



\* 文章编号:1000-5811(2017)01-0178-05

# 基于 AdaBoost-SVM 的葡萄酒品质 分类模型优化设计

杨 云<sup>1</sup>, 卢美静<sup>1</sup>, 穆天红<sup>2</sup>

(1. 陕西科技大学 电气与信息工程学院, 陕西 西安 710021; 2. 青海省农牧业市场信息中心, 青海 西宁 810008)

**摘 要:**针对传统葡萄酒品质分类中低品质类葡萄酒样本识别率低的问题,提出一种基于集成支持向量机的葡萄酒品质分类优化算法. 首先,通过“一对多”支持向量机实现多分类;其次,把支持向量机作为基分类器,反复训练支持向量机分类样本,通过 AdaBoost 得到多个支持向量机基分类器组合的强分类器,运用 AdaBoost 算法动态调整样本权值,适当提高低品质类样本权重,使低品质类中错判的样本代价增大,从而改进不平衡样本分类性能;最后,以 Wine Quality 数据集为研究对象,建立以多分类器优化集成为核心的葡萄酒品质分类模型. 仿真结果表明,与传统的 SVM 算法相比,所提方法显著提高了低品质类葡萄酒分类精度.

**关键词:**分类; 支持向量机; 集成学习; 葡萄酒品质; 不平衡数据

中图分类号:TP391

文献标志码: A

DOI:10.19481/j.cnki.issn1000-5811.2017.01.032

## Optimal design of wine quality classification model based on AdaBoost-SVM

YANG Yun<sup>1</sup>, LU Mei-jing<sup>1</sup>, MU Tian-hong<sup>2</sup>

(1. College of Electrical and Information Engineering, Shaanxi University of Science &amp; Technology, Xi'an 710021, China; 2. Qinghai Agriculture and Animal Husbandry Market Information Center, Xining 810008, China)

**Abstract:** Focused on the issue that traditional classification algorithms for wine quality classification have a low recognition rate to low-quality wines, an optimization algorithm based on ensemble Support Vector Machine (SVM) was proposed. Firstly, multi-class was accomplished by 1-against-the rest SVM; Secondly, SVM was repeatedly trained as weaker classifier and a strong classifier was gotten by grouping a number of base classifiers based on SVM. The sample weight were dynamically adjusted by using AdaBoost algorithm, the sample weight of low quality were appropriately increased, and then the cost of misjudge samples was also increased for improving classification performance of unbalanced datasets; Finally, the wine quality datasets of UCI database was taken as research object, the classification

\* 收稿日期:2016-08-19

基金项目:陕西省科技厅社会发展科技攻关计划项目(2015SF277, 2016SF-444); 陕西省科技厅科学技术研究发展计划项目(2014K15-03V06); 西安市科技计划项目(NC1403(2), NC1319(1))

作者简介:杨 云(1965—),女,山东青岛人,教授,博士,研究方向:嵌入式应用、材料计算机应用

model of wines quality was established that using multi-classifiers optimal integration as the core. The simulation results show that compared with the standard SVM algorithm, classification accuracy of low quality wine was significantly improved based on AdaBoost-SVM.

**Key words:** classification; support vector machine; ensemble learning; wine quality; unbalanced data

## 0 引言

随着我国市场经济的蓬勃发展和人们对葡萄酒消费需求的迅速增加,葡萄酒行业正处于一个高速发展的时期<sup>[1]</sup>.葡萄酒品质一直以来都是消费者以及生产者最关心的问题,尤其是对于中高档的葡萄酒来说,葡萄酒品质高低近乎决定了其价值的高低<sup>[2]</sup>.传统葡萄酒品质分类方法主要通过感官分析法,但这种方法易受到评酒人员嗜好、经验等因素的影响,通常使得评定存在一定主观性和不确定性<sup>[3]</sup>.因此有学者研究基于数据挖掘的葡萄酒品质分类方法.最初,Cortez 等<sup>[4]</sup>提出采用支持向量分类算法建立葡萄酒品质分类模型.之后,徐海涛<sup>[5]</sup>在 Cortez 工作基础上进行了优化,提出采用改进的近似支持向量机方法进行葡萄酒品质分类,其目标是改进其分类精度;刘延玲<sup>[6]</sup>提出基于一种新的 Hopfield 神经网络分类模型,使得该模型能够在较短收敛时间内实现葡萄酒品质分类.

