
Multi-Granulation Decision-Theoretic Rough Set Based on Maximal Consistent Relation

Fan Bingbing, Li Jin, Chen Xicheng, Liu Mengbo, Gu Jinghao, Liu Ming

Graduate College Air Force Engineering University, Xi'an, China

Email address:

fb19940103@163.com (Fan Bingbing), 179239647@qq.com (Li Jin), 1139899855@qq.com (Chen Xicheng), 619736846@qq.com (Liu Mengbo), jinghaogukgd@163.com (Gu Jinghao)

To cite this article:

Fan Bingbing, Li Jin, Chen Xicheng, Liu Mengbo, Gu Jinghao, Liu Ming. Multi-Granulation Decision-Theoretic Rough Set Based on Maximal Consistent Relation. *Science Discovery*. Vol. 6, No. 4, 2018, pp. 290-297. doi: 10.11648/j.sd.20180604.20

Received: June 19, 2018; **Accepted:** August 1, 2018; **Published:** August 10, 2018

Abstract: In incomplete information system, combining the advantages of maximal consistent relation and multiparticle theory, this paper proposed the multi-granulation decision-theoretic rough set based on maximal consistent relation based on consistent relation. Firstly, this paper define variable precision maximal consistent relation and dual-variable maximal consistent relation respectively for two kinds of incomplete information systems with different value types. Then, this paper establish optimistic and pessimistic multi-granulation decision-theoretic rough set model by replacing the equivalence relation with the maximal consistent relation in multi-granulation decision-theoretic rough set. Finally, it is proved that the maximum compatible relationship can improve the classification accuracy effectively based on model of optimistic maximal consistent relation, and this paper prove that the robustness of the classification can be improved by multiple classification thresholds at Multi-granulation.

Keywords: Maximal Consistent Relation, Multi-Granulation, Decision-Theoretic Rough Set, Classification Accuracy

基于极大相容关系的多粒度决策粗糙集

范兵兵, 李进, 陈西成, 刘梦波, 顾竞豪, 刘明

空军工程大学研究生学院, 西安市, 中国

邮箱:

fb19940103@163.com (范兵兵), 179239647@qq.com (李进), 1139899855@qq.com (陈西成), 619736846@qq.com (刘梦波), jinghaogukgd@163.com (顾竞豪)

摘要: 在不完备信息系统中, 在相容关系的基础上, 结合极大相容关系及多粒度理论的优点, 提出一种基于极大相容关系的多粒度决策粗糙集模型。首先, 针对两种取值类型不同的不完备信息系统, 分别定义了变精度极大相容关系及双精度极大相容关系。其次, 在多粒度决策粗糙集的模型上以极大相容关系代替等价关系建立乐观和悲观多粒度决策粗糙集模型。最后, 以乐观极大相容关系下的模型进行试验, 经证明极大相容关系能有效提高分类精度, 经UCI数据集实验验证, 多粒度下的多个分类阈值可提高分类的鲁棒性。

关键词: 极大相容关系, 多粒度, 决策粗糙集, 分类精度

1. 引言

现实生活中由于人们对数据的理解及获取条件限制等原因，导致所获得的信息系统往往存在缺失的情况，此时的信息系统我们定义为不完备信息系统。此外，由于客观事物的复杂性不同以及人的认识模糊程度不同，信息会以具体数字、区间数或者模糊数的形式表现出来。决策者想从这种取值类型不同的不完备信息系统中提取出对决策有用的信息面临很大挑战。Pawlak粗糙集理论[1]是数据挖掘知识的一个重要工具，对于数据集的分类、属性约简和消除数据冗余具有良好的处理效果，然而Pawlak粗糙集仅适用于完备的信息系统，因此，广大学者针对不完备信息系统下的粗糙集理论进行了广泛研究，文献[2]提出一种容差关系来代替等价关系使粗糙集能适用于不完备信息系统的挖掘。文献[3]提出一种应用于不完备区间信息系统的双精度容差关系，使得分类精度得以提高。文献[4]对相容关系进行改进提出一种极大相容关系，使得近似分类精度进一步提高。

决策粗糙集[5]最初由Yao Y Y提出，Yao在长期研究粗糙集理论过程中发现，概率粗糙集模型中利用两个参数以及隶属度函数将论域分成三个区域：正域、负域、边界域，因此提出了一种符合人们认知模式的三只决策，在此基础上，Yao结合贝叶斯风险理论，以最小决策风险为准则提出了决策粗糙集模型。针对决策粗糙集改进，中国学者Qian结合多粒度和决策粗糙集的优势，提出多粒度决策粗糙集[6]的模型，试图从不同粒度空间来进行数据集的决策分类。

本文提出基于极大相容关系的多粒度决策粗糙集，首先，针对容差关系进行改进，并针对两种取值类型不同的不完备信息系统，分别定义了变精度极大相容关系及双精度极大相容关系。并以极大相容关系代替等价关系建立乐观和悲观多粒度决策粗糙集模型。最后，经证明极大相容关系能有效提高分类的集合近似精度，通过UCI数据集实验验证，多粒度下的多个分类阈值可有效提高分类的鲁棒性。

