山西大学学报(自然科学版)48(1):77－88,2025
Journal of Shanxi University(Nat. Sci. Ed. )

**DOI**:10.13451/j.sxu.ns.2024150

# 基于自信息和模糊邻域条件熵的特征选择方法

徐久成[1,2],段江豪[1,2*],牛武林[1,2],张杉[1,2],白晴[1,2]
(1.河南师范大学 计算机与信息工程学院,河南 新乡 453007;
2.智慧商务与物联网技术河南省工程实验室,河南 新乡 453007)

**摘　要**:针对模糊邻域粗糙集的特征选择方法通常仅考虑下近似中的分类信息,而忽略上近似和边界域中的分类信息这一问题,本文提出了一种基于自信息和模糊邻域条件熵的特征选择算法。首先,结合下近似、上近似和边界域提出了三种自信息不确定性测度,并将三种自信息相结合提出了相似自信息。其次,在信息论视角下,给出了模糊邻域条件熵的不确定性度量,并将其与相似自信息相结合,提出了更为全面的特征评价函数,用于衡量特征子集分类信息的不确定性,并基于此利用最大相关最小冗余技术设计特征选择算法。最后,通过在数据集上进行对比实验,实验结果表明所提算法能有效处理上近似和边界域中的分类信息;且所提算法在两个分类器下其平均分类精度,在低维数据集中分别提高了2.55%和4.15%,在高维数据集中分别提高了0.83%和2.54%。

**关键词**:模糊邻域粗糙集;自信息;不确定性度量;模糊邻域熵;模糊邻域条件熵
**中图分类号**:TP181　　**文献标志码**:A　　**文章编号**:0253-2395(2025)01-0077-12

# Feature Selection Method Based on Self-information and Fuzzy Neighborhood Conditional Entropy

XU Jiucheng[1,2], DUAN Jianghao[1,2*], NIU Wulin[1,2], ZHANG Shan[1,2], BAI Qing[1,2]
(1. *College of Computer and Information Engineering, Henan Normal University, Xinxiang* 453007*, China;*
2. *Engineering Lab of Intelligence Business & Internet of Things of Henan Province, Xinxiang* 453007*, China*)

**Abstract:** The feature selection method for fuzzy neighborhood rough sets usually only considers the classification information in the approximation, but cannot evaluate the classification information in the approximation and boundary domains. In this paper, we propose a feature selection algorithm based on self-information measure and fuzzy neighborhood conditional entropy. Firstly, three measures of self-information uncertainty are proposed by combining the lower approximation, the upper approximation and the boundary domain, and the similarity self-information is proposed by combining the three types of self-information. Secondly, from the perspective of information theory, the uncertainty measure of fuzzy neighborhood conditional entropy is given, and combined with similar self-information, a more comprehensive feature evaluation function is proposed to measure the uncertainty of feature subset classification information, and based on this, the feature selection algorithm is designed by using the maximum correlation and minimum redundancy technology. Finally, through comparative experiments on the dataset, the results show that the proposed algorithm can effectively process the classification information in the approximation and boundary domains; and under the two classifiers of the proposed algorithm, its average classification accuracy is improved by 2.55% and 4.15%, respectively, in the low-dimensional data set compared with the existing algorithms, and is improved by 0.83% and 2.54%, respectively, in the high-dimensional data set.

**Key words:** fuzzy neighborhood rough set; self-information; uncertainty measure; fuzzy neighborhood entropy; fuzzy neighborhood conditional entropy

**收稿日期**:2024-06-20;**接受日期**:2024-10-24
**基金项目**:国家自然科学基金(61976082;62076089;62002103)
**作者简介**:徐久成(1963－),男,河南洛阳人,博士,教授,研究方向为粒计算、数据挖掘和生物信息学。E-mail:xjc@htu.edu.cn
* **通信作者**:段江豪(DUAN Jianghao),E-mail:15903076989@163.com
**引文格式**:徐久成,段江豪,牛武林,等.基于自信息和模糊邻域条件熵的特征选择方法[J].山西大学学报(自然科学版),2025,48(1):77-88. DOI:10.13451/j.sxu.ns.2024150.

## 0 引言

目前,人们在数据挖掘和机器学习中获得的信息和数据正在迅速扩大,信息的不确定性有所增加,数据之间的关系也变得更加复杂。从模糊和不精确的数据中获取有用的信息已成为数据挖掘领域的一个重大挑战[1]。特征选择作为数据预处理的一项重要技术,其从原始的特征集合中筛选出具有最佳区分能力的特征子集,旨在消除不相关和冗余的特征,以降低特征的维度和减少特征的空间[2-3],目前已被广泛应用于模式识别、数据挖掘和机器学习等领域[4-5]。因此,特征选择在实际应用和理论研究方面具有重要的实际意义和学术价值。

粗糙集的概念最初是由Pawlak[6]提出的,其作为一种处理数据模糊和不确定性的数学方法,已成功应用于特征选择[7-9]。然而,Pawlak提出的粗糙集是基于一般的等价关系,这限制了该模型仅适用于离散数据处理。然而,包含离散数据和连续数据的混合数据在实际生产和生活中广泛存在。因此,当应用经典的粗糙集模型处理连续数据时,需要对数据离散化,但离散化过程会导致信息的丢失[10-11]。为克服这一问题,许多学者扩展了粗糙集模型。Hu等[12]采用邻域关系代替等价关系构造邻域粗糙集模型,广泛应用于混合数据的处理。Dubois等[13]将粗糙集和模糊集相结合,提出了模糊粗糙集模型,其允许数据之间在某种程度上可以区分,能够处理不确定的信息,解决了经典粗糙集处理数值型数据的局限性。Qian等[14]研究了悲观的多粒度粗糙集决策模型,克服了大多数模型由于单一等价关系而限制其应用的缺点。Wang等[15]提出的模糊邻域粗糙集模型解决特征选择问题,该模型的优点是不仅可以有效处理连续性数据,而且可用参数化的模糊关系来描述模糊信息的粒度,降低了样本被误分类的可能性。受此启发,本文利用模糊邻域粗糙集,进一步解决特征选择问题。

特征选择的关键在于构建一个有效的特征评价函数,以衡量特征子集的分类性能[16]。在模糊邻域粗糙集中依赖度是一个广泛使用的特征评价函数。Xu等[17]应用依赖度构建了相对依赖互补互信息的拟合模糊粗糙集特征选择方法,解决了模糊相似性和模糊粗糙近似易受数据分布影响的问题。Zhou等[18]应用依赖度提出一种面向高维类不平衡数据的在线流特征选择方法,以在线方式处理类不平衡数据。Ma等[19]研究了一种动态的基于依赖度的无监督模糊粗糙集特征选择算法,使得时间效率上得以提高。

然而,针对上述以及大部分的特征选择算法利用依赖度这一评价函数时,仅利用下近似的分类信息,忽略了上近似和边界域中的分类信息,而Wang等[3]表明上近似中的不一致实例同时具有分类信息,Shu[20]等表明可将边界样本看作是含有噪声的正域。此外,Shannon[21]提出的自信息是一类能有效刻画随机变量的不确定性的测度。因此,本文为考虑上近似和边界域中的分类信息,首先结合上近似定义了模糊度,结合边界域定义了近似精度,并将依赖度、模糊度和近似精度分别引入自信息的概念中构建了积极自信息、消极自信息和相对自信息,并为综合考虑下近似、上近似和边界域中的分类信息,故将三种自信息相结合构建了相似自信息,解决了上近似和边界域中信息没有利用的问题。其次,将代数观构建的相似自信息与信息观定义的模糊邻域条件熵相结合,其在双视角下构建了更为全面的特征评价函数,用于衡量特征子集分类信息的不确定性,此弥补了仅从单一视角下构建特征评价函数的不足。最后,基于上述理论设计了一种特征选择算法(Feature Selection Method Based on Self-information and Fuzzy Neighborhood Conditional Entropy,FS-SIFNCE),其实验结果表明,所提算法相较于其他算法能有效提高数据的分类性能。

