

## 动态加权条件互信息的特征选择算法

张俐<sup>\*①②</sup> 陈小波<sup>③</sup>

<sup>①</sup>(江苏理工学院计算机工程学院 常州 213001)

<sup>②</sup>(北京邮电大学可信分布式计算与服务教育部重点实验室 北京 100876)

<sup>③</sup>(中国人民银行常州市中心支行 常州 213001)

**摘要:** 特征选择是机器学习、自然语言处理和数据挖掘等领域中数据预处理阶段必不可少的步骤。在一些基于信息论的特征选择算法中,存在着选择不同参数就是选择不同特征选择算法的问题。如何确定动态的非先验权重并规避预设先验参数就成为一个急需解决的问题。该文提出动态加权的最大相关性和最大独立性(WMRI)的特征选择算法。首先该算法分别计算新分类信息和保留类别信息的平均值。其次,利用标准差动态调整这两种分类信息的参数权重。最后,WMRI与其他5个特征选择算法在3个分类器上,使用10个不同数据集,进行分类准确率指标(fmi)验证。实验结果表明,WMRI方法能够改善特征子集的质量并提高分类精度。

**关键词:** 特征选择; 分类信息; 平均值; 标准差; 动态加权

中图分类号: TN911.7

文献标识码: A

文章编号: 1009-5896(2021)10-3028-07

DOI: 10.11999/JEIT200615

## Feature Selection Algorithm for Dynamically Weighted Conditional Mutual Information

ZHANG Li<sup>①②</sup> CHEN Xiaobo<sup>③</sup>

<sup>①</sup>(College of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China)

<sup>②</sup>(Key Laboratory of Trustworthy Distributed Computing and Service (Ministry of Education), Beijing University of Posts and Telecommunications, Beijing 100876, China)

<sup>③</sup>(The People's Bank of China, Changzhou Branch, Changzhou 213001, China)

**Abstract:** Feature selection is an essential step in the data preprocessing phase in the fields of machine learning, natural language processing and data mining. In some feature selection algorithms based on information theory, there is a problem that choosing different parameters means choosing different feature selection algorithms. How to determine the dynamic, non-a priori weights and avoid the preset a priori parameters become an urgent problem. A Dynamic Weighted Maximum Relevance and maximum Independence (WMRI) feature selection algorithm is proposed in this paper. Firstly, the algorithm calculates the average value of the new classification information and the retained classification information. Secondly, the standard deviation is used to dynamically adjust the parameter weights of these two types of classification information. At last, WMRI and the other five feature selection algorithms use ten different data sets on three classifiers for the fmi classification metrics validation. The experimental results show that the WMRI method can improve the quality of feature subsets and increase classification accuracy.

**Key words:** Feature selection; Classification information; Average value; Standard deviation; Dynamic weighting

收稿日期: 2020-07-23; 改回日期: 2021-02-05; 网络出版: 2021-03-19

\*通信作者: 张俐 zhangli\_3913@163.com

基金项目: 国家科技基础性工作专项(2015FY111700-6), 江苏理工学院博士科研基金(KYY19042)

Foundation Items: The National Science and Technology Basic Work Project (2015FY111700-6), The Doctoral Research Fund of Jiangsu University of Technology (KYY19042)

### 1 引言

在过去几十年里,新型计算机和互联网技术正在以前所未有的速度产生着大量高维数据<sup>[1,2]</sup>。在这些高维数据中包含着许多无关和冗余特征。因为不相关和冗余特征不仅会增加模型训练时间而且也使得模型的可解释变得很差。如何处理这些不相关和冗余特征是数据分析和知识发现中所面临的重大

挑战。特征选择不同于其他数据降维技术(如特征提取)<sup>[3]</sup>, 它可以删除无关和冗余特征, 保留相关原始物理特征, 从而降低数据维数。这样有利于提高数据质量和分类性能, 并使得模型的训练时间大幅缩小而且也使得模型的可解释性变得更强<sup>[4,5]</sup>。

通常特征选择技术又分为分类依赖型<sup>[6]</sup>(包装器方法和嵌入式方法)和分类器无关型(过滤式方法)。基于信息论的过滤式特征选择方法优点<sup>[7-11]</sup>为: (1) 它可以直接从数据中提取有价值的知识, 而且这些知识对于问题真正的解决又起到至关重要的作用。(2) 它的计算成本低且与具体分类器无关。(3) 目前该方法应用领域广泛, 包括基因表达数据、文本分类和网络入侵检测等多个领域。因此基于信息论的过滤式特征选择方法逐渐成为特征选择技术的研究热点<sup>[12-16]</sup>。