2. 预备知识

定义1[3] 假设 $IS = (U, AT, V, f)$ 为一信息系统，其中 $U = \{x_1, x_2, \dots, x_n\}$ 为有限对象集合，元素 x_i 为每个对象； $AT = \{c_1, c_2, \dots, c_m\}$ 为有限属性集合，元素 c_i 为每个属性； f 为信息函数，表示 U 与 AT 之间的关系集，且 $\forall x_i \in U$ ， $f(x_i, c_i)$ 表示 x_i 在属性 c_i 上的取值； V 表示全体值域的集合。当存在 $f(x_i, c_i) = *$ ，其中 $*$ 表示缺失值，此时称信息系统为不完备信息系统；根据不完备信息系统中值域的取值类型不同，对不完备信息系统进一步细化，若 $f(x_i, c_i) = [a_i^k, b_i^k]$ ，称信息系统为不完备区间值信息系统，若 $f(x_i, c_i) = a_i$ ，称信息系统为不完备具体值信息系统。

定义2设 (U, AT, V, f) 为一不完备信息系统， $c_k \in AT$ ， $x_i, x_j \in U$ ，则对象 x_i, x_j 相对于属性 c_k 的相似度定义为

(1)信息系统为不完备具体值信息系统时：

$$S_{ij}^k = \begin{cases} 0, & |a_i - a_j| \geq \frac{\delta}{2} \\ 1, & |a_i - a_j| \leq \frac{\delta}{2} \vee \\ & f(x_i, c_k) = * \vee f(x_j, c_k) = * \end{cases} \quad (1)$$

式中 δ 为信息系统中属性 c_k 取值的方差。

(2)信息系统为不完备区间值信息系统时：

$$S_{ij}^k = \begin{cases} 0, & |a_i^k - a_j^k| + |b_i^k - b_j^k| \geq t \\ 1, & f(x_i, c_k) = * \vee f(x_j, c_k) = * \\ & \frac{t - |a_i^k - a_j^k| + |b_i^k - b_j^k|}{t} \end{cases} \quad (2)$$

其他式中： $t = u \bullet \max\{|a_i^k - a_j^k| + |b_i^k - b_j^k|\}$ ， $u \in [0.5, 1]$ 。

定义3设 (U, AT, V, f) 为一不完备信息系统， $B \subseteq AT$ ， $x_i, x_j \in U$ ， $0 < \alpha \leq 1$ ，称式 (3)、(4)

(1)信息系统为不完备具体值信息系统时：

$$SR_B = \frac{|\{c_k \in B\} | S_{ij}^k = 1|}{|B|} \quad (3)$$

(2)信息系统为不完备区间值信息系统时：

$$SR_B^\alpha = \frac{|\{c_k \in B\} | S_{ij}^k \geq \alpha|}{|B|} \quad (4)$$

为 x_i, x_j 在属性集 B 下的相似率。

定义4 设 (U, AT, V, f) 为一不完备信息系统，对 $c_k \in AT$ ， α 是区间值信息系统下的相似度阈值，不完备具体值、区间值信息系统下的容差关系分别为：

$$\Gamma_{AT} = \{(x_i, x_j) \in U \times U | S_{ij}^k = 1, \forall a_k \in AT\} \quad (5)$$

$$\Gamma_{AT}^\alpha = \{(x_i, x_j) \in U \times U | S_{ij}^k \geq \alpha, \forall a_k \in AT\} \quad (6)$$

Γ_{AT} 、 Γ_{AT}^α 满足自反性、对称性，不具有传递性。

定义5 对于一不完备信息系统 $IS = (U, AT, V, f)$ ， $c_k \in B$ ， $B \subseteq AT$ ， α 是区间值信息系统下的相似度阈值， $X \subseteq U$ ，则不完备具体值、区间值信息系统下的 X 的上、下近似为：

$$\begin{cases} \underline{\Gamma}_B(X) = \{x_i \in U \mid \underline{\Gamma}_B(x_i) \subseteq X\} \\ \overline{\Gamma}_B(X) = \{x_i \in U \mid \underline{\Gamma}_B(x_i) \cap X \neq \emptyset\} \end{cases} \quad (7)$$

$$\begin{cases} \Gamma_B^\alpha(X) = \{x_i \in U \mid \Gamma_B^\alpha(x_i) \subseteq X\} \\ \overline{\Gamma}_B^\alpha(X) = \{x_i \in U \mid \Gamma_B^\alpha(x_i) \cap X \neq \emptyset\} \end{cases} \quad (8)$$

定义6给定不完备具体值信息系统 $IS = (U, AT, V, f)$ 。若 $M \subseteq U$ 对任意的 $x_i, x_j \in M$ 均满足 $(x_i, x_j) \in \Gamma_{AT}$ ，则称 M 是 K 中关于属性集 AT 的相容类。若对任意的 $x_m \in U - M_{AT}(x_i)$ 必存在 $x_j \in M_{AT}(x_i)$ 使得 $(x_j, x_m) \notin \Gamma_{AT}$ 。此时，称 $M_{AT}(x_i)$ 为 IS 中关于属性集 AT 的极大相容类。

定义7类似地，对于不完备区间值信息系统 $IS = (U, AT, V, f)$ ， α 是相似度阈值。若 $M \subseteq U$ 对任意的 $x_i, x_j \in M$ 均满足 $(x_i, x_j) \in \Gamma_{AT}^\alpha$ ，则称 M 是 K 中关于属性集 AT 的相容类。若对任意的 $x_m \in U - M_{AT}^\alpha(x_i)$ 必存在 $x_j \in M_{AT}^\alpha(x_i)$ 使得 $(x_j, x_m) \notin \Gamma_{AT}^\alpha$ 。此时，称 $M_{AT}^\alpha(x_i)$ 为 IS 中关于属性集 AT 的极大相容类。