## 1 相关工作

本节主要介绍了模糊邻域粗糙集、自信息和模糊邻域熵的基本知识。

**定义1**[22] 模糊邻域系统由以下四部分组成,简记为$DS=\{U, C, D, \delta\}$。

(1)$U=\{x_1, x_2, \cdots, x_n\}$是被称为论域的样本集合;

(2)$C=\{c_1, c_2, \cdots, c_m\}$是用于描述每个样

本的条件特征集合;

(3)$D$是决定每个样本类别的决策特征,能将$U$划分为互不相交的$t$个样本子集$U/D=\{D_1, D_2, \cdots, D_t\}$;

(4)$\delta$是邻域半径参数,邻域半径的大小决定了该邻域范围的扩展程度。

**定义** 2[17] 设$B\subseteq C$是论域$U$上的一个属性集,则$B$在$U$上诱导出一个模糊二元关$R_B$。如果满足以下条件,则称$R_B$是一个模糊相似关系:

(1)自反性:$R_B(x, x)=1, \forall x\in U$;

(2)对称性:$R_B(x, y)=R_B(y, x), \forall x, y\in U$。

**定义** 3[17] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$a\in C$,模糊邻域半径$\delta(0<\delta<1)$用于描述样本的相似性,则对任意$x, y\in U$,两个样本$x$和$y$之间关于$a$的模糊邻域相似关系表示为:

$$R_a^{\delta}(x, y)=\begin{cases}0, & R_a(x, y)<\delta\\ R_a(x, y), & R_a(x, y)\geqslant\delta\end{cases}。\tag{1}$$

两个样本$x$和$y$关于特征$a$的模糊邻域相似矩阵表示为$[x]_a(y)=R_a(x, y)$。

**定义** 4[17] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$B\subseteq C$,$a\in B$,对任意的$x, y\in U$,$x$关于$B$的模糊相似矩阵为$[x]_B(y)=\min_{a\in B}\left([x]_a(y)\right)$,$x$关于$B$的参数化模糊邻域信息粒表示为:

$$\alpha_B(x)=[x]_B(y)=\begin{cases}0, & R_B(x, y)<\delta\\ R_B(x, y), & R_B(x, y)\geqslant\delta\end{cases}。\tag{2}$$

**定义** 5[17] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$U=\{x_1, x_2, \cdots, x_n\}$,$U/D=\{D_1, D_2, \cdots, D_t\}$,由$D$导出的样本的模糊决策表示为:$FD=\{FD_1^{\mathrm{T}}, FD_2^{\mathrm{T}}, \cdots, FD_t^{\mathrm{T}}\}$,样本决策的模糊等价类$FD_j=\{FD_j(x_1), FD_j(x_2), \cdots, FD_j(x_n)\}$,$j=1, 2, \cdots, t$,当$l=1, 2, \cdots, n$时,$FD_j(x_l)$是$FD_j$上$x_l\in U$的隶属度,可表示为:

$$FD_j(x_l)=\frac{\left|[x_l]_C(y)\cap D_j\right|}{\left|[x_l]_C(y)\right|}。\tag{3}$$

其中$[x_l]_C(y)$是模糊相似类,$FD_i^{\mathrm{T}}$是$FD_i$的转置。

**定义** 6[17] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$\alpha_B(x_i)$是$x_i\in U$的参数化模糊邻域信息粒,特征子集$B\subseteq C$,样本子集$x_i\in X\subseteq U$相对于$B$的上近似$\overline{N_B}(X)$和下近似$\underline{N_B}(X)$分别表达为:

$$\overline{N_B}(X)=\left\{x_i\in U|\alpha_B(x_i)\cap X\neq\varnothing\right\},\tag{4}$$

$$\underline{N_B}(X)=\left\{x_i\in U|\alpha_B(x_i)\subseteq X\right\}。\tag{5}$$

**定义** 7[22] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,特征子集$B\subseteq C$,$U/D=\{D_1, D_2, \cdots, D_t\}$决策特征$D$相对于$B$的上近似$\overline{N_B}(D)$和下近似$\underline{N_B}(D)$分别表达为:

$$\overline{N_B}(D)=\bigcup_{i=1}^{t}\overline{N_B}(D_i),\tag{6}$$

$$\underline{N_B}(D)=\bigcup_{i=1}^{t}\underline{N_B}(D_i)。\tag{7}$$

**定义** 8[20] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,特征子集$B\subseteq C$,正域和边界域分别定义为:

$$P_B(D)=\underline{N_B}(D),\tag{8}$$

$$F_B(D)=\overline{N_B}(D)-\underline{N_B}(D)。\tag{9}$$

则决策特征$D$相对于$B$的依赖度表达为:

$$r_B(D)=\frac{\left|P_B(D)\right|}{|U|}。\tag{10}$$

**定义** 9[21] 测度$I(x)$是由Claude Shannon提出的,表示随机变量$x$的不确定性,如果满足下列四个条件,则称$I(x)$是$x$的自信息。

(1)非负性:$I(x)\geqslant 0$;

(2)若$q(x)=1$,则有$I(x)=0$;

(3)若$q(x)\to 0$,则有$I(x)\to\infty$;

(4)单调性:若$q(x)\leqslant q(y)$,则$I(x)\geqslant I(y)$。

其中$q(x)$表示$x$发生概率。

**定义** 10[17] 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$B\subseteq C$,关于$B$的模糊邻域熵表示为:

$$E(B)=-\frac{1}{|U|}\sum_{i=1}^{n}\log\frac{|\alpha_B(x_i)|}{|U|}。\tag{11}$$

其中$\frac{|\alpha_B(x_i)|}{|U|}$表示样本邻域在全集中的比例。

## 2 模糊邻域自信息和熵度量模型

本节阐述四种自信息的相关定理,并将提出的相似自信息与模糊邻域条件熵相结合,建立了模糊邻域自信息和熵度量模型。

### 2.1 相似自信息

**定义** 11 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, 特征子集 $B\subseteq C$, $U/D=\{D_1, D_2, \cdots, D_t\}$,则决策特征 $D$ 相对于 $B$ 的依赖度表达为:

$$r_B(D)=\frac{\left|\underline{N_B}(D)\right|}{|U|}。\tag{12}$$

**定理** 1 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, $U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集 $B_1\subseteq B_2\subseteq C$,则有 $r_{B_1}(D)\leqslant r_{B_2}(D)$。

**证明** 由公式(2)知,当 $B_1\subseteq B_2$ 时,有 $\alpha_{B_1}(x_i)\supseteq\alpha_{B_2}(x_i)$, 所以, $\underline{N_{B_1}}(D_i)\subseteq\underline{N_{B_2}}(D_i)$, $\underline{N_{B_1}}(D)\subseteq\underline{N_{B_2}}(D)$,$\left|\underline{N_{B_1}}(D)\right| /|U|\leqslant\left|\underline{N_{B_2}}(D)\right|/|U|$。根据公式(12)可得 $r_{B_1}(D)\leqslant r_{B_2}(D)$。

**定义** 12 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, 特征子集 $B\subseteq C$, $U/D=\{D_1, D_2, \cdots, D_t\}$,则决策特征 $D$ 相对于 $B$ 的模糊度表达为:

$$m_B(D)=\frac{\left|\overline{N_B}(D)\right|}{|U|}。\tag{13}$$

**定理** 2 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, $U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集 $B_1\subseteq B_2\subseteq C$,则有 $m_{B_1}(D)\geqslant m_{B_2}(D)$。

