

## 利用混杂核模糊互补信息选择特征

袁 钟 陈红梅 王志红 李天瑞

(西南交通大学计算机与人工智能学院 成都 611756)

(综合交通大数据应用技术国家工程实验室(西南交通大学) 成都 611756)

([yuanzhong2799@foxmail.com](mailto:yuanzhong2799@foxmail.com))

## Exploiting Hybrid Kernel-Based Fuzzy Complementary Mutual Information for Selecting Features

Yuan Zhong, Chen Hongmei, Wang Zhihong, and Li Tianrui

(School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756)

(National Engineering Laboratory of Integrated Transportation Big Data Application Technology (Southwest Jiaotong University), Chengdu 611756)

**Abstract** Fuzzy rough set theory is currently receiving a lot of attention in the fields of data mining and machine learning. The theory provides an effective tool to overcome the discretization problem and can be applied directly to numerical or mixed attribute data. In the fuzzy rough set model, fuzzy relations are defined to measure the similarity between objects and numerical attribute values no longer need to be discretized. The theory has been successfully applied to many fields such as attribute reduction, rule extraction, cluster analysis and outlier detection. Information entropy has been introduced into fuzzy rough set theory for the representation of fuzzy and uncertainty information, resulting in different forms of fuzzy uncertainty measures such as fuzzy information entropy, fuzzy complementary entropy, and fuzzy mutual information. However, most of the proposed fuzzy mutual information on decisions is non-monotonic, which may lead to a non-convergent learning algorithm. To this end, the fuzzy complementary mutual information on decisions is defined based on the hybrid kernel fuzzy complementary entropy, which is shown to vary monotonically with features. Then, the feature selection method is explored by using the hybrid kernel-based fuzzy complementary mutual information and a corresponding algorithm is designed. Experimental results show that the proposed algorithm can select fewer features and maintain or improve the classification accuracy in most cases.

**Key words** fuzzy rough set theory; hybrid kernel; complementary entropy; uncertainty measure; feature selection

**摘 要** 模糊粗糙集理论目前在数据挖掘和机器学习等领域受到了广泛的关注. 该理论提供了一种能克服离散化问题的有效工具, 并能直接应用于数值或混合属性数据. 在模糊粗糙集模型中, 定义模糊关系来测量对象之间的相似性, 数值属性值不再需要离散化. 模糊粗糙集理论已经被成功应用于许多领域, 如属性约简、规则提取、聚类分析和离群点检测. 信息熵被引入到模糊粗糙集理论进行模糊和不确定信息的表示, 产生了不同形式的模糊不确定性度量, 如模糊信息熵、模糊补熵和模糊互信息等. 然而, 大部分所提关于决策的模糊互信息都是非单调的, 这可能导致一个不收敛的学习算法. 为此, 基于混杂核模糊补熵, 定义

收稿日期: 2021-12-24; 修回日期: 2022-08-08

基金项目: 国家自然科学基金项目(61976182, 62076171, 61976245, 61876157); 四川省重点研发项目(2020YFG0035); 四川省科技成果转移转化示范项目(2022ZHC0005)

This work was supported by the National Natural Science Foundation of China (61976182, 62076171, 61976245, 61876157), the Sichuan Key Research and Development Project (2020YFG0035), and Sichuan Science and Technology Achievements Transfer and Transformation Demonstration Project (2022ZHC0005).

通信作者: 陈红梅([hmchen@swjtu.edu.cn](mailto:hmchen@swjtu.edu.cn))

了关于决策的模糊互补信息,证明了其随特征呈单调性变化.进而,利用混杂核模糊互补信息探索特征选择方法并且设计了相关的算法.实验结果展示了在大多数情况下所提算法可以选取更少的特征且能保持或提高分类准确率.

**关键词** 模糊粗糙集理论;混杂核;补熵;不确定性度量;特征选择

**中图法分类号** TP18

特征选择是一种有效的预处理方法,已经广泛应用于数据挖掘、机器学习和模糊识别等领域<sup>[1-6]</sup>.其目的是剔除冗余特征以提高学习任务的泛化能力.特征选择方法的关键步骤是如何构建一个用于评价特征重要性的函数<sup>[7]</sup>.这个函数可以被使用来从原始高维数据中选择一个最小的特征子集,使得数据类的概率分布尽可能地接近使用原始特征得到的分布.目前,已经有许多特征评价函数被提出,如一致性<sup>[8]</sup>、依赖度<sup>[9]</sup>和不确定性度量<sup>[10]</sup>等.

模糊粗糙集理论首先由 Dubois 等人<sup>[11]</sup>提出,它是一种能处理模糊和不确定数据的有效数学工具.模糊粗糙集理论的基本思想是使用模糊相似关系来诱导模糊信息粒,这使得它克服了经典粗糙集理论处理数值属性数据需要做离散化处理的问题.因而,该理论可以直接应用于数值或混合特征数据.模糊粗糙集理论主要包括模型的扩展和应用.受原始模糊粗糙集模型的启发,提出了一系列模糊粗糙集的扩展<sup>[12-14]</sup>.此外,模糊粗糙集理论也已经被应用于很多领域.特征选择是模糊粗糙集理论最重要的应用之一,已经提出了有许多基于模糊粗糙集理论的特征选择方法<sup>[15-18]</sup>.这些方法可以大致分为基于模糊依赖度的<sup>[15]</sup>、基于模糊不确定性度量的<sup>[16]</sup>和模糊区分矩阵的方法<sup>[17]</sup>.

信息熵起源于信息论,它能够建立一种衡量不确定性的有效机制.针对模糊关系的重要性,Yager<sup>[19]</sup>首先将信息熵的概念引入模糊相似关系中,然后讨论了模糊信息的不确定性测度.至此,在模糊粗糙集中提出了不同形式的模糊不确定性度量,如模糊信息熵<sup>[20]</sup>、模糊补熵<sup>[21]</sup>、模糊粗糙熵<sup>[22]</sup>和模糊邻域决策熵<sup>[23]</sup>等.许多模糊不确定性度量已经被应用到特征选择<sup>[10,16,24-25]</sup>.例如,Hu 等人<sup>[20]</sup>提出了在模糊近似空间和模糊概率近似空间中计算信息熵的广义公式并将模糊条件熵应用于混合属性约简;Qian 等人<sup>[21]</sup>将补熵引入模糊粗糙集,并提出模糊补熵的定义,但所提的模糊补熵尚未应用到属性约简中.Zhao 等人<sup>[26]</sup>在模糊粗糙集中提出了一种基于任意模糊关系的补熵模型,并将该信息熵应用于特征选择.然而,由模糊不确定性度量所定义的模糊互信息大部分都是非单

调的,这可能导致一个不收敛的学习算法.此外,模糊不确定性度量的定义也不能很好地反映系统的不确定性.其原因是在计算模糊熵时,采用交运算来计算模糊相似关系.这样的计算策略可能会降低高维样本空间中模糊相似关系的辨别能力<sup>[27]</sup>.因此,它可能不能真实反映样本之间的关系.