现有研究虽然能够取得较高分类准确率,但没有考虑样本数据分布不平衡的特点,忽略了低品质类葡萄酒样本分类精度.实际在葡萄酒品质分类过程中,高品质葡萄酒被误判为低品质葡萄酒所产生的成本远小于忽视低品质问题而造成的损失.本文采用与上述文献同样的数据集,充分考虑葡萄酒样本数据分布,以提高低品质类葡萄酒识别率为目标,提出一种基于集成学习改进的支持向量机葡萄酒品质分类优化模型.提取样本数据特征,建立数据分析模型,通过实验验证,该方法可以准确识别葡萄酒品质类别,有效提高低品质葡萄酒分类精度.

## 1 不平衡数据分类

不平衡数据是指样本类之间呈现不相等的分布,其中某些类的样本数远远多于其他类的样本数.通常标准的分类算法是偏向样本数量多的类,没有将数据分布考虑在内并且使用总体精度作为训练目标<sup>[7]</sup>.在最坏的情况下,少样本类会被视为多样本类的异常点而被忽略,从而导致少样本类分类性能下降.

随着集成学习理论体系的不断发展完善,集成学习方法通过使用不同权重加权的投票策略,有效解决了非平衡数据集分类问题<sup>[8]</sup>.对于多类不平衡数据集分类问题,可以通过分解策略将多类分类问题转化为更容易解决的二分类问题<sup>[9]</sup>.文中提出采用一对多支持向量机方法,在一类样本与剩余的多类样本之间构造分类平面,从而达到多类识别的目的.

## 2 理论基础

### 2.1 支持向量机原理

支持向量机(Support Vector Machine, SVM)是 Vapnik 等提出的一种应用于二类平衡数据集分类的有效算法,其核心思想是基于结构风险最小原理寻找最优分类平面<sup>[10]</sup>.如图 1 所示,以二元线性可分问题为例,所谓最优分类面,就是能将两类样本点正确分类,且要使两类的分类空隙最大<sup>[11]</sup>.线性分类平面描述如式(1)所示.

$$\omega^T \cdot x + b = 0, x \in R^d \quad (1)$$

式(1)中: $\omega$  和  $b$  分别表示权重向量和最优超平面偏移.

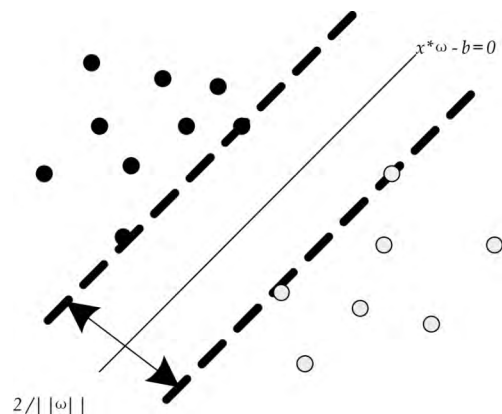


图1 最优分类超平面示意图

对于非线性分类问题,支持向量机通过引入核函数,通过一个非线性函数  $\phi(\cdot)$  将训练数据集  $X$  映射到某个高维线性特征空间,在变换空间求最优分类面,并得到分类器的决策函数<sup>[12]</sup>.采用适当的核函数  $K(x_i, x_j)$ ,并使  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ ,把优化问题中的所有点积运算都运用核函数运算代替,得到对偶最优化问题见式(2).

$$\max W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (2)$$

$$\text{s. t. } 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \quad (i=1, 2, \dots, n)$$

采用拉格朗日乘子法, 得到最优判别函数见式(3).

$$f(x) = \text{sign} \left[ \sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right] \quad (3)$$

## 2.2 AdaBoost 算法

集成学习的基本思想是对同一问题通过集成多个基分类器共同决策的机器学习技术, 通过调用简单的分类算法, 获得多个不同的基分类器, 并使用某种策略将基分类器组合成一个强分类器<sup>[13]</sup>. Boosting 算法<sup>[14]</sup>是被广泛使用的集成学习算法, AdaBoost 算法<sup>[15]</sup>是 Boosting 算法的代表算法, 其核心原则是反复修改训练样本集权重分布拟合一系列基分类器. 在每次迭代过程中, 增大错分类样本权重, 减小正确分类样本权重, 使得基分类器聚焦在那些很难分的样本上, 根据调整权重后的训练样本重新进行学习, 这使得分类器在下次迭代过程中更关注被错分的样本, 循环迭代此过程, 所有弱基分类器的预测通过加权多数投票法, 最终得到一个强分类器<sup>[16]</sup>. 该方法对非平衡数据集分类有利.