**证明** 由公式(2)知,当 $B_1\subseteq B_2$ 时,有 $\alpha_{B_1}(x_i)\supseteq\alpha_{B_2}(x_i)$, 所以, $\overline{N_{B_1}}(D_i)\supseteq\overline{N_{B_2}}(D_i)$, $\overline{N_{B_1}}(D)\supseteq\overline{N_{B_2}}(D)$,$\left|\overline{N_{B_1}}(D)\right|/|U|\geqslant\left|\overline{N_{B_2}}(D)\right|/|U|$。根据公式(13)可得,$m_{B_1}(D)\geqslant m_{B_2}(D)$。

**定义** 13 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, 特征子集 $B\subseteq C$, $U/D=\{D_1, D_2, \cdots, D_t\}$,则决策特征 $D$ 相对于 $B$ 的近似精度表达为:

$$u_B(D)=1-\frac{\left|\overline{N_B}(D)-\underline{N_B}(D)\right|}{|U|}。\tag{14}$$

**定理** 3 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, $U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集 $B_1\subseteq B_2\subseteq C$,则有 $u_{B_1}(D)\leqslant u_{B_2}(D)$。

**证明** 当 $B_1\subseteq B_2$ 时,由定理1和定理2可知,$\underline{N_{B_1}}(D)\subseteq\underline{N_{B_2}}(D)$,$\overline{N_{B_1}}(D)\supseteq\overline{N_{B_2}}(D)$,故

$$1-\left(\left|\overline{N_{B_1}}(D)-\underline{N_{B_1}}(D)\right|/|U|\right)\leqslant 1-\left(\left|\overline{N_{B_2}}(D)-\underline{N_{B_2}}(D)\right|/|U|\right),$$

根据公式(14)可得,$u_{B_1}(D)\leqslant u_{B_2}(D)$。

**定义** 14 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, 特征子集 $B\subseteq C$, $U/D=\{D_1, D_2, \cdots, D_t\}$,$B$ 的积极自信息表达为:

$$I_B^r(D)=-\left(1-r_B(D)\right)\log r_B(D)。\tag{15}$$

**定理** 4 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, $U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集 $B_1\subseteq B_2\subseteq C$,则有 $I_{B_1}^r(D)\geqslant I_{B_2}^r(D)$。

**证明** 当 $B_1\subseteq B_2$ 时,由定理1可知,$r_{B_1}(D)\leqslant r_{B_2}(D)$,且 $r_{B_1}(D)$ 和 $r_{B_2}(D)\in[0,1]$,所以, $1-r_{B_1}(D)\geqslant 1-r_{B_2}(D)$, $-\log r_{B_1}(D)\geqslant -\log r_{B_2}(D)$, 因此, $-(1-r_{B_1}(D))\log r_{B_1}(D)\geqslant -(1-r_{B_2}(D))\log r_{B_2}(D)$。根据公式(15)可得,$I_{B_1}^r(D)\geqslant I_{B_2}^r(D)$。

评注:$I_B^r(D)$利用下近似中的分类信息来评价 $B$ 中分类信息的不确定性,其值越小表明不确定性越小。当 $r_B(D)=0$,即 $\left|\underline{N_B}(D)\right|=0$,表明了特征子集 $B$ 无法区分不同类别的关键特征和属性,具有较差的分类能力且 $I_B^r(D)\to\infty$。当 $r_B(D)=1$,即 $\left|\underline{N_B}(D)\right|=|U|$,表明了特征子集 $B$ 可以区分不同类别的关键特征和属性,具有较好的分类能力且 $I_B^r(D)=0$。

**定义** 15 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, 特征子集 $B\subseteq C$, $U/D=\{D_1, D_2, \cdots, D_t\}$,$B$ 的消极自信息表达为:

$$I_B^m(D)=-\left(1-m_B(D)\right)\log m_B(D)。\tag{16}$$

**定理** 5 给定模糊邻域系统 $DS=\{U, C, D, \delta\}$, $U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集 $B_1\subseteq B_2\subseteq C$,则有 $I_{B_1}^m(D)\leqslant I_{B_2}^m(D)$。

**证明** 当 $B_1\subseteq B_2$ 时,由定理2可知 $m_{B_1}(D)\geqslant m_{B_2}(D)$,且 $m_{B_1}(D)$ 和 $m_{B_2}(D)\in[0,1]$,所以 $1-m_{B_1}(D)\leqslant 1-m_{B_2}(D)$, $-\log m_{B_1}(D)\leqslant -\log m_{B_2}(D)$, 因此 $-(1-m_{B_1}(D))\log m_{B_1}(D)\leqslant -(1-m_{B_2}(D))\log m_{B_2}(D)$。根据公式(16)可

得,$I^m_{B_1}(D)\leqslant I^m_{B_2}(D)$。

评注:$I^m_B(D)$利用上近似中的分类信息来评价$B$中分类信息的不确定性,其值越小表明不确定性越小。当$m_B(D)=0$,即$\left|\overline{N_B}(D)\right|=0$,表明了特征子集$B$可能无法区分不同类别的关键特征和属性,具有较差的分类能力且$I^m_B(D)\to\infty$。当$m_B(D)=1$,即$\left|\overline{N_B}(D)\right|=|U|$,表明了特征子集$B$可能可以区分不同类别的关键特征和属性,具有较好的分类能力且$I^m_B(D)=0$。

**定义** 16 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,特征子集$B\subseteq C$,$U/D=\{D_1, D_2, \cdots, D_t\}$,$B$的相对自信息表达为:

$$I^u_B(D)=-\left(1-u_B(D)\right)\log u_B(D)。\tag{17}$$

**定理** 6 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集$B_1\subseteq B_2\subseteq C$,则有$I^u_{B_1}(D)\geqslant I^u_{B_2}(D)$。

**证明** 当$B_1\subseteq B_2$时,由定理3可知$u_{B_1}(D)\leqslant u_{B_2}(D)$,且$u_{B_1}(D)$和$u_{B_2}(D)\in[0,1]$,所以,$1-u_{B_1}(D)\geqslant 1-u_{B_2}(D)$,$-\log u_{B_1}(D)\geqslant -\log u_{B_2}(D)$,因此,$-(1-u_{B_1}(D))\log u_{B_1}(D)\geqslant -\left(1-u_{B_2}(D)\right)\log u_{B_2}(D)$。根据公式(17)得,$I^u_{B_1}(D)\geqslant I^u_{B_2}(D)$。

评注:$I^u_B(D)$利用边界域中的分类信息的近似精度来评价$B$中分类信息的不确定性,其值越小表明不确定性越小。当$u_B(D)=0$,即$\left|\overline{N_B}(D)-\underline{N_B}(D)\right|=|U|$,表明了特征子集$B$无法区分不同类别的关键特征和属性,具有较差的分类能力且$I^u_B(D)\to\infty$。当$u_B(D)=1$,即$\left|\overline{N_B}(D)-\underline{N_B}(D)\right|=0$,表明了特征子集$B$可以区分不同类别的关键特征和属性,具有较好的分类能力且$I^u_B(D)=0$。

然而,$I^r_B(D)$、$I^m_B(D)$和$I^u_B(D)$仅分别代表下近似、上近似和边界域中的分类信息来评价特征子集$B$中分类信息的不确定性,因此这三种不确定性测度是片面的。所以结合了$I^r_B(D)$、$I^m_B(D)$和$I^u_B(D)$提出相似自信息来综合确定$B$中分类信息的不确定性。