核方法把低维空间中的非线性学习问题转化为高维空间中的线性学习问题,该方法可以弥补模糊关系计算中交运算带来的不足.为此,文献<sup>[28]</sup>利用核函数生成的模糊关系粒化论域,进而构造了不同的基于核的模糊粗糙集.进一步,文献<sup>[29-30]</sup>研究了基于核模糊粗糙集的属性约简.文献<sup>[31]</sup>提出了一种基于混杂核的模糊补熵,并将其应用于无监督属性约简.然而,该文献并未定义关于决策的混杂核模糊互补信息.

基于上述讨论,本文基于混杂核模糊补熵,定义关于决策的混杂核模糊互补信息.进一步,基于所提关于决策的混杂核模糊互补信息,构建了一种考虑内外重要度的混杂特征选择评价函数.进而设计了基于混杂核模糊互补信息(kernel-based fuzzy complementary mutual information, KFCMI)的特征选择算法.最后,数据实验结果表明在大多数情况下所提算法可以选取更少的特征且能保持或提高分类准确率.

## 1 预备知识

在模糊信息理论中,数据表定义为一个信息系统  $IS = \langle U, A \rangle$ .其中:  $U = \{x_1, x_2, \dots, x_n\}$  是一个非空有限对象集;  $A$  是非空有限属性集合.

设  $\mathcal{A}: U \rightarrow [0, 1]$  是一个映射,则称  $\mathcal{A}$  是  $U$  上的模糊集.  $\forall x \in U$ ,  $\mathcal{A}(x)$  称为  $x$  对  $\mathcal{A}$  的隶属度.论域  $U$  上全体模糊集的集合,记为  $\mathcal{F}(U)$ .设  $\mathcal{A}, \mathcal{B} \in \mathcal{F}(U)$ ,对任意  $x \in U$ ,一些运算定义为:

- 1)  $\mathcal{A} \subseteq \mathcal{B} \Leftrightarrow \mathcal{A}(x) \leq \mathcal{B}(x)$ ;
- 2)  $(\mathcal{A} \cup \mathcal{B})(x) = \max\{\mathcal{A}(x), \mathcal{B}(x)\}$ ;
- 3)  $(\mathcal{A} \cap \mathcal{B})(x) = \min\{\mathcal{A}(x), \mathcal{B}(x)\}$ ;
- 4)  $\mathcal{A}^c(x) = 1 - \mathcal{A}(x)$ .

一个关于  $U$  的模糊关系  $\mathcal{R}$  定义为  $U \times U$  的一个模

糊集, 即  $\mathcal{R}: U \times U \rightarrow [0, 1], \forall x, y \in U$ , 如果满足: 自反性( $\mathcal{R}(x, x) = 1$ )和对称性( $\mathcal{R}(x, y) = \mathcal{R}(y, x)$ ), 则称  $\mathcal{R}$  是  $U$  上的模糊相似关系.

信息熵被引入模糊粗糙集理论进行相关的不确定性测量与表示, 产生了不同形式<sup>[21-22, 32]</sup>. Qian 等人<sup>[21]</sup>提出了一种形式的模糊补熵. 设  $\mathcal{R}$  是  $U$  上的一个模糊关系且  $[x_i]_{\mathcal{R}}$  表示由  $\mathcal{R}$  生成的广义模糊类. 关于  $\mathcal{R}$  的模糊补熵定义为

$$E(\mathcal{R}) = \frac{1}{|U|} \sum_{i=1}^n \left( 1 - \frac{|[x_i]_{\mathcal{R}}|}{|U|} \right). \quad (1)$$

式(1)所示的模糊补熵通过交运算来定义模糊信息粒. 这样的策略可能会降低高维样本空间中模糊相似度的差异<sup>[27]</sup>. 因此, 模糊信息粒可能不能真实反映对象的粒结构. 此外, 由上述模糊补熵定义的模糊补互信息是非单调的. 为此, 本文定义关于决策的混杂核模糊补互信息, 进而构建了一种考虑内外重要度的有监督特征选择评价函数.

## 2 利用混杂核模糊补互信息选择特征

### 2.1 混杂核函数

假设  $U = \{x_1, x_2, \dots, x_n\}$ ,  $E \subseteq A = \{a_1, a_2, \dots, a_m\}$  且  $\forall x_i \in U$  和  $a \in A$ ,  $a(x_i)$  表示  $x_i$  在  $a$  下的取值.

现在生活中存在大量的分类、数值和混合属性数据. 对于分类属性子集  $E_c \subseteq A$ , 利用匹配核<sup>[30]</sup>来计算  $x_i$  和  $x_j$  关于  $E_c$  的模糊相似关系, 其定义为

$$k_{E_c}(x_i, x_j) = \begin{cases} 0, & \exists a \in E_c, a(x_i) \neq a(x_j), \\ 1, & \forall a \in E_c, a(x_i) = a(x_j). \end{cases} \quad (2)$$

对于数值属性子集  $E_n \subseteq A$ , 高斯核<sup>[31]</sup>被使用来抽取  $x_i$  和  $x_j$  关于  $E_n$  的模糊相似关系, 其计算方式为

$$k_{E_n}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|_{E_n}^2}{2\sigma^2}\right). \quad (3)$$

其中  $\|x_i - x_j\|_{E_n}$  是  $x_i$  和  $x_j$  之间关于  $E_n$  的欧氏距离;  $\sigma$  用于调整基于混杂核的模糊近似空间的粒度, 它反映了人们对数值属性中数据噪声的容忍程度.

**定义 1.** 对于混杂属性  $E = E_c \cup E_n \subseteq C$ , 混杂核函数  $k_{E_c \cup E_n}(x_i, x_j)$  通过如下代数乘积的方式计算<sup>[31]</sup>:

$$k_E(x_i, x_j) = k_{E_c \cup E_n}(x_i, x_j) = k_{E_c}(x_i, x_j) \times k_{E_n}(x_i, x_j). \quad (4)$$

**性质 1.** 对  $\forall E \subseteq A$ , 有

$$k_E(x_i, x_j) = \prod_{b \in E} k_b(x_i, x_j).$$

**性质 2.** 如果  $E_1 \subseteq E_2 \subseteq C$ , 则  $k_{E_1} \supseteq k_{E_2}$ .

容易看到上述混杂核函数满足自反和对称的性质. 因此, 通过上述这些核函数计算的关系是模糊相

似关系.

### 2.2 不确定性度量

设  $U = \{x_1, x_2, \dots, x_n\}$ ,  $\forall E \subseteq A$ ,  $E$  可以诱导一个  $U$  上的混杂核模糊相似关系  $k_E$ .  $k_E$  可以用模糊关系矩阵  $M_{k_E} = (r_{ij}^{k_E})_{n \times n}$  来表示, 其中  $r_{ij}^{k_E} = k_E(x_i, x_j)$ , 每一行  $(r_{i1}^{k_E}, r_{i2}^{k_E}, \dots, r_{in}^{k_E})$  表示一个模糊集. 显然,  $\forall E_1, E_2 \subseteq A$ , 有  $M_{k_{E_1 \cup E_2}} = M_{k_{E_1}} \otimes M_{k_{E_2}}$ , 其中  $\otimes$  表示代数乘积, 即  $(r_{ij}^{k_{E_1 \cup E_2}})_{n \times n} = (r_{ij}^{k_{E_1}} \times r_{ij}^{k_{E_2}})_{n \times n}$ .