## 3 AdaBoost-SVM 分类算法设计

运用 AdaBoost-SVM 分类算法进行集成分类器设计, 主要有两个问题, 一是核函数及参数选择; 二是改善标准支持向量机对非平衡数据集分类效果不佳问题. AdaBoost-SVM 算法是把基于径向基核函数(RBF)的 SVM 作为 AdaBoost 的基分类器, 并应用网格搜索法<sup>[17]</sup>进行参数优化. 主要的设计思路: (1) 基于随机采样, 由原始样本  $D$  产生  $m$  个训练子集; (2) 训练子集由  $D_i (i=1, 2, \dots, m)$  产生子分类器  $\text{SVM}_i$ ; (3) 采用加权多数投票法组合得到强分类器. 具体实现过程如图 2 所示.

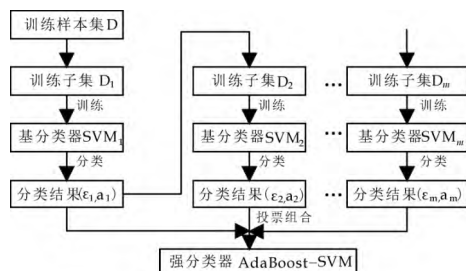


图 2 AdaBoost-SVM 算法流程图

AdaBoost-SVM 具体算法描述如下:

输入: 训练集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ; 基分类器 RBFSVM, 训练循环次数  $T$ .

Step1: 初始化. 训练集权值分布  $D_t(i) = 1/N$ , 循环次数  $t=1$ ;

Step2: 参数选择. 利用网格搜索方法在  $D$  上进行十折交叉验证法求得 SVM 基分类器建模最优参数  $(C, \sigma)$ ;

Step3: 训练基分类器.

Step3.1: For  $t=1, 2 \dots T$ ;

Step3.2: 通过具有权值分布  $D_t(i)$  的训练集得到第  $t$  次弱学习器  $H_t = L(D, D_t)$ ;

Step3.3: 根据式(4)计算  $H_t$  的错误率  $\epsilon_t$ ;

$$\epsilon_t = \sum_{i=1}^m D_t(i) (H_t(x_i) \neq y_i) \quad (4)$$

Step3.4:

IF:  $0 < \epsilon_t \leq 0.5$ , 根据式(5)计算权重值  $\alpha_t$ , 根据式(6)更新样本权重;

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} \quad (5)$$

$$D_{t+1}(i) = \frac{D_t(i)}{|D_t|} \cdot \begin{cases} \exp(-\alpha_t), & h_t(x_i) = y_i \\ \exp(\alpha_t), & h_t(x_i) \neq y_i \end{cases} \quad (6)$$

$t++$ ; 跳转至 Step3.2 继续执行.

ELSE: 跳转至步骤 4, 结束循环.

Step4: 构建基分类器的线性组合. 训练  $T$  轮后得到  $T$  组弱分类器  $H_t(x)$ , 由  $T$  组弱分类器组合得到强分类器  $H(X)$  如式(7)所示.

$$\text{输出: } H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t H_t(x) \right) \quad (7)$$

## 4 AdaBoost-SVM 葡萄酒品质分类算法实现

基于 AdaBoost-SVM 的葡萄酒品质分类系统如图 3 所示, 包括了分类器设计的核心模块. 如上文所述, 葡萄酒品质分类属于多类分类问题, 文中采用一对余类支持向量机 (1-against-the rest SVM) 方法将其转换为多个二类分类, 在某类样本与剩余多类样本间构造决策平面, 从而达到多类分类的目的. 如图 3 所示, AdaBoost-SVM 分类器: AdaBoost-SVM<sub>1</sub> 表示低等级 (类 0) 和中 (类 1)、高 (类 2) 等级间的分类器; AdaBoost-SVM<sub>2</sub> 表示中等级 (类 1) 和高等级 (类 2) 间的分类器.