**定义** 17 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,特征子集$B_1\subseteq B_2\subseteq C$,$U/D=\{D_1, D_2, \cdots, D_t\}$,$B$的相似自信息表达为:

$$I^n_B(D)=I^r_B(D)+I^m_B(D)+I^u_B(D)。\tag{18}$$

**定理** 7 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集$B_1\subseteq B_2\subseteq C$,则有相似自信息$I^n_B(D)$不满足单调性。

**证明** 当$B_1\subseteq B_2$时,由公式(2)知,$\alpha_{B_1}(x_i)\supseteq\alpha_{B_2}(x_i)$,由定理4、定理5和定理6可知,$I^r_{B_1}(D)\geqslant I^r_{B_2}(D)$,$I^m_{B_1}(D)\leqslant I^m_{B_2}(D)$,$I^u_{B_1}(D)\geqslant I^u_{B_2}(D)$,所以相似自信息$I^n_{B_1}(D)$与$I^n_{B_2}(D)$间的关系是不确定的,即相似自信息不满足单调性。

### 2.2 模糊邻域条件熵

**定义** 18 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$U=\{x_1, x_2, \cdots, x_n\}$,$U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集$B\subseteq C$,决策特征$D$关于特征子集$B$的模糊邻域条件熵表达为:

$$\mathrm{FNE}(D|B)=-\frac{1}{|U|}\sum_{i=1}^{n}\sum_{j=1}^{t}\frac{\left|\alpha_B(x_i)\cap FD_j\right|}{\left|\alpha_B(x_i)\right|}\times\log\frac{\left|\alpha_B(x_i)\cap FD_j\right|}{\left|\alpha_B(x_i)\right|}。\tag{19}$$

其中$\left|\alpha_B(x_i)\cap FD(x_i)\right|$表示$\alpha_B(x_i)$的隶属度不大于$FD(x_i)$非零值的个数。

**定义** 19 给定模糊邻域系统$DS=\{U, C, D, \delta\}$,$U=\{x_1, x_2, \cdots, x_n\}$,$U/D=\{D_1, D_2, \cdots, D_t\}$,特征子集$B\subseteq C$,关于$B$和$D$基于相似自信息的模糊邻域条件熵表达为:

$$\mathrm{IFNE}(D|B)=-I^n_B(D)\times\frac{1}{|U|}\sum_{i=1}^{n}\sum_{j=1}^{t}\frac{\left|\alpha_B(x_i)\cap FD_j\right|}{\left|\alpha_B(x_i)\right|}\log\frac{\left|\alpha_B(x_i)\cap FD_j\right|}{\left|\alpha_B(x_i)\right|}。\tag{20}$$

其中$I^n_B(D)$是$DS$中的相似自信息。

$I^n_B(D)$是从代数视角出发表示模糊邻域自信息测度,$\mathrm{FNE}(D|B)$是从信息视角表示模糊邻域条件熵。因此,$\mathrm{IFNE}(D|B)$能同时从代数视角和信息视角下进行不确定性度量。

## 3 基于自信息和模糊邻域条件熵的特征选择算法设计

苗夺谦和王国胤等研究表明[23-24]对于分类性能较差的数据,满足单调性的特征选择算法并不能取得很好的分类效果。定理7表明相似自信息具有非单调性。因而,提出的基于自信息和模糊邻域条件熵的特征选择算法还可以克服基于单调测度所设计算法的不足。

**定义** 20 给定模糊邻域系统$DS=\{U, C,$

$D, \delta\}$,特征子集$B \subseteq C$,若满足以下条件,则称$B$是$C$相对于$D$的约简:

(1) $\mathrm{IFNE}(D|B) = \mathrm{IFNE}(D|C)$;

(2) $\mathrm{IFNE}(D|B) > \mathrm{IFNE}(D|B-\{c_i\}), \forall c_i \in B$。

其中,式(1)通过约简子集,可以达到与整个数据集相同的分类结果。式(2)约简子集保留了与整个数据集相同的分类准确性,但拥有更少的属性,没有冗余。

**定义** 21 给定模糊邻域系统$DS = \{U, C, D, \delta\}$,特征子集$B \subseteq C$,对于任何属性子集$c_i \in C - B$,则属性$c_i$相对于$D$的属性重要度定义为:

$$\mathrm{Sig}(c_i, B, D) = \mathrm{IFNE}(D|B) - \mathrm{IFNE}(D|B \cup \{c_i\})。 \quad (21)$$

根据公式(21),$\mathrm{Sig}(c_i, B, D)$代表了$c_i$提供的分类信息量。通过定义20从条件特征集合中挑选出能够提供最大分类信息的特征子集,以此作为特征集合的约简子集。依据上述内容,提出一种基于自信息和模糊邻域条件熵的特征选择算法(FS-SIFNCE),算法描述见算法1。

**算法**1 基于自信息和模糊邻域条件熵的特征选择算法(FS-SIFNCE)。

输入:模糊邻域决策系统$DS = \{U, C, D, \delta\}$

输出:特征约简子集$red$

步骤1. 初始化$red = \varnothing, B = C - red$;

步骤2. While $B \neq \varnothing$ do

步骤3. For each $c \in B$

步骤4. 计算$\mathrm{Sig}(c, red, D)$;

步骤5. end for

步骤6. 选出特征$c_i = \max \mathrm{Sig}(c, red, D)$;

步骤7. if $\mathrm{Sig}(r, red, D) > 0$

步骤8. Let $red = \{red \cup c_i\}$

步骤9. $B = C - c_i$

步骤10. else

步骤11. break

步骤12. end if

步骤13. end while

步骤14. for each $r \in red$

步骤15. 计算$\mathrm{Sig}(r, red, D)$

步骤16. if $\mathrm{Sig}(r, red, D) \leqslant 0$

步骤17. Let $red = \{red - r\}$

步骤18. end if

步骤19. end for

步骤20. 返回约简子集$red$。

假设数据存储是$n \times m$维度矩阵,其中$n$是实例数,$m$是条件特征数.该算法对复杂度影响较大的是参数化模糊邻域粒的计算,时间复杂度约为$O(mn)$。步骤2—13的最佳时间复杂度是$O(mn^2)$,它最坏的时间复杂度近似$O(mn^3)$。步骤14—19的时间复杂度为$O(n)$。假设所选特征子集的大小为$r$,则模糊邻域粒的时间复杂度为$O(rn)$,在算法中大多数情况下$n$远大于$r$,所以FS-SIFNCE总的时间复杂度为$O(mn)$。

## 4 实验及结果分析

本节通过对实验结果的分析,证明所提算法在能剔除冗余属性的同时提高了分类精度。

### 4.1 实验准备

为验证FS-SIFNCE算法对于提高数据集的分类性能是有效的,选取了8个公共数据集作为实验对象,包括4个低维数据集(Wine, Wisconsin Prognostic Breast Cancer (WPBC), Wisconsin Diagnostic Breast Cancer (WDBC), Heart)和4个高维数据集(Colon, Breast, Lung, Diffuse Large B-cell Lymphoma(DLBCL))。表1给出了数据集的详细信息。

**表1** 8个数据集信息

**Table 1** Information of eight datasets

| 数据集 | 样本数 | 特征数 | 类别数 |
|---|---|---|---|
| Wine | 178 | 13 | 3 |
| WPBC | 194 | 32 | 2 |
| Heart | 270 | 13 | 2 |
| WDBC | 569 | 30 | 2 |
| DLBCL | 77 | 5 649 | 2 |
| Lung | 181 | 12 533 | 2 |
| Colon | 62 | 2 000 | 2 |
| Breast | 84 | 9 216 | 5 |

本文所有实验均在Windows 10操作系统,Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz和8.0 GB RAM环境下进行,并使用Matlab 2016a实验和完成所有对比实验。实验分析了

FS-SIFNCE算法在所选8个数据集上的特征子集大小,以及约简数据集在K-近邻(K-Nearest Neighbors,KNN)和分类和回归树(Classification and Regression Tree,CART)分类器下的分类精度,分类器中的参数均使用默认值。为保证本文的实验具有一致性,对所有的数据集均采用十折交叉验证获取最后的实验数据。