3. 相容关系拓展

通过定义5发现，基于相容关系进行划分的上下近似类似于Pawlak粗糙集中的等价关系，两者存在相同的问题，分类的粒度粗、抗噪声能力差、冗余度高，定义4的相容关系在进行不完备信息系统的分类时，条件过于苛刻，导致较大的误分率，为了提高分类精度，在此基础上定义的极大相容关系分类精度仍然不高，具有改进空间。因此，我们引入变精度极大相容类及双精度极大相容类。

定义8设 $IS = (U, AT, V, f)$ 为一不完备具体值信息系统，对 $c_k \in B, B \subseteq AT, \beta$ 是相似率阈值， $\beta \in (0, 1]$ ， $X \subseteq U$ ，在不完备具体值信息系统下，基于相似率的变精度容差关系为：

$$\begin{aligned} \Gamma_B^\beta &= \{(x_i, x_j) \in U \times U \mid (SR_B(x_i, x_j) \geq \beta, f(x_i, c_k) \\ &\neq * \wedge f(x_j, c_k) \neq *) \vee f(x_i, c_k) = * \vee f(x_j, c_k) = * \\ &, \forall c_k \in B\} \end{aligned} \quad (9)$$

显然 Γ_B^β 满足自反性，对称性，不具有传递性。若 M_B^β 其中任意两个对象满足关于 B 在关系 Γ_B^β 相容，且 $\forall x_i \in U - M_B^\beta, \exists x_j \in M_B^\beta$ ，使得 $(x_i, x_j) \notin \Gamma_B^\beta$ ，称 M_B^β 为 IS 关于属性集 B 的变精度极大相容类。

定义9设 $IS = (U, AT, V, f)$ 为一不完备区间值信息系统，对 $c_k \in B, B \subseteq AT, \alpha, \beta$ 分别是相似度、相似率阈值， $\alpha \in (0, 1], \beta \in (0, 1]$ ， $X \subseteq U$ ，在不完备具体值信息系统下，双精度容差关系为：

$$\begin{aligned} \Gamma_B^{\alpha\beta} &= \{(x_i, x_j) \in U \times U \mid (SR_B^{\alpha\beta}(x_i, x_j) \geq \beta, f(x_i, c_k) \\ &\neq * \wedge f(x_j, c_k) \neq *) \vee f(x_i, c_k) = * \vee f(x_j, c_k) = * \\ &, \forall c_k \in B\} \end{aligned} \quad (10)$$

显然 $\Gamma_B^{\alpha\beta}$ 满足自反性，对称性，不具有传递性。若 $M_B^{\alpha\beta}$ 其中任意两个对象满足关于 B 在关系 $\Gamma_B^{\alpha\beta}$ 相容，且 $\forall x_i \in U - M_B^{\alpha\beta}, \exists x_j \in M_B^{\alpha\beta}$ ，使得 $(x_i, x_j) \notin \Gamma_B^{\alpha\beta}$ ，称 $M_B^{\alpha\beta}$ 为 IS 关于属性集 B 的双精度极大相容类。

定义10给定不完备区间值信息系统 $IS = (U, AT, V, F)$ ，对 $a_k \in AT, \alpha, \beta$ 分别是相似度、相似率阈值， $X \subseteq U$ ，则 X 的下近似与上近似分别为：

$$\begin{aligned} \underline{APR}_{AT}^\alpha(X) &= \{x_i \in U \mid M_{AT}^\alpha(x_i) \subseteq X\} \\ \overline{APR}_{AT}^\alpha(X) &= \{x_i \in U \mid M_{AT}^\alpha(x_i) \cap X \neq \emptyset\} \\ \underline{APR}_{AT}^{\alpha\beta}(X) &= \{x_i \in U \mid M_{AT}^{\alpha\beta}(x_i) \subseteq X\} \\ \overline{APR}_{AT}^{\alpha\beta}(X) &= \{x_i \in U \mid M_{AT}^{\alpha\beta}(x_i) \cap X \neq \emptyset\} \end{aligned} \quad (11)$$

定理1设 $m_B^\alpha(x_i), M_B^\alpha(x_i), M_B^{\alpha\beta}(x_i)$ 分别为属性集 B 中关于 x_i 的包含的相容类、极大相容类、双精度极大相容类的集合（当 α 等于1时，即为变精度情况），则对任意的 $x_i \in U, \forall B \subseteq AT$ ，有 $\Gamma_B^\alpha(x_i) = \cup\{m_B^\alpha(x_i) \in M_B^\alpha(x_i)\}$ ， $APR_B^\alpha(x_i) = \cup\{M_B^\alpha(x_i) \in M_B^{\alpha\beta}(x_i)\}$ 。

证明：由定义7以及 $\Gamma_B^\alpha(x_i) = \{x_j \in U \mid (x_i, x_j) \in \Gamma_B^\alpha\}$ 可知， $\forall x_j \in \Gamma_B^\alpha(x_i)$ 存在 $x_j \in \cup\{m_B^\alpha(x_i) \in M_B^\alpha(x_i)\}$ ，即 $\Gamma_B^\alpha(x_i) \subseteq \cup\{m_B^\alpha(x_i) \in M_B^\alpha(x_i)\}$ ；又 $\forall x_j \in m_B^\alpha(x_i)$ ， $m_B^\alpha(x_i) \in M_B^\alpha(x_i)$ ，即 x_i, x_j 满足容差关系 Γ_B^α ，因此 $x_j \in \Gamma_B^\alpha(x_i)$ ，进而 $\cup\{m_B^\alpha(x_i) \in M_B^\alpha(x_i)\} \subseteq \Gamma_B^\alpha(x_i)$ 。综上所述， $\Gamma_B^\alpha(x_i) = \cup\{m_B^\alpha(x_i) \in M_B^\alpha(x_i)\}$ 。又因为变精度极大相容类是双精度极大相容类的一种特殊情况，即 $\beta = 1$ 时为变精度极大相容类，因此有 $\forall M_B^\alpha(x_i) \in M_B^{\alpha\beta}(x_i)$ ，又 $APR_B^\alpha(x_i) = \cup\{M_B^\alpha(x_i)\}$ ，因此 $APR_B^\alpha(x_i) = \cup\{M_B^\alpha(x_i) \in M_B^{\alpha\beta}(x_i)\}$ 。

