

基于差别矩阵和 mRMR 的分步优化特征选择算法

樊鑫¹ 陈红梅²

1 西南交通大学信息科学与技术学院 成都 611756

2 西南交通大学云计算与智能技术高校重点实验室 成都 611756

(fanxinchn@foxmail.com)



摘要 分类问题普遍存在于现代工业生产中。在进行分类任务之前,利用特征选择筛选有用的信息,能够有效地提高分类效率和分类精度。最小冗余最大相关算法(mRMR)考虑最大化特征与类别的相关性和最小化特征之间的冗余性,能够有效地选择特征子集;但该算法存在中后期特征重要度偏差大以及无法直接给出特征子集的问题。针对该问题,文中提出了结合邻域粗糙集差别矩阵和 mRMR 原理的特征选择算法。根据最大相关性和最小冗余性原则,利用邻域熵和邻域互信息定义了特征的重要度,以更好地处理混合数据类型。基于差别矩阵定义了动态差别集,利用差别集的动态演化有效去除冗余属性,缩小搜索范围,优化特征子集,并根据差别矩阵判定迭代截止条件。实验选取 SVM,J48,KNN 和 MLP 作为分类器来评价该特征选择算法的性能。在公共数据集上的实验结果表明,与已有算法相比,所提算法的平均分类精度提升了 2% 左右,同时在特征较多的数据集上能够有效地缩短特征选择时间。所提算法继承了差别矩阵和 mRMR 的优点,能够有效地处理特征选择问题。

关键词: 特征选择;邻域粗糙集;差别矩阵;mRMR

中图分类号 TP301.6

Stepwise Optimized Feature Selection Algorithm Based on Discernibility Matrix and mRMR

FAN Xin¹ and CHEN Hong-mei²

1 School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China

2 Key Laboratory of Cloud Computing and Intelligent Technology, Southwest Jiaotong University, Chengdu 611756, China

Abstract Classification is a common problem in modern industrial production. The classification efficiency and classification accuracy can be improved effectively by using feature selection to filter useful information before classifying tasks. Considering the maximum correlation between features and class and the minimum redundancy among features, the minimal-redundancy-maximal-relevance(mRMR) algorithm can effectively select feature subset. However, two problems exist in the algorithm mRMR, i. e., the relatively large deviation of the importance of the features in the middle and later stages of mRMR, and the feature subset is not given directly. A novel algorithm based on the principle of mRMR and discernibility matrix of neighborhood rough set was proposed, to solve these problems. The significance of the feature is defined by employing neighborhood entropy and neighborhood mutual entropy based on the principle of the minimal redundancy and maximal relevance, which can deal the mixed type date better. Dynamic discernibility set is defined based on the discernibility matrix. The dynamic evolution of the discernibility set is utilized as the policy to delete redundant features and narrows search range. The optimized feature subset is given when the iteration is stop by the stop condition given by discernibility matrix. In this paper, SVM, J48, KNN and MLP were selected as classifiers to evaluate the performance of the feature selection algorithm. The experimental results on the public datasets show that the average classification accuracy of the proposed algorithm is about 2% more than that of previous algorithm, and the proposed algorithm can effectively shorten the feature selection time on the data set with more features. Therefore, the proposed algorithm inherits the advantages of discernibility matrix and MRMR, and can effectively deal with feature selection problems.

Keywords Feature selection, Neighborhood rough set, Discernibility matrix, mRMR

1 引言

特征选择的目标是从原始数据中选择一组能够描述原始

数据的特征子集,同时移除不相关特征和噪声特征,进而提高分类预测的效果^[1]。随着信息技术的发展,生产、生活各个领域采集的数据维度迅速增加,从多维数据中发掘有用的信

收稿日期:2018-12-14 返修日期:2019-04-29 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61572406)

This work was supported by the National Natural Science Foundation of China (61572406).

通信作者:陈红梅(hmchen@swjtu.edu.cn)

息变得尤为重要。特征选择已被应用于许多领域,如基因序列分析、文本分析、图像处理等^[2-4]。标准的基因表达数据可能有成千上万个特征,而其中许多特征可能与其他特征相关。此时,相互关联的特征对分类目标不提供额外的信息。

通常,特征选择算法可以分为过滤式(Filter)、包裹式(Wrapper)和嵌入式(Embedded)^[5-7]。过滤式算法根据数据本身的信息评价特征的优劣,因此其独立于分类器。然而,包裹式算法依靠具体分类器的分类效果评价特征子集。一般情况下,包裹式算法的最终分类效果优于过滤式算法,但同时也存在计算量相对较大、特征子集过拟合等问题。此外,嵌入式算法通常与特定的分类算法结合,在分类的同时选择特征,因此缺乏通用性。鉴于过滤式算法具有通用性和计算量相对较小的特点,本文主要研究过滤式算法。

过滤式方法通过特征重要度排序的方法来进行特征选择。学者们提出了不同的特征重要度量方法,如距离度量、依赖性度量、一致性度量、信息量度量以及分类错误率度量等。信息论中的互信息能够较为准确地评价两个特征之间的相互依赖性,基于信息论的过滤式特征选择方法得到了很好的应用。最小冗余最大相关算法是由 Peng 等提出的一种重要的基于信息论的过滤式特征选择算法^[8],并已成功应用于多个领域。学者们对 mRMR 算法进行了研究和改进。Estevez 等提出了将度量特征间的相关性归一化的归一化互信息特征选择算法^[9]。Sun 等结合类别相关度,将 mRMR 扩展到了多标签学习,同时为每个特征分配不同的权重,借助凸优化来优化特征权重,从而获得特征排序^[10]。Wang 等提出了独立分类信息,该分类信息能够表达候选特征提供的独有分类信息^[11]。

Pawlak 提出的粗糙集理论是一种刻画不完整性和不确定性数据的数学工具^[12]。特征选择(属性约简)是粗糙集理论的重要应用之一,如基于正域的属性约简、基于差别矩阵的属性约简、结合信息论的属性约简等^[13-15]。针对数值数据的分析处理,Dubois 等提出了模糊粗糙集^[16],胡清华等提出了基于邻域关系的邻域粗糙集^[17]。为有效获得所有约简和最小约简,Skowron 提出了差别矩阵(又称为辨识矩阵或区分矩阵)^[18]。基于差别矩阵的特征选择是粗糙集特征选择的重要方法之一,通过差别矩阵可以获得所有的约简,但约简的计算是一个 NP 问题。因此,学者们提出了多种基于差别矩阵的启发式搜索方法^[19-20]。针对差别矩阵的空间复杂度问题,Yang 等提出了压缩树的概念。压缩树使用树的一条路径代表差别矩阵的元素,从而减小元素的存储空间^[21]。

在结合信息论和粗糙集理论来提高特征选择的性能方面,学者们也做了相应的研究工作。Hu 等提出了基于邻域熵的最小冗余最大依赖算法^[22]。Lin 等在邻域熵的基础上,将邻域熵推广到了多标签学习中^[23]。