假设第  $k$  类样本看做正类 ( $j=0, 1, 2$ ), 而将其余  $k-1$  类样本合并看做负类, 通过二分 SVM 方法求出一个决策函数如式(8)所示:

$$f_j(x) = \omega_j \cdot \phi(x) + b = \sum_{i=1}^m a_i^j y_i K(x_i, x) + b_j \quad (8)$$

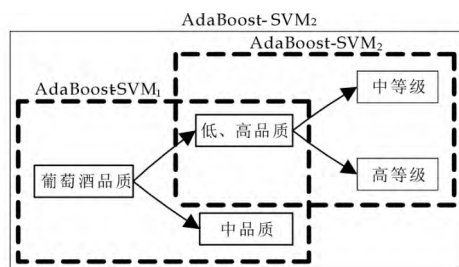


图 3 葡萄酒品质分类系统框图

## 5 仿真实验

### 5.1 数据采集及预处理

实验数据来自 UCI 数据库<sup>[18]</sup>中的 Wine Quality 数据集,该数据集共 4 898 个样本数据,每一个样本由 11 个理化指标(特征变量)和 1 个感官指标(目标变量)组成,具体的理化指标统计量如表 1 所示。感官指标指的是品酒师的感官评价结果,也就是葡萄酒的品质等级,按照十分制的评分标准将葡萄酒分为 0(最差)到 10(最优),样本数据共 7 个等级:3~9。由于数据集中有些类别样本数过少,为了方便研究,文中依据感官评价结果将葡萄酒品质分为 3 个等级,具体等级划分标准以及每个等级的样本分布如表 2 所示,三类样本的非平衡比例约为 1:20:6。

表 1 葡萄酒理化指标

Attribute/(units)	Mean	Min	Max	SD
酒石酸/(g/dm <sup>3</sup> )	6.86	3.80	14.20	0.84
醋酸/(g/dm <sup>3</sup> )	0.29	0.08	1.10	0.10
柠檬酸/(g/dm <sup>3</sup> )	0.33	0.00	1.66	0.12
残留糖/(g/dm <sup>3</sup> )	6.39	0.60	65.80	5.07
氯化钠/(g/dm <sup>3</sup> )	0.05	0.01	0.35	0.02
游离 SO <sub>2</sub> /(mg/dm <sup>3</sup> )	35.30	2	289	17.00
总 SO <sub>2</sub> /(mg/dm <sup>3</sup> )	138.4	9	440	42.49
密度/(g/cm <sup>3</sup> )	0.99	0.99	1.04	0.003
pH	3.19	2.72	3.82	0.15
硫酸钾/(g/dm <sup>3</sup> )	0.49	0.22	1.08	0.11
酒精/(vol. %)	10.51	8.00	14.20	1.23

表 2 等级划分及每个等级的样本分布情况

感官评分	<5	>=5, <7	>=7
品质等级	低等级	中等级	高等级
类别号	0	1	2
样本分布	183	3 655	1 060

如表 1 所示,酒石酸的测量值介于[3.8, 14.2] g/dm<sup>3</sup> 之间,而总 SO<sub>2</sub> 的测量值则介于[9, 400] mg/dm<sup>3</sup> 之间,为了提高最后分类的准确率,对样本数据集进行归一化预处理。本文采用最大最小归一化法,将所有属性归一到[0, 1]之间,计算方法见式(9):

$$x_{\text{normalization}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

在 Python 中使用 StandardScaler(). fit\_transform(Data)函数对数据集归一化处理,如某

样本酒石酸初始值为 7.0, 则  $x_{\text{normalization}} = (7.0 - 3.80)/(14.20 - 3.80) = 0.31$ 。

### 5.2 模型构建

经过数据预处理后,根据建模样本数据建立基于 AdaBoost-SVM 算法的葡萄酒品质分类模型。在建立分类模型过程中,为了防止过拟合,样本数据集分为训练数据集和测试数据集如表 3 所示。

表 3 样本数据

类标签	训练集	测试集	总计
类 0	123	60	183
类 1	2 951	704	3 655
类 2	844	216	1 060
总计	3 918	980	4 898

在训练支持向量机的时候,选取高斯径向基核(RBF)作为支持向量机核函数,包括两个核参数:  $C$  和  $\sigma$ 。选择适当的  $C$  和  $\sigma$ , 对于支持向量的性能是至关重要的,过低的参数  $C$  使决策表面光滑,而过高的  $C$  会出现训练样本过拟合现象。文中使用网格搜索法选取有效的参数( $C, \sigma$ )。设  $C$  和  $\sigma$  的搜索范围分别为  $[10^{-1}, 10^3]$  和  $[10^{-5}, 10^5]$ , 可以尝试  $5 \times 11 = 55$  种参数组合,选择使得分类器的正确率保持在 50%~60% 之间的参数组合,经过多次实验当( $C, \sigma$ )取(1, 0.01)时训练得到的分类器可以满足系统算法的需要。