由混杂核模糊关系  $k_E$  引起的包含  $x_i$  的模糊信息粒子定义为  $[x_i]_{k_E} = \frac{r_{i1}^{k_E}}{x_1} + \frac{r_{i2}^{k_E}}{x_2} + \dots + \frac{r_{in}^{k_E}}{x_n} = (r_{i1}^{k_E}, r_{i2}^{k_E}, \dots, r_{in}^{k_E})$ . 显然,  $[x_i]_{k_E}$  是一个关于  $E$  的模糊集, 有  $[x_i]_{k_E}(x_j) = k_E(x_i, x_j) = r_{ij}^{k_E}$ . 模糊集  $[x_i]_{k_E}$  的基数定义为  $|[x_i]_{k_E}| = \sum_{j=1}^n r_{ij}^{k_E} = \sum_{j=1}^n k_E(x_i, x_j)$ . 显然, 可以得到  $1 \leq |[x_i]_{k_E}| \leq n$ . 在不引起混淆的情况下, 用  $E$  来代替  $k_E$ .

**定义 2.**<sup>[31]</sup> 关于  $k_E$  的模糊补熵定义为

$$CH(E) = CH(k_E) = \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_E^c|}{|U|}, \quad (5)$$

易得到  $0 \leq CH(E) \leq 1 - \frac{1}{|U|}$ .

**性质 3.** 设  $E_1 \subseteq E_2 \subseteq A$ , 有

$$CH(E_1) \leq CH(E_2). \quad (6)$$

**性质 4.** 给定  $E \subseteq A$  是含数值属性的属性子集. 设  $k_{E^1}$  和  $k_{E^2}$  分别由  $\sigma_1$  和  $\sigma_2$  在  $E$  下计算得到的模糊关系. 如果  $\sigma_1 \leq \sigma_2$ , 则有

$$CH(k_{E^1}) \geq CH(k_{E^2}). \quad (7)$$

**证明.** 给定  $E \subseteq A$  是含数值属性的属性子集, 如果  $\sigma_1 \leq \sigma_2$ , 则有  $\exp\left(-\frac{\|x_i - x_j\|_E^2}{2\sigma_1^2}\right) \leq \exp\left(-\frac{\|x_i - x_j\|_E^2}{2\sigma_2^2}\right)$ , 从而  $k_{E^1} \subseteq k_{E^2}$ . 因此,  $CH(k_{E^1}) \geq CH(k_{E^2})$ . 证毕.

从性质 4 可以看出, 在相同的属性下, 较大的核参数  $\sigma$  会导致更粗的粒度.  $\sigma$  的取值越大, 则任意一对对象之间的相似程度也越大. 在这种情况下, 很难将任意的对象与其他对象区分开来. 结果, 模糊补熵越低.

当  $A = C \cup D$  且  $C \cap D = \emptyset$  时, 该信息系统被称为决策系统 (decision system, DS), 其中  $C$  表示条件属性集,  $D$  表示决策属性集. 基于上述模糊补熵,  $\forall B \subseteq C$ , 以下给出关于决策属性集  $D$  的模糊补联合熵、模糊补条件熵和模糊补互信息的概念.

**定义 3.**  $B$  和  $D$  的模糊补联合熵定义为

$$CH(B, D) = \frac{1}{|U|} \sum_{i=1}^n \frac{|([x_i]_B \otimes [x_i]_D)^c|}{|U|}. \quad (8)$$

**定义 4.**  $D$  关于  $B$  的模糊补条件熵定义为

$$CH(D|B) = \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_B \otimes [x_i]_D^c|}{|U|}. \quad (9)$$

性质 5.  $CH(D|B) = CH(B, D) - CH(B)$ .

证明. 根据定义 2 和定义 3, 有

$$\begin{aligned} CH(B, D) - CH(B) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{|([x_i]_B \otimes [x_i]_D)^c| - |[x_i]_B^c|}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{\sum_{j=1}^n (1 - r_{ij}^B \cdot r_{ij}^D) - \sum_{j=1}^n (1 - r_{ij}^B)}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{\sum_{j=1}^n (r_{ij}^B - r_{ij}^B \cdot r_{ij}^D)}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{\sum_{j=1}^n [r_{ij}^B \cdot (1 - r_{ij}^D)]}{|U|} \right) &= CH(D|B). \end{aligned}$$

证毕.

性质 6. 如果  $B_1 \subseteq B_2 \subseteq C$ , 则  $CH(D|B_1) \geq CH(D|B_2)$ .

证明. 设  $B_1 \subseteq B_2 \subseteq C$ , 由性质 2, 有  $k_{B_1} \supseteq k_{B_2}$ . 所以有  $r_{ij}^{B_1} - r_{ij}^{B_2} \geq 0$ .

$$\begin{aligned} CH(D|B_1) - CH(D|B_2) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{|[x_i]_{B_1} \otimes [x_i]_D^c| - |[x_i]_{B_2} \otimes [x_i]_D^c|}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{\sum_{j=1}^n r_{ij}^{B_1} \cdot (1 - r_{ij}^D) - \sum_{j=1}^n r_{ij}^{B_2} \cdot (1 - r_{ij}^D)}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{\sum_{j=1}^n [(r_{ij}^{B_1} - r_{ij}^{B_2}) \cdot (1 - r_{ij}^D)]}{|U|} \right) &\geq 0. \end{aligned}$$

因此, 如果  $B_1 \subseteq B_2 \subseteq C$ , 则  $CH(D|B_1) \geq CH(D|B_2)$ .

证毕.

定义 5.  $B$  和  $D$  的模糊补互信息定义为

$$CMI(B; D) = \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_B^c \otimes [x_i]_D^c|}{|U|}. \quad (10)$$

性质 7. 设  $B \subseteq C$ , 有 3 个等式成立:

- 1)  $CMI(B; D) = CMI(D; B)$ ;
- 2)  $CMI(B; D) = CH(D) - CH(D|B)$ ;
- 3)  $CMI(B; D) = CH(B) + CH(D) - CH(B, D)$ .

证明. 由定义 5, 显然有  $CMI(B; D) = CMI(D; B)$ , 因此性质 7 的等式 1) 成立. 因为

$$\begin{aligned} CMI(B; D) &= \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_B^c \otimes [x_i]_D^c|}{|U|} = \\ \frac{1}{|U|} \sum_{i=1}^n \frac{\sum_{j=1}^n [(1 - r_{ij}^B)(1 - r_{ij}^D)]}{|U|} &= \\ \frac{1}{|U|} \sum_{i=1}^n \frac{\sum_{j=1}^n (1 - r_{ij}^D) - \sum_{j=1}^n r_{ij}^B \cdot (1 - r_{ij}^D)}{|U|} &= \\ \frac{1}{|U|} \sum_{i=1}^n \left( \frac{|[x_i]_D^c| - |[x_i]_B \otimes [x_i]_D^c|}{|U|} \right) &= \\ \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_D^c|}{|U|} - \frac{1}{|U|} \sum_{i=1}^n \frac{|[x_i]_B \otimes [x_i]_D^c|}{|U|} &= \\ CH(D) - CH(D|B). \end{aligned}$$

所以, 性质 7 的等式 2) 成立. 由性质 5, 易证性质 7 的等式 3) 成立. 证毕.

