

多准则赋权排序与C-SVM相结合的特征选择算法

孙 勤^{1,2}, 蒋艳凰^{1,2}, 胡 维², 张 毅², 高 峰³

SUN Qin^{1,2}, JIANG Yanhuang^{1,2}, HU Wei², ZHANG Yi², GAO Feng³

1. 国防科学技术大学 高性能计算国家重点实验室, 长沙 410073

2. 国防科学技术大学 计算机学院, 长沙 410073

3. 中国人民解放军 91550 部队

1. State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, China

2. College of Computer, National University of Defense Technology, Changsha 410073, China

3. Unit 91550 of PLA, China

SUN Qin, JIANG Yanhuang, HU Wei, et al. Feature selection algorithm based on multi-criterion ranking and C-SVM. Computer Engineering and Applications, 2018, 54(3): 125-130.

Abstract: Large-scaled and multi-dimension in data mining which may increase the storage of system, lead to the waste of time and low-accuracy precision. A new feature selection approach mCRC based on multi-criterion ranking and C-SVM is introduced in this paper towards the defects of present feature selection such as low accuracy and undeterminable amounts of optimal features. mCRC computes the dependencies between features and labels through mutual information and class distances, meanwhile deletes irrelevant and completely redundant attributes, and then all features based on dependencies and redundancies are ranked. For the sake of obtaining the optimal subset, it uses C-SVM classifiers to filtrate the sorted features via Sequential Forward Floating Selection (SFFS). Compared with traditional algorithm, mCRC algorithm achieves high accuracy with less training time, it's able to provide a strong guarantee for quick and efficient massive data mining.

Key words: feature selection; maximum relevance minimum redundancy; multi-criterion ranking; C-SVM; Sequential Forward Floating Selection(SFFS)

摘 要: 数据挖掘中所获取的数据维数多, 常常导致数据存储所需容量大, 知识挖掘所需时间长, 预测正确率不高等问题, 特征选择是解决上述问题的重要方法之一。针对现有特征选择算法最佳特征个数难以确定及分类准确率有待进一步提高等问题, 提出一种同时考虑相关性和冗余度的多准则赋权排序的算法(mCRC), mCRC 结合两种准则同时对特征进行排序, 并利用 C-SVM 对按重要性降序排好的特征采用顺序前向浮动搜索得出最佳特征子集。实验结果表明, mCRC 算法与单独基于互信息或类别可分性赋权排序的特征选择方法相比能在更短的时间内获得分类性能更好的最佳特征子集, 为快速并高效地对数据集进行挖掘提供了有力保障。

关键词: 特征选择; 最大相关最小冗余; 多准则赋权排序; C-支持向量机; 顺序前向浮动搜索

文献标志码: A **中图分类号:** TP181 **doi:** 10.3778/j.issn.1002-8331.1608-0326

1 引言

特征选择是数据挖掘中特征降维的重要方式之一^[1], 通过特征选择能适当去除无关和冗余特征, 降低系

统的存储负担和不必要的模型训练开销, 同时能提高学习算法的性能。与特征提取相比, 特征选择是从原始特征集中选出一组最能替代原始特征的子集达到降维的

基金项目: 国家自然科学基金(No.61120106005, No.61272141)。

作者简介: 孙勤(1992—), 女, 硕士研究生, 主要研究领域为机器学习, E-mail: sunqin14@nudt.edu.cn; 蒋艳凰(1976—), 女, 博士, 副研究员, 主要研究领域为机器学习、高性能计算; 胡维(1982—), 男, 博士研究生, 主要研究领域为机器学习; 张毅(1994—), 男, 硕士研究生, 主要研究领域为网络安全; 高峰(1982—), 男, 工程师, 主要研究领域为机器学习。

收稿日期: 2016-08-15 **修回日期:** 2016-09-28 **文章编号:** 1002-8331(2018)03-0125-06

CNKI 网络优先出版: 2017-03-22, <http://kns.cnki.net/kcms/detail/11.2127.TP.20170322.1930.026.html>