常见基于信息论的特征选择算法<sup>[5,17-19]</sup>可分为两类。最小化冗余特征的算法(maxMIFS<sup>[8]</sup>, MRMR<sup>[20]</sup>, CIFE<sup>[8]</sup>, CMIM<sup>[8]</sup>和JMI<sup>[10]</sup>等)和最大化新分类信息的算法(DCSF<sup>[12]</sup>和JMIM<sup>[16]</sup>)。maxMIFS和MRMR通过去除特征之间冗余特征来提高最优特征子集( $S$ )整体识别质量。但是它们却忽视两个特征与类标签之间的冗余性问题。因此, 产生许多经典的多信息去除冗余性的算法, 例如JMI, CIFE和CMIM等。然而它们却忽视最大化新分类信息来提高 $S$ 集合整体识别的质量。随着特征选择算法的发展, 如何将减少冗余的特征选择算法和最大化新分类信息的特征选择算法进行融合逐渐成为研究的新热点。代表性的算法有MRI<sup>[13]</sup>, CFR<sup>[14]</sup>和DISR<sup>[21]</sup>等。以上这些算法都是基于信息论特征选择框架<sup>[7]</sup>的具体实现。Brown等人<sup>[7]</sup>认为选择不同的参数就是选择不同的特征选择算法。它们存在的问题是参数设置过大还是过小都会对特征选择过程造成影响, 即存在对无关特征和冗余特征的忽略与误判。

在大数据环境下, 针对数据多样性和高维性的特点, 寻找一种动态的非预先设置参数的特征选择方法就成为目前需要解决的问题。本文提出一种新的过滤式特征选择算法(Weighted Maximum Relevance and maximum Independence, WMRI)。本文主要贡献为: (1) 利用条件互信息衡量特征与类标签之间的相关性以及特征之间冗余性; (2) 提出通过均值和标准差来动态调节新分类信息和保留类别信息的权重与平衡问题; (3) 通过对10个基准数据集进行实验对比, 实验结果表明, 该算法(WMRI)优于其他特征选择算法(DCSF, MRI, CFR, IG-RFE<sup>[15]</sup>和JMIM)。

## 2 WMRI算法的提出

Brown等人<sup>[7]</sup>提出基于信息论特征选择框架, 具体为

$$J(f_k) = I(f_k; C) - \beta \sum_{f_{\text{sel}} \in S} I(f_k; f_{\text{sel}}) + \lambda \sum_{f_{\text{sel}} \in S} I(f_k; f_{\text{sel}}|C) \quad (1)$$

其中, 设 $F$ 是原始特征集合,  $|S|$ 是最优特征子集数,  $S \subset F$ ,  $J(\cdot)$ 代表评估标准,  $f_k$ 表示候选特征,  $f_{\text{sel}}$ 表示已选特征,  $f_{\text{sel}} \in S, f_k \in F - S$ ,  $C$ 表示类标签集合,  $|C|$ 是类标签数。

Wang等人<sup>[13]</sup>在Brown的研究基础上提出MRI算法, 具体评估标准为

$$J_{\text{MRI}}(f_k) = I(C; f_k) + \sum_{f_{\text{sel}} \in S} \{I(C; f_k|f_{\text{sel}}) + I(C; f_{\text{sel}}|f_k)\} \propto I(f_k; C) - \frac{2}{|S|+1} \sum_{f_{\text{sel}} \in S} [I(f_k; f_{\text{sel}}) - I(f_k; f_{\text{sel}}|C)] \quad (2)$$

从式(2)中可知, 在MRI算法中, 独立分类信息由新分类信息项 $I(C; f_k|f_{\text{sel}})$ 与保留类别信息项 $I(C; f_{\text{sel}}|f_k)$ 构成, 并且这两种分类信息同等重要, 存在问题是在实际中 $I(C; f_k|f_{\text{sel}})$ 与 $I(C; f_{\text{sel}}|f_k)$ 之间存在差异性。同时, 结合式(1)和式(2)又可知该算法存在预先设置参数 $\beta$ 和 $\lambda$ 的问题, 即 $\lambda = \beta = \frac{2}{|S|+1}$ 。

那么, 如何在不增加计算量和复杂度的情况下, 动态区分新分类信息和保留类别信息之间的重要程度。以适应在大数据环境下, 数据多样性和高维性的特点, 并提高 $S$ 集合整体数据的质量, 就成为目前在特征选择领域中需要研究的一个问题。