本文基于邻域粗糙集的邻域熵和邻域互信息,定义了属性重要度和动态差别集,由差别集动态演化逐步去除冗余属性,缩小候选特征子集的搜索范围,从而给出基于差别矩阵的分步优化特征选择算法,并通过实验证明了本文方法的有效性。

2 相关理论

本节对最大相关最小冗余算法、经典粗糙集及邻域粗糙

集的基本概念进行了介绍。

2.1 mRMR 方法

从信息论的角度看,特征选择的目标就是选择特征子集 S ,使得 S 与目标属性 d 之间具有最大依赖度(Max-Dependency),即互信息 $MI(S;d)$ 最大。离散随机变量 X 和 Y 的互信息 $MI(X;Y)$ 定义为:

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

实际应用中,基于最大依赖度的特征选择实现较困难,因为随着特征的增加,正确地估计概率(概率密度)较难,从而导致互信息不能被正确地估计。为此,Peng 等提出了基于最大相关最小冗余算法来近似计算最大依赖度^[8]。mRMR 使用特征子集 S 中的所有特征 f_i 与决策属性 d 的互信息的平均值来代替最大依赖度:

$$\max D(S,d), D(S,d) = \frac{1}{|S|} \sum_{f_i \in S} MI(f_i,d) \quad (2)$$

通过最大相关度准则选择出来的特征很有可能具有较多的冗余特征,因此在最大相关度准则的基础上加入如下最小冗余准则:

$$\min R(S), R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} MI(f_i, f_j) \quad (3)$$

mRMR 结合了以上两种约束,Peng 等定义了同时考虑 D 和 R 的算子:

$$\max \Phi(D,R), \Phi(D,R) = D - R \quad (4)$$

若现在已经选出了 $m-1$ 个特征,则按式(5)从剩下的特征集合 $X - S_{m-1}$ 中选择第 m 个特征加入特征子集 S 。

$$\max_{f_j \in X - S_{m-1}} [MI(f_j;d) - \frac{1}{|S_{m-1}|} \sum_{f_i \in S_{m-1}} MI(f_j; f_i)] \quad (5)$$

虽然 mRMR 在很多数据上能够取得不错的效果,但是其只考虑了冗余在数量上的大小,忽视了有些特征可能拥有的少量独有信息。因此,Peng 等在 mRMR 算法后还运用了包裹式的前向选择或反向排除的方法来提升特征子集的质量。除此之外,mRMR 方法还有可能过早地选入不相关的特征,而过晚地加入某些有用的特征。例如,当某一候选特征与已选特征具有较高的冗余度,但是它却对一些对象具有独有的区分能力时,mRMR 将会低估该属性的重要性。同时,某一属性与已选属性集提供的信息完全冗余,但因其相关度较高,mRMR 仍旧会将其选入特征子集。如已知 a_i 和 a_j 两个属性表达完全一致的信息,并且与 d 的相关度很高,在搜索进程的前期就选择了 a_i ;在算法搜索的中后期,因为式(5)中表达冗余的分量的权重会随着选择属性数量 $m-1$ 的增加而减小,最终 a_j 的属性重要度会逐渐变大,最终导致错误地选择了 a_j 。本文将借助粗糙集理论的差别矩阵来改善以上问题。

2.2 经典粗糙集理论

在经典粗糙集模型中^[12],四元组 $DIS = \langle U, CUD, V, f \rangle$ 为决策信息系统,其中 $U = \cup \{x_i \mid (1 \leq i \leq n)\}$ 为非空有限论域, C 为条件属性集(特征集), D 为决策属性集, V 为属性值域, $f: U \times A \rightarrow V$ 称为信息函数,它为每个对象的每个属性赋予一个信息值。为简化处理,本文中 $D = \{d\}$ 。 $f(x,a)$ 表示对象 x 为属性 a 下的属性值。 $\forall E \subseteq C$,对象间通过属性子集 E 上的属性取值形成对论域 U 的划分,即 U/E 。由属性子集诱导论域 U 上的等价关系 R 。与对象 x_i 在属性子集 E 上

具有相同属性值的对象集构成 x_i 的等价类 $[x_i]_E$, 这些等价类被称为信息粒或初等概念。论域 U 和等价关系 R 构成的二元组 (U, R) 被称为近似空间, 通过一对初等概念可以近似描述论域中的任意子集。给定任意子集 $X, B \subseteq C, X$ 可以通过由 B 诱导的等价类来近似描述, 即通过上近似集 \overline{BX} 和下近似集 \underline{BX} 来近似刻画。下近似集和上近似集的定义如下:

$$\underline{BX} = \{[x]_B \mid [x]_B \subseteq X, x \in U, [x]_B \in U/B\} \quad (6)$$

$$\overline{BX} = \{[x]_B \mid [x]_B \cap X \neq \emptyset, x \in U, [x]_B \in U/B\} \quad (7)$$

下近似集是包含在 X 内的信息粒的并集, 上近似集是与 X 相交不为空的信息粒的并集, 两者的差集称为边界。

$$BND_B(X) = \overline{BX} - \underline{BX} \quad (8)$$

对于属性集合 $B \subseteq C$, 其相对于决策属性集 D 的正域的定义如下:

$$POS_B(D) = \bigcup_{X \in U/D^-} \underline{BX} \quad (9)$$

2.3 邻域粗糙集理论

Pawlak 经典粗糙集模型可以直接处理符号型数据, 但不能直接处理数值型数据。为了弥补这一缺点, Hu 等提出了邻域粗糙集模型^[17], 将经典粗糙集中的等价关系推广到邻域关系。

定义 1(邻域关系) 已知 $DIS = \langle U, C \cup \{d\}, V, f \rangle$ 为决策信息系统, 属性子集 $B \subseteq C$, 则属性子集 B 下的邻域关系定义为:

$$N_B = \{(x_i, x_j) \in U \times U \mid \Delta^B(x_i, x_j) \leq \delta, 0 \leq \delta < 1\} \quad (10)$$

其中, Δ^B 表示对象间在属性子集 B 上的距离, 它可以是任意距离函数, 如欧氏距离、1 范数和无穷范数等。

定义 2(邻域) 已知 $DIS = \langle U, C \cup \{d\}, V, f \rangle$ 为决策信息系统, 给定任意样本 $\forall x_i \in U$, 属性子集 $B \subseteq C, N_B$ 为论域 U 下的邻域关系, x_i 的邻域表示为:

$$\delta_B(x_i) = \{x_j \mid x_j \in U, (x_i, x_j) \in N_B\} \quad (11)$$

定义 3(邻域关系 N_B 下的近似集) 对于任意对象的子集 $X \subseteq U, B \subseteq C$, 基于邻域关系 N_B 的上下近似集定义为:

$$\underline{N_B}X = \{x_i \mid x_i \in U, \delta_B(x_i) \subseteq X\} \quad (12)$$

$$\overline{N_B}X = \{x_i \mid x_i \in U, \delta_B(x_i) \cap X \neq \emptyset\} \quad (13)$$