目的,它不改变原始集合的物理特性,能在降维的同时真正降低系统的存储负担和时间开销。

根据特征选择是否与后续学习算法相关,可将其分为两类:过滤式(Filter)^[2]和封装式(Wrapper)^[3]特征选择算法。过滤式特征选择独立于学习过程,它直接利用原始数据集的统计特性进行评估从而选出合适的子集,其效率高,但评估准确度低;使用过滤式方法力求某种评价准则最优^[4],而对于不同的特征集,很难确定最适合的评价准则。目前已有的基于过滤式算法的评价标准有信息度量^[5-6]、距离度量^[6-8]、一致度量等^[8-9]。不同评价方法的复杂度和分类精度各不相同,不同的准则着重于不同的层面反应特征的优劣,单一的准则往往无法全面评估特征集的好坏^[10]。封装式算法直接利用学习算法的准确率评价特征子集的好坏,其评估准确率高但是运行速度慢,特征越多,运行速度越慢。

针对两类特征选择算法各自存在的优缺点,本文提出一种过滤式和封装式相组合的特征选择方法 mCRC (Feature Selection Algorithm Based on multi-Criterion-Ranking and C-SVM),使用两种准则同时对特征评估赋权,剔除完全冗余和不相关特征,得到重要程度权重降序排列的特征子集,再通过 C-SVM 分类器^[11]采用顺序前向浮动搜索算法(Sequential Forward Floating Selection, SFFS)^[12]搜索分类精度高的特征,得到最佳特征子集,从而补充特征选择两大类算法存在的不足。

2 相关定义

设特征空间 D 由 n 个特征组成,这 n 个特征组成原始特征集 $F = \{f_1, f_2, \dots, f_n\}$, D 中的样本分成 m 类。在有监督特征选择中通常以特征集 F 为研究对象,以 F 中各元素与类别的相关程度以及 F 中各元素之间的冗余度综合衡量某个特征的好坏。而不同的评价准则也会带来差异较大的衡量结果。

2.1 互信息

互信息^[5-6]是用来度量两个变量的统计相关性的重要方法,也是目前特征选择中普遍运用的评价准则。互信息的计算公式为:

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (1)$$

其中, $H(X)$ 和 $H(Y)$ 为 X 和 Y 的信息熵, $H(X, Y)$ 为两者的联合熵。设 $p(x, y)$ 为 (X, Y) 的联合概率,则:

$$H(X) = -\sum_{i=1}^n p(x_i) \times \lg p(x_i) \quad (2)$$

$$H(X, Y) = -\sum_{i=1}^n p(x_i, y_i) \times \lg p(x_i, y_i) \quad (3)$$

类别与各特征之间的互信息为:

$$I(C, f_i) = H(C) + H(f_i) - H(C, f_i) \quad (4)$$

特征两两之间的互信息为:

$$I(f_i, f_j) = H(f_i) + H(f_j) - H(f_i, f_j) \quad (5)$$

通过公式(2)可以计算特征与类别之间的关联程度, I 值越大,特征与类别标签的相关性程度越高; $I=0$ 时特征与类别不相关,可以剔除。

通过公式(3)可以计算特征两两之间的相关性, I 值越大,说明特征之间的冗余度越高;当 f_i 和 f_j 完全冗余时 $H(f_i) = H(f_j) = H(f_i, f_j)$, 则二者保留其一即可。

2.2 距离度量

距离度量也可以称作类别可分性度量^[7-8]。对于有监督分类,不同类别间距越大类别的相似程度越低,可区分的概率越大,相同类别之间距离越小相似性越大,可区分性越小,分类准确率越高。基于距离度量的特征子集评价能有效提高小样本与线性不可分数据集上特征选择的能力。特征选择应该选择类间间距大且类内间距小的特征。