### 4.2 Fisher Score用于高维数据集初步降维

在高维数据中,不是所有的特征都对于解决问题或任务都有意义或贡献。降维可帮助识别重要的特征,从而提高模型的性能并减少计算成本。故将Fisher Score引入FS-SIFNCE算法中。根据Fisher Score中的特征评价准则的原理,特征得分越高,其对应特征的辨别能力就越强,因此根据不同数据特征的Fisher Score,选出得分前$k$的特征作为候选子集,并将其分割为10个不同维度的特征集合,然后采用KNN和CART分类器来评估10个不同维度下的约简数据集的分类精度。最后,根据分类精度,选择具有最高精度的约简数据集作为后续研究的输入数据。

图1直观地展示分类精度不会随着维度的增加而增高。故针对不同的数据集要结合具体的维数和精度,将降维后的数据集作为分类器的输入。因此,对于数据集Colon的最佳分类精度所对应的候选特征维度为300维,对于数据集Breast的最佳分类精度所对应的候选特征维度为250维,类似的,对于数据集Lung和DLBCL分别取200维和100维。

### 4.3 选取模糊邻域半径

模糊邻域半径的选择会影响邻域关系的形成,进而影响上下近似集合。因此合理选择模糊邻域半径至关重要。表2展示了8个数据集在两个分类器上所获得的最佳模糊邻域半径的具体数值。图2和图3展现了8个数据集在两个分类器上的分类效果以及所选特征的个数。

### 4.4 低维数据集实验结果分析

为评估FS-SIFNCE算法在低维数据上的有效性,本节研究对比了6种不同的特征选择算法(其中包括:一种基于直觉模糊正区域寻找直觉模糊决策系统约简的启发式算法(A Heuristic Algorithm for Finding a Reduct of an Intuitionistic Fuzzy Decision System Based on Intuitionistic Fuzzy Positive Region,IFPR)[25]、基于邻域自信息的特征选择算法(Feature Selection Based on Neighborhood Self-information,NSI)[3]、一种基于悲观邻域多粒度依赖联合熵的属性约简算法(A Pessimistic Neighborhood Multi-granulation Dependency Joint Entropy-based Attribute Reduction Algorithm,PDJE-AR)[1]、一种基于邻域容差依赖联合熵的特征选择算法(A Feature Selection Algorithm Based on the Neighborhood Tolerance Dependency Joint Entropy,FSNTDJE)[26]、一种在拟合模糊粗糙集模型中利用相对依赖补互信息的启发式算法(A Heuristic Algorithm Using Relative Dependency Complement Mutual Information in the Fitting Fuzzy Rough

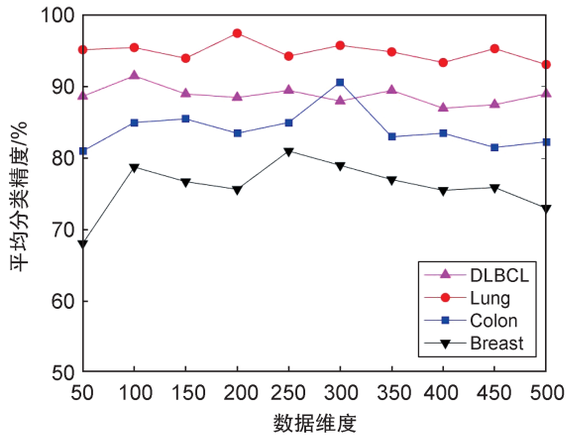


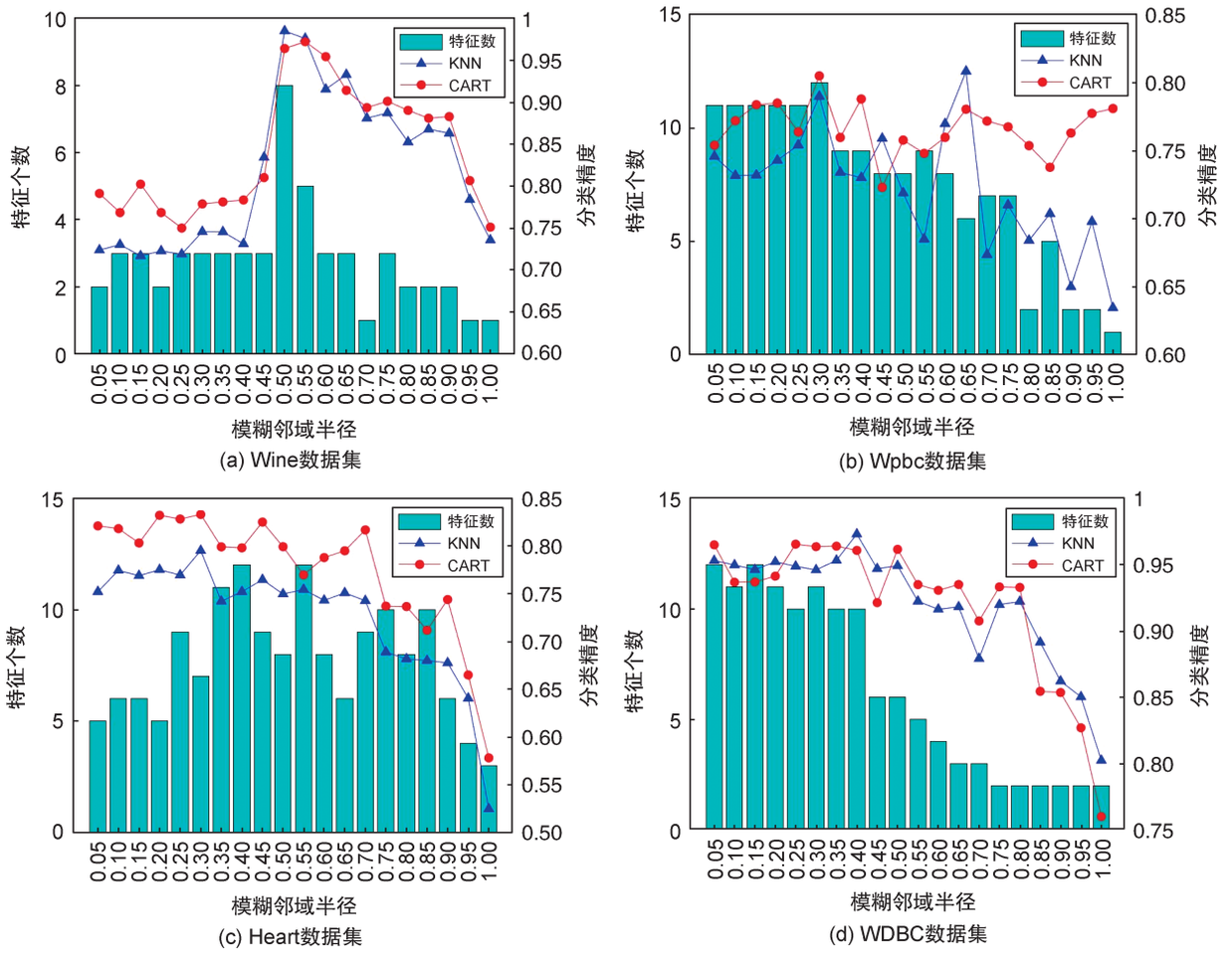

**图1** 50—500维度下4个高维数据集在KNN和CART分类器下的平均分类精度

**Fig. 1** The average classification accuracy of four high-dimensional datasets under KNN and CART classifiers in 50-500 dimensions