本文提出一种新的过滤式特征选择算法(WMRI)。该方法通过引入标准差方法来分别计算 $I(C; f_k|f_{\text{sel}})$ 与 $I(C; f_{\text{sel}}|f_k)$ 之间的权重。因为标准差<sup>[2]</sup>是一种常见的测量系统稳定程度的度量方法。标准差值越高, 表示分散度越高; 反之亦然。因此, 通过标准差可以动态平衡新分类信息项与保留类别信息项之间的重要程度。WMRI算法评估标准具体为

$$J_{\text{WMRI}}(f_k) = I(C; f_k) + (1 - \alpha) \times \sum_{f_{\text{sel}} \in S} I(C; f_k|f_{\text{sel}}) + (1 - \beta) \times \sum_{f_k \in F-S} I(C; f_{\text{sel}}|f_k) \quad (3)$$

从式(3)可以得出,  $\alpha$ 和 $\beta$ 可以分别动态测量新分类信息项 $I(C; f_k|f_{\text{sel}})$ 与保留类别信息项 $I(C; f_{\text{sel}}|f_k)$

的重要程度。通过这样，WMRI算法可以解决 $I(C; f_k|f_{sel})$ 项与 $I(C; f_{sel}|f_k)$ 项之间平衡和权重问题。其中，式(3)中 $\alpha$ 和 $\beta$ 分别由式(5)和式(7)表示，它的伪代码如表1所示。

$$\mu_1 = \frac{1}{|S|} \sum_{i=1}^{|S|} I(C; f_k|f_{sel}) \quad (4)$$

$$\alpha = \sqrt{\frac{1}{|S|} \sum_{i=1}^{|S|} (I(C; f_k|f_{sel}) - \mu_1)^2} \quad (5)$$

$$\mu_2 = \frac{1}{|S|} \sum_{i=1}^{|S|} I(C; f_{sel}|f_k) \quad (6)$$

$$\beta = \sqrt{\frac{1}{|S|} \sum_{i=1}^{|S|} (I(C; f_{sel}|f_k) - \mu_2)^2} \quad (7)$$

从式(3)可以知道，WMRI算法与MRI算法相类似，都采用前向顺序搜索特征子集。通过表1可知，WMRI算法主要分为3部分。第1部分(第(1)~(6)行)主要包括：(1) 初始化 $S$ 集合和计数器 $k$ ；(2) 计算集合 $F$ 中每个特征的互信息，选择出最大

的特征 $f_k$ ，将该特征 $f_k$ 从 $F$ 集合中删除，并将特征 $f_k$ 加入 $S$ 集合，这时候候选特征 $f_k$ 变成已选特征 $f_{sel}$ 。第2部分(第(7)~(15)行)主要是分别计算 $I(C; f_k|f_{sel})$ ， $I(C; f_{sel}|f_k)$ ， $\mu_1$ ， $\alpha$ ， $\mu_2$ 和 $\beta$ 的值。在第3部分(第(16)~(20)行)，根据式(3)的选择标准，选择出最大 $J_{WMRI}(f_k)$ 值所对应的特征 $f_k$ ，并将该特征 $f_k$ 存入 $S$ 并从 $F$ 中删除该特征 $f_k$ ，然后一直循环到用户指定的阈值 $K$ 就停止循环。

WMRI算法包括2个“for”循环和1个“while”循环。因此，WMRI算法的时间复杂性是 $O(Kmn)$ ( $K$ 代表已选特征数， $n$ 代表所有特征数， $m$ 代表所有样本数， $K \ll n$ )。WMRI算法复杂性高于MRI算法，IG-RFE算法，CFR算法，JMIM算法和DCSF算法。主要原因在于WMRI算法还需计算 $\mu_1$ ， $\alpha$ ， $\mu_2$ 和 $\beta$ 的值。

### 3 实验分析与讨论

#### 3.1 数据集描述

为了验证所提出WMRI算法的有效性，在实验中使用10个不同数据集进行验证。这些数据集来自不同的领域，同时它们可以在UCI<sup>[13]</sup>和ASU<sup>[19]</sup>中找到。这些数据集包括手写数字数据(Semeion和Mfeat-kar)、文字数据(CANE-9)、语音数据(Isolet)、图像数据(COIL20和USPS)、生物学数据(WPBC和ALLAML)和其他类数据(Madelon和Musk2)。更详细的描述可以在表2中找到。

#### 3.2 实验环境设置

在实验中，使用K近邻(KNN)<sup>[19]</sup>、决策树(C4.5)<sup>[13]</sup>和随机森林(RandomForest)<sup>[22]</sup>来评估不同的特征选择算法。本文的实验环境是Intel-i7处理器，使用8 GB内存，仿真软件是Python2.7。