设第 j 维特征 f_j 包含 N 个样本,这 N 个样本的平均值为 $avg(j)$, 则:

$$avg(j) = \frac{1}{N} \sum_{i=1}^N x_i(j) \quad (6)$$

其中 $x_i(j)$ 为第 t 个样本的第 j 维特征。

设第 c' 类样本个数为 M , 则第 c' 类样本的第 j 维特征的特征均值为 $avg_{c'}(j)$, 则

$$avg_{c'}(j) = \frac{1}{M} \sum_{i=1}^M x_i^{c'}(j) \quad (7)$$

其中 $x_i^{c'}(j)$ 为第 c' 类第 t 个样本的第 j 维特征,则第 c' 类样本的第 j 维特征的类内距离为:

$$\sigma_{c'}^2(j) = \sum_{i=1}^M |x_i^{c'}(j) - avg_{c'}(j)|^2 \quad (8)$$

第 j 维特征的类间距离为:

$$\sigma_{\text{类间}}^2(j) = \sum_{c'=1}^m |avg(j) - avg_{c'}(j)|^2 \quad (9)$$

则第 j 维特征 f_j 的类别可分性度量可表示为:

$$W(f_j) = \frac{\sigma_{\text{类间}}^2(j)}{\sum_{c'=1,2,\dots,m} \sigma_{c'}^2(j)} \quad (10)$$

可用 $W(f_j)$ 来衡量特征的分类能力。 $W(f_j)$ 越大,特征的分类准确率越高。

2.3 C-支持向量机

在机器学习中,C-支持向量机(C-SVM)^[11]是一种用于分类的监督式学习算法,其基本思想是样本点在线性不可分的情况下,引入松弛变量 ξ 和从输入空间 R^n 到 Hilbert 空间 Hil 的变换:

$$\Phi: \chi \subset R^n \rightarrow \text{Hil}$$

$$x \rightarrow \chi = \Phi(x)$$

对于两类问题 $y_i \in \{-1, 1\}, i = 1, 2, \dots, N$, C-SVM 算法将训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

映射为:

$$\bar{T} = \{(\chi_1, y_1), \dots, (\chi_N, y_N)\} = \{(\Phi(x_1), y_1), \dots, (\Phi(x_N), y_N)\}$$

这样便得到原始问题:

$$\min_{\omega, b, \xi} \frac{1}{2} \|\omega\|^2 + \delta \sum_{i=1}^N \xi_i$$

$$\text{s.t. } y_i [\omega \cdot \Phi(x_i) + b] \geq 1 - \xi_i, 1 \leq i \leq N, \xi_i \geq 0 \quad (11)$$

其中 δ 为惩罚参数, δ 越大表示对错误分类的惩罚越大且 $\delta > 0$ 。

对于多分类问题, C-SVM 通常转化为多个两类问题求解。C-SVM 作为迄今为止最小错分率和最大泛化能力的分类器, 应用于特征选择能用来分辨出分类正确率高的特征。因此, 本文采用 C-SVM 作为分类器对特征集分类, 筛选出分类精确度高的特征, 从而得到最佳特征子集。

3 mCRC 特征选择算法

过滤式特征选择中的互信息和距离两种度量方法各有优缺点, 若将两种评价准则结合起来共同对特征进行评价能够综合各方法的优势, 能获得比单个评价准则更高的精度。基于以上阐述, 本文提出了多准则赋权排序的过滤式算法和 C-SVM 互补的特征选择算法 mCRC。