在性质 7 的等式 2) 中,  $B$  和  $D$  的模糊补互信息为  $D$  的模糊补熵减去  $D$  关于  $B$  的模糊补条件熵, 因此  $B$  和  $D$  的模糊补互信息反映了  $B$  和  $D$  共同含有的模糊补信息量, 体现了 2 个属性集  $B$  和  $D$  之间相关性程度, 这与其他形式互信息熵的结论是完全一致的.

性质 8. 如果  $B_1 \subseteq B_2 \subseteq C$ , 则  $CMI(B_1; D) \leq CMI(B_2; D)$ .

证明. 设  $B_1 \subseteq B_2 \subseteq C$ , 则有  $k_{B_1} \supseteq k_{B_2}$ . 由性质 6 和性质 7 的等式 2), 易证  $CMI(B_1; D) \leq CMI(B_2; D)$ . 证毕.

性质 8 表明模糊补互信息与特征子集大小的单调性变化. 这些性质对设计一个启发式搜索算法是至关重要的. 因为这保证了向已选取的特征子集中添加候选特征不会减少新特征子集的信息. 因此, 上述定义的模糊补互信息可以用来作为特征选择的标准.

### 2.3 属性重要度

特征选择的关键问题是建立属性质量的评价函数.  $B$  和  $D$  的模糊补互信息体现了  $B$  和  $D$  之间相关性程度. 因此, 基于模糊补互信息, 分别构建了 2 种属性重要度评价函数.

定义 6. 对  $\forall b \in B$ , 属性  $b$  在  $B$  中相对于  $D$  的内重要度定义为

$$\text{sig}^{\text{in}}(b, B, D) = CMI(B; D) - CMI(B - \{b\}; D). \quad (11)$$

定义 7. 对  $\forall b \in C - B$ , 属性  $b$  在  $B$  中相对于  $D$  的外重要度定义为

$$\text{sig}^{\text{out}}(b, B, D) = CMI(B \cup \{b\}; D) - CMI(B; D). \quad (12)$$

由性质 8,  $\text{sig}^{\text{in}}(a, B, D)$  越高, 表示  $a$  越重要.

当数据集含有数值属性时, 从式 (3) 和式 (4) 可以看到当  $k_{E_c}(x_i, x_j) = 1$  但  $k_{E_n}(x_i, x_j) \rightarrow 0$ , 有  $k_{E_c \cup E_n}(x_i, x_j) \rightarrow$



0且 $k_{E_c \cup E_n}(x_i, x_j) \neq 0$ . 因此, 不能直接利用保持特征子集与决策的模糊补互信息不变来定义属性约简; 而是利用允许模糊补互信息在小的扰动之内变化的思想来定义属性约简, 即可以引入一个参数 $\varepsilon > 0$ 来定义 $\varepsilon$ -核属性和属性约简.

如果 $\text{sig}^{\text{in}}(c, C, D) > \varepsilon$ , 属性 $c$ 是 $\varepsilon$ -必不可少的, 也就是说,  $c$ 是一个 $\varepsilon$ -核属性.

**定义 8.**  $C$  的核属性集定义为

$$\text{Core}_C(D) = \{c \in C \mid \text{sig}^{\text{in}}(c, C, D) > \varepsilon\}. \quad (13)$$

基于核属性的启发式属性约简算法可以通过将选定的属性逐步添加到核属性集中来找到一个属性约简. 下面是基于模糊补互信息的约简定义.

**定义 9.** 设 $B \subseteq C$ ,  $B$  称为  $C$  的一个 $\varepsilon$ -约简, 如果  $B$  满足 2 个条件:

- 1)  $\text{CMI}(C; D) - \text{CMI}(B; D) \leq \varepsilon$ ;
- 2)  $\forall b \in B, \text{CMI}(B; D) - \text{CMI}(B - \{b\}; D) > \varepsilon$ .

其中: 条件 1) 保证所选属性子集与整个属性集在 $\varepsilon$ -误差之内具有相同的区分能力; 条件 2) 通过删除选定属性子集中的每个属性, 确保其所有属性在 $\varepsilon$ -误差之内都是不可缺少的.

## 2.4 相关算法

基于定义 6 和定义 7 的 2 种重要度, 设计一种用于特征选择的启发式算法并分析其运行的复杂度. 算法 1 以空集为起点, 首先通过内重要度计算核属性集; 然后以核属性集为基础, 每次计算全部剩余属性的属性重要度, 从中选择重要度值最大的属性加入核属性集中, 直到所有条件属性集和约简集的决策模糊补互信息之差小于等于 $\varepsilon$ .

**算法 1.** KFCMI 算法.

输入:  $DS = \langle U, C \cup D \rangle$ ,  $|C| = m$ , 阈值 $\sigma, \varepsilon$ ;

输出: 一个约简集 ( $Red$ ).

- ①  $Red \leftarrow \emptyset, Label \leftarrow 1, B \leftarrow C - Red$ ;
- ② for  $k \leftarrow 1$  to  $m$  do
- ③ 计算模糊关系矩阵  $M_{c_k}$ ;
- ④ end for
- ⑤ for  $k \leftarrow 1$  to  $m$  do
- ⑥ 通过定义 6 计算  $\text{sig}^{\text{in}}(c_k, C, D)$ ;
- ⑦ if  $\text{sig}^{\text{in}}(c_k, C, D) > \varepsilon$  then
- ⑧  $Red \leftarrow Red \cup \{c_k\}$ ;
- ⑨  $B \leftarrow C - \{c_k\}$ ;
- ⑩ end if
- ⑪ end for

- ⑫ while  $\text{CMI}(D; C) - \text{CMI}(D; Red) > \varepsilon$  do
- /\*令  $B = \{c_{k_1}, c_{k_2}, \dots, c_{k_h}\}$ \*/
- ⑬ for  $l \leftarrow 1$  to  $|B|$  do
- ⑭ 通过定义 7 计算  $\text{sig}^{\text{out}}(c_{k_l}, Red, D)$ ;
- ⑮ end for
- ⑯ 选择  $c_{k_r}$ , 使得  $\text{sig}^{\text{out}}(c_{k_r}, Red, D)$  具有最大值;
- ⑰  $Red \leftarrow Red \cup \{c_{k_r}\}, B \leftarrow B - \{c_{k_r}\}$ ;
- ⑱ end while
- ⑲ if  $|Red| == m$  then
- ⑳ return  $Red \leftarrow Red(1 : m - 1)$ ;
- ㉑ else
- ㉒ return  $Red$ ;
- ㉓ end if

在算法 1 中, 步骤②~④的循环次数为 $m$ , 步骤③的循环次数为 $n \times n$ , 步骤⑤~⑪的循环次数为 $m$ , 步骤⑫~⑱的循环次数为 $h$ . 从而, 算法 1 总的循环次数为 $m \times n \times n + m + h$ . 因此, 在最坏的情况下, 算法 1 的时间复杂度为 $O(mn^2)$ .