实验由3个部分组成。第1部分是数据预处理。为保证实验的公正性，整个实验过程采用6折交叉验证方法进行验证，就是将实验数据集均匀分成

表1 WMRI算法的伪代码

输入：原始特征集合 $F = \{f_1, f_2, \dots, f_n\}$ ，类标签集合 $C$ ，阈值 $K$
输出：最优特征子集 $S$
(1) 初始化： $S = \phi, k = 0$
(2) for $k=1$ to $n$
(3) 计算每个特征与标签的互信息值 $I(C; f_k)$
(4) $J_{WMRI}(k) = \arg \max(I(C; f_k))$
(5) Set $F \leftarrow F \setminus \{f_k\}$
(6) Set $S \leftarrow \{f_k\}$
(7) while $k \leq K$
(8) for each $f_k \in F$ do
(9) 分别计算新分类信息项 $I(C; f_k f_{sel})$ 与保留类别信息项 $I(C; f_{sel} f_k)$ 的值
(10) 根据式(4)，计算新分类信息项的平均值 $\mu_1$
(11) 根据式(5)，计算新分类信息项的标准方差 $\alpha$
(12) 根据式(6)，计算保留类别信息项的平均值 $\mu_2$
(13) 根据式(7)，计算保留类别信息项的标准方差 $\beta$
(14) 根据式(3)，更新 $J_{WMRI}(f_k)$
(15) end for
(16) 根据 $J_{WMRI}(f_k)$ 评估标准，寻找最优的候选特征 $f_k$
(17) Set $F \leftarrow F \setminus \{f_k\}$
(18) Set $S \leftarrow \{f_k\}$
(19) $k=k+1$
(20) end while

表2 数据集描述

序号	数据集名称	特征数	样本数	类别数	数据来源
1	musk2	167	476	2	UCI
2	madelon	500	2600	2	ASU
3	ALLAML	7129	72	2	ASU
4	CNAE-9	857	1080	9	UCI
5	mfeat-kar	65	2000	10	UCI
6	USPS	256	9298	10	ASU
7	semeion	257	1593	10	ASU
8	COIL20	1024	1440	20	ASU
9	wpbc	34	198	22	UCI
10	Isolet	617	1560	26	ASU

6等份，5份作为训练数据集，1份作为测试数据集。第2部分是特征子集的生成。在实验中，采用不同特征选择方法生成特征子集。特征子集的规模设为30。第3部分是特征子集评价。在这个部分中，用fmi来评估分类器在特征子集上的分类准确率。分类准确率是指正确分类的样本数占样本总数的比例。设TP(True Positive)指正类判定为正类的个数；FP(False Positive)指负类判定为正类的个数；TN(True Negative)指负类判定为负类的个数；FN(False Negative)指正类判定为负类的个数。sen, prc和fmi定义分别为

$$\text{sen} = \frac{\frac{1}{|C|} \sum_{i=1}^{|C|} \text{TP}_i}{\frac{1}{|C|} \sum_{i=1}^{|C|} \text{TP}_i + \frac{1}{|C|} \sum_{i=1}^{|C|} \text{FN}_i} \quad (8)$$

$$\text{prc} = \frac{\frac{1}{|C|} \sum_{i=1}^{|C|} \text{TP}_i}{\frac{1}{|C|} \sum_{i=1}^{|C|} \text{TP}_i + \frac{1}{|C|} \sum_{i=1}^{|C|} \text{FP}_i} \quad (9)$$

$$\text{fmi} = 2 \times \frac{\text{prc} \times \text{sen}}{\text{prc} + \text{sen}} \quad (10)$$

### 3.3 实验结果与讨论

表3—表5分别选择KNN, C4.5和Random Forest这3种分类器，同时以fmi分类准确率作为评价指标对WMRI, IG-RFE, CFR, JMIM, DCSF和MRI进行统计分析。表中每行中最大值用黑体字标识。命名为“平均值”的所在行表示平均fmi值。通过使用“+”，“=”和“-”表示WMRI算法分别“优于”、“等于”和“差于”其他特征选择算法。命名为“W/T/L”的所在行，

表 3 KNN分类器的平均分类准确率fmi(%)

数据集	WMRI	IG-RFE	CFR	JMIM	DCSF	MRI
madelon	<b>79.807</b>	55.037(+)	69.422(+)	76.654(+)	58.465(+)	77.999(+)
USPS	<b>84.372</b>	52.601(+)	55.796(+)	84.372(=)	48.376(+)	62.11(+)
COIL20	<b>87.014</b>	79.236(+)	79.306(+)	83.542(+)	79.236(+)	77.569(+)
musk2	<b>74.002</b>	68.065(+)	70.172(+)	70.592(+)	71.019(+)	69.766(+)
CNAE-9	<b>76.667</b>	74.722(+)	76.667(=)	52.315(+)	76.667(=)	76.667(+)
mfeat-kar	95.961	<b>96.11(-)</b>	95.905(+)	94.459(+)	95.908(+)	95.905(+)
ALLAML	<b>73.873</b>	72.29(+)	68.774(+)	73.873(=)	68.211(+)	62.403(+)
wpbc	<b>36.298</b>	29.482(+)	33.427(+)	27.814(+)	33.423(+)	36.298(=)
Isolet	<b>57.628</b>	47.756(+)	46.795(+)	45.705(+)	51.154(+)	48.91(+)
semeion	<b>75.196</b>	64.981(+)	68.989(+)	55.015(+)	73.501(+)	68.728(+)
平均值	<b>74.082</b>	64.028	66.525	66.434	65.596	67.636
W/T/L		9/0/1	9/1/0	8/2/0	9/1/0	9/1/0