定理2给定不完备信息系统 $IS = (U, AT, V, f)$, α 、 β 分别是相似度、相似率阈值, $\forall B \subseteq AT, X \subseteq U$, 则:

$$\frac{|\underline{APR}_B^{\alpha\beta}(X)|}{|\overline{APR}_B^{\alpha\beta}(X)|} \geq \frac{|\underline{APR}_B^\alpha(X)|}{|\overline{APR}_B^\alpha(X)|} \geq \frac{|\underline{\Gamma}_B^\alpha(X)|}{|\overline{\Gamma}_B^\alpha(X)|} \quad (12)$$

证明: 设 $M^\alpha(B)$ 为信息系统上属性集 B 所决定的所有极大相容类的集合。由于 $\overline{\Gamma}_B^\alpha(X) = \{x_i \in U \mid \Gamma_B^\alpha(x_i) \cap X \neq \emptyset\} = \bigcup \{\Gamma_B^\alpha(x_i) \mid x_i \in X\}$, 由定理1可知 $\overline{\Gamma}_B^\alpha(X) = \bigcup \{\bigcup \{M_B^\alpha(x_i) \in m_B^\alpha(x_i)\} \mid x_i \in X\} = \bigcup \{M_B^\alpha(x_i) \in M^\alpha(B) \mid M_B^\alpha(x_i) \cap X \neq \emptyset\} = \overline{APR}_B^\alpha(X)$; $\forall x_i \in \underline{\Gamma}_B^\alpha(X)$, 有 $\Gamma_B^\alpha(x_i) \subseteq X$, 由于 $\Gamma_B^\alpha(x_i) = \bigcup \{M_B^\alpha(x_i) \in m_B^\alpha(x_i)\}$, 故 $M_B^\alpha(x_i) \subseteq X$, 从而 $x_i \in \underline{APR}_B^\alpha(X)$, 则 $\underline{\Gamma}_B^\alpha(X) \subseteq \underline{APR}_B^\alpha(X)$ 。所以, $\frac{|\underline{APR}_B^\alpha(X)|}{|\overline{APR}_B^\alpha(X)|} \geq \frac{|\underline{\Gamma}_B^\alpha(X)|}{|\overline{\Gamma}_B^\alpha(X)|}$ 。同理

我们可以证得:

$$\frac{|\underline{APR}_B^{\alpha\beta}(X)|}{|\overline{APR}_B^{\alpha\beta}(X)|} \geq \frac{|\underline{APR}_B^\alpha(X)|}{|\overline{APR}_B^\alpha(X)|} \quad \text{。 综合上述}$$

$$\frac{|\underline{APR}_B^{\alpha\beta}(X)|}{|\overline{APR}_B^{\alpha\beta}(X)|} \geq \frac{|\underline{APR}_B^\alpha(X)|}{|\overline{APR}_B^\alpha(X)|} \geq \frac{|\underline{\Gamma}_B^\alpha(X)|}{|\overline{\Gamma}_B^\alpha(X)|} \quad \text{。}$$

采用变精度、双精度极大相容类对粗糙上、下近似定义较采用改进相似度容差类定义粗糙下近似与上近似有更高的粗糙集近似度。因此, 在不完备信息系统上建立决策粗糙集模型时采用极大相容类对粗糙集近似算子定义。

4. 基于极大相容关系多粒度决策粗糙集模型

4.1. 多粒度决策粗糙集模型

多粒度概念是粒计算理论中的核心, 中国学者钱宇华、梁吉业基于多粒度的思想, 将决策粗糙集中采用单个粒空间描述问题的方式拓展为采用多个粒空间描述问题的方式, 提出了多粒度决策粗糙集模型。一般来说, 多粒度决策粗糙集模型存在两种形式, 一种称为乐观多粒度决策粗糙集模型, 另一种称为悲观多粒度决策粗糙集模型。

将信息系统定义为一个四元组 $IS = (U, AT, V, f)$, 其中 U 是非空有限的对象集合, $AT = C \cup D$ 是非空有限的属性集合, 其中 C 为条件属性, D 为决策属性; $V = \bigcup_{a \in AT} V_a$ 是属性的值域, 其中有 $f_a : U \times a \rightarrow V_a (\forall x \in U, a \in AT)$ 。在信息系统 IS 中, 称 $[x]_{R_i}$ 为等价关系 R_i 下元素 x 的等价类。