3.1 特征排序与筛选

通过互信息和距离两种度量准则同时计算原始特征集 S 中各特征与类别 C 的相关性, 求得各特征的均值权重 $\overline{W}(f_i)$, 在此过程可删除互信息为零, 即与类别不相关的特征, 得到去掉不相关特征且含均值权重的特征子集 U ; 对于集合 U , 本文设定一个初始空集 Q , 从 U 每次选择一个特征放入 Q 中。在第一步, 使得

$$g_1 = \arg \max_{1 \leq i \leq n} \{\overline{W}(f_i)\} \quad (12)$$

$g_j \in Q$, 对于 U 中未被选中的特征, 取任意 g_j 若 $H(f_i) = H(g_j) = H(f_i, g_j)$, 则 f_i 与 g_j 完全冗余, 从 U 中删除 f_i , 否则, 对于 f_i 取互信息最大值 $I_{\max}(f_i, g_j)$ 表示其与子集 Q 的冗余度, 通过最大相关最小冗余标准来评估特征的重要性, 并选择当前重要程度最高的特征放入集合 Q 中, 评价公式为:

$$J(f_i) = \frac{\overline{W}(f_i)}{I_{\max}(f_i, g_j)} \quad (13)$$

其中 Q 的第 l (l 为 Q 的大小) 个特征被选依据为:

$$g_l = \arg \max_{1 \leq i \leq n} \{J(f_i)\} \quad (14)$$

3.2 C-SVM 封装及最佳特征子集的确定

在过滤排序阶段, mCRC 算法删除了不相关的和完

全冗余的特征, 并将特征进行重要性降序排序, 使得在封装阶段可以直接利用 C-SVM 对过滤阶段的排序结果采用 SFSS 策略选择最佳特征子集, 在搜索之前设定阈值 λ ($\lambda > 0$, 且 $\lambda = 1$ 时退化为顺序前向搜索), 当搜索到某个特征时子集分类精度上升, 则将该特征加入最佳子集中, 若搜索到某个特征之后的连续 λ 个特征都使得子集的分类精度均不大于这个特征的精度, 则停止搜索。这种方式能够较快地确定特征选择的结果。相比传统的 Wrapper 算法利用分类器对整个原始特征集进行评估再做特征选择, mCRC 算法节省了特征选择所需的时间。另过滤器排序部分多准则同时评估与单准则评估相比大大提高了权重的可靠性。

3.3 mCRC 算法描述

根据以上描述, 给出多准则赋权排序与 C-SVM 相结合的特征选择算法 (mCRC) 的伪代码描述:

输入: 训练集 F , 阈值 λ ($\lambda > 1$)

输出: 最佳特征子集 A

1. initialization; $U = \emptyset$; $Q = \emptyset$; $R = \emptyset$; $k = 0$; $count = 0$

2. for each $f_i \in F$ do

3. calculate $I(C, f_i)$, $W(f_i)$ by (4) and (10);

4. if $I \neq 0$;

5. $U \leftarrow f_i$; $\overline{W}(f_i) = \frac{1}{2}(I(C, f_i) + W(f_i))$;

6. find g_1 by (12); $g_1 \in Q$; $U = U - g_1$;

7. for all $f_i \in U$

8. for all $g_r \in Q$, calculate $I(f_i, g_r)$ by (5);

9. if $H(f_i) = H(g_j) = H(f_i, g_j)$; then

10. delete f_i ;

11. else calculate $J(f_i)$ by (13);

12. find g_l by (14); $g_l \in Q$; $U = U - g_l$;

13. if $U \neq \emptyset$; Skip to step 7;

14. for $i = 1$ to l // l is the size of Q ;

15. $\epsilon_i = \text{learn}(A \cup g_i)$; // ϵ is the accuracy

16. if $\epsilon_{i-1} < \epsilon_i$; $A = A \cup g_i$;

17. else $\epsilon_i \leq \epsilon_{i-1}$; $count++$;

18. if $count \geq \lambda$; break;

19. return A

4 实验设置与分析

本文采用 UCI^[13] 机器学习知识库中的数据集进行测试, 表 1 给出了实验所用数据集的基本信息。实验环境为: Intel Core i5-6400 2.70 GHz CPU, 8 GB 内存, Win10 64 位操作系统, Matlab R2010a。

