 计算 $Sig_{outer}^{\Delta}(R_i, RED, d)$;

6.2 $R_{max} \leftarrow \arg \max_{R_i} Sig_{outer}^{\Delta}(R_i, RED, d)$;

6.3 $Q \leftarrow Q - \{R_{max}\}$;

6.4. $RED \leftarrow RED \cup R_{max}$;

7. 输出粒度约简集合 RED。

计算各个粒度下对论域 U 做 $|AT|$ 个划分的时间复杂度 $O(|AT| \cdot |U|^2)$, 因此计算下近似的时间复杂度是 $O(|AT| \cdot |U|^2)$

(见算法 1、算法 2),则计算由 AT 表示 $Y_k \in U/d(k=1,2,\cdots, r)$ 的下近似的时间复杂度为 $O((r+1)|AT||U|^2)$ 。因此,算法 3 的时间复杂度为: $O((r+1)|AT||U|^2|U|^2 + \sum_{i=1}^{|AT|} (|AT| - i+1)(r+1)|AT||U|^2)$ 。

5 实例分析

本节将多粒度决策粗糙集的粒度约简算法应用于企业风险投资问题,以说明该算法的可行性和有用性。考虑该企业 10 个在线的投资方案,分别在 5 个网站上进行评估,表 1—表 5 是分别从 5 个站点得到的关于公司投资方案的评估表数据。将整个评估表看作一个完备决策信息系统,论域 $U=\{x_i, i \in \{1, \cdots, 10\}\}$ 表示 10 个投资方案,条件属性集 $C=\{a_1, a_2, a_3, a_4, a_5\}$ 表示给出的评价指标,各个站点选择合适的评价指标来进行评价,其中评价标准被分成 3 类,即 $e=\{1, 2, 3\}$,分别表示{优,中,差},决策属性 d 表示方案是否通过评估。这里假定每个站点对于方案评估的风险损失相等,给出相应的风险损失矩阵,如表 6 所列。在评估过程中,将每个站点的评估看成一个粒度空间,在多粒度决策空间下考虑该问题。分别采用乐观融合策略和悲观融合策略来分别进行方案评估,利用本文提出的多粒度决策粗糙集的粒度约简算法对其进行粒度约简,找到最小评估站点集。

表 1 站点 1 的评估表

方案	a_1	a_2	a_3
x_1	1	2	2
x_2	1	2	2
x_3	1	2	2
x_4	2	1	1
x_5	3	2	2
x_6	2	1	1
x_7	3	2	2
x_8	3	2	2
x_9	3	2	3
x_{10}	3	2	3

表 2 站点 2 的评估表

方案	a_1	a_3	a_4
x_1	1	2	1
x_2	1	2	1
x_3	2	3	3
x_4	1	2	1
x_5	2	3	3
x_6	2	1	2
x_7	2	3	3
x_8	2	1	2
x_9	2	1	2
x_{10}	3	3	2

表 3 站点 3 的评估表

方案	a_2	a_3	a_5
x_1	1	1	2
x_2	2	2	3
x_3	2	2	3
x_4	1	1	2
x_5	2	2	3
x_6	1	1	2
x_7	3	3	2
x_8	2	1	1
x_9	3	3	2
x_{10}	3	3	2

表 4 站点 4 的评估表

方案	a_1	a_2	a_3	a_4
x_1	2	2	2	1
x_2	2	2	2	1
x_3	2	2	2	1
x_4	1	1	3	2
x_5	2	2	2	1
x_6	1	3	2	1
x_7	1	3	2	1
x_8	1	3	2	1
x_9	3	2	3	2
x_{10}	3	2	3	2

表 5 站点 5 的评估表

方案	a_1	a_2	a_5
x_1	1	2	2
x_2	2	2	3
x_3	1	2	2
x_4	1	2	2
x_5	3	2	2
x_6	1	2	2
x_7	2	2	3
x_8	3	2	2
x_9	3	3	1
x_{10}	3	3	1

表 6 损失函数表

	X	X^c
a_P	0	8
a_B	2	2
a_N	6	0

粒度集 $AT=\{R_1, R_2, R_3, R_4, R_5\}, U/d=\{\{x_1, x_2, x_4, x_6, x_8\}, \{x_3, x_5, x_7, x_9, x_{10}\}\}$, 即 $Y_1=\{x_1, x_2, x_4, x_6, x_8\}, Y_2=\{x_3, x_5, x_7, x_9, x_{10}\}$ 。

对于每一个站点,计算对应的粒结构下的划分:

$U/R_1=\{\{x_1, x_2, x_3\}, \{x_4, x_6\}, \{x_5, x_7, x_8\}, \{x_9, x_{10}\}\}$

$U/R_2=\{\{x_1, x_2, x_4\}, \{x_3, x_5, x_7\}, \{x_6, x_8, x_9\}, \{x_{10}\}\}$

$U/R_3=\{\{x_1, x_4, x_6\}, \{x_2, x_3, x_5\}, \{x_8\}, \{x_7, x_9, x_{10}\}\}$

$U/R_4=\{\{x_1, x_2, x_3, x_5\}, \{x_4\}, \{x_6, x_7, x_8\}, \{x_9, x_{10}\}\}$

$U/R_5=\{\{x_1, x_3, x_4, x_6\}, \{x_2, x_7\}, \{x_5, x_8\}, \{x_9, x_{10}\}\}$

根据损失函数计算阈值 $\alpha=0.75$,在当前阈值下分别计算乐观多粒度决策下的近似分布和悲观多粒度决策下的近似分布。

$\mu_{AT}^{P,a}(d)=\{\{x_4\}, \{x_{10}\}\}$

$\mu_{AT}^{O,a}(d)=\{\{x_1, x_2, x_3, x_4, x_6, x_8\}, \{x_3, x_5, x_7, x_9, x_{10}\}\}$

接着,根据定义 8 计算每个站点的内部重要度。

$Sig_{inner}^P(R_1, AT, d)=0.1-0.1=0$

$Sig_{inner}^P(R_2, AT, d)=0.15-0.1=0.05$

$Sig_{inner}^P(R_3, AT, d)=0.1-0.1=0$

$Sig_{inner}^P(R_4, AT, d)=0.1-0.1=0$

$Sig_{inner}^P(R_5, AT, d)=0.1-0.1=0$

根据这些粒度内部重要度的数值,选择 $R=\{R_2\}$ 作为启发式粒度约简算法的起点粒度集。

计算 $\gamma_{\{R_2\}}^{P,a}(d)=0.35$,判定 $\gamma_{\{R_2\}}^{P,a}(d) \neq \gamma_{AT}^{P,a}(d)$,根据定义 9 计算粒度的外部重要度来获得下一个站点。

$Sig_{outer}^P(R_1, R_2, d)=0.35-0.1=0.25$

$Sig_{outer}^P(R_3, R_2, d)=0.35-0.2=0.15$

$Sig_{outer}^P(R_4, R_2, d)=0.35-0.1=0.25$

$Sig_{outer}^P(R_5, R_2, d)=0.35-0.15=0.2$

选择将重要度最大的站点加入到约简粒度集中,得:

$\gamma_{\{R_2, R_1\}}^{P,a}(d)=\gamma_{\{R_2, R_4\}}^{P,a}(d)=\gamma_{AT}^{P,a}(d)$

因此, $RED^P=\{R_1, R_2\}$ 或者 $\{R_2, R_4\}$ 即为悲观多粒度决策粗糙集的粒度约简集。

同理计算乐观多粒度决策粗糙集粒结构的约简集。

$Sig_{inner}^O(R_1, AT, d)=0.55-0.55=0$

$Sig_{inner}^O(R_2, AT, d)=0.55-0.4=0.15$

$Sig_{inner}^O(R_3, AT, d)=0.55-0.5=0.05$

$Sig_{inner}^O(R_4, AT, d)=0.55-0.55=0$

$Sig_{inner}^O(R_5, AT, d)=0.55-0.5=0.05$

选择 $\{R_2, R_3, R_5\}$ 作为启发式粒度约简算法的起点粒度集。

$\gamma_{\{R_2, R_3, R_5\}}^{O,a}(d)=\gamma_{AT}^{O,a}(d)=0.55$,此时满足 $\mu_{\{R_2, R_3, R_5\}}^{O,a}(d)=\{\{x_1, x_2, x_3, x_4, x_6, x_8\}, \{x_3, x_5, x_7, x_9, x_{10}\}\}$, 即 $\mu_{\{R_2, R_3, R_5\}}^{O,a}(d)=\mu_{AT}^{O,a}(d)$,因此 $RED^O=\{R_2, R_3, R_5\}$ 为乐观多粒度决策粗糙集的粒度约简集。

实例表明,在计算多粒度决策粗糙集的粒度约简时,利用损失函数得到概率阈值,将近似分布质量的变化作为粒度重要度的度量,分别得到了乐观/悲观多粒度决策粗糙集意义下的粒度约简集,最终求得的粒度约简与原始多粒度空间具有同样的决策能力。在本实例中每个站点的评估互相独立的情况下,乐观多粒度意义下,站点2、站点3和站点5的评估构成了最终的粒度约简,而站点1和站点4的评估可以忽略。悲观多粒度意义下,站点1和站点2的评估或者站点2和站点4的评估构成了最终的粒度约简。

结束语 本文在泛化的多粒度决策粗糙集模型的基础上进一步分析了多个粒度空间在乐观和悲观的融合策略下的概率融合关系,推导出基于最大条件概率和最小条件概率的粗糙集近似表示,并且基于此构建了乐观多粒度决策粗糙集模型和悲观多粒度决策粗糙集模型。基于该模型提出了基于 α -下近似分布的粒度约简算法,并通过实例验证了算法的有效性。针对多粒度决策粗糙集的粒度选择问题的研究为多粒度粗糙集方法在解决实际的风险决策问题时提供了新的思路和方法。在多粒度决策粗糙集中的粒选择以及进一步的决策分析都将是以后研究工作的重点。