## 3 实 验

在这部分, 为了验证所提算法的可行性和有效性, 将本文所提算法与基于模糊信息熵(fuzzy information entropy, FIE)<sup>[17]</sup>的算法、基于模糊粗糙集的特征选择(fuzzy rough-based feature selection, FRFS)<sup>[33]</sup>、基于模糊区分矩阵的属性约简(fuzzy discernibility matrix-based attribute reduction, FDMAR)<sup>[34]</sup>、适应模糊粗糙集(fitting fuzzy rough sets, FFRS)<sup>[18]</sup>和基于区分邻域数的特征选择(discernible neighborhood counting-based feature selection, DNCFS)<sup>[35]</sup>进行了分类实验的对比分析.

### 3.1 实验准备

实验使用了 15 个数据集, 它们是从 UCI 机器学习库<sup>①</sup>中挑选出来的. 这些数据集的基本信息如表 1 所示. 对于一些数据集中存在的缺失值, 本文采用最大概率值法来填补缺失值. 此外, 所有的数值属性值通过最小-最大标准化归一化为区间  $[0, 1]$ .

分类回归树(classification and regression tree, CART)、朴素贝叶斯(naive Bayes, NB)和  $k$ -最近邻( $k$ -nearest neighbor,  $k$ NN)被使用来评估这些对比算法的分类效果. 所有分类实验通过 10 折交叉验证来实施, 分类准

① <http://archive.ics.uci.edu/ml>

Table 1 Basic Information of Datasets

表 1 数据集的基本信息

数据集	对象数	特征数	决策类	数据类型
Ann (Annealing)	798	38	5	混杂
Arrh (Arrhythmia)	452	279	13	混杂
Autos	205	25	6	混杂
Credit (Credit approval)	690	15	2	混杂
Ecoli	336	7	8	数值
Glass	214	10	6	数值
Heart	270	13	2	混杂
Move (Movement libras)	360	90	15	数值
Park (Parkinsons)	195	22	2	数值
Sick	3 772	29	2	混杂
Spam (Spambase)	4 601	57	2	数值
Wave (Waveform)	5 000	21	3	数值
WDBC (Wisconsin Diagnostic Breast Cancer)	569	31	2	数值
Wine	178	13	3	数值
WPBC (Wisconsin Prognostic Breast Cancer)	198	34	2	数值

确率的平均值和标准差作为最终结果。

在实验中, KFCMI 算法中有 2 个参数 $\sigma$ 和 $\varepsilon$ . 引入参数 $\sigma$ 控制样本模糊相似度, 这对算法的性能有很大的影响; 而参数 $\varepsilon$ 为算法的停止条件. 对于给定的数据集, 如果参数 $\varepsilon$ 的值变小, 所选特征数不会减少. 一般来说, 不同的 $\sigma$ 和 $\varepsilon$ 会导致不同的分类准确率. 因此, 通过调整参数值, 使 $\sigma$ 在 0~1 之间变化, 步长为 0.02, 且 $\varepsilon \in \{10^{-3}, 10^{-4}\}$ , 为每个数据集选择一个最优特征子集. 在 FFRS 算法中引入了 2 个参数 $\varepsilon$ 和 $\lambda$ . 遵循文献[19]的参数设置, 以 0.05 的步长将 $\varepsilon$ 设置为 0.1~0.5 之间的值, 以 0.1 的步长将 $\lambda$ 设置为 0.1~0.6 之间的值. DNCFS 算法涉及邻域半径参数 $\delta$ , 其调节范围为 [0, 1], 步长为 0.02. 所有实验结果都是在最高分类准确率的情况下给出的.

### 3.2 实验结果

表 2 给出了不同算法下所选特征的平均数. 从表 2 可见, KFCMI 算法大多数情况下取得了较小的

Table 2 Average Number of Features Selected by Different Algorithms

表 2 不同算法所选特征的平均数

数据集	原始特征数	FIE	FRFS	FDMAR	FFRS	DNCFS	KFCMI
Ann	38	15	1	17	11.7	15.7	11.7
Arrh	279	40	1	142	36.3	122.7	125.3
Autos	25	17	14	20	4.0	10.7	9.7
Credit	15	14	14	14	8.0	8.7	7.3
Ecoli	7	7	7	7	7.0	5.3	6.0
Glass	10	10	9	10	8.7	2.0	3.7
Heart	13	13	11	13	6.0	9.7	7.0
Move	90	69	39	72	66.3	71.7	64.3
Park	22	14	13	20	13.7	4.7	8.0
Sick	29	26	1	26	16.0	11.7	10.0
Spam	57	53	56	57	55.0	52.0	48.7
Wave	21	21	21	21	21.0	18.3	19.3
WDBC	31	22	25	31	22.7	13.0	5.7
Wine	13	13	11	13	11.7	5.0	5.0
WPBC	34	28	20	34	28.3	7.0	6.3
平均值	45.6	24.1	16.2	33.1	21.1	23.9	22.5

注: 黑体数值表示没有去除任何冗余特征.

特征平均数; FIE, FRFS, FDMAR 算法在一些数据集上得到的是整个属性集, 这说明 FIE, FRFS, FDMAR 算法在一些数据集上无法有效地去除冗余特征. 此外, 对于平均值而言, KFCMI 算法小于 FIE, FDMAR, DNCFS 算法, 但略大于 FRFS 和 DNCFS 算法. 表 3~5 分别展示了原始数据和基于这 6 种算法的约简数据的分类准确率.

从表 3~5 中可以看到, 算法 KFCMI 在所有数据集上都能提高或保持原始数据的分类准确率. 在表 3~5 中 45 条记录中, KFCMI 算法有 29 条记录实现了最佳分类准确率. 然而, 对于 FIE, FRFS, FDMAR, FFRS, DNCFS 算法, 分别仅有 2, 0, 0, 8, 8 条记录实现了最佳分类准确率. 更多的是, KFCMI 算法的平均分类准确率在 3 种分类算法上均优于其他所有特征选择算法.

Table 3 Classification Accuracy of Different Algorithms Based on CART Algorithm

表 3 基于 CART 算法的不同算法分类准确率

%

数据集	原始数据	FIE	FRFS	FDMAR	FFRS	DNCFS	KFCMI
Ann	92.08±0.33	92.19±0.41	76.19±0.00	92.23±0.51	91.25±0.60	<b>92.63±0.50</b>	92.42±0.50
Arrh	65.15±1.56	64.98±0.79	54.20±0.00	64.31±1.77	53.34±0.93	66.02±1.27	<b>66.06±1.82</b>
Autos	74.00±3.98	75.76±1.83	75.32±1.78	75.90±2.63	51.02±2.58	76.54±1.74	<b>77.90±1.90</b>
Credit	80.43±0.80	81.25±1.25	81.30±1.05	81.14±0.86	69.93±1.36	85.36±0.66	<b>85.43±0.90</b>
Ecoli	80.68±1.53	<b>81.88±1.09</b>	80.45±1.68	80.51±1.52	<b>81.88±1.86</b>	81.73±1.17	<b>81.88±1.36</b>
Glass	98.50±0.43	98.69±0.37	98.50±0.30	98.55±0.41	98.32±0.33	98.83±0.33	<b>98.88±0.39</b>