定义 4(邻域熵和邻域互信息)^[24] 已知 $DIS = \langle U, C \cup \{d\}, V, f \rangle$ 为决策信息系统, 属性子集 $S \subseteq C, E \subseteq C, S \cap E = \emptyset$, $\delta_S(x_i)$, $\delta_E(x_i)$ 和 $\delta_{S \cup E}(x_i)$ 分别为 S, E 和 $S \cup E$ 诱导的邻域关系下 x_i 的邻域, 则邻域熵 $NH_\delta(S)$ 和邻域互信息 $NMI_\delta(E; S)$ 的定义如下:

$$NH_\delta(S) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|\delta_S(x_i)|}{|U|} \quad (14)$$

$$NMI_\delta(E; S) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|\delta_E(x_i)| |\delta_S(x_i)|}{|U| |\delta_{S \cup E}(x_i)|} \quad (15)$$

目前基于粗糙集的特征选择算法主要可以分为两类:

1) 基于正域的算法; 2) 基于差别矩阵的算法。本文主要应用差别矩阵进行特征选择。

2.4 差别矩阵

给定一个信息系统 $DIS = \langle U, C \cup \{d\}, V, f \rangle$, 其差别矩阵 \mathbf{M} 是一个 $|U| \times |U|$ 的矩阵, 矩阵 (i, j) 位置上的元素 $m_{i,j}$ 的定义如下:

$$m_{i,j} = \begin{cases} \{a \in C \mid f(x, a) \neq f(y, a)\}, & f(x, d) \neq f(y, d) \\ \emptyset, & \text{其他} \end{cases} \quad (16)$$

差别矩阵对应的区分函数定义为:

$$\Delta^* = \bigwedge_{i=1}^{|U|} \bigwedge_{j=1}^{|U|} \left(\bigvee_{k=1}^{|m_{ij}|} a_k, a_k \in m_{ij} \right) \quad (17)$$

区分函数的极小析取范式中所含的所有合取式是属性集 C 的所有约简。

对于基于邻域粗糙集的差别矩阵 \mathbf{M} , 其 (i, j) 位置上的元素 $m_{i,j}$ 的定义如下:

$$m_{ij} = \begin{cases} \{a \in C \mid \Delta^{[a]}(x_i, x_j) > \delta\}, & f(x, d) \neq f(y, d) \\ \emptyset, & \text{其他} \end{cases} \quad (18)$$

定义 5(约简) 给定一个决策表 $DIS = \langle U, C \cup \{d\}, V, f \rangle$, 其差别矩阵为 \mathbf{M} , $\forall m_{i,j} \in \mathbf{M}, m_{i,j} \neq \emptyset, \exists R \in C, R \cap m_{i,j} \neq \emptyset$, 那么 R 是 C 关于 d 的约简。

3 算法思路及流程

基于差别矩阵和 mRMR 准则, 我们给出了改进的特征选择算法。

3.1 搜索策略

启发式的搜索方法是重要的特征选择算法之一, 它采用前向搜索或者反向排除的策略等启发式的搜索方法获取特征子集。但是, 我们发现启发式策略指导下, 候选集合内可能存在一些不必要的特征, 这些特征不仅影响搜索结果的质量, 还会降低搜索的速度。而差别矩阵具有较好分辨不必要属性的能力, 这种能力能够有效地辅助搜索算法找到更优的特征子集。为了有效地处理混合数据, 我们将采用基于邻域粗糙集模型的差别矩阵。为了描述启发式搜索过程中记录已选择属性的临时集合, 我们将其定义为残缺约简。

定义 6(残缺约简) $Red(C)$ 为 C 的所有约简集合, 对于 $\forall A \subseteq C$, 如果 $\exists R \in Red(C)$ 使得 $A \subseteq R$, 则称 A 为 C 关于 d 的残缺约简。

随着启发式搜索过程的推进, 残缺约简中的属性逐步增加, 显然其包含的区别信息也在不断增加, 同时整个训练集中需要进一步区分的对象也不断减少。对于这些需要进一步区分的对象, 或许能够通过增加选择的属性来进行区分。借助差别矩阵, 我们能够通过集合直观地将这些具有区分能力的属性和它们能够区分的对象呈现出来。因此, 我们更进一步地定义了动态差别集。

定义 7(动态差别集) 已知 $DIS = \langle U, C \cup \{d\}, V, f \rangle$ 为决策信息系统, 其差别矩阵为 \mathbf{M} , $\forall A \subseteq C, DIS$ 关于 A 的动态差别集的定义如下:

$$DDS_A = \{m_{ij} \mid A \cap m_{ij} = \emptyset, m_{ij} \in \mathbf{M}\} \quad (19)$$

定理 1 已知 $t-1$ 时刻选择出的特征子集 $A \subseteq C$ 及 A 的动态差别集 DDS_A , t 时刻选择的特征为 a_t , 则 t 时刻的候选子集 $DDS_{A+\{a_t\}} = \{m_{ij} \mid a_t \notin m_{ij}, m_{ij} \in DDS_A\}$ 。

证明: 根据动态差别集的定义, $DDS_{A+\{a_t\}}$ 的元素满足 $(A + \{a_t\}) \cap m_{ij} = \emptyset$ 且 $m_{ij} \in \mathbf{M}, (A + \{a_t\}) \cap m_{ij} = \emptyset$ 等价于 $A \cap m_{ij} = \emptyset$ 且 $\{a_t\} \cap m_{ij} = \emptyset$ 。即 $DDS_{A+\{a_t\}}$ 的元素满足 $a_t \notin m_{ij}, m_{ij} \in DDS_A$ 。

定义 8(动态候选集) 已知 $DIS = \langle U, CU\{d\}, V, f \rangle$ 为决策信息系统, 其差别矩阵为 \mathbf{M} , $\forall A \subseteq C$, DIS 关于 A 的动态候选集 DCS 的定义如下:

$$DCS_A = \cup m_{ij} (m_{ij} \in DDS_A) \quad (20)$$

定理 2 对于任意的残缺约简 A , $\forall a_j \in C - A$ 且 $a_j \notin DCS_A$, a_j 是 C 关于 d 的残缺约简 A 的冗余属性。

证明: 因为 $DDS_A = \{m_{ij} | A \cap m_{ij} = \emptyset, m_{ij} \in \mathbf{M}\}$, $DCS_A = \cup m(m \in DDS_A)$, 所以 $DCS_A \cap m_{ij} \neq \emptyset (A \cap m_{ij} = \emptyset)$, $(A \cup DCS_A) \cap m_{ij} \neq \emptyset$. 根据定义 5, $A \cup DCS_A$ 是 C 关于 d 的约简. 因此, $\forall a_j \in C - A$ 且 $a_j \notin DCS_A$, a_j 是 C 关于 d 的残缺约简 A 的冗余属性。

通过定理和推论, 我们发现在增量选择的搜索方法 (first-order) 中每选择一个属性, 更新动态差别集, 并计算动态候选集, 就可以优化下一轮特征选择过程的候选特征集合. 确定候选集后, 再通过传统的属性重要度方法来最终选定本轮的最佳属性. 通过这样的方式, 可以将 mRMR 这类增量选择的属性搜索方法和差别矩阵相结合, 互相取长补短. 根据粗糙集理论, 新得到的候选特征集合的每个特征均具有区分某些样本对象的能力, 并且这些能力是已选择特征所不具有的. 因此, 通过计算动态差别集, 不但可以保证每次选择特征具有新的识别信息, 而且可以减少新一轮搜索特征的范围, 从而提高特征选择算法的效率. 显然, 当动态差别集内的元素为空时, 不存在候选子集, 因此不再进一步搜索特征, 输出已选特征子集。