表 4 C4.5分类器的平均分类准确率fmi(%)

数据集	WMRI	IG-RFE	CFR	JMIM	DCSF	MRI
madelon	<b>79.886</b>	53.653(+)	69.305(+)	65.302(+)	57.808(+)	78.694(+)
USPS	75.94	71.756(+)	76.198(-)	75.66(+)	72.057(+)	<b>77.908(-)</b>
COIL20	<b>78.403</b>	68.889(+)	71.944(+)	72.014(+)	78.333(+)	71.944(+)
musk2	66.42	63.485(+)	61.612(+)	<b>66.433(-)</b>	64.499(+)	61.981(+)
CNAE-9	<b>81.481</b>	80.833(+)	81.481(=)	62.593(+)	81.481(=)	81.481(=)
mfeat-kar	<b>85.058</b>	84.563(+)	84.863(+)	81.612(+)	84.914(+)	84.915(+)
ALLAML	<b>74.815</b>	69.619(+)	59.606(+)	73.553(+)	70.523(+)	59.499(+)
wpbc	<b>35.483</b>	27.942(+)	31.991(+)	29.547(+)	31.757(+)	35.176(+)
Isolet	<b>53.141</b>	41.667(+)	50.513(+)	39.936(+)	52.372(+)	52.244(+)
semeion	<b>72.311</b>	65.974(+)	67.614(+)	55.03(+)	70.435(+)	67.355(+)
平均值	<b>70.258</b>	62.854	65.518	62.183	66.455	67.129
W/T/L		10/0/0	8/1/1	9/0/1	9/1/0	8/1/1

表5 Random Forest分类器的平均分类准确率fmi(%)

数据集	WMRI	IG-RFE	CFR	JMIM	DCSF	MRI
madelon	<b>81.308</b>	53.882(+)	70.499(+)	64.306(+)	60.227(+)	81.04(+)
USPS	<b>84.414</b>	79.038(+)	82.791(+)	84.091(+)	79.662(+)	84.018(+)
COIL20	<b>88.681</b>	80.972(+)	82.847(+)	80.278(+)	84.653(+)	82.222(+)
musk2	<b>69.379</b>	65.138(+)	67.886(+)	68.695(+)	67.229(+)	68.917(+)
CNAE-9	<b>82.222</b>	81.667(+)	81.852(+)	62.407(+)	81.944(+)	82.037(+)
mfeat-kar	<b>89.53</b>	83.666(+)	83.812(+)	89.024(+)	87.581(+)	88.52(+)
ALLAML	<b>84.683</b>	72.504(+)	69.367(+)	80.73(+)	71.824(+)	64.132(+)
wdbc	<b>45.655</b>	43.239(+)	45.496(+)	44.105(+)	45.059(+)	44.067(+)
Isolet	<b>60.641</b>	48.782(+)	55.833(+)	46.538(+)	60.128(+)	58.013(+)
semeion	<b>78.722</b>	69.169(+)	73.203(+)	59.113(+)	76.269(+)	73.33(+)
平均值	<b>76.524</b>	67.806	71.359	67.929	71.458	72.63
W/T/L		10/0/0	10/0/0	10/0/0	10/0/0	10/0/0

分别表示WMRI算法与其他特征选择算法的胜/平/负的次数。

从表3可以得出, WMRI算法在10个数据集的平均fmi值是最高(74.082%)。同时, WMRI分别优于IG-RFE, CFR, JMIM, DCSF和MRI为9, 9, 8, 9和9次。在表4中, WMRI算法在10个数据集的平均fmi值也是最高(70.258%)。同时, WMRI分别优于IG-RFE, CFR, JMIM, DCSF和MRI为10, 8, 9, 9和8次。在表5中, WMRI算法在10个数据集的平均fmi值也是最高(76.524%)。同时, WMRI分别优于IG-RFE, CFR, JMIM, DCSF和MRI为10, 10, 10, 10和10次。