当  $C=10, \sigma=0.01$  时,使用训练数据集训练支持向量机模型。构建 2 个子分类器, AdaBoost-SVM<sub>1</sub> 和 AdaBoost-SVM<sub>2</sub>。建立模型之后,首先使用训练数据集进行回判,然后使用测试数据集进行测试,分类结果如表 4 和表 5 所示。对于训练数据集和测试数据集来说, AdaBoost-SVM<sub>1</sub> 和 AdaBoost-SVM<sub>2</sub> 子分类器的分类准确率均达到 90% 以上。最后,将子分类器的分类结果通过加权多数投票法得到 AdaBoost-SVM 的分类结果混淆矩阵如表 5 所示。混淆矩阵对角线上元素表示每个等级类被正确分类的样本数,第一行说明类 0 的 60 个样本有 44 个分类正确,16 个错分为类 1。最终,测试数据集共有 892 个样本被正确分类, AdaBoost-SVM 分类器取得了 91.02%(892/980)的准确率。

表 4 AdaBoost-SVM 子分类器分类结果

子分类器	训练样本数	测试样本数	训练正确率/%	测试正确率/%
AdaBoost-SVM <sub>1</sub>	3 918	980	93.02	90.03
AdaBoost-SVM <sub>2</sub>	998	245	92.18	91.37

表 5 AdaBoost-SVM 分类结果混淆矩阵

实际类别	分类类别			召回率/%
	0	1	2	
0	44	16	0	73.33
1	0	704	0	100
2	0	40	176	81.48

### 5.3 对比分析

为了比较 AdaBoost-SVM 的分类性能,基于同样样本数据集,以及相同的参数值( $C, \sigma$ ),采用标准 SVM 算法再次进行分类实验,分类结果如表 6 和表 7 所示.对比发现,表 6 和表 7 基于 AdaBoost-SVM 算法的分类结果均优于表 4 和表 5 基于标准 SVM 算法的分类结果.并且从实验的结果上可以看出,基于 AdaBoost-SVM 分类器各等级的分类正确率都有了明显的提升,例如类 0 的分类准确率由原来的 46.67% 提高到优化后的 73.33%.同理,由表 7 的混淆矩阵计算可得,最终 SVM 分类器的分类准确率为 85.51%(838/980).

表 6 SVM 子分类器分类结果

子分类器	训练样本数	测试样本数	训练准确率/%	测试准确率/%
SVM <sub>1</sub>	3 918	980	80.97	80.11
SVM <sub>2</sub>	998	245	84.32	82.72

表 7 SVM 分类器分类结果混淆矩阵

实际类别	分类类别			召回率/%
	0	1	2	
0	29	31	0	46.67
1	44	673	31	95.60
2	13	67	136	53.70

非平衡数据集分类器常用正类覆盖率(True Positive Rate, TPR)和负类覆盖率(False Positive Rate, FPR)来衡量其分类性能,正类覆盖率是指在葡萄酒样本中正确识别出低品质葡萄酒的概率,负类覆盖率是指在非低品质样本中分类结果是否定的比例,分别表示为公式(10)、(11):

$$TP/(TP+FN) \quad (10)$$

$$FN/(TP+TN) \quad (11)$$

式(10)、(11)中,其中 TP 为正确的肯定、FN 为错误的否定、FP 为错误的肯定, TN 为正确的否定.

以 TPR 为纵轴, FPR 为横轴,绘制标准支持向量机与集成支持向量机的葡萄酒品质分类器的 ROC 曲线,分别绘制在如图 4~5 中,曲线下面积记为 AUC, AUC 越大则模型分类效果越好.当且仅当  $AUC > 0.5$  and  $AUC \leq 1$ ,分类器才是有价值的.ROC 分析结果显示,标准支持向量机 AUC 约为  $0.85 \pm 0.045$ ,表现为 82.80% 的 TPR 和 100% 的 FPR,集成支持向量机 AUC 约为  $0.89 \pm 0.045$ ,表现为 86.72% 的 TPR 和 100% 的 FPR,说明基于 AdaBoost-SVM 的葡萄酒品质分类器有良好的分类性能.