3.2 属性重要度

最大相关性和最小冗余性原则能够有效地应用于属性重要度的排序. 在实际应用中, 数据往往是既有符号属性又有数值属性, 而传统的互信息只能在符号型属性上计算, 邻域熵和邻域互信息可以有效地定义混合数据的相关性. 结合邻域熵和邻域互信息的属性重要度的定义如下。

定义 9(基于邻域互信息的属性重要度) 已知 $DIS = \langle U, CU\{d\}, V, f \rangle$ 为决策信息系统, $a_i \in A, A \subseteq C$, 其中 $f(a_i, x_i) (a_i \in C, x_i \in U)$ 可以为数值类型或符号类型. $NMI_\delta(a_j; d)$ 为属性 a_j 和决策属性 d 间的互信息, $NMI_\delta(a_i; a_j)$ 为属性 a_i 和 a_j 间的邻域互信息, 则基于邻域互信息属性 a_j 关于 A 的属性重要度的定义如下:

$$SIG_A(a_j) = NMI_\delta(a_j; d) - \frac{1}{|A|} \sum_{a_i \in A} NMI_\delta(a_j; a_i) \quad (21)$$

其中, 当 $A = \emptyset$ 时, $\frac{1}{|A|} \sum_{a_i \in A} NMI_\delta(a_j; a_i) = 0$.

对于混合类型数据, 数值型数据离散化后可以使用经典的互信息计算方法, 则基于离散互信息的属性重要度的定义如下。

定义 10(基于离散互信息的属性重要度) 已知 $DIS = \langle U, CU\{d\}, V, f \rangle$ 为决策信息系统, $a_i \in A, A \subseteq C$, 其中 $f(a_i, x_i) (a_i \in C, x_i \in U)$ 为符号类型. $MI(a_j; d)$ 为属性 a_j 和决策属性 d 间的互信息, $MI(a_i; a_j)$ 为属性 a_i 和 a_j 间的互信息, 则基于离散互信息属性 a_j 关于 A 的属性重要度的定义如下:

$$SIG_A(a_j) = MI(a_j; d) - \frac{1}{|A|} \sum_{a_i \in A} MI(a_j; a_i) \quad (22)$$

其中, 当 $A = \emptyset$ 时, $\frac{1}{|A|} \sum_{a_i \in A} MI(a_j; a_i) = 0$.

3.3 存储策略

首先, 在算法运行过程中始终只保留最新的动态差别集, 由于差别集不包含重复元素, 因此能够一定程度上节省空间. 其次, 通过对多个数据集的分析发现, 直接计算并保存完整的差别矩阵的方法存在大量的空间浪费. 实验分析表明, 在可以计算出最为重要的第一个属性 a^1 后, 直接计算并保存关于 $\{a^1\}$ 的动态差别集能够极大地节省存储空间, 如图 1 所示。

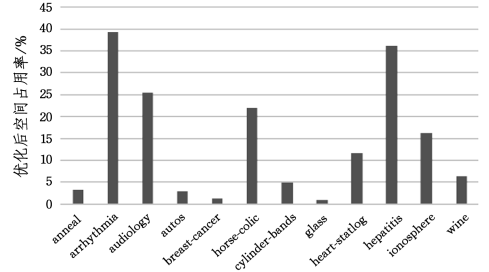


图 1 $\{a^1\}$ 的动态差别集相对于差别矩阵的空间节省率
Fig. 1 Space saving rate of dynamic discernibility set of $\{a^1\}$ relative to discernibility matrix

为了进一步降低算法的空间复杂度, 我们直接构造已经删除了包含最重要的第一个属性的差别集, 并且定义初始动态差别集。

定义 11 已知 $DIS = \langle U, CU\{d\}, V, f \rangle$ 为决策信息系统, $\forall x, y \in U$, DIS 关于 $\{a^1\}$ 的初始动态差别集 DDS^1 的元素定义为:

$$m_{ij}^1 = \begin{cases} \emptyset, & f(x, d) = f(y, d) \\ \emptyset, & \Delta^{(a^1)}(x_i, x_j) \leq \delta \\ \{a \in C - \{a^1\} | \Delta^{(a^1)}(x_i, x_j) > \delta\}, & \text{其他} \end{cases} \quad (23)$$

其中, a^1 为根据属性重要度选择出的第一个属性。

3.4 特征选择算法

通过上述分析, 结合差别矩阵的 mRMR 算法的总体思路如下: (1) 通过选择的属性重要度函数评价所有属性, 并选择最重要的属性作为第一个属性 a^1 , 同时将其加入约简集合; (2) 计算关于 a^1 的动态差别集, 并将其作为初始动态差别集; (3) 当动态差别集不为空时, 循环执行如下操作: 1) 根据当前动态差别集计算候选属性集合; 2) 计算候选集合中各属性的重要度, 并将最重要属性加入约简; 3) 更新动态差别集. 当动态差别集为空时, 停止搜索并输出约简. 算法流程图如图 2 所示. 具体算法如算法 1 所示。

算法 1 分步优化特征选择算法 (SOFS)

输入: $IS = \langle U, CU\{d\}, V, f \rangle$ 和 δ/δ 是控制邻域大小的阈值
输出: 特征子集 R

1. $R \leftarrow \emptyset, DDS \leftarrow \emptyset, a_{sel} \leftarrow \emptyset$
2. $a^1 \leftarrow \text{BestAtt}(R, C)$
3. $R \leftarrow R \cup \{a^1\}$
4. for $i = 1$ to $|U|$
5. for $j = 1$ to $|U|$
6. compute m_{ij}^1
7. $DDS \leftarrow DDS \cup \{m_{ij}^1\}$
8. end for
9. end for
10. while $DDS \neq \emptyset$

```

11. DCS ← CandiSet(DDS)
12. asel ← BestAtt(R, DCS)
13. R ← R ∪ {asel}
14. for each mij ∈ DDS
15.     if asel ∈ mij then
16.         DDS ← DDS - {mij}
17.     end if
18. end for
19. end while
20. return R
    
```

Function CandiSet(DDS)

```

1. DCS ← ∅
2. for each mij ∈ DDS
3.     DCS ← DCS ∪ mij
4. end for
5. return DCS
    
```

Function BestAtt(A, DCS)

```

1. SIGmax ← 0
2. for each a ∈ DCS
3.     compute SIGA(a)
4.     if SIGA(a) > SIGmax then
5.         abest ← a
6.         SIGmax ← SIGA(a)
7.     end if
8. end for
9. return abest
    
```