**表2** KNN与CART分类器上8个数据集的最佳模糊邻域半径值

**Table 2** The best fuzzy neighborhood radius values for eight datasets on KNN and CART classifiers

| 数据集 | KNN | CART |
|---|---|---|
| Wine | 0.5 | 0.5 |
| WPBC | 0.6 | 0.3 |
| Heart | 0.3 | 0.3 |
| WDBC | 0.4 | 0.25 |
| DLBCL | 0.15 | 0.15 |
| Lung | 0.45 | 0.5 |
| Colon | 0.5 | 0.5 |
| Breast | 0.05 | 0.15 |




注:柱状青色对应左侧$y$轴,蓝色和红色线条对应右侧$y$轴。

**图2** 不同模糊邻域半径下4个低维数据集的约简特征子集大小和分类精度

**Fig. 2** The feature reduction subset size and classification accuracy of four low- dimensional datasets with different fuzzy neighborhood radius

Set,FNRDCI)[17]和基于最大相关最小冗余的自适应邻域粗糙集粗糙互信息的特征选择算法(Maximum Relevance Minimum Redundancy-based Feature Selection Using Rough Mutual Information in Adaptive Neighborhood Rough Sets,FSRMI)[22])在4个低维数据集(Wine、WPBC、Heart、WDBC)的分类结果。表3展示了7种算法在4个低维数据集上所选特征的数量。而表4和表5分别展示了这4种算法在KNN和CART分类器下的分类精度。其中所选特征数和分类精度均采用十折交叉验证得到。加粗的内容代表了对比中的最佳结果。

根据表3可知,FSRMI算法所选择的特征数量均值最小,而FS-SIFNCE算法紧随其后,说明其具有比较好的约简能力。根据表3和表4中的KNN分类器数据,FSRMI算法在Heart数据集上的分类能力表现出良好的性能,但该算法在该数据集上所选的特征数和平均分类精度均弱于FS-SIFNCE算法,其中前者所选特征数多出1.4且平均分类精度低2.8%。这也表明FS-SIFNCE算法在低维数据集上性能是稳定的。另在WPBC数据集上,FS-SIFNCE算法在

**表3** 7种算法在4个低维数据集上选择特征的个数

**Table 3** The number of features selected by seven algorithms on four low-dimensional datasets

| 算法 | Wine | WPBC | Heart | WDBC | Mean |
|---|---|---|---|---|---|
| IFPR | 6.2 | 10.4 | 8.4 | 17.1 | 10.5 |
| NSI | 8.0 | 8.7 | 6.9 | 7.6 | 7.8 |
| PDJE-AR | **5.7** | 12.5 | 9.6 | 15.7 | 10.9 |
| FSNTDJE | 7.2 | 7.6 | 8.1 | 11.4 | 8.6 |
| FNRDCI | 6.9 | 6.2 | 7.3 | 9.0 | 7.4 |
| FSRMI | 6.2 | **5.4** | 6.4 | **3.8** | **5.5** |
| FS-SIFNCE | 7.5 | 7.2 | **5.0** | 6.0 | 6.4 |

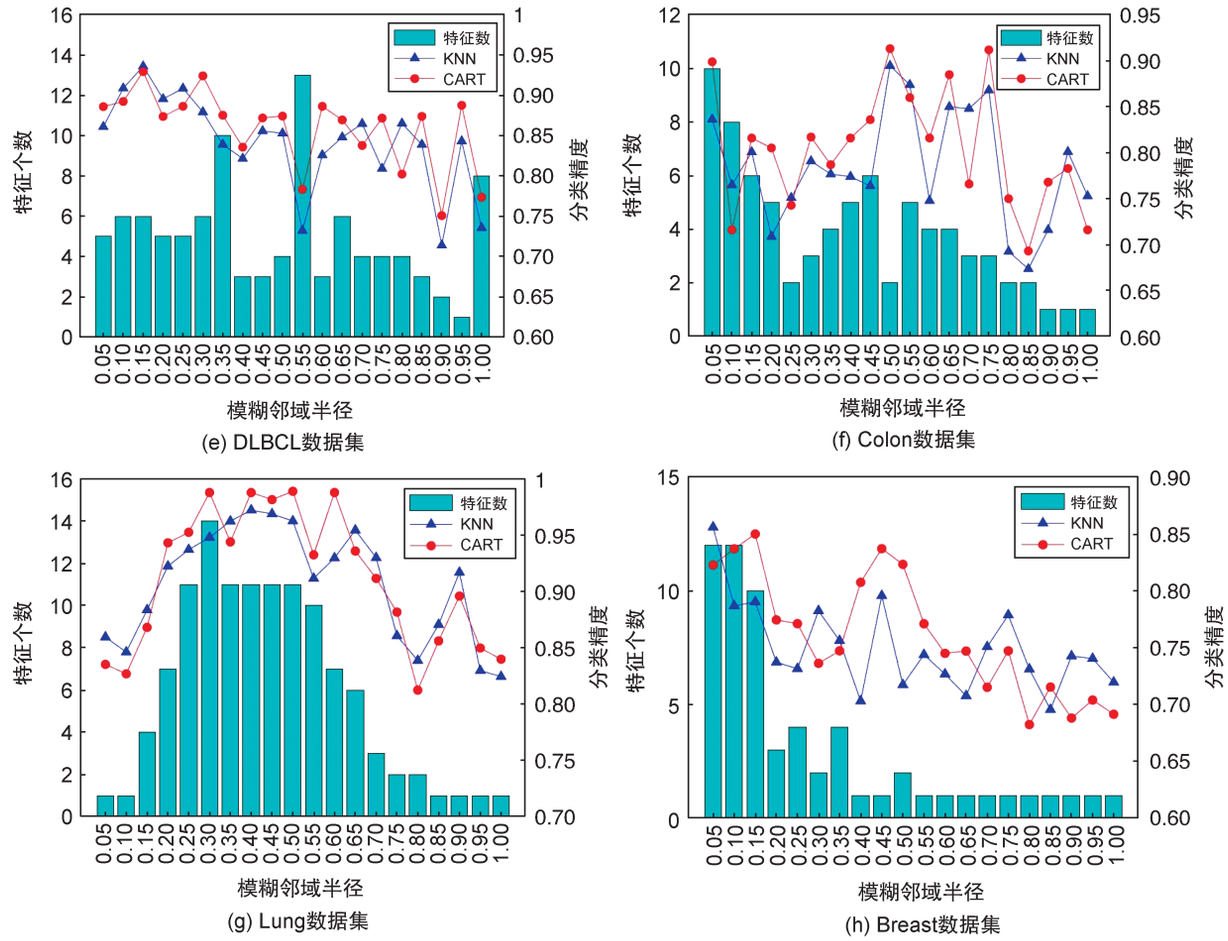



**图3** 不同模糊邻域半径下4个高维数据集的约简特征子集大小和分类精度

**Fig. 3** The feature reduction subset size and classification accuracy of four high- dimensional datasets with different fuzzy neighborhood radii

**表4** 使用KNN分类器时,7种算法在4个低维数据集上的分类精度对比

**Table 4** Comparison of classification accuracy of seven algorithms on four low-dimensional datasets when using KNN classifier

| 算法 | Wine | WPBC | Heart | WDBC | Mean |
|---|---|---|---|---|---|
| IFPR | 0.942 9 | 0.684 0 | 0.788 9 | 0.931 1 | 0.836 7 |
| NSI | 0.936 1 | 0.767 0 | **0.814 0** | 0.956 0 | 0.868 3 |
| PDJE-AR | 0.949 4 | 0.757 6 | 0.803 7 | 0.956 1 | 0.866 7 |
| FSNTDJE | 0.906 0 | 0.737 0 | 0.774 0 | 0.947 0 | 0.841 0 |
| FNRDCI | 0.952 3 | 0.756 7 | 0.810 6 | 0.933 7 | 0.863 3 |
| FSRMI | 0.951 0 | 0.759 0 | 0.804 1 | 0.949 0 | 0.865 8 |
| FS-SIFNCE | **0.984 9** | **0.808 4** | 0.795 1 | **0.973 1** | **0.890 4** |