续表 3

数据集	原始数据	FIE	FRFS	FDMAR	FFRS	DNCFS	KFCMI
Heart	76.89±2.38	76.22±1.65	75.26±1.59	76.26±1.68	<b>82.44±1.04</b>	79.11±2.30	81.07±1.12
Move	66.11±2.90	66.69±2.00	67.17±2.62	65.22±1.41	64.19±1.60	66.67±2.36	<b>67.64±2.44</b>
Park	85.59±1.67	86.72±1.99	86.82±2.20	86.10±2.20	<b>88.31±1.51</b>	87.33±1.64	87.79±1.41
Sick	98.83±0.12	98.49±0.11	93.88±0.00	98.79±0.10	98.34±0.08	98.86±0.08	<b>98.87±0.08</b>
Spam	91.80±0.36	91.62±0.49	91.71±0.18	91.83±0.27	91.98±0.27	91.92±0.28	<b>92.16±0.28</b>
Wave	75.23±0.45	75.36±0.38	75.27±0.48	75.11±0.62	<b>75.63±0.49</b>	75.06±0.60	75.60±0.37
WDBC	92.46±0.76	93.06±0.70	92.86±0.65	92.02±0.53	93.30±0.74	93.90±0.79	<b>94.22±0.63</b>
Wine	89.61±1.57	90.51±1.11	91.57±1.91	89.66±1.48	90.90±1.09	93.03±1.22	<b>94.04±0.66</b>
WPBC	66.21±3.25	67.63±2.83	67.58±1.87	68.79±2.18	69.55±2.43	73.08±3.88	<b>73.69±2.23</b>
平均值	82.24±1.47	82.74±1.13	80.54±1.09	82.43±1.21	80.03±1.13	84.00±1.25	<b>84.51±1.07</b>

注：黑体值为最高的分类准确率。

Table 4 Classification Accuracy of Different Algorithms Based on NB Algorithm

表 4 基于 NB 算法的不同算法分类准确率

%

数据集	原始数据	FIE	FRFS	FDMAR	FFRS	DNCFS	KFCMI
Ann	75.98±0.26	76.02±0.41	76.19±0.00	76.10±0.12	<b>76.29±0.23</b>	76.22±0.10	76.19±0.31
Arrh	62.83±0.59	66.55±0.93	54.20±0.00	60.69±0.76	62.43±0.62	66.97±0.80	<b>68.14±0.42</b>
Autos	60.20±2.08	65.32±1.67	68.34±0.96	63.22±1.32	43.61±1.28	68.73±1.83	<b>70.83±1.07</b>
Credit	66.74±0.39	67.73±0.46	67.56±0.45	67.65±0.69	72.14±0.34	84.86±0.24	<b>86.26±0.34</b>
Ecoli	84.46±0.61	84.38±0.86	84.85±0.45	84.26±0.68	<b>85.00±0.68</b>	84.76±0.44	84.91±0.37
Glass	90.70±0.87	90.47±0.55	92.20±1.15	90.61±0.64	91.54±0.60	<b>98.55±0.46</b>	98.41±0.45
Heart	79.63±0.78	80.04±0.54	79.63±1.00	79.70±0.46	<b>82.41±0.53</b>	82.22±0.43	82.30±0.55
Move	68.03±1.47	67.25±1.12	64.50±1.29	67.17±1.60	66.58±1.14	68.17±1.23	<b>68.42±1.78</b>
Park	74.56±0.65	78.87±0.90	81.38±0.87	74.41±0.61	82.51±1.20	85.74±0.83	<b>85.95±0.36</b>
Sick	93.75±0.03	93.72±0.05	93.88±0.00	93.71±0.04	93.85±0.04	96.70±0.06	<b>96.78±0.04</b>
Spam	54.21±0.12	53.81±0.13	54.33±0.16	54.36±0.19	54.33±0.17	54.17±0.22	<b>56.04±0.17</b>
Wave	80.71±0.10	80.68±0.08	80.73±0.06	80.69±0.06	80.76±0.06	81.50±0.08	<b>81.57±0.09</b>
WDBC	93.71±0.55	94.07±0.17	92.63±0.32	93.74±0.37	92.29±0.31	95.13±0.22	<b>96.80±0.22</b>
Wine	97.98±0.60	97.70±0.41	97.87±0.24	97.47±0.48	96.74±0.36	<b>98.03±0.30</b>	98.03±0.30
WPBC	65.71±2.29	67.22±1.36	68.18±2.01	66.06±2.10	67.88±1.85	<b>78.99±0.90</b>	78.99±1.07
平均值	76.61±0.76	77.59±0.64	77.10±0.60	76.65±0.67	76.56±0.63	81.38±0.54	<b>81.97±0.50</b>

注：黑体值为最高的分类准确率。

Table 5 Classification Accuracy of Different Algorithms Based on kNN Algorithm

表 5 基于 kNN 算法的不同算法分类准确率

%

数据集	原始数据	FIE	FRFS	FDMAR	FFRS	DNCFS	KFCMI
Ann	90.53±0.62	90.78±0.64	76.19±0.00	90.71±0.46	90.24±0.52	<b>91.29±0.36</b>	90.79±0.65
Arrh	54.29±0.66	54.60±0.88	0.66±0.00	52.94±0.84	51.35±0.78	55.44±0.84	<b>56.66±0.65</b>
Autos	74.20±0.90	70.83±1.56	75.37±1.08	72.98±1.15	51.80±1.07	<b>82.93±0.92</b>	82.49±1.44
Credit	80.51±0.54	80.26±0.94	80.67±0.57	80.51±0.49	68.23±0.78	<b>81.35±0.40</b>	81.33±0.67
Ecoli	80.60±0.48	80.89±0.86	80.89±0.52	81.16±0.44	<b>81.28±0.89</b>	80.95±0.61	81.19±0.50
Glass	90.89±0.67	91.31±0.59	91.64±0.60	91.26±0.99	91.12±0.66	98.27±0.44	<b>98.50±0.48</b>
Heart	75.41±0.84	75.74±1.25	78.96±0.94	75.89±0.83	74.04±0.85	79.59±0.79	<b>80.26±1.00</b>
Move	85.97±0.82	86.25±0.77	84.92±0.94	85.97±0.53	86.72±0.58	<b>87.08±0.79</b>	87.00±0.67
Park	95.74±0.59	95.28±0.68	95.08±0.60	95.74±0.73	96.46±0.38	96.87±0.78	<b>97.03±0.47</b>
Sick	96.17±0.08	96.43±0.12	93.88±0.00	96.19±0.14	96.28±0.14	96.69±0.14	<b>96.99±0.12</b>
Spam	90.93±0.09	90.80±0.19	90.84±0.25	90.90±0.09	90.99±0.16	90.69±0.23	<b>91.04±0.14</b>
Wave	77.29±0.24	77.22±0.11	77.34±0.20	77.25±0.20	77.47±0.27	77.12±0.28	<b>77.50±0.17</b>
WDBC	95.40±0.41	<b>95.99±0.26</b>	95.45±0.23	95.76±0.46	95.85±0.24	95.62±0.31	95.98±0.17
Wine	95.11±0.46	94.94±0.00	95.84±0.39	95.28±0.39	96.69±0.32	98.65±0.47	<b>98.88±0.26</b>
WPBC	69.24±0.97	71.52±1.70	73.13±1.54	70.20±1.47	72.88±1.31	74.85±0.92	<b>75.10±1.22</b>
平均值	83.48±0.56	83.52±0.70	79.39±0.52	83.52±0.61	81.43±0.60	85.83±0.55	<b>86.05±0.58</b>

注：黑体值为最高的分类准确率。

综上, KFCMI 算法实现了较优的分类准确率. KFCMI 算法可以选出相对较小的特征子集并提高或保持分类准确率. 而这 15 个数据集包括数值和混合属性数据集, 因此, KFCMI 算法适用于多种属性类型的特征选择.