为了使实验结果更具统计意义, 采用 5 次交叉验证^[14]实验, 同时实验采用 LibSVM 工具箱^[15], C-SVM 的核函数采用径向核函数 RBF, 并采用网格搜索来选取最优惩罚因子 δ 和核函数参数 g 。

表1 实验数据集基本情况表

数据集	特征总数	样本个数	类别数	数据类型
breast	9	277	2	连续
glass	9	214	2	连续
dna	180	2 000	3	离散
vehicle	18	846	4	连续
somar	60	208	2	连续
segment	18	2 310	7	离散/连续
ionosphere	34	351	2	离散/连续
landsat	36	2 000	6	连续

$$\text{RBF: } K(x, x') = \exp(-\|x - x'\|^2) / g^2,$$
$$\delta \in \{2^{-5}, \dots, 2^5\}, g \in \{2^{-5}, \dots, 2^5\}$$

实验分为两部分,第一部分,分别对mCRC算法、基于互信息排序的mRMR算法^[5]及基于类别可分性排序的CSB算法^[7]的排序部分进行性能比较,用三种算法将特征进行排序后再用C-SVM对去除了完全冗余和完全不相关并按重要性降序排好的所有特征逐一进行子集分类精度测试,检验各算法能否识别重要特征。图1给出

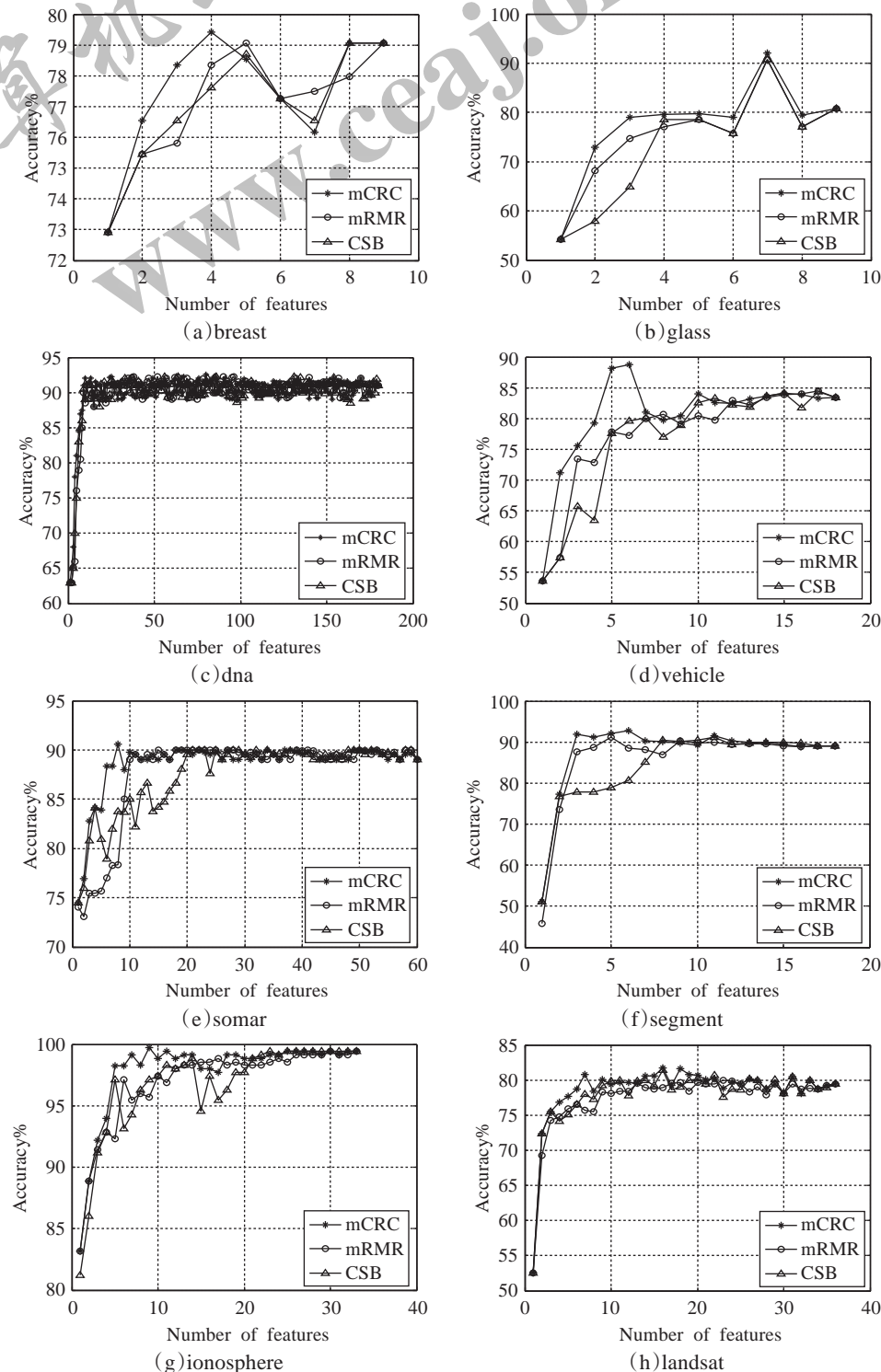


图1 mCRC、mRMR、CSB在各数据集上排序性能比较图

了三种算法的排序阶段在各数据集上的实验结果。

由图 1 可知,对于表 1 中的 8 个数据集,三种算法都表现出随着特征数目的增多,分类精度先不断上升,随后保持相对稳定或下降状态,表明通过 mCRC、mRMR、CSB 三种算法各自的过滤排序部分对特征进行重要性排序之后可以通过少量特征就达到与全集相同甚至高于全集的分类率,说明三种特征选择算法的过滤排序部分都基本能够识别重要特征。