注:最后一列Mean为每个算法在4个低维数据集上的分类精度的平均值。

**表5** 使用CART分类器时,7种算法在4个低维数据集上的分类精度对比

**Table 5** Comparison of classification accuracy of seven algorithms on four low-dimensional datasets when using CART classifier

| 算法 | Wine | WPBC | Heart | WDBC | Mean |
|---|---|---|---|---|---|
| IFPR | 0.872 7 | 0.665 8 | 0.748 1 | 0.929 6 | 0.804 1 |
| NSI | 0.937 0 | 0.771 0 | 0.778 1 | 0.912 3 | 0.849 6 |
| PDJE-AR | 0.910 1 | 0.707 1 | 0.800 0 | 0.940 2 | 0.839 4 |
| FSNTDJE | 0.803 0 | 0.687 0 | 0.765 0 | 0.912 0 | 0.791 8 |
| FNRDCI | 0.925 2 | 0.708 0 | 0.826 6 | 0.942 0 | 0.850 5 |
| FSRMI | 0.919 0 | 0.736 3 | **0.833 0** | 0.945 0 | 0.858 3 |
| FS-SIFNCE | **0.972 2** | **0.805 0** | **0.833 0** | **0.965 5** | **0.893 9** |

KNN分类器中显示出最佳的分类性能,相对于对比算法,其精度高出5.4%至18.0%。根据表2和表5中的CART分类器数据,得知FS-SIFNCE算法在CART分类器下,所选数据集的分类精度均超过了对比算法。具体而言,平均精度高出其他对比算法4.1%至12.9%。此外,在Wine数据集上,FSRMI算法在CART分类器中显示出最佳的分类性能,并且其精度相对于对比算法高出3.8%至21%。这也表明FS-SIF-

NCE算法所选择的特征与决策之间存在着强烈的相关性。总之,FS-SIFNCE算法在所选低维数据上具有较好的分类能力和较强的约简能力。

### 4.5 高维数据集实验结果分析

为评估FS-SIFNCE算法在高维数据上的有效性,研究对比了4种不同的特征选择算法(FSNTDJE[26]、IFPR[25]、FSRMI[22]、FNRDCI[17])在4个高维数据集(DLBCL、Lung、Colon、Breast)的分类结果。具体地,表6展示了5种算法在4个高维数据集上所选特征的数量。而表7和表8分别展示了这4种算法在KNN和CART分类器下的分类精度。其中所选特征数和分类精度均采用十折交叉验证得到。需要注意的是,加粗的内容代表了对比中的最佳结果。

**表6** 5种算法在4个高维数据集上选择特征的个数

**Table 6** The five algorithms select the number of features on four high-dimensional datasets

| 算法 | DLBCL | Lung | Colon | Breast | Mean |
|---|---|---|---|---|---|
| FSNTDJE | 6.3 | 9.0 | 11.5 | **7.0** | 8.5 |
| IFPR | 12.4 | 8.6 | 6.0 | 9.5 | 9.1 |
| FSRMI | 7.3 | 8.8 | 8.5 | 9.2 | 8.5 |
| FNRDCI | **4.2** | 8.0 | 8.1 | 13.6 | 8.5 |
| FS-SIFNCE | 6.0 | **7.0** | **2.3** | 17.0 | **8.1** |

**表7** 使用KNN分类器时,5种算法在4个高维数据集上的分类精度对比

**Table 7** Comparison of the classification accuracy of five algorithms on four high-dimensional datasets when using the KNN classifier

| 算法 | DLBCL | Lung | Colon | Breast | Mean |
|---|---|---|---|---|---|
| FSNTDJE | 0.805 0 | 0.970 3 | 0.774 0 | 0.714 0 | 0.815 8 |
| IFPR | 0.948 0 | 0.968 5 | 0.757 4 | 0.783 3 | 0.864 3 |
| FSRMI | **0.954 0** | 0.960 1 | 0.873 0 | 0.813 0 | 0.900 0 |
| FNRDCI | 0.930 7 | **0.980 9** | 0.872 1 | 0.845 0 | 0.907 2 |
| FS-SIFNCE | 0.936 1 | 0.972 2 | **0.894 7** | **0.856 0** | **0.914 8** |

根据表6可知,算法FSNTDJE和FNRDCI分别在数据集Breast和DLBCL上选择了最少特征数分别为7.0和4.2,算法FS-SIFNCE分别在数据集Lung和Colon均取得了最小的特征数分别为7.0和2.3。整体而言,算法FS-SIFNCE在平均选取特征数量上低于其余算法,表明所提算法可有效的消除冗余特征。

根据表7和表8可知在Colon数据集上,

**表8** 使用CART分类器时,5种算法下4个高维数据集上的分类精度对比

**Table 8** Comparison of the classification accuracy of five algorithms on four high-dimensional datasets when using the KNN classifier

| 算法 | DLBCL | Lung | Colon | Breast | Mean |
|---|---|---|---|---|---|
| FSNTDJE | 0.805 0 | 0.962 9 | 0.750 0 | 0.691 0 | 0.802 2 |
| IFPR | 0.870 0 | 0.934 1 | 0.763 2 | 0.731 0 | 0.824 6 |
| FSRMI | 0.889 0 | 0.953 2 | 0.883 0 | **0.865 0** | 0.897 6 |
| FNRDCI | 0.902 3 | 0.964 9 | 0.848 0 | 0.801 7 | 0.879 2 |
| FS-SIFNCE | **0.929 5** | **0.988 9** | **0.913 3** | 0.850 0 | **0.920 4** |

FS-SIFNCE算法在KNN和CART两个分类器中显示出最佳的分类性能,精度分别为89.47%和91.33%,分别高于对比算法2.5%~18.1%和3.4%~21.8%,且选取特征数量最小。根据表6—7和表8,可知虽FNRDCI算法在DLBCL数据集上具有良好的约简能力,但在两个分类器下的平均精度均低于FS-SIFNCE算法。所以FNRDCI算法并不能像FS-SIFNCE算法一样,保持强约简能力的同时具有高分类能力。此外,根据表7和表8可知算法FS-SIFNCE在KNN和CART分类器下的平均分类精度均最高,分别为91.48%和92.04%。这也揭示了FS-SIFNCE算法可以提高高维数据的分类结果。总之,FS-SIFNCE算法的平均所选特征数量最小,且平均分类精度最高,表明所提算法在减少冗余特征的同时能有效地提高分类精度。

## 5 结论

针对在模糊邻域粗糙集中的重要特征评价函数依赖度仅利用了下近似中的信息,忽略了上近似和边界域中存在的有效信息这一问题,本文提出了一种基于自信息和模糊邻域条件熵的特征选择方法。首先用自信息这一度量工具,定义了相近自信息,使得上近似和边界域中的有效信息得以利用。其次,利用条件熵的概念引入到模糊邻域中提出模糊邻域条件熵,其可通过特征子集的变化选择出对决策属性不确定性最小的特征子集,并针对大多特征评价函数仅从代数视角或信息视角构建的问题,将相近自信息与模糊邻域条件熵相结合,在双视角下提出了更为全面的特征评价函数,用于衡量特征子集分类信息的不确定性,并基于此利用

了最大相关最小冗余技术设计特征选择算法。最后通过实验及分析验证所提 FS-SIFNCE 算法相对于其他算法提高了分类精度和特征选择的效果。今后的工作将主要集中在进一步优化模糊邻域半径的设置,以提高算法的时间效率。

**参考文献:**


[1] SUN L, WANG L Y, DING W P, *et al*. Neighborhood Multi-granulation Rough Sets-based Attribute Reduction Using Lebesgue and Entropy Measures in Incomplete Neighborhood Decision Systems[J]. *Knowl Based Syst*, 2020, **192**: 105373. DOI: 10.1016/j.knosys.2019.105373.