### 3.3 实验参数

为了分析本文所提算法 KFCMI 的性能对参数的敏感性, 3 种分类算法的所选特征数及分类准确率随参数的变化如图 1 所示. 显然, 可以看到不同参数可能导致不同的特征数及分类准确率. 下面从所选特

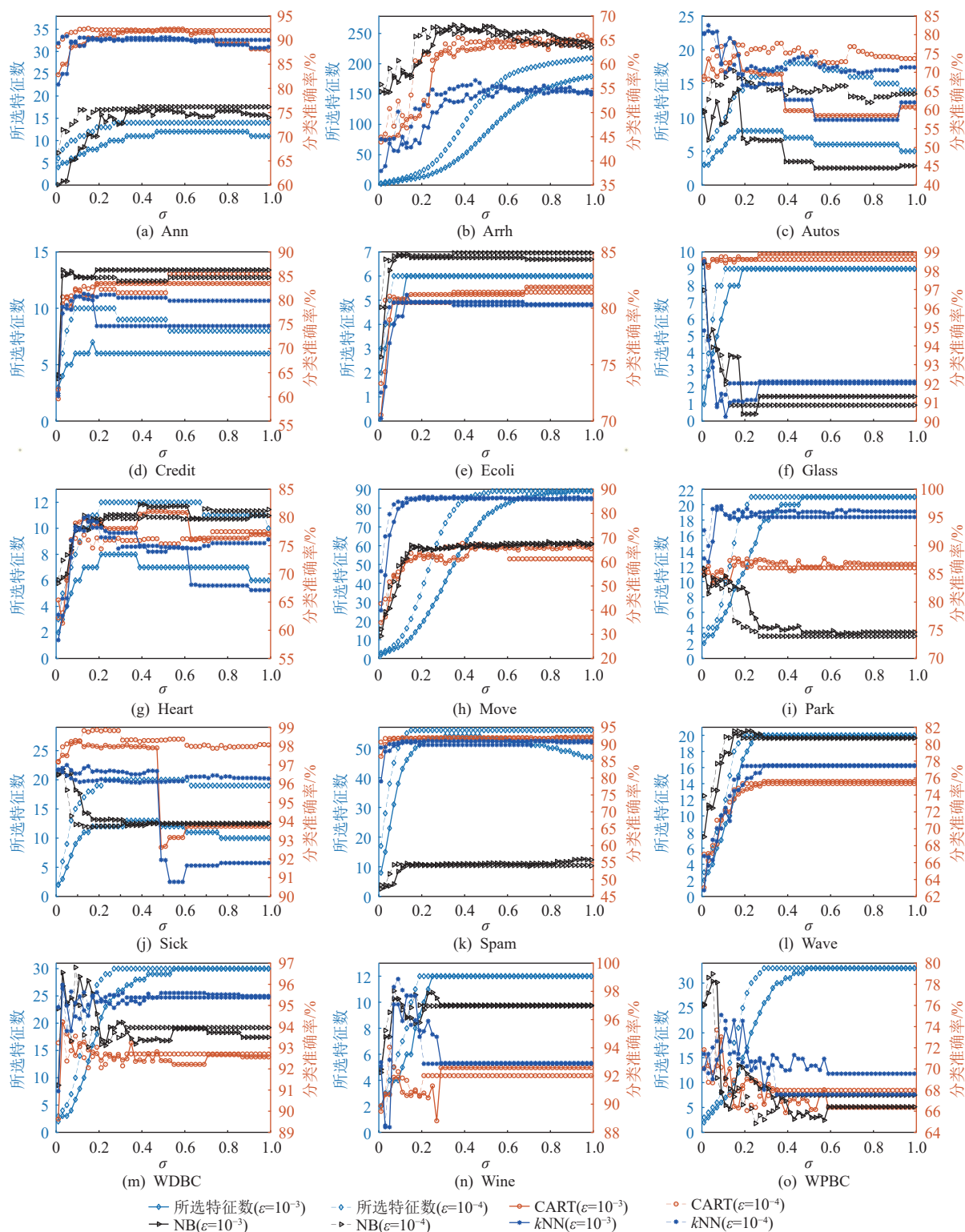


Fig. 1 Average classification accuracy varies with parameters

图 1 平均分类准确率随参数的变化



征数和分类准确率 2 方面来进行分析。

1) 就特征数而言. 通过图 1 可以看到, 在大多数数据集上, 随着 $\sigma$ 的增加, 所选特征数均呈递增的趋势, 最后趋于平衡. 例如, Arrh, Ecoli, Move 等数据集. 然而, 对于 Autos, Credit, Heart 等数据集, 所选特征数随着 $\sigma$ 的增加而增加, 然后减小并开始波动, 最后并未趋于平稳.

2) 就分类准确率而言. 从图 1 可以看到大多数数据集可以在多个参数 $\sigma$ 值下取得最优值. 对于每个数据集, 可以根据图 1 选择合适的 $\sigma$ 值来获得较优的分类准确率. 此外, 从 3 种分类算法的分类准确率变化曲线来看, 可以看到 3 种分类算法获得的结果基本上一致.

通过上述分析可以看到, 实验性能对参数 $\sigma$ 具有一定的敏感性. 因此, 在所提的算法中调节参数是很有必要的. 但是, 在合适的参数值的条件下, KFCMI 算法在大多数情况下也可以获得较优的结果.

综上, KFCMI 算法对于数值和混合属性分类算法的特征选择是可行且有效的.

## 4 结 论

本文基于混杂核函数定义了关于决策的模糊补互信息, 并证明了其关于特征呈单调性变化. 进而提出了基于混杂核模糊补互信息的特征选择方法, 且设计了相应的启发式算法. 最后, 在 15 个实际数据集上对所提算法与现有算法进行了实验对比分析. 实验结果表明所提算法是混合特征选择的一种有效方案. 在将来的工作中, 可以进一步考虑特征的交互性.