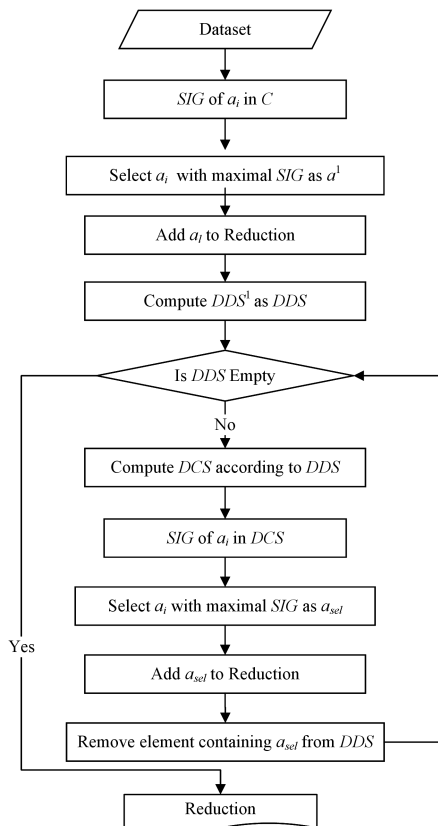


图2 算法流程图

Fig. 2 Flow chart of algorithm

4 实验及结果分析

由于我们在计算属性重要度时分别采用了基于邻域互信

息和基于离散互信息的两种度量方式,为简化表示,其结果分别记为 SOFS-N 和 SOFS-D。将 SOFS-N 和 SOFS-D 分别与 Rerank 算法、FARNeM 算法进行了分类精度的比较,其中 Rerank 算法采用了基于 mRMR 重要度的实现^[25-26]。我们还将提出的算法与原始的 mRMR 算法进行了对比,最后对算法的空间复杂度进行了研究。

4.1 数据集与参数设置

4.1.1 数据集

实验中的数据来源于 UCI 机器学习库,数据的基本信息如表 1 所列。为了保障结果的准确性并避免特殊类型数据的影响,这些数据集涵盖了数值数据和符号数据,同时包含二分类和多分类问题。由于部分测试的特征选择算法并不支持数值数据,在进行特征选择时本文采用了基于最小描述长度原则(MDLP)的离散化方法将数值数据转化为符号属性。

表 1 数据集信息

Table 1 Information of datasets

Data Set	Instances	Numerical	Categorical	Class
anneal	898	6	32	6
arrhythmia	452	206	73	16
audiology	226	0	69	24
breast-cancer	286	0	9	2
cylinder-bands	540	18	19	2
german_credit	1000	7	13	2
glass	214	9	0	7
heart-statlog	270	13	0	2
hepatitis	155	6	13	2
horse-colic	368	7	15	2
ionosphere	351	34	0	2
vehicle	846	18	0	4
vote	435	0	16	2
wdbc	569	30	0	2
wine	178	13	0	3

4.1.2 参数设置

本文算法采用邻域关系进行信息粒化,因此需要输入邻域参数 δ ,同时 FARNeM 方法作为对比方法也需要设置该参数。本文取一组 δ 参数进行实验,并取各自的最佳结果为算法的最终结果。类似于 Warpper 方法, δ 取值分别为 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45。由于对比的 Rerank 方法是 Warppr 方法,因此对比结果具有公平性。SOFS-N 中属性重要度的计算采用了邻域熵,本文采用无穷范数作为其距离函数。将分类精度作为算法的评价指标,采用 10 折交叉验证进行测试,并采用 Wilcoxon 符号秩检验来验证实验结论。此外,为了探究粗糙集理论的引入对原始 mRMR 算法的影响,4.2.2 节将提出的算法与原始的 mRMR 算法进行对比。

4.1.3 实验方案

实验中先采用特征选择算法对原始数据进行处理,之后使用特征选择处理过的数据训练分类模型。为了保证可靠性,模型训练采用 10 折交叉验证。实验中采用了 4 组分类器,分别是支持向量机(SVM)、J48 决策树、近邻算法 kNN ($k=3$) 和多层感知机(MLP)分类器。实验平台为 Ubuntu 15.10,使用 Java 1.8 编程,分类算法使用 Weka 3.8 提供的 API。

4.2 实验结果与分析

下文从分类精度、差别矩阵对特征选择的影响和参数敏

感性这几方面对实验结果进行分析。

4.2.1 分类精度的比较

为了给出全面的分析,本文给出了不同特征选择方法在4个不同类型分类器上的分类精度。SVM在根本上是一种线性分类器,J48是一种树形结构的分类器,kNN是基于最近邻的分类器,MLP是一种人工神经网络。其中,SVM和MLP都只能直接使用在数值数据上,但是符号数据需要转换成一组(0,1)二元组属性才能被使用。J48在分类过程中需要对数值属性进行划分。各表所列测试结果中每个数据集上的最优结果用加粗字体表示。

(1)SVM 分类结果

运用SVM进行分类的结果如表2所列。从表2可以看出,SOFS-D算法和SOFS-N算法在SVM上的平均分类精度均达到了76.8%和77%,相比其他方法提高了2%。

表2 特征选择算法在SVM上的分类精度的比较

Table 2 Accuracy comparison of feature selection algorithms on SVM

Data Set	Origin	Rerank	FARNeM	SOFS-D	SOFS-N
anneal	91.871	84.410	91.314	91.203	91.425
arrhythmia	54.204	55.752	56.637	57.522	56.637
audiology	44.248	60.177	53.540	60.619	61.504
breast-cancer	70.629	71.678	70.280	70.280	70.280
cylinder-bands	60.741	58.519	60.000	62.222	61.111
german_credit	71.800	71.000	71.100	73.100	73.100
glass	47.664	49.533	49.065	47.664	48.131
heart-statlog	83.704	82.963	83.704	84.074	84.074
hepatitis	80.645	82.581	81.935	81.290	81.935
horse-colic	84.239	81.250	—	84.239	84.783
ionosphere	86.895	88.319	89.459	87.464	89.744
vehicle	61.111	50.946	63.239	62.648	63.475
vote	95.632	95.632	—	95.632	95.632
wdbc	94.903	95.079	95.431	95.782	95.606
wine	97.753	98.315	97.753	98.315	98.315
Average	75.069	75.077	74.112	76.804	77.050

(2)J48 分类结果

J48分类器的测试结果如表3所列。可以看出,在J48分类器上,SOFS-D算法和SOFS-N算法的分类精度也提高到了82%以上,说明较好的符号属性也得到了保留。

表3 特征选择算法在J48上的分类精度的比较

Table 3 Accuracy comparison of feature selection algorithms on J48

Data Set	Origin	Rerank	FARNeM	SOFS-D	SOFS-N
anneal	98.441	96.771	98.664	98.441	98.552
arrhythmia	64.381	69.248	61.726	69.690	68.363
audiology	77.876	76.991	73.451	77.876	76.991
breast-cancer	75.524	72.727	74.476	75.524	75.524
cylinder-bands	72.037	65.556	69.444	70.556	70.741
german_credit	70.500	71.100	73.200	73.200	71.700
glass	66.822	69.159	68.224	67.757	69.159
heart-statlog	76.667	80.741	80.370	78.519	80.370
hepatitis	83.871	80.645	80.645	84.516	83.871
horse-colic	85.326	81.522	—	85.598	85.598
ionosphere	91.453	91.738	91.738	92.877	92.308
vehicle	72.459	68.440	72.931	73.404	72.813
vote	96.322	95.632	—	96.322	96.322
wdbc	93.146	93.322	94.903	94.376	94.552
wine	93.820	93.820	94.382	94.944	94.944
Average	81.243	80.494	79.550	82.240	82.121