通过对表3—表5分析可以得出, 不同分类器表现出的分类结果也不相同。但是, WMRI算法的平均fmi值都是最高。这说明WMRI算法优于其他特征选择算法(IG-RFE, CFR, JMIM, DCSF和MRI)。

为了进一步观察特征子集对fmi值的影响, 图1, 图2和图3分别给出部分不同数据集的fmi性能比较。从图1、图2和图3可以看出, 当数据的维数不

断增加时, 由于WMRI算法通过平均值和标准差动态调整新分类信息项 $I(C; f_k|f_{sel})$ 与保留类别信息项 $I(C; f_{sel}|f_k)$ 的重要程度。结果显示, WMRI算法明显优于MRI算法。例如在图1(b)和图2(b)中, JMIM算法优于MRI算法, 而WMRI算法优于JMIM算法。图3(a)和图3(b), DCSF算法优于MRI算法, 而WMRI算法优于DCSF算法。这些充分说明WMRI算法对分类效果的提升非常明显。并且, WMRI明显优于IG-RFE, CFR, JMIM, DCSF和MRI。

#### 4 结束语

本文提出一种基于过滤式的特征选择方法: 动态加权的最大相关和最大独立分类特征选择算法(WMRI)。该算法旨在解决新分类信息和保留类别信息不同重要度的问题并提高特征子集的数据质量。为了全面验证WMRI算法的有效性, 实验在10个数据集上进行。通过使用分类器(KNN, C4.5和Random Forest)和分类准确率指标(fmi)全面评估所选特征子集的质量。实验结果证明WMRI明显