**作者贡献声明:** 袁钟负责算法思路和实验方案的提出, 以及完成实验并撰写论文; 陈红梅提出指导意见并修改论文; 王志红负责论文实验分析; 李天瑞提出指导意见并修改论文.

## 参 考 文 献

- [1] Liang Jiye, Wang Feng, Dang Chuangyin, et al. A group incremental approach to feature selection applying rough set technique[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 26(2): 294–308
- [2] An Shuang, Hu Qinghua, Pedrycz W, et al. Data-distribution-aware fuzzy rough set model and its application to robust classification[J]. *IEEE Transactions on Cybernetics*, 2015, 46(12): 3073–3085
- [3] Dai Jianhua, Chen Jiaolong. Feature selection via normative fuzzy information weight with application in biological data classification[J]. *Applied Soft Computing*, 2020, 92(7): 106299
- [4] Sun Lin, Wang Lanying, Ding Weiping, et al. Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2021, 29(1): 19–33
- [5] Wang Chongzhong, Huang Yang, Ding Weiping, et al. Attribute reduction with fuzzy rough self-information measures[J]. *Information Sciences*, 2021, 49(5): 68–86
- [6] Yao Sheng, Xu Feng, Zhao Peng, et al. Intuitionistic fuzzy entropy Feature selection algorithm based on adaptive neighborhood space rough set model[J]. *Journal of Computer Research and Development*, 2018, 55(4): 802–814 (in Chinese)  
(姚晟, 徐风, 赵鹏, 等. 基于自适应邻域空间粗糙集模型的直觉模糊熵特征选择[J]. *计算机研究与发展*, 2018, 55(4): 802–814)
- [7] Wang Chongzhong, Huang Yang, Shao Mingwen, et al. Feature selection based on neighborhood self-information[J]. *IEEE Transactions on Cybernetics*, 2020, 50(9): 4031–4042
- [8] Dash M, Liu Huan. Consistency-based search in feature selection[J]. *Artificial Intelligence*, 2003, 151(1/2): 155–176
- [9] Hu Qinghua, Xie Zongxia, Yu Daren. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation[J]. *Pattern Recognition*, 2007, 40(12): 3509–3521
- [10] Wang Chongzhong, Huang Yang, Shao Mingwen, et al. Uncertainty measures for general fuzzy relations[J]. *Fuzzy Sets and Systems*, 2019, 360(4): 82–96
- [11] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. *International Journal of General System*, 1990, 17(2/3): 191–209
- [12] Mi Jusheng, Zhang Wenxiu. An axiomatic characterization of a fuzzy generalization of rough sets[J]. *Information Sciences*, 2004, 160(1/4): 235–249
- [13] Wang Jinbo, Wu Weizhi. Intuitionistic fuzzy rough sets over two universes[J]. *Fuzzy Systems and Mathematics*, 2021, 35(6): 1–13 (in Chinese)  
(王金波, 吴伟志. 双论域上的直觉模糊粗糙集[J]. *模糊系统与数学*, 2021, 35(6): 1–13)
- [14] Yeung D S, Chen Degang, Tsang E C C, et al. On the generalization of fuzzy rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2005, 13(3): 343–361
- [15] Jensen R, Shen Qiang. Fuzzy-rough attribute reduction with application to Web categorization[J]. *Fuzzy Sets and Systems*, 2004, 141(3): 469–485
- [16] Hu Qinghua, Yu Daren, Xie Zongxia. Information-preserving hybrid data reduction based on fuzzy-rough techniques[J]. *Pattern Recognition Letters*, 2006, 27(5): 414–423
- [17] Chen Degang, Hu Qinghua, Yang Yongping. Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets[J]. *Information Sciences*, 2011, 181(23): 5169–5179
- [18] Wang Chongzhong, Qi Yali, Shao Mingwen, et al. A fitting model for feature selection with fuzzy rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2017, 25(4): 741–753
- [19] Yager R R. Entropy measures under similarity relations[J]. *International Journal of General System*, 1992, 20(4): 341–358

- [20] Hu Qinghua, Yu Daren, Xie Zongxia, et al. Fuzzy probabilistic approximation spaces and their information measures[J]. *IEEE Transactions on Fuzzy Systems*, 2006, 14(2): 191–201
- [21] Qian Yuhua, Liang Jiye, Wu Weizhi, et al. Information granularity in fuzzy binary GrC model[J]. *IEEE Transactions on Fuzzy Systems*, 2010, 19(2): 253–264
- [22] Xu Jiucheng, Wang Yun, Xu Keqiang, et al. Feature genes selection using fuzzy rough uncertainty metric for tumor diagnosis[J]. *Computational and Mathematical Methods in Medicine*, 2019, 27(1): 1–9
- [23] Fan Yunrui, Zhang Xianrong, Yang Jilin. Uncertainty measurement of decision-entropies based on fuzzy neighborhood rough sets[J]. *Computer Engineering and Design*, 2021, 42(5): 1300–1306 (in Chinese)  
(樊雲端, 张贤勇, 杨霖琳. 模糊邻域粗糙集的决策熵不确定性度量[J]. *计算机工程与设计*, 2021, 42(5): 1300–1306)
- [24] Zhang Xiao, Mei Changlin, Chen Degang, et al. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy[J]. *Pattern Recognition*, 2016, 56(8): 1–15
- [25] Lin Yaojin, Hu Qinghua, Liu Jinghua, et al. Streaming feature selection for multilabel learning based on fuzzy mutual information[J]. *IEEE Transactions on Fuzzy Systems*, 2017, 25(6): 1491–1507
- [26] Zhao Junyang, Zhang Zhili, Han Chongzhao, et al. Complement information entropy for uncertainty measure in fuzzy rough set and its applications[J]. *Soft Computing*, 2015, 19(7): 1997–2010
- [27] Wang Chongzhong, Huang Yang, Shao Mingwen, et al. Fuzzy rough set-based attribute reduction using distance measures[J]. *Knowledge-Based Systems*, 2019, 164(1): 205–212
- [28] Hu Qinghua, Zhang Lei, Chen Degang, et al. Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications[J]. *International Journal of Approximate Reasoning*, 2010, 51(4): 453–471
- [29] Hu Qinghua, Yu Daren, Pedrycz W, et al. Kernelized fuzzy rough sets and their applications[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2010, 23(11): 1649–1667
- [30] Hu Qinghua, Zhang Lingjun, Zhou YuCan, et al. Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2017, 26(1): 226–238
- [31] Yuan Zhong, Chen Hongmei, Yang Xiaoling, et al. Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction[J]. *Knowledge-Based Systems*, 2021, 231(11): 107398
- [32] 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
- [33] Jensen R, Shen Qiang. New approaches to fuzzy-rough feature selection[J]. *IEEE Transactions on Fuzzy Systems*, 2009, 17(4): 824–838
- [34] Chen Degang, Zhao Suyun. Local reduction of decision system with fuzzy rough sets[J]. *Fuzzy Sets and Systems*, 2010, 161(13): 1871–1883
- [35] Yang Yanyan Song Shiji, Chen Degang, et al. Discernible neighborhood counting based incremental feature selection for heterogeneous data[J]. *International Journal of Machine Learning and Cybernetics*, 2019, 11(8): 1115–1127



**Yuan Zhong**, born in 1991. PhD. His main research interests include granular computing and knowledge discovery.

袁 钟, 1991 年生. 博士. 主要研究方向为粒计算与知识发现.



**Chen Hongmei**, born in 1971. PhD, professor, PhD supervisor. Member of CCF. Her main research interests include granular computing, rough set, and data mining.

陈红梅, 1971 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为粒计算、粗糙集和数据挖掘.



**Wang Zhihong**, born in 1993. PhD candidate. Her main research interests include granular computing, rough set, and data mining.

王志红, 1993 年生. 博士研究生. 主要研究方向为粒计算、粗糙集和数据挖掘.



**Li Tianrui**, born in 1969. PhD, professor, PhD supervisor. Member of CCF. His main research interests include granular computing, rough set, and data mining.

李天瑞, 1969 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为粒计算、粗糙集和数据挖掘.