定义11在信息系统 IS 中, 假定 R_1, R_2, \dots, R_m 是 m 个粒结构, 对于 $\forall X \subseteq U$, 定义乐观多粒度决策粗糙集的下近似和上近似分别为:

$$\begin{aligned} \underline{\sum_{i=1}^m R_i^o(X)} &= \left\{ x : \left(P(X \mid [x]_{R_1}) \geq \alpha_1 \vee P(X \mid [x]_{R_2}) \geq \alpha_2 \vee \dots \vee P(X \mid [x]_{R_m}) \geq \alpha_m \right), x \in U \right\} \\ \overline{\sum_{i=1}^m R_i^o(X)} &= U - \left\{ x : \left(P(X \mid [x]_{R_1}) \leq \beta_1 \vee P(X \mid [x]_{R_2}) \leq \beta_2 \vee \dots \vee P(X \mid [x]_{R_m}) \leq \beta_m \right), x \in U \right\} \end{aligned} \quad (13)$$

定义12在信息系统 IS 中, R_1, R_2, \dots, R_m 是 m 个粒结构, 对于 $\forall X \subseteq U$, 定义悲观多粒度决策粗糙集的下近似和上近似分别为:

$$\begin{aligned} \underline{\sum_{i=1}^m R_i^p(X)} &= \left\{ x : \left(P(X \mid [x]_{R_1}) \geq \alpha_1 \wedge P(X \mid [x]_{R_2}) \geq \alpha_2 \wedge \dots \wedge P(X \mid [x]_{R_m}) \geq \alpha_m \right), x \in U \right\} \\ \overline{\sum_{i=1}^m R_i^p(X)} &= U - \left\{ x : \left(P(X \mid [x]_{R_1}) \leq \beta_1 \wedge P(X \mid [x]_{R_2}) \leq \beta_2 \wedge \dots \wedge P(X \mid [x]_{R_m}) \leq \beta_m \right), x \in U \right\} \end{aligned} \quad (14)$$

因此下近似 $\underline{\sum_{i=1}^m R_i^p(X)}$ 和上近似 $\overline{\sum_{i=1}^m R_i^p(X)}$,

设有限论域为 $U = \{x_1, x_2, \dots, x_n\}$, $X \subseteq U$; 设 R_1, R_2, \dots, R_m 是 m 个粒结构, 同时它是 AT 的 m 个条件属性集, 决策行为集合为 $\gamma = \{\gamma_P, \gamma_N, \gamma_B\}$, 此三类决策行为分别表示接受某事物采取的决策行为, 拒绝某事物采取的决策行为, 待定的决策行为。设 $\lambda_{PP}^{(i)}$ 表示 $x \in X$ 时在粒度空间 R_i 下采取决策行为 γ_P 的风险损失, $\lambda_{PN}^{(i)}$ 表示 $x \in \sim X$ 时在粒度空间 R_i 下采取决策行为 γ_P 的风险损失, 依此类推得到其余四类损失分别为: $\lambda_{NP}^{(i)}$ 、 $\lambda_{NN}^{(i)}$ 、 $\lambda_{BP}^{(i)}$ 、 $\lambda_{BN}^{(i)}$ 。

因而在一定的粒度空间 R_i 下采取决策行为 $\gamma_P, \gamma_N, \gamma_B$ 的期望损失分别为:

$$\begin{cases} R(\gamma_P \mid [x]_{R_i}) = \lambda_{PP}^{(i)} P(X \mid [x]_{R_i}) + \lambda_{PN}^{(i)} P(\sim X \mid [x]_{R_i}) \\ R(\gamma_N \mid [x]_{R_i}) = \lambda_{NP}^{(i)} P(X \mid [x]_{R_i}) + \lambda_{NN}^{(i)} P(\sim X \mid [x]_{R_i}) \\ R(\gamma_B \mid [x]_{R_i}) = \lambda_{BP}^{(i)} P(X \mid [x]_{R_i}) + \lambda_{BN}^{(i)} P(\sim X \mid [x]_{R_i}) \end{cases} \quad (15)$$

以贝叶斯风险最小理论，最终可解得令 α_i 、 β_i 、 η_i 分别为：

$$\alpha_i = \frac{\lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}}{\lambda_{BP}^{(i)} - \lambda_{PP}^{(i)} + \lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}} \quad (16)$$

$$\beta_i = \frac{\lambda_{BN}^{(i)} - \lambda_{NN}^{(i)}}{\lambda_{NP}^{(i)} - \lambda_{BP}^{(i)} + \lambda_{BN}^{(i)} - \lambda_{NN}^{(i)}} \quad (17)$$

$$\eta_i = \frac{\lambda_{PN}^{(i)} - \lambda_{NN}^{(i)}}{\lambda_{NP}^{(i)} - \lambda_{PP}^{(i)} + \lambda_{PN}^{(i)} - \lambda_{NN}^{(i)}} \quad (18)$$

本节将利用变精度、双精度极大相容关系代替等价关系建立不完备信息系统下的多粒度决策粗糙集，对不完备信息系统决策粗糙集进行拓展。由于变精度极大相容关系是双精度极大相容关系的一种特殊情况，因此本节以双精度极大相容关系为例提出基于双精度极大相容关系的多粒度决策粗糙集模型。

在不完备区间值信息系统 $IS = (U, AT, V, f)$ 中，称 $[M_B^{\alpha\beta}(x)]_{R_i}$ 为等价关系 R_i 下元素 x 的双精度极大相容类。

定义13在信息系统 IS 中，假定 R_1, R_2, \dots, R_m 是 m 个粒结构，对于 $\forall X \subseteq U$ ，定义乐观多粒度决策粗糙集的下近似和上近似分别为：