同时几乎在所有数据集上 mCRC 都可以在比其余两种算法少的特征数目下达到甚至高于初始数据集的分类准确率。但是,从图 1 中也可以看出,三种算法并不是随特征数据一直递增到某个精度后再稳定或下降,在精度上升过程中个别特征的加入会使得算法的性能突然下降再上升,说明排序过程存在某些误差,因此,本文不用顺序前向搜索算法来获取最佳特征子集,而是通过采用设置一个 λ 值的顺序前向浮动搜索的方法来得到最佳特征子集,从而避免采用顺序前向搜索时精度在上升过程中因个别特征的加入降低特征集的分类性能而停止搜索达不到最佳性能的情况。 λ 值不宜设的过大,过大的 λ 值相当于在整个特征集上进行顺序浮动前向搜索而失去的特征排序的意义。

实验第二部分采用 mCRC、mRMR、CSB 三种特征选择算法的排序部分分别对特征排序后再通过 C-SVM 和设阈值 λ 的 SFFS 进行性能比较。 λ 取 2 和 3 进行两次实验,实验结果如表 2、表 3 所示。

表 2 中 λ 等于 2 时的 SFFS 结果显示:mCRC 在 6 个特征集上有规模最小的最佳属性子集,在 7 个数据集上

的最佳属性子集的分类性能高于其余两种算法;mRMR 在 2 个数据集上有最小约减的最佳特征子集,但相比其余两种算法,mRMR 在所有数据集上的最佳特征子集的分类性能最低;CSB 在 3 个数据集上有最小约减的最佳属性子集,相比其余二者,CSB 在 1 个数据集上最佳属性子集的分类精度最高。从效率来看,mCRC 在 4 个数据集上有最高效率,而 mRMR 只有一个,CSB 只有两个。

表 3 取 λ 等于 3 时的 SFFS 结果与表 2 相比精度总体稍有变化,在多个数据集上 mCRC 的精度仍然最高,说明 λ 变化,mCRC 的分类性能仍然最优;相比表 2,表 3 中 λ 取 3 使得各算法在个别数据集上对低精度的特征个数容忍度稍微增大,但在所有数据集上运行时间都有所增长,由此得出对于效率要求高的学习算法, λ 的取值应尽量小。总体来看,在表 3 中,mCRC 无论在从最佳子集规模、分类精度还是运行时间来看,mCRC 算法都占优势。