(3)KNN 分类结果

KNN分类器的测试结果如表4所列。同为基于邻域思想的SOFS-N算法在kNN算法上的表现较好,在几乎所有数据集上都有提升,尤其在arrhythmia,hepatitis和ionosphere上分类精度均高出了原始分类3%以上。

表4 特征选择算法在kNN上的分类精度的比较

Table 4 Accuracy comparison of feature selection algorithms on kNN

Data Set	Origin	Rerank	FARNeM	SOFS-D	SOFS-N
anneal	96.882	96.214	98.552	98.330	98.330
arrhythmia	57.743	65.708	59.071	65.044	63.274
audiology	68.584	68.142	64.602	70.354	69.912
breast-cancer	73.776	72.028	73.776	73.776	73.776
cylinder-bands	69.259	63.148	73.148	71.667	71.111
german_credit	73.300	70.900	72.300	72.600	73.200
glass	71.963	73.364	72.430	72.430	71.963
heart-statlog	78.889	78.889	81.111	81.481	80.000
hepatitis	81.290	79.355	81.935	82.581	84.516
horse-colic	81.250	79.076	—	83.696	82.880
ionosphere	86.610	87.179	89.744	90.883	92.877
vehicle	71.513	59.574	70.922	71.986	71.986
vote	92.644	94.713	—	92.874	93.103
wdbc	96.837	95.431	97.364	97.188	97.364
wine	94.944	97.191	96.629	97.191	96.629
Average	79.699	78.727	79.353	81.472	81.395

(4)MLP 分类结果

MLP分类器的测试结果如表5所列。从表5也可以看出,本文算法的平均分类精度分别为82.7%和83%,高于其他方法,说明提出的算法能够找出较好的数值属性。

表5 特征选择算法在MLP上的分类精度的比较

Table 5 Accuracy comparison of feature selection algorithms on MLP

Data Set	Origin	Rerank	FARNeM	SOFS-D	SOFS-N
anneal	98.998	94.989	99.109	99.220	99.109
arrhythmia	67.257	67.920	62.389	67.478	65.929
audiology	83.186	76.549	72.124	79.204	79.204
breast-cancer	64.685	73.077	66.084	66.084	67.133
cylinder-bands	75.185	65.556	71.667	73.704	74.259
german_credit	71.500	73.500	72.400	72.500	74.200
glass	67.757	62.150	68.224	66.822	69.626
heart-statlog	78.148	78.519	80.000	81.481	81.852
hepatitis	80.000	81.290	81.935	82.581	83.226
horse-colic	80.435	80.163	—	84.511	84.239
ionosphere	91.168	91.738	92.593	93.447	92.877
vehicle	81.678	67.730	81.206	84.161	83.924
vote	94.713	94.943	—	95.172	94.483
wdbc	96.309	96.661	97.540	96.661	97.540
wine	97.191	99.438	98.315	98.876	98.315
Average	81.881	80.281	80.276	82.794	83.061

综上所述,本文算法优于Rerank算法。使用mRMR作为重要度的Rerank方法虽然在mRMR方法的基础上通过再排序选择了更优的属性,但是没有从根本上解决mRMR在后期对属性重要度估计不够准确的问题,尤其是对一些具有少

量独特分类信息属性的低估,从而可能选择完全冗余但相关度较高的属性。因此,本文方法优于 Rerank 方法。

FARNeM 算法在 horse-colic 和 vote 数据集上不能得出约简,这是因为这两个数据集上任何单一属性导出的正域大小均为 0,所以无法启动算法。另外,由于其属性重要度的度量并未考虑属性冗余,因此可能选择冗余属性,导致分类精度低于本文算法。实验结果表明,粗糙集理论与 mRMR 原则的结合有效提升了选择的属性的质量。

另外,SOFS-D 在 SVM 分类器上的表现比 SOFS-N 差,这是由于 SOFS-D 的重要度是通过离散化后的数据计算的,数值属性的信息有一定程度的损失,而 SVM 又恰好是一个适合数值数据的分类器。SOFS-D 在 J48 分类器上的表现较好,由于 J48 分类器本身也只能处理符号属性,因此 SOFS-D 中对数据离散后再计算重要度的方法与分类器较为契合,因此取得了较高的精度。

(4) 统计显著性分析

为了更好地说明本文算法的有效性,我们采用 Wilcoxon 符号秩检验进行统计显著性分析。Wilcoxon 符号秩检验是一种非参数检验,若检验值小于给定的显著水平,则有充分的理由拒绝原假设,而采取备择假设。这里的原假设为原始数据或其他特征选择算法得出的分类精度优于本文算法,而备择假设为本文算法较优。因此,当检验值小于给定的显著水平时,拒绝原假设,同时说明本文算法具有有效性。Wilcoxon 符号秩检验的检验结果如表 6 和表 7 所列。

从检验结果来看,在显著水平 $\alpha=0.05$ 的条件下,SOFS-N 方法显著优于原始数据、Rerank 方法和 FARNeM 方法。在显著水平 $\alpha=0.1$ 的条件下,SOFS-D 方法也显著优于原始数据、Rerank 方法和 FARNeM 方法。另外,我们也可以发现 SOFS-N 方法和 SOFS-D 方法的表现其实非常相近,因为它们之间彼此的检验结果都较大,说明两者之间没有明显的优劣之分。因此,在计算能力不足时,适合采用计算量较小的 SOFS-D 方法,这样不会对结果造成太大影响。

表 6 DNMFs 的 Wilcoxon 符号秩检验结果

Table 6 Wilcoxon rank sum test results of SOFS-N

Classifier	Origin	Rerank	FARNeM	SOFS-D
SVM	0.00159	0.01802	0.00674	0.15403
J48	0.02273	0.00321	0.01010	0.74756
KNN	0.00105	0.00754	0.02731	0.69494
MLP	0.02768	0.02063	0.00126	0.22563

表 7 NMFs 的 Wilcoxon 符号秩检验结果

Table 7 Wilcoxon rank sum test results of SOFS-D

Classifier	Origin	Rerank	FARNeM	SOFS-N
SVM	0.00674	0.05399	0.07439	0.86890
J48	0.01469	0.00623	0.01191	0.28166
KNN	0.00159	0.00347	0.01055	0.33307
MLP	0.03186	0.02582	0.00718	0.79278

4.2.2 动态候选集对原始 mRMR 算法的影响

为了进一步探究粗糙集理论方法的引入对原始 mRMR 方法的影响,我们还绘制了算法选择特征个数与分类精度曲线图,如图 3、图 4 所示,图中给出了 SOFS-N 和 SOFS-D 的邻域参数。