参考文献

- [1] PAWLAK Z. Rough Set[J]. International Journal of Computer and Information Science, 1982, 11: 341-356.
- [2] PAWLAK Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177: 3-27.
- [3] 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: 42-67.
- [4] ZIARKO W. Variable precision rough set model[J]. Journal of Computer System Sciences, 1993, 46(1): 39-59.
- [5] SLEZAK D, ZIARKO W. The investigation of the Bayesian rough set model[J]. International Journal of Approximate Reasoning, 2005, 40: 81-91.
- [6] YAO Y Y, WONG S K M. A decision theoretic framework for approximating concepts[J]. International Journal of Man-machine Studies, 1992, 37: 793-809.
- [7] YAO Y Y. Three-way decision with probabilistic rough sets[J]. Information Sciences, 2010, 180: 341-353.
- [8] LIN T Y. Granular computing on binary relations II: Rough set representations and belief functions [M] // Rough Sets and Knowledge Discovery. 1998: 122-140.
- [9] YAO Y Y, SHE Y H. Rough set models in multigranulation spaces[J]. Information Sciences, 2016, 327: 40-56.
- [10] QIAN Y H, LIANG J Y, YAO Y Y, et al. MGRS: a multi-granulation rough set[J]. Information Sciences, 2010, 180: 949-970.
- [11] QIAN Y H, LI S Y, LIANG J Y, et al. Pessimistic rough set based decisions: A multigranulation fusion strategy[J]. Information Sciences, 2014, 264(20): 196-210.
- [12] QIAN Y H, ZHANG H, SANG Y L, et al. Multigranulation decision-theoretic rough sets[J]. International Journal Approximate Reason, 2014, 55: 225-237.
- [13] SANG Y L, QIAN Y H. A Granular Space Reduction Approach to Pessimistic Multi-Granulation Rough Sets[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(3): 361-366. (in Chinese)
- [14] 桑妍丽, 钱宇华. 一种悲观多粒度粗糙集中的粒度约简算法[J]. 模式识别与人工智能, 2012, 25(3): 361-366.
- [15] ZHANG M, TANG Z M, XU W Y, et al. Variable Multi-granulation Rough Set Model[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(4): 709-720. (in Chinese)
- [16] 张明, 唐振民, 徐维艳, 等. 可变多粒度粗糙集模型[J]. 模式识别与人工智能, 2012, 25(4): 709-720.
- [17] MENG H L, MA Y Y, XU J C. Granularity Reduction of Variable Precision Pessimistic Multi-granulation Rough Set Based on Granularity Entropy of Lower Approximate Distribution[J]. Computer Science, 2016, 43(2): 83-86. (in Chinese)
- [18] 孟慧丽, 马媛媛, 徐久成. 基于下近似分布粒度熵的变精度悲观多粒度粗糙集粒度约简[J]. 计算机科学, 2016, 43(2): 83-86.
- [19] YANG X B, QI Y, YU H L, et al. Updating multigranulation rough approximations with increasing of granular structures [J]. Knowledge-Based Systems, 2014, 64: 59-69.
- [20] ZHAI Y J, ZHANG H. Reduction of Variable Precision Multi-granulation Rough Sets[J]. Journal of Jinling Institute of Technology, 2013, 29(4): 1-8. (in Chinese)
- [21] 翟永建, 张宏. 变精度多粒度粗糙集的约简研究[J]. 金陵科技学院学报, 2013, 29(4): 1-8.
- [22] ZHANG W X, MI J S, WU W Z. Knowledge Reductions in Inconsistent Information Systems[J]. Chinese Journal of Computers, 2003, 26(1): 12-18. (in Chinese)
- [23] 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简[J]. 计算机学报, 2003, 26(1): 12-18.
- [24] line generation for Knowledge-base entities[C] // Proc. of KDD. 2015.
- [25] BORDINO I, MEJOVA Y, LALMAS M. Penguins in Sweaters, or Serendipitous Entity Search on user-generated Content[C] // Proc. of CIKM. 2013.
- [26] METZGER S, SCHENKEL R, SYDOW M. QBES: query by entity examples[C] // Proc. of CIKM. 2013.
- [27] YU X, MA H, HSU B J, et al. On Building Entity Recommender Systems Using User Click Log and Freebase Knowledge[C] // Proc. of WSDM. 2014.
- [28] BI P, MA H, HSU B J. Learning to Recommend Related Entities to Search Users[C] // Proc. of WSDM. 2015.

(上接第192页)

- [10] MILCHEVSKI D. Entity recommendation based on Wikipedia [D]. University of Saarland, Germany, 2013.
- [11] BLANCO R, CAMBAZOGLU B B, MIKA P, et al. Entity Recommendations in Web Search[C] // Proc. of ISWC. 2013: 33-48.
- [12] YU X, REN X, SUN Y Z, et al. Personalized Entity recommendation: a heterogeneous information network approach [C] // Proc. of WSDM. 2014.
- [13] MILCHEVSKI D, BERBERICH K. X-REC: Cross-category entity recommendation[C] // Proc. of IliX. 2014.
- [14] ALTHOFF K, DONG X L, MURPHY K. TimeMachine: Time-