表 2、表 3 表明无论在效率、最佳特征子集大小还是分类精度上,mCRC 的性能都优于 mRMR 和 CSB,说明基于多准则赋权排序选择的最佳特征子集不仅规模小,泛化能力好,而且运行效率高。对于时间性要求很高的学习算法,数据处理的速度稍有提高都能大大提高学习算法的性能。而对于所有学习算法,数据集的分类精度越高越好。本文提出的 mCRC 算法能兼顾快速识别重要特征、得出小规模的最佳特征子集的同时减轻计算机的存储负担。

表 2 λ 取 2 时 mCRC、mRMR、CSB 的性能比较

数据集	原始特征数	被选特征数			分类正确率/%			运行时间/s		
		mCRC	mRMR	CSB	mCRC	mRMR	CSB	mCRC	mRMR	CSB
breast	9	4	5	4	79.422 4	79.061 4	78.700 4	22.400 6	23.393 3	17.904 1
glass	9	6	7	7	92.990 7	85.046 7	87.850 5	31.475 2	32.000 2	32.614 6
dna	180	9	9	12	93.782 2	92.363 8	91.942 0	11 880.870 0	11 899.550 0	11 897.990 0
vehicle	18	6	4	11	88.770 7	78.014 2	84.633 6	328.465 6	281.100 4	661.947 9
somar	60	6	8	4	92.134 6	89.115 4	84.134 6	80.514 2	94.319 4	81.206 2
Segment	18	3	5	8	91.978 0	91.098 9	93.406 6	92.362 8	160.280 0	180.094 0
ionosphere	34	6	8	5	96.866 1	96.832 4	96.832 4	47.757 9	70.101 5	41.103 2
landsat	36	9	9	10	98.621 6	97.368 4	97.493 7	319.757 5	287.404 5	353.246 3

表 3 λ 取 3 时 mCRC、mRMR、CSB 的性能比较

数据集	原始特征数	被选特征数			分类正确率/%			运行时间/s		
		mCRC	mRMR	CSB	mCRC	mRMR	CSB	mCRC	mRMR	CSB
breast	9	4	5	5	79.422 4	79.061 4	79.061 4	28.170 9	27.911 5	30.078 7
glass	9	6	7	7	92.990 7	85.046 7	87.850 5	31.517 8	32.003 0	32.557 6
dna	180	9	9	12	93.782 2	92.363 8	91.942 0	11 910.870 0	11 958.460 0	10 897.780 0
vehicle	18	8	5	12	97.163 1	88.770 7	91.725 8	616.847 2	439.656 3	815.903 4
Somar	60	6	8	6	92.134 6	89.115 4	95.673 1	81.838 8	96.741 8	92.909 0
Segment	18	3	5	8	91.978 0	91.098 9	93.406 6	107.308 1	242.967 8	197.278 1
ionosphere	34	7	9	5	99.430 2	96.832 4	96.832 4	64.868 5	74.569 9	44.557 2
landsat	36	9	9	10	98.621 6	97.368 4	97.493 7	336.832 3	323.058 1	380.877 2

5 结论

本文提出了一种基于两种准则赋权排序的过滤式和C-SVM混合的特征选择方法,该方法使用两种准则同时评估特征与类别的相关度及特征间的冗余度来对特征进行赋权排序,并使用C-SVM对按重要性降序排列的特征子集通过设容忍度阈值 λ 的顺序前向浮动搜索策略(SFFS)搜索最优特征子集。实验结果表明,mCRC能够识别重要特征,相较于单准则赋权排序,mCRC能在相对较短的时间内得到分类精度更高的特征子集,并较大力度地降低了数据集的特征数,为快速并高效地对海量及高维数据进行挖掘提供了有力保障。