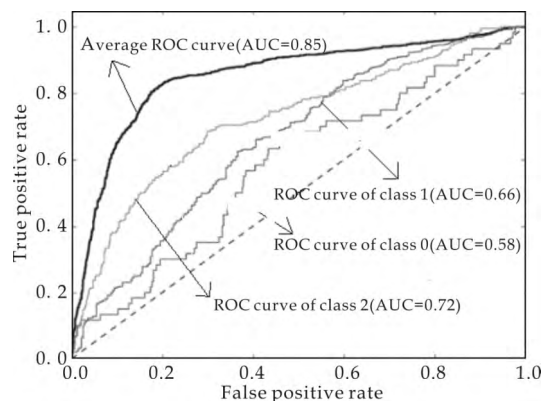


图 4 ROC Curve of SVM

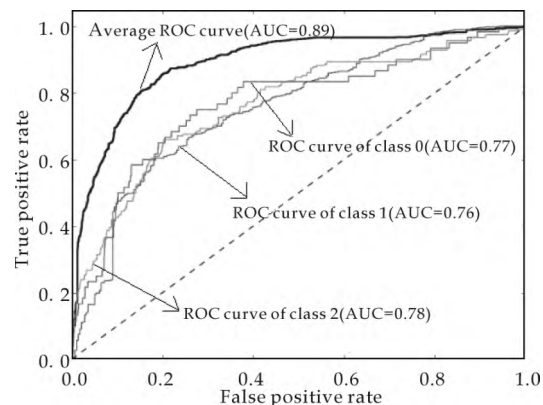


图 5 ROC Curve of AdaBoost-SVM

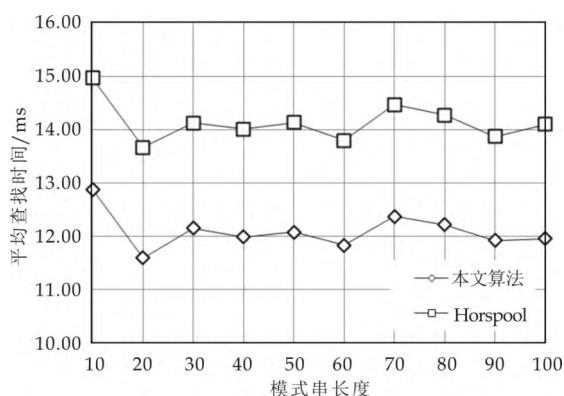
## 6 结论

基于 AdaBoost-SVM 分类算法的葡萄酒品质分类模型,克服了现有葡萄酒品质分类方法存在的缺点,有效解决了非平衡葡萄酒样本数据分类效果不佳的问题.文中所提算法与标准 SVM 算法分类结果进行对比分析,分析结果表明文中算法有效地改善了标准 SVM 对非平衡数据集的分类效果,使得低品质类葡萄酒样本识别率得到了很大提高的同时,整体样本分类性能以及泛化能力也得到了提升.因此该方法也可以被集成到支持决策系统中,辅助提高酿酒师的酿酒速度以及质量.

## 参考文献

- [1] 何 瑜. 中国葡萄酒产业竞争力研究[D]. 杨凌: 西北农林科技大学, 2014.
- [2] Baker A K, Ross C F. Sensory evaluation of impact of wine matrix on red wine finish: A preliminary study[J]. Sensory Studies, 2014, 29(2): 139-148.
- [3] 邵志芳. 葡萄酒品质分析方法研究进展[J]. 中国酿造, 2015, 34(4): 17-20.

(下转第 187 页)



(b) 模式串长度为 10—100

图 6 两种算法在 DNA 序列上的平均匹配时间

本文算法对基本模式串匹配算法 BM 的坏字符策略进行改进,实现了任一位置不匹配时,均有大于或等于同为坏字符策略实现匹配的 Horspool 算法的移动量.算法的基本思想简单、容易实现,在不同性质的实验数据上测试结果有效.

#### 4 结论

本文提出了一种有效的模式串匹配算法.当不匹配出现时,按照 BM 算法的坏字符策略计算当前已经匹配的每个字符的移动量为参考量,并从多个参考量中选择一个最大值作为模式串当前的移动量,任意位置的移动量均可以通过预处理计算得到.实验结果表明:在不匹配出现时,本文算法获得的模式串移动量总是最大的,从而有效地提高了模

式串的匹配效率.模式匹配算法已被广泛地应用于入侵检测和情报检索等系统中,并随着信息技术的不断发展,对于检索效率提出了更高的要求,高效率的匹配算法的提出具有极大的现实意义.本文作者的下一步工作将本文算法思想和其它模式串匹配技术相结合应用于病毒检测及 DNA 检测中.