由于 SOFS-N 和 SOFS-D 是存在停止条件的,曲线最多只能到达停止条件时的属性个数(这里的最大属性个数以 10 折交叉中选择属性最少的一次验证为准),而 mRMR 曲线则能达到全部属性个数。由图 3、图 4 可知,SOFS-N 和 SOFS-D 所选择的前几个特征的质量总体上优于 mRMR。这对于 Filter 特征选择方法是有意义的,因为 Filter 特征选择方法往往需要借助交叉验证来确定最终的特征个数,而本文算法能够在前期选择出高质量的特征,这意味着最终能够得出更为精简且高质量的特征。此外,本文算法能够自行停止并排除掉冗余的特征,减少了交叉验证所需计算的特征集的数量,减小了计算量。

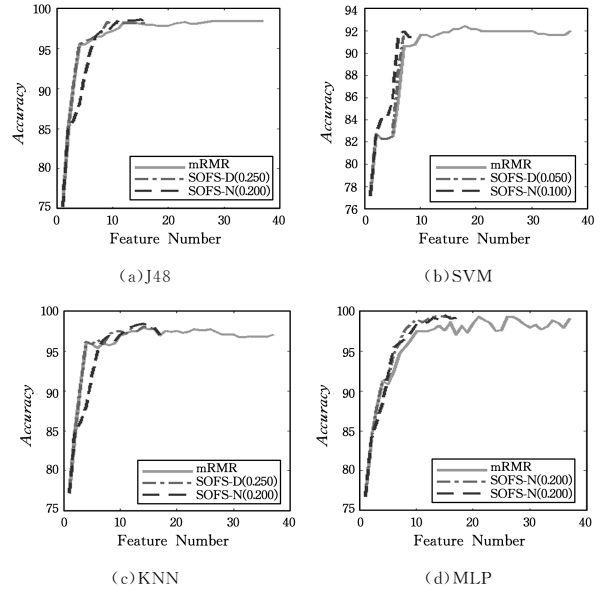


图 3 anneal 数据集上分类精度与特征个数的关系

Fig. 3 Relation between classification accuracies and number of feature on data set anneal

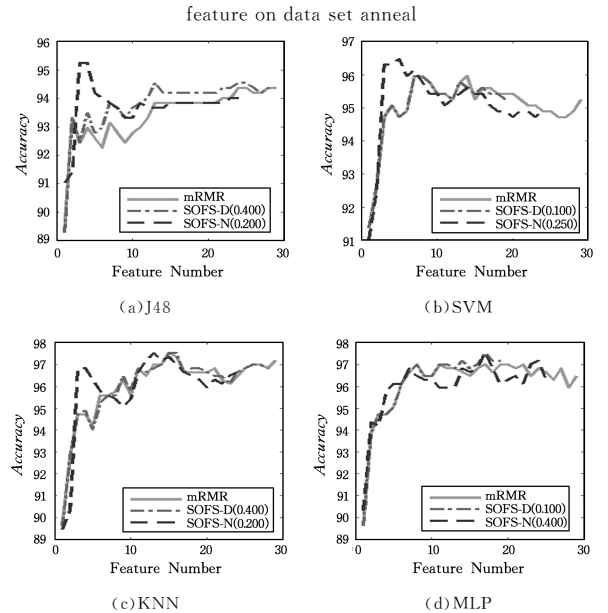


图 4 wdbc 数据集上分类精度与特征个数的关系

Fig. 4 Relation between classification accuracies and number of feature on data set wdbc

4.2.3 时间复杂度分析

动态候选集的引入,必然会规避掉一些不必要属性重要

度的计算,但同时动态差别集本身又需要消耗一定的时间。为了探究算法的时间复杂度,我们对本文算法与原始的mRMR算法及其采用领域熵计算的mRMR(以下简称mRMR-N)进行了算法运行时间的对比,并给出了应用动态差别集后算法的加速比(SpeedUp),如表8和表9所列。

从表8和表9可以看出,随着数据集特征数量的增加,本文算法的速度相比原始mRMR有了较明显的提升,这主要得益于动态候选集能限制候选属性范围并给出适当的搜索停止时机。通常情况下,越多的特征意味着越多的冗余特征,而本文算法能够有效规避冗余特征,从而缩短算法的运行时间。同时,在特征较少的情况下,本文算法耗时反而有一定程度的增加,显然这是因为动态差别集计算增加了额外的时间,而特征较少的数据集冗余特征也较少,动态候选集的效果并不明显。因此,本文算法更适合高维数据。

表8 SOFS-D与mRMR的算法时间复杂度的对比

Table 8 Comparison of time complexity between SOFS-D and

Data Set	mRMR			Feature Count
	mRMR/ms	SOFS-D/ms	SpeedUp	
breast-cancer	8	44	0.182	11
glass	13	12	1.083	11
heart-statlog	13	29	0.448	15
wine	10	8	1.250	15
vowel	164	205	0.800	15
vote	26	54	0.481	18
vehicle	162	493	0.329	20
hepatitis	15	14	1.071	21
german_credit	171	1166	0.147	22
horse-colic	71	201	0.353	24
autos	61	177	0.345	27
wdbc	297	284	1.046	32
ionosphere	493	215	2.293	36
cylinder-bands	364	1517	0.240	39
anneal	4292	4218	1.018	40
audiology	808	597	1.353	71
arrhythmia	62357	6303	9.893	281

表9 SOFS-N与mRMR-N的算法时间复杂度的对比

Table 9 Comparison of time complexity between SOFS-D and

mRMR-N

Data Set	mRMR-N			Feature Count
	mRMR-N/ms	SOFS-N/ms	SpeedUp	
breast-cancer	123	163	0.755	11
glass	84	94	0.894	11
heart-statlog	334	420	0.795	15
wine	108	72	1.500	15
vowel	4331	4413	0.981	15
vote	1797	1506	1.193	18
vehicle	7062	6341	1.114	20
hepatitis	266	242	1.099	21
german_credit	11949	12973	0.921	22
horse-colic	2177	2507	0.868	24
autos	814	379	2.148	27
wdbc	10643	6599	1.613	32
ionosphere	5083	2180	2.332	36
cylinder-bands	18344	12194	1.504	39
anneal	92075	32728	2.813	40
audiology	23672	3233	7.322	71
arrhythmia	4350689	167987	25.899	281