4.2. 极大相容关系下的多粒度决策粗糙集模型

$$\begin{aligned} \underline{\sum_{i=1}^m M_B^{\alpha\beta}(X)} &= \left\{ x : \left(P(X | [M_B^{\alpha\beta}(x)]_{R_i}) \geq \alpha_1 \vee \dots \vee P(X | [x]_{R_m}) \geq \alpha_m \right), x \in U \right\} \\ \overline{\sum_{i=1}^m M_B^{\alpha\beta}(X)} &= U - \left\{ x : \left(P(X | [M_B^{\alpha\beta}(x)]_{R_i}) \leq \beta_1 \vee \dots \vee P(X | [M_B^{\alpha\beta}(x)]_{R_m}) \leq \beta_m \right), x \in U \right\} \end{aligned} \quad (19)$$

定义14在信息系统 IS 中， R_1, R_2, \dots, R_m 是 m 个粒结构，对于 $\forall X \subseteq U$ ，定义悲观多粒度决策粗糙集的下近似和上近似分别为：

$$\begin{aligned} \underline{\sum_{i=1}^m M_B^{\alpha\beta}(X)} &= \left\{ x : \left(P(X | [M_B^{\alpha\beta}(x)]_{R_i}) \geq \alpha_1 \right. \right. \\ &\quad \left. \wedge \dots \wedge P(X | [x]_{R_m}) \geq \alpha_m \right\}, x \in U \\ \overline{\sum_{i=1}^m M_B^{\alpha\beta}(X)} &= U - \left\{ x : \left(P(X | [M_B^{\alpha\beta}(x)]_{R_i}) \leq \beta_1 \wedge \dots \wedge P(X | [M_B^{\alpha\beta}(x)]_{R_m}) \leq \beta_m \right), x \in U \right\} \end{aligned} \quad (20)$$

因此下近似 $\underline{\sum_{i=1}^m M_B^{\alpha\beta}(X)}$ 和上近似 $\overline{\sum_{i=1}^m M_B^{\alpha\beta}(X)}$ ，

设有限论域为 $U = \{x_1, x_2, \dots, x_n\}$, $X \subseteq U$ ；设 R_1, R_2, \dots, R_m 是 m 个粒结构，同时它是 AT 的 m 个条件属性集，决策行为集合为 $\gamma = \{\gamma_P, \gamma_N, \gamma_B\}$ ，此三类决策行为分别表示接受某事物采取的决策行为，拒绝某事物采取的决策行为，待定的决策行为。设 $\lambda_{PP}^{(i)}$ 表示 $x \in X$ 时在粒度空间 R_i 下采取决策行为 γ_P 的风险损失， $\lambda_{PN}^{(i)}$ 表示 $x \in \sim X$ 时在粒度空间 R_i 下采取决策行为 γ_P 的风险损失，依此类推得到其余四类损失分别为： $\lambda_{NP}^{(i)}$ 、 $\lambda_{NN}^{(i)}$ 、 $\lambda_{BP}^{(i)}$ 、 $\lambda_{BN}^{(i)}$ 。

因而在一定的粒度空间 R_i 下采取决策行为 $\gamma_P, \gamma_N, \gamma_B$ 的期望损失分别为：

$$\begin{cases} R(\gamma_P | [M_B^{\alpha\beta}(x)]_{R_i}) = \lambda_{PP}^{(i)} P(X | [M_B^{\alpha\beta}(x)]_{R_i}) + \lambda_{PN}^{(i)} P(\sim X | [M_B^{\alpha\beta}(x)]_{R_i}) \\ R(\gamma_N | [M_B^{\alpha\beta}(x)]_{R_i}) = \lambda_{NP}^{(i)} P(X | [M_B^{\alpha\beta}(x)]_{R_i}) + \lambda_{NN}^{(i)} P(\sim X | [M_B^{\alpha\beta}(x)]_{R_i}) \\ R(\gamma_B | [M_B^{\alpha\beta}(x)]_{R_i}) = \lambda_{BP}^{(i)} P(X | [M_B^{\alpha\beta}(x)]_{R_i}) + \lambda_{BN}^{(i)} P(\sim X | [M_B^{\alpha\beta}(x)]_{R_i}) \end{cases} \quad (21)$$

其中 $P(X | M_B^{\alpha\beta}(x)) = \frac{|M_B^{\alpha\beta}(x) \cap X|}{|M_B^{\alpha\beta}(x)|}$ ，以贝叶斯风险最小理论，最终可解得令 α_i 、 β_i 、 η_i 分别为：

$$\alpha_i = \frac{\lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}}{\lambda_{BP}^{(i)} - \lambda_{PP}^{(i)} + \lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}} \quad (22)$$

$$\beta_i = \frac{\lambda_{BN}^{(i)} - \lambda_{NN}^{(i)}}{\lambda_{NP}^{(i)} - \lambda_{BP}^{(i)} + \lambda_{BN}^{(i)} - \lambda_{NN}^{(i)}} \quad (23)$$

$$\eta_i = \frac{\lambda_{PN}^{(i)} - \lambda_{NN}^{(i)}}{\lambda_{NP}^{(i)} - \lambda_{PP}^{(i)} + \lambda_{PN}^{(i)} - \lambda_{NN}^{(i)}} \quad (24)$$

5. 实验分析

为了检验第三节中提出的基于变精度、双精度极大相容关系的多粒度决策粗糙集模型具有一定的优越性，从UCI数据库中选择4个数据集进行仿真实验，所选数据集信息如下表1所示

表1 数据描述。

数据集	样本个数	条件属性	决策类别
breast	699	10	2
iris	150	4	3
page-blocks	5473	10	5
statlog	690	14	2