本文中两种准则对特征进行赋权后采用平均值得到特征的最终权重,这样面临某个准则对某一特征集的评估很好,而另一准则对此特征集的评估效果很差,取平均权重会使得效果较差的准则拉低效果较好的准则的评估性能,如表2中的数据集segment。下一步,计划采取某种方式对两种准则进行加权,使得对某个数据集评估效果较好的准则占较重权重,评估效果较差的准则占较小权重,从而使得本文提出的mCRC特征选择算法更具适应性。

参考文献:

- [1] Kantardzic M, 坎塔尔季奇, 王晓海, 等. 数据挖掘: 概念, 模型, 方法和算法[M]. 北京: 清华大学出版社, 2013.
- [2] Saeys Y, Inza I, Larrañaga P. WLD: review of feature selection techniques in bioinformatics[J]. Bioinformatics, 2007, 23(19): 2507-2517.
- [3] John G H, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem[C]//11th International Conference on Machine Learning, 1994: 121-129.
- [4] Li Y, Lu B L. Feature selection based on loss-margin of nearest neighbor classification[J]. Pattern Recognition, 2009, 42(9): 1914-1921.
- [5] 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.
- [6] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述[J]. 控制与决策, 2012, 27(2): 161-166.
- [7] 张岐龙, 单甘霖, 段修生, 等. 基于特征空间中类别可分性判据的特征选择[J]. 火力与指挥控制, 2010, 35(6): 118-120.
- [8] 任双桥, 傅耀文, 黎湘, 等. 基于分类间隔的特征选择算法[J]. 软件学报, 2008, 19(4): 842-850.
- [9] Dash M, Liu H. Consistency-based search in feature selection[J]. Artificial Intelligence, 2003, 151(1/2): 155-176.
- [10] 李勇明, 张素娟, 曾孝平, 等. 轮询式多准则特征选择算法的研究[J]. 系统仿真学报, 2009, 21(7): 2010-2013.
- [11] 邓乃扬, 田英杰. 数据挖掘中的新方法: 支持向量机[M]. 北京: 科学出版社, 2004.
- [12] Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection[J]. Pattern Recognition Letters, 1994, 15(11): 1119-1125.
- [13] Blake C. UCI repository of machine learning databases[C]// Neural Information Processing Systems, 1998.
- [14] Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation[J]. Journal of Machine Learning Research, 2004, 5(3): 1089-1105.
- [15] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 389-396.
- [12] 李勇平. 基于改进粒子群神经网络的电信业务预测模型研究[D]. 广州: 华南理工大学, 2009.
- [13] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals, and Systems, 1989, 4(2): 303-314.
- [14] 杨帆, 胡春平, 颜学封. 基于蚁群系统的参数自适应粒子群算法及其应用[J]. 控制理论与应用, 2010, 27(11): 1479-1488.
- [15] Xie Shanyi, Zhai Ruicong, Liu Xianhu, et al. Self-adaptive genetic algorithm and fuzzy decision based multi-objective optimization in microgrid with DGs[J]. Open Electrical and Electronic Engineering Journal, 2016, 10(1): 46-57.
- [16] 李宏. 自适应粒子群优化算法及其在图像分割中的应用[D]. 辽宁大连: 大连理工大学, 2006.
- [17] Dubey H M, Pandit M, Panigrahi B K. A biologically inspired modified flower pollination algorithm for solving economic dispatch problem in modern power system[J]. Cognitive Computation, 2015, 7(5): 1-15.
- [18] 吴亮红, 王耀南, 袁小芳, 等. 基于快速自适应差分进化算法的电力系统经济负荷分配[J]. 控制与决策, 2013, 28(4): 557-562.
- [19] University of California Irvine. UCI machine learning repository[EB/OL]. (2016-09-12). <http://archive.ics.uci.edu/ml/>.

(上接56页)