4.2.4 参数敏感性分析

为了探究参数对算法的影响,我们研究了参数 δ 取不同

值时算法在不同分类器上的分类精度,实验结果如图5和图6所示。

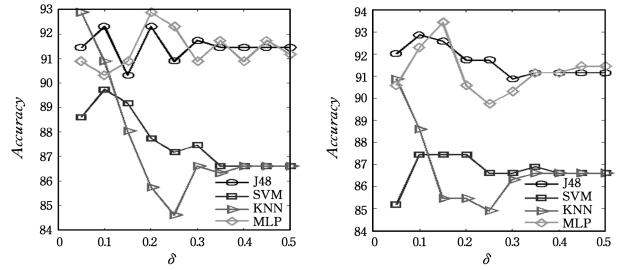


图5 ionosphere数据集上分类精度对 δ 的敏感性

Fig. 5 Sensitivity of classification accuracies on δ of data

det ionosphere

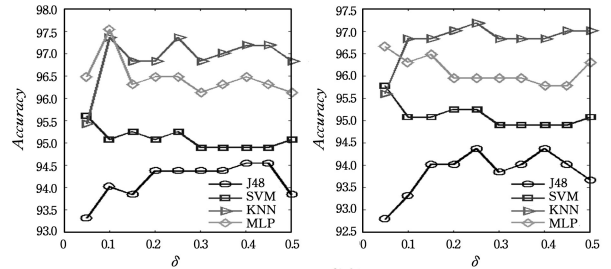


图6 wdbc数据集上分类精度对 δ 的敏感性

Fig. 6 Sensitivity of classification accuracies on δ of data set wdbc

从图5和图6可以看出,对于不同的分类器,同一算法取得最大精度时的 δ 值也有所不同,这是由于不同分类器对单个属性和数据集属性个数的敏感程度不同导致的。从图中可以看出,SVM分类器更加偏好 δ 取值较小的特征选择结果。对比不同数据集下的 δ 取值与精度的关系可以发现,不同数据的最佳 δ 取值也有所不同。另外,从图中还可以观察到,当 δ 取值偏大时,选择出的特征的质量偏低。因此,实际使用本文算法时不应选择过大的 δ 。

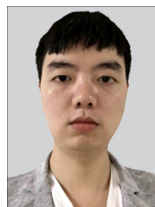
以上实验结果表明,本文提出的特征选择方法能够有效地提高分类质量。同时,通过对数据的分析也证明了本文提出的初始动态差别集的空间复杂度相比于传统差别矩阵有了大幅降低。

结束语 结合mRMR方法和差别矩阵的优点,本文提出了SOFS-D和SOFS-N特征选择方法;利用差别矩阵能够排除无用信息的性质,提出了动态差别集和动态候选集,缩小了候选特征集合的搜索范围,从而在最小冗余的基础上进一步排除冗余特征,加速启发式搜索进程,并确定约简大小;定义了初始动态差别集来解决差别矩阵空间复杂度高的问题。实验结果验证了所提算法的有效性。本文算法尽管在多数情况下具有较好的效果,但存在一些局限性,如邻域熵计算较为复杂,如何快速且准确地计算两个特征的相关度将是下一步的研究方向。

参考文献

- [1] CHANDRASHEKAR G, SAHIN F. A survey on feature selection methods[J]. Computers and Electrical Engineering, 2014, 40(1):16-28.

- [2] XE J Y, WANG M Z, ZHOU Y, et al. Differentially expressed gene selection algorithms for unbalanced gene datasets [J]. Chinese Journal of Computers, 2019, 42(6): 1232-1250.
- [3] WANG D, NIE F, HUANG H. Feature Selection via Global Redundancy Minimization [J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(10): 2743-2755.
- [4] TASKIN G, KAYA H, BRUZZONE L. Feature selection based on high dimensional model representation for hyperspectral images [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2918-2928.
- [5] SHEIKHPOUR R, SARRAM M A, GHARAGHANI S, et al. A Survey on semi-supervised feature selection methods [J]. Pattern Recognition, 2016, 64(C): 141-158.
- [6] MAFARJA M, MIRJALILI S. Whale optimization approaches for wrapper feature selection [J]. Applied Soft Computing Journal, 2018, 62: 441-453.
- [7] VIOLA M, SANGIOVANNI M, TORALDO G, et al. A generalized eigenvalues classifier with embedded feature selection [J]. Optimization Letters, 2017, 11(2): 299-311.
- [8] PENG H, LONG F, DING C. Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226-1238.
- [9] ESTEVEZ P A, TESMER M, PEREZ C A, et al. Normalized mutual information feature selection [J]. IEEE Transactions on Neural Networks, 2009, 20(2): 189-201.
- [10] SUN Z, ZHANG J, DAI L, et al. Neurocomputing Mutual information based multi-label feature selection via constrained convex optimization [J]. Neurocomputing, 2019, 329: 447-456.
- [11] WANG J, WEI J M, YANG Z, et al. Feature Selection by Maximizing Independent Classification Information [J]. IEEE Transactions on Knowledge & Data Engineering, 2017, 29(4): 828-841.
- [12] PAWLAK Z. Rough sets [J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356.
- [13] LIU Y, CHEN Y, TAN K, et al. Maximum relevance, minimum redundancy band selection based on neighborhood rough set for hyperspectral data classification [J]. Measurement Science and Technology, 2016, 27(12): 125501.
- [14] YONG L, WENLIANG H, YUNLIANG J, et al. Quick attribute reduct algorithm for neighborhood rough set model [J]. Information Sciences, 2014, 271: 65-81.
- [15] JIANG Y, YU Y. Minimal attribute reduction with rough set based on compactness discernibility information tree [J]. Soft Computing, Springer Berlin Heidelberg, 2016, 20(6): 2233-2243.
- [16] DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets [J]. International Journal of General Systems, 1990, 17(2/3): 191-209.
- [17] HU Q H, YU D R, XIE Z X. Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation [J]. Journal of Software, 2008, 19(3): 640-649.
- [18] SKOWRON A, RAUSZER C. The Discernibility Matrices and Functions in Information Systems [M] // Intelligent Decision Support. Dordrecht: Springer, 1992: 331-362.
- [19] DAI J. Rough set approach to incomplete numerical data [J]. Information Sciences, 2013, 241: 43-57.
- [20] LI M, SHANG C, FENG S, et al. Quick attribute reduction in inconsistent decision tables [J]. Information Sciences, 2014, 254: 155-180.
- [21] YANG M, YANG P. A novel condensing tree structure for rough set feature selection [J]. Neurocomputing, 2008, 71(4/5/6): 1092-1100.
- [22] HU Q, ZHANG L, ZHANG D, et al. Measuring relevance between discrete and continuous features based on neighborhood mutual information [J]. Expert Systems with Applications, 2011, 38(9): 10737-10750.
- [23] LIN Y, HU Q, LIU J, et al. Multi-label feature selection based on neighborhood mutual information [J]. Applied Soft Computing, 2016, 38: 244-256.
- [24] HU Q, ZHANG L, ZHANG D, et al. Measuring relevance between discrete and continuous features based on neighborhood mutual information [J]. Expert Systems with Applications, 2011, 38(9): 10737-10750.
- [25] BERMEJO P, OSSA L DE, GAMEZ J A. Knowledge-Based Systems Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking [J]. Knowledge-Based Systems, 2012, 25(1): 35-44.
- [26] HU Q, YU D, XIE Z. Neighborhood classifiers [J]. Expert Systems with Applications, 2008, 34(2): 866-876.



FAN Xin, born in 1994, postgraduate. His main research interests include the areas of data mining and pattern recognition.



CHEN Hong-mei, born in 1971, Ph.D., professor, Ph.D supervisor, is member of China Computer Federation (CCF). Her main research interests include rough set, granular computing and intelligent information processing.