表1为完整的实验数据集，为了构造出不完备信息系统，在实验前我们随机剔除信息系统中的数据，然后我们分别以极大相容关系决策粗糙集模型和变精度极大相容关系乐观多粒度决策粗糙集模型进行分类。

为了研究多粒度下建立的模型优势，实验过程以随机剔除20%数据的不完备数据集为分类对象，分别分别求解极大相容关系决策粗糙集模型和变精度极大相容关系乐观多粒度决策粗糙集模型下的接受域、拒绝域的近似精度整个决策域的近似精度，以近似精度来表示分类的可信度。实验过程选取正态分布生成代价矩阵，由于正确分类下没有代价，因此， $\lambda_{PP} = 0$ ， $\lambda_{NN} = 0$ 。单粒度的情况，每个粒度的分类阈值均选择 α ；多粒度的情况下，每个粒度选

取不同的阈值 α_i ，但满足 $\frac{\alpha_i}{i} = \alpha$ 。实验结果如表2所示，

数据表明多粒度下的多个阈值使决策更具有稳定性，对比可信度可以发现变精度极大相容关系乐观多粒度决策粗糙集模型分类精度更高，分类精度提高5%~10%。

此外，实验分析多粒度下模型的鲁棒性，即多粒度下的模型对决策代价矩阵重数的敏感性，选取1000个代价矩阵、50为基数递增对多粒度下模型分类的敏感性进行研究，分别求出page-blocks数据集在乐观、悲观、保守、激进决策和平均语义下的近似精度，结果如图1所示，从图1中可以发现，多粒度下的多个阈值决策有很强的鲁棒性。

表2 数据集实验结果。

	单粒度下的模型			多粒度下的模型		
	可信度	接受可信度	拒绝可信度	可信度	接受可信度	拒绝可信度
breast	0.74464	0.80703	0.79888	0.74930	0.90703	0.89888
	0.74930	0.80703	0.79622	0.74375	0.90703	0.89888
	0.74464	0.80703	0.79888	0.74930	0.90703	0.89888
	0.74930	0.90703	0.79622	0.74375	0.90703	0.89888
	0.74464	0.80703	0.79888	0.74930	0.90703	0.89888
Iris	0.75667	0.77900	0.71875	0.75380	0.88763	0.81875
	0.79224	0.7888	0.71875	0.83466	0.89888	0.81875
	0.76042	0.79888	0.71875	0.83466	0.89888	0.81875
	0.78849	0.77900	0.71875	0.75380	0.88763	0.81875
	0.75667	0.78763	0.71875	0.78849	0.88763	0.81875
	0.61417	0.74727	0.72869	0.67196	0.84893	0.88304
page-blocks	0.62225	0.78104	0.77520	0.84438	0.85007	0.88304
	0.61455	0.78104	0.72869	0.90824	0.85007	0.88304
	0.62187	0.74727	0.57520	0.60810	0.74893	0.88304
	0.61851	0.74893	0.58304	0.62187	0.74893	0.88304
	0.78427	0.78456	0.79888	0.78627	0.88548	0.80935
statlog	0.88054	0.72697	0.72597	0.83504	0.89602	0.80935
	0.78778	0.72697	0.79888	0.95741	0.89602	0.80935
	0.87702	0.78456	0.72597	0.66390	0.88548	0.80935
	0.84514	0.78548	0.70935	0.87702	0.88548	0.80935