#### 参考文献

- [1] 范洪博,姚念民.一种高速精确单模式串匹配算法[J].计算机研究与发展,2009,46(8):1341-1348.
- [2] R. Boyer, J. Moore. A fast string searching algorithm[J]. Communications of the ACM, 1977, 20(10): 762-772.
- [3] R. N. Horspool. Practical fast searching in strings[J]. Software-Practice & Experience, 1980, 10(6): 501-506.
- [4] M. Sunday. A very fast substring search algorithm[J]. Communications of the ACM, 1990, 33(8): 132-142.
- [5] Zhao X, Li S, Yang Y, et al. A new substring searching algorithm[J]. Ieice Transactions on Information & Systems, 2014, E97-D(7): 1893-1896.
- [6] 胡金柱,熊春秀,舒江波,等.一种改进的字符串模式匹配算法[J].模式识别与人工智能,2010,23(1):103-106.
- [7] 李明月,张善卿,陆剑锋,等.一种改进的 Sunday 匹配算法[J].杭州电子科技大学学报,2015,35(1):93-96.
- [8] 韩光辉,曾 诚. Boyer-Moore 串匹配算法的改进[J].计算机应用,2014,34(3):865-868.
- [9] Sun Kelei. Fast algorithm for pattern matching in intrusion detection system[J]. Journal of Anhui University of Science and Technology (Natural Science), 2006, 26(3): 52-55.
- [10] G. Navarro, M. Raffinot. Flexible pattern matching in strings[M]. Cambridge: Cambridge University Press, 2002:19-22.

【责任编辑:蒋亚儒】

(上接第 182 页)

- [4] Paulo Cortez, Antonio Cerdeira, Fernando Almeida, et al. Modeling wine preferences by data mining from physico-chemical properties[J]. Decision Support Systems, 2009, 47(4): 547-553.
- [5] 徐海涛.改进的近似支持向量机在葡萄酒质量鉴定中的应用[J].安徽农业科学,2010,38(29):16105-16106.
- [6] 刘延玲.新的 Hopfield 神经网络分类器在葡萄酒质量评价中的应用[J].价值工程,2012,35(2):181-182.
- [7] Jose A. Seas, Bartosz Krawczyk, Michal Wozniak. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets[J]. Pattern Recognition, 2016, 3(12): 164-178.
- [8] 黄久玲.面向失衡数据集的集成学习分类方法及其应用研究[D].黑龙江:哈尔滨理工大学,2015.
- [9] Zhong liang Zhang, Bartosz Krawczyk, Salvador Garcia, et al. Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced[J]. Knowledge-Based Systems, 2016, 5(48): 251-263.
- [10] 顾燕萍,赵文杰,吴占松.最小二乘支持向量机鲁棒回归算法研究[J].清华大学学报(自然科学版),2015,55(4): 396-402.
- [11] 袁兴梅,杨 明,杨 杨.一种面向不平衡数据的结构化

SVM 集成分类器[J].模式识别与人工智能,2013,26(3):315-320.

- [12] 吕 锋,李 翔,杜文霞.基于 MultiBoost 的集成支持向量机分类方法及其应用[J].控制与决策,2015,30(1): 81-85.
- [13] 李 勇,刘战东,张海军.不平衡数据的集成分类算法综述[J].计算机应用研究,2014,31(5):1287-1291.
- [14] 李秋洁,茅耀斌.基于数据重平衡的 AUC 优化 Boosting 算法[J].自动化学报,2013,39(9):1467-1475.
- [15] Ebenezer Owusu, Yong Zhao Zhan, Qi Rong Mao. An SVM-AdaBoost-based face detection system[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2014, 26(4): 477-491.
- [16] 李 垒,任越美.基于改进 AdaBoost 集成学习的空间目标识别[J].计算机系统应用,2015,32(8):202-205.
- [17] 魏 峻.一种有效的支持向量机参数优化算法[J].计算机技术与发展,2015,25(12):97-100,104.
- [18] Paulo Cortez. Center for machine learning and intelligent systems[DB/OL]. <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>, 2009-10-07.

【责任编辑:陈 佳】