[2] XIA S Y, WU S L, CHEN X X, *et al*. GRRS: Accurate and Efficient Neighborhood Rough Set for Feature Selection[J]. *IEEE Trans Knowl Data Eng*, 2023, **35**(9): 9281−9294. DOI: 10.1109/TKDE.2022.3222447.

[3] WANG C Z, HUANG Y, SHAO M W, *et al*. Feature Selection Based on Neighborhood Self-information[J]. *IEEE Trans Cybern*, 2020, **50**(9): 4031−4042. DOI: 10.1109/TCYB.2019.2923430.

[4] HUANG Z H, LI J J. Multi-level Granularity Entropies for Fuzzy Coverings and Feature Subset Selection[J]. *Artif Intell Rev*, 2023, **56**(10): 12171−12200. DOI: 10.1007/s10462-023-10479-3.

[5] 孙敏杰. 基于模糊邻域熵的特征基因选择方法研究[D]. 广州: 广州大学, 2023.
SUN M J. Research on Feature Gene Selection Method Based on Fuzzy Neighborhood Entropy[D]. Guangzhou: Guangzhou University, 2023.

[6] PAWLAK Z. Rough Sets[J]. *Int J Comput Inf Sci*, 1982, **11**(5): 341−356. DOI: 10.1007/bf01001956.

[7] YUAN Z, CHEN H M, ZHANG P F, *et al*. A Novel Unsupervised Approach to Heterogeneous Feature Selection Based on Fuzzy Mutual Information[J]. *IEEE Trans Fuzzy Syst*, 2022, **30**(9): 3395−3409. DOI: 10.1109/TFUZZ.2021.3114734.

[8] 冉虹, 候婷, 贺龙雨, 等. 基于模糊邻域系统的模糊粗糙集模型[J]. 计算机科学, 2022, **49**(S2): 330−334. DOI: 0.11896/jsjkx.211100224.
RAN H, HOU T, HE L Y, *et al*. Fuzzy Rough Set Model Based on Fuzzy Neighborhood System[J]. *Comput Sci*, 2022, **49**(S2): 330−334. DOI: 0.11896/jsjkx.211100224.

[9] SHI Z Q, XIE S R, LI L Q. Generalized Fuzzy Neighborhood System-based Multigranulation Variable Precision Fuzzy Rough Sets with Double TOPSIS Method to MADM[J]. *Inf Sci*, 2023, **643**: 119251. DOI: 10.1016/j.ins.2023.119251.

[10] GOU H Y, ZHANG X Y. Feature Selection Based on Double-hierarchical and Multiplication-optimal Fusion Measurement in Fuzzy Neighborhood Rough Sets[J]. *Inf Sci*, 2022, **618**: 434−467. DOI: 10.1016/j.ins.2022.10.133.

[11] 庞婧, 姚炳学, 李令强. 基于模糊邻域系的a-粗糙集及其应用[J]. 模糊系统与数学, 2022, **36**(5): 158−165.
PANG J, YAO B X, LI L Q. The A-rough Set Based on Fuzzifying Neighborhood Systems and Its Applications [J]. *Fuzzy Syst Math*, 2022, **36**(5): 158−165.

[12] HU Q H, YU D R, LIU J F, *et al*. Neighborhood Rough Set Based Heterogeneous Feature Subset Selection[J]. *Inf Sci*, 2008, **178**(18): 3577−3594. DOI: 10.1016/j.ins.2008.05.024.

[13] DUBOIS D, PRADE H. Rough Fuzzy Sets and Fuzzy Rough Sets[J]. *Int J Gen Syst*, 1990, **17**(2/3): 191−209. DOI: 10.1080/03081079008935107.

[14] QIAN Y H, LI S Y, LIANG J Y, *et al*. Pessimistic Rough Set Based Decisions: A Multigranulation Fusion Strategy[J]. *Inf Sci*, 2014, **264**: 196−210. DOI: 10.1016/j.ins.2013.12.014.

[15] WANG C Z, SHAO M W, HE Q, *et al*. Feature Subset Selection Based on Fuzzy Neighborhood Rough Sets [J]. *Knowl Based Syst*, 2016, **111**: 173−179. DOI: 10.1016/j.knosys.2016.08.009.

[16] CHEN Y Y, CHEN Y M. Feature Subset Selection Based on Variable Precision Neighborhood Rough Sets [J]. *Int J Comput Intell Syst*, 2021, **14**(1): 572. DOI: 10.2991/ijcis.d.210106.003.

[17] XU J C, MENG X R, QU K L, *et al*. Feature Selection Using Relative Dependency Complement Mutual Information in Fitting Fuzzy Rough Set Model[J]. *Appl Intell*, 2023, **53**(15): 18239−18262. DOI: 10.1007/s10489-022-04445-9.

[18] ZHOU P, HU X G, LI P P, *et al*. Online Feature Selection for High-dimensional Class-imbalanced Data[J]. *Knowl Based Syst*, 2017, **136**: 187−199. DOI: 10.1016/j.knosys.2017.09.006.

[19] 马磊, 罗川, 李天瑞, 等. 基于模糊粗糙集的无监督动态特征选择算法[J]. 计算机应用, 2023, **43**(10): 3121−3128. DOI: 10.11772/j.issn.1001-9081.2022101543.
MA L, LUO C, LI T R, *et al*. Fuzzy-rough Set Based Unsupervised Dynamic Feature Selection Algorithm[J]. *J Comput Appl*, 2023, **43**(10): 3121−3128. DOI: 10.11772/j.issn.1001-9081.2022101543.

[20] SHU W H, LI S H, QIAN W B. A Composite Entropy-based Uncertainty Measure Guided Attribute Reduction for Imbalanced Mixed-type Data[J]. *J Intell Fuzzy Syst*, 2024, **46**(3): 7307-7325. DOI: 10.3233/JIFS-237211.

[21] Shannon C E. A Mathematical Theory of Communication[J]. *Bell Syst Tech J*, 1948, **27**(4): 623−656. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

[22] QU K L, XU J C, HAN Z Q, *et al*. Maximum Rel-

evance Minimum Redundancy-based Feature Selection Using Rough Mutual Information in Adaptive Neighborhood Rough Sets[J]. *Appl Intell*, 2023, **53**(14): 17727−17746. DOI: 10.1007/s10489-022-04398-z.

[23] 苗夺谦, 胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展, 1999, **36**(6): 681−684.
MIAO D Q, HU G R. A Heuristic Algorithm for Reduction of Knowledge[J]. *J Comput Res Dev*, 1999, **36**(6): 681−684.

[24] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, **25**(7): 759−766. DOI: 10.3321/j.issn: 0254-4164.2002.07.013.
WANG G Y, YU H, YANG D C. Decision Table Reduction Based on Conditional Information Entropy[J]. *Chin J Comput*, 2002, **25**(7): 759−766. DOI: 10.3321/j.issn: 0254-4164.2002.07.013.

[25] TAN A H, WU W Z, QIAN Y H, *et al*. Intuitionistic Fuzzy Rough Set-based Granular Structures and Attribute Subset Selection[J]. *IEEE Trans Fuzzy Syst*, 2019, **27**(3): 527−539. DOI: 10.1109/TFUZZ.2018.2862870.

[26] SUN L, WANG L Y, QIAN Y H, *et al*. Feature Selection Using Lebesgue and Entropy Measures for Incomplete Neighborhood Decision Systems[J]. *Knowl Based Syst*, 2019, **186**: 104942. DOI: 10.1016/j.knosys.2019.104942.