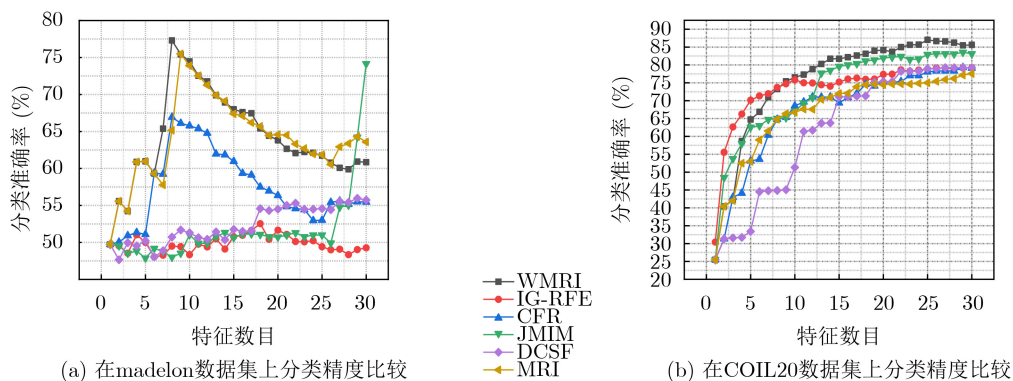


图1 KNN在高维数据集上的性能比较

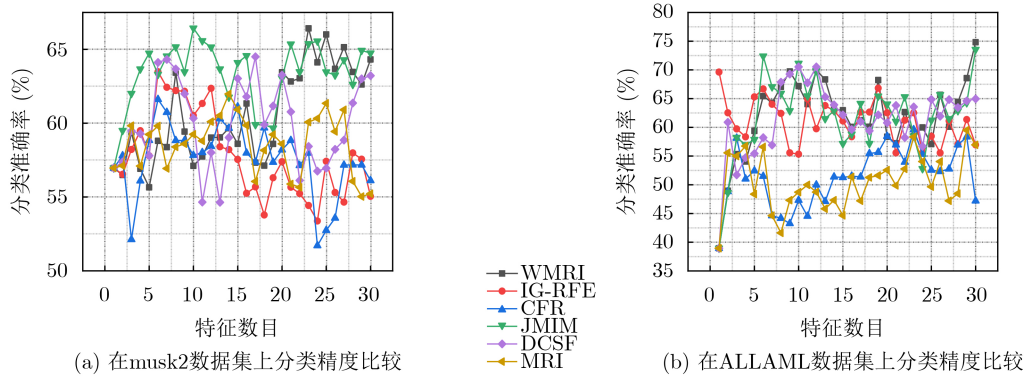


图2 C4.5在高维数据集上的性能比较

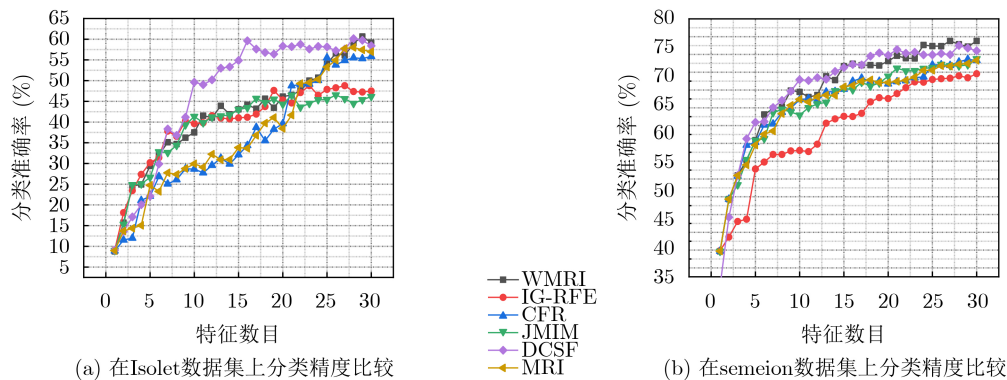


图3 Random Forest在高维数据集上的性能比较

优于MRI, CFR, JMIM, DCSF和IG-RFE等5种特征选择算法, 但WMRI算法有时也会导致特征选择的结果不理想。未来的工作包括进一步改进新分类信息项和保留类别信息项的动态平衡问题并优化WMRI算法, 同时在更广泛的领域中验证所提出的方法。

### 参考文献

- [1] CHE Jinxing, YANG Youlong, LI Li, *et al.* Maximum relevance minimum common redundancy feature selection for nonlinear data[J]. *Information Sciences*, 2017, 409/410: 68-86. doi: [10.1016/j.ins.2017.05.013](https://doi.org/10.1016/j.ins.2017.05.013).
- [2] CHEN Zhijun, WU Chaozhong, ZHANG Yishi, *et al.* Feature selection with redundancy-complementariness dispersion[J]. *Knowledge-Based Systems*, 2015, 89: 203-217. doi: [10.1016/j.knosys.2015.07.004](https://doi.org/10.1016/j.knosys.2015.07.004).
- [3] 张天骐, 范聪聪, 葛宛苑, 等. 基于ICA和特征提取的MIMO信号调制识别算法[J]. *电子与信息学报*, 2020, 42(9): 2208-2215. doi: [10.11999/JEIT190320](https://doi.org/10.11999/JEIT190320).  
ZHANG Tianqi, FAN Congcong, GE Wanying, *et al.* MIMO signal modulation recognition algorithm based on ICA and feature extraction[J]. *Journal of Electronics & Information Technology*, 2020, 42(9): 2208-2215. doi: [10.11999/JEIT190320](https://doi.org/10.11999/JEIT190320).
- [4] ZHANG Yishi, ZHANG Qi, CHEN Zhijun, *et al.* Feature assessment and ranking for classification with nonlinear sparse representation and approximate dependence analysis[J]. *Decision Support Systems*, 2019, 122: 113064.1-113064.17. doi: [10.1016/j.dss.2019.05.004](https://doi.org/10.1016/j.dss.2019.05.004).
- [5] ZENG Zilin, ZHANG Hongjun, ZHANG Rui, *et al.* A novel feature selection method considering feature interaction[J]. *Pattern Recognition*, 2015, 48(8): 2656-2666. doi: [10.1016/j.patcog.2015.02.025](https://doi.org/10.1016/j.patcog.2015.02.025).
- [6] 赵湛, 韩璐, 方震, 等. 基于可穿戴设备的日常压力状态评估研究[J]. *电子与信息学报*, 2017, 39(11): 2669-2676. doi: [10.11999/JEIT170120](https://doi.org/10.11999/JEIT170120).  
ZHAO Zhan, HAN Lu, FANG Zhen, *et al.* Research on daily stress detection based on wearable device[J]. *Journal of Electronics & Information Technology*, 2017, 39(11): 2669-2676. doi: [10.11999/JEIT170120](https://doi.org/10.11999/JEIT170120).
- [7] BROWN G, POCOCK A, ZHAO Mingjie, *et al.* Conditional likelihood maximisation: A unifying framework for information theoretic feature selection[J]. *The Journal of Machine Learning Research*, 2012, 13(1): 27-66.
- [8] MACEDO F, OLIVEIRA M R, PACHECO A, *et al.* Theoretical foundations of forward feature selection methods based on mutual information[J]. *Neurocomputing*, 2019, 325: 67-89. doi: [10.1016/j.neucom.2018.09.077](https://doi.org/10.1016/j.neucom.2018.09.077).
- [9] GAO Wanfu, HU Liang, ZHANG Ping, *et al.* Feature selection by integrating two groups of feature evaluation criteria[J]. *Expert Systems with Applications*, 2018, 110:

- 11–19. doi: [10.1016/j.eswa.2018.05.029](https://doi.org/10.1016/j.eswa.2018.05.029).
- [10] 肖利军, 郭继昌, 顾翔元. 一种采用冗余性动态权重的特征选择算法[J]. 西安电子科技大学学报, 2019, 46(5): 155–161. doi: [10.19665/j.issn1001-2400.2019.05.022](https://doi.org/10.19665/j.issn1001-2400.2019.05.022).  
XIAO Lijun, GUO Jichang, and GU Xiangyuan. Algorithm for selection of features based on dynamic weights using redundancy[J]. *Journal of Xidian University*, 2019, 46(5): 155–161. doi: [10.19665/j.issn1001-2400.2019.05.022](https://doi.org/10.19665/j.issn1001-2400.2019.05.022).
- [11] WANG Xinzhen, GUO Bing, SHEN Yan, *et al.* Input feature selection method based on feature set equivalence and mutual information gain maximization[J]. *IEEE Access*, 2019, 7: 151525–151538. doi: [10.1109/ACCESS.2019.2948095](https://doi.org/10.1109/ACCESS.2019.2948095).
- [12] GAO Wanfu, HU Liang, and ZHANG Ping. Class-specific mutual information variation for feature selection[J]. *Pattern Recognition*, 2018, 79: 328–339. doi: [10.1016/j.patcog.2018.02.020](https://doi.org/10.1016/j.patcog.2018.02.020).
- [13] WANG Jun, WEI Jinmao, YANG Zhenglu, *et al.* Feature selection by maximizing independent classification information[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(4): 828–841. doi: [10.1109/TKDE.2017.2650906](https://doi.org/10.1109/TKDE.2017.2650906).
- [14] GAO Wanfu, HU Liang, ZHANG Ping, *et al.* Feature selection considering the composition of feature relevancy[J]. *Pattern Recognition Letters*, 2018, 112: 70–74. doi: [10.1016/j.patrec.2018.06.005](https://doi.org/10.1016/j.patrec.2018.06.005).
- [15] LIN Xiaohui, LI Chao, REN Weijie, *et al.* A new feature selection method based on symmetrical uncertainty and interaction gain[J]. *Computational Biology and Chemistry*, 2019, 83: 107149. doi: [10.1016/j.compbiolchem.2019.107149](https://doi.org/10.1016/j.compbiolchem.2019.107149).
- [16] BENNASAR M, HICKS Y, and SETCHI R. Feature selection using joint mutual information maximisation[J]. *Expert Systems with Applications*, 2015, 42(22): 8520–8532. doi: [10.1016/j.eswa.2015.07.007](https://doi.org/10.1016/j.eswa.2015.07.007).
- [17] LYU Hongqiang, WAN Mingxi, HAN Jiuqiang, *et al.* A filter feature selection method based on the maximal information coefficient and Gram-Schmidt orthogonalization for biomedical data mining[J]. *Computers in Biology and Medicine*, 2017, 89: 264–274. doi: [10.1016/j.combiomed.2017.08.021](https://doi.org/10.1016/j.combiomed.2017.08.021).
- [18] SHARMIN S, SHOYAIB M, ALI A A, *et al.* Simultaneous feature selection and discretization based on mutual information[J]. *Pattern Recognition*, 2019, 91: 162–174. doi: [10.1016/j.patcog.2019.02.016](https://doi.org/10.1016/j.patcog.2019.02.016).
- [19] SUN Guanglu, LI Jiabin, DAI Jian, *et al.* Feature selection for IoT based on maximal information coefficient[J]. *Future Generation Computer Systems*, 2018, 89: 606–616. doi: [10.1016/j.future.2018.05.060](https://doi.org/10.1016/j.future.2018.05.060).
- [20] PENG Hanchuan, LONG Fuhui, and 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. doi: [10.1109/TPAMI.2005.159](https://doi.org/10.1109/TPAMI.2005.159).
- [21] MEYER P E, SCHRETTTER C, and BONTEMPI G. Information-theoretic feature selection in microarray data using variable complementarity[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2008, 2(3): 261–274. doi: [10.1109/JSTSP.2008.923858](https://doi.org/10.1109/JSTSP.2008.923858).
- [22] SPEISER J L, MILLER M E, TOOZE J, *et al.* A comparison of random forest variable selection methods for classification prediction modeling[J]. *Expert Systems with Applications*, 2019, 134: 93–101. doi: [10.1016/j.eswa.2019.05.028](https://doi.org/10.1016/j.eswa.2019.05.028).
- 张 俐: 男, 1977年生, 博士, 副教授, 研究方向为特征选择和机器学习.
- 陈小波: 女, 1980年生, 硕士, 工程师, 研究方向为特征选择和机器学习.

责任编辑: 马秀强