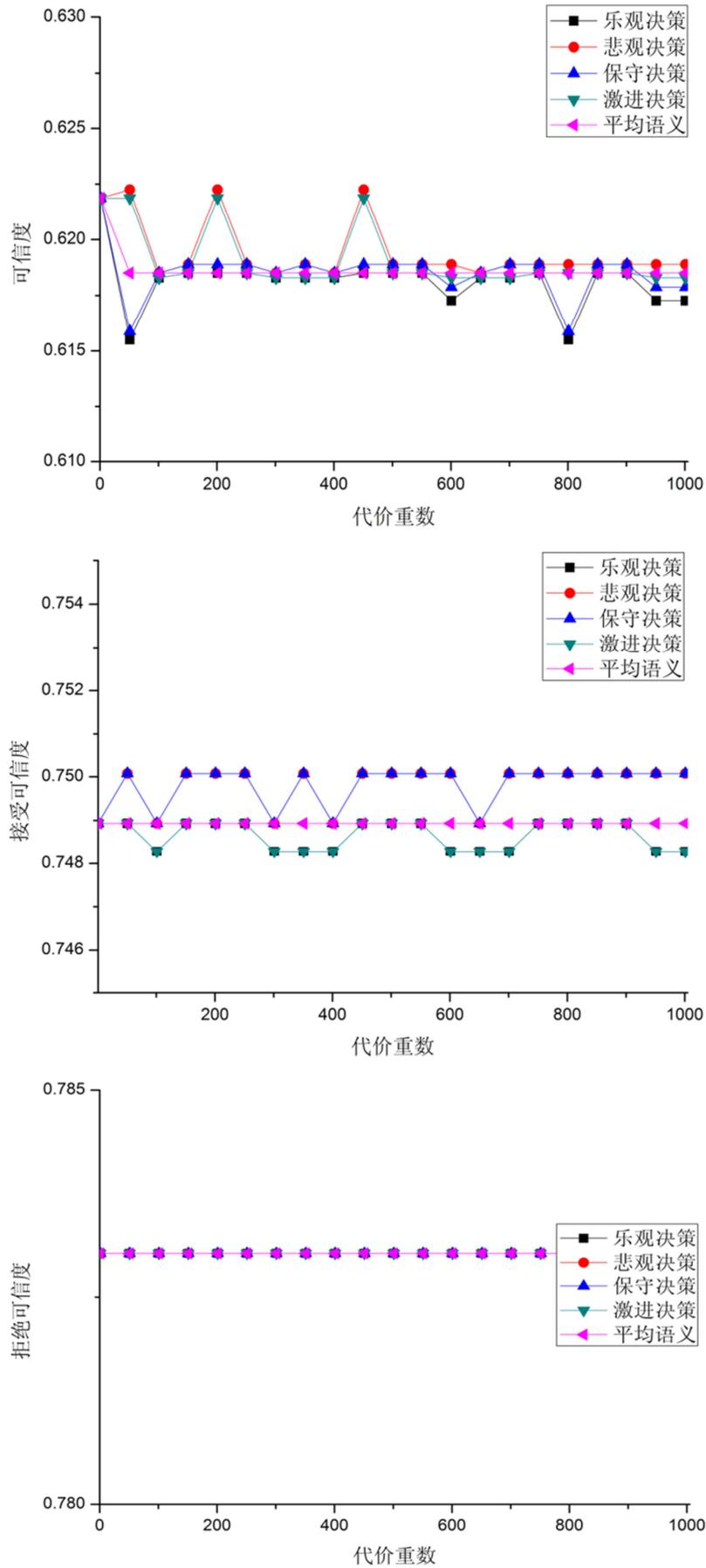


图1 多粒度下的模型鲁棒性分析。

6. 结论

本文提出基于极大相容关系的多粒度决策粗糙集, 首先, 针对容差关系进行改进, 并针对两种取值类型不同的不完备信息系统, 分别定义了变精度极大相容关系及双精度极大相容关系。并以极大相容关系代替等价关系建立乐观和悲观多粒度决策粗糙集模型。最后, 经证明极大相容关系能有效提高分类的集合近似精度, 通过UCI数据集实验验证, 多粒度下的多个分类阈值可有效提高分类的鲁棒性。

参考文献

- [1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5):341-356.
- [2] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1-4):39-49.
- [3] 曾玲,付敏,何普彦.双精度容差关系的粗糙集拓展模型及约简[J].江苏大学学报:自然科学版,2014,35(3):312-318.
- [4] Zhang xin, Li Xuwu. Decision-Theoretic Rough Set Based on Incomplete Interval-Valued Information System[J]. Application Research of Computers,2017,34(1):110-113.
- [5] Yao Y Y. Decision-theoretic rough set models[J].LNAI4481,2007:1-12.
- [6] Qian Y H, Liang J Y, Yao Y Y, et al. MGRS: a multi-granulation rough set[J]. Information Sciences, 2010, 180(6):949-970.
- [7] Decui Liang, Dun Liu, Witold Pedrycz, Pei Hu. Triangular fuzzy decision-theoretic rough sets[J]. International Journal of Approximate Reasoning, 2013,54:1087-1106.
- [8] Yiyu Yao, Yan Zhao. Attribute Reduction in Decision-Theoretic Rough Set Models[J]. Information Science, 2008,178(17):3356-3373.
- [9] Xiuyi Jia, Wenhe Liao, Zhenmin Tang, Lin Shang. Minimum cost attribute reduction in decision-theoretic rough set models[J]. Information Science,2013,219:151-167.
- [10] Liu D, Li TR, Li HX.A multiple-category classification approach with decision-theoretic rough sets. Fundamenta Informaticae, 2012, 115 (2-3) : 173-188.
- [11] Li Feng, Miao Duoqian, Liu Chaihui , et al. Image segmentation algorithm based on the decision-theoretic rough set model[J].CAAI Transactions on Intelligent Systems,2014,9(2): 143- 147.
- [12] Yang Xiaoping, Yao Jingtao. A Multi-agent Decision-Theoretic Rough Set Model [J]. Lecture Notes in Computer Science, 2010, 6401(2):711-718.
- [13] Yang Xiaoping, Yao Jingtao.Modeling Multi-agent Three-way Decisions with Decision-theoretic Rough Sets[J]. Fundamenta Informaticae, 2012, 115(2-3): 157-171.
- [14] Ma Xinbin, Ju Hengrong, Yang Xibei, et al. Multi-cost based decision-theoretic rough sets in incomplete information systems[J].Journal of Nanjing University (Natural Sciences), 2015, 51(2): 335-342.
- [15] Dou Huili, Yang Xibei, Song Xiaoning, et al. Decision-theoretic rough set: A multicost strategy[J]. Knowledge-Based Systems, 2016(91):71-83.
- [16] Yu Hong, Wang Guoying, Yao Yiyu. Current research and future perspectives on decision-theoretic rough sets[J] Chinese journal of computers, 2015,38(8):1628-1639.
- [17] Li Feng, Miao Duoqian, Liu Chaihui , et al. Image segmentation algorithm based on the decision-theoretic rough set model[J].CAAI Transactions on Intelligent Systems,2014,9(2): 143- 147.
- [18] Lafferty,J, McCallum A, Pereira F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]//Proc of the 18th International Conference on Machine Learning. Sanfrancisco:Morgan Kaufmann,2001:282-289.

作者简介



范兵兵 (1994-), 男,江西临川人, 硕士研究生, 主要研究方向为网络与信息安全



李进 (1971-), 男, 陕西西安人, 博士副教授, 研究方向为地空导弹武器系统总体和嵌入式系统。