

基于机器学习算法的缺损米粉块在线快速检测

On line fast detection of defective rice flour based on machine learning algorithm

谭卢敏

冯新刚

TAN Lu-min FENG Xin-gang

(江西理工大学应用科学学院, 江西 赣州 341000)

(College of Applied Science, Jiangxi University of Technology, Ganzhou, Jiangxi 341000, China)

摘要:目的:实现缺损米粉块的快速在线检测。方法:提出运用机器学习算法对缺损米粉块检测数据进行分析。通过相机对米粉块进行非接触式数据采集,图像上传及处理后,获取米粉块的轮廓周长和面积、近似轮廓的周长和面积、近似轮廓点数、轮廓外接圆半径 6 个特征数据,依据米粉块样本数据特点,运用支持向量机(SVM)分类算法对米粉块的多特征数据组成的样本集进行分析。结果:通过与 5 种算法测试对比,GBDT 分类算法平均准确率 89%,用时 1.10 s;KNN 分类算法平均准确率 88%,用时 0.23 s;Logistic Regression 分类算法平均准确率 88%,用时 0.68 s;Random Forest 分类算法平均准确率 87%,用时 0.47 s;tree 分类算法平均准确率 87%,用时 0.084 s;SVM 分类算法检测平均准确率最高,达 95%,平均用时最短,为 0.000 97 s。结论:利用 SVM 分类算法进行米粉块缺损检测准确率高、用时短,适用于生产线的在线检测。

关键词:机器学习;支持向量机;米粉块;缺损检测

Abstract: **Objective:** To realize the rapid on-line detection of defective rice flour. **Methods:** Non-contact data acquisition of rice flour blocks through cameras, image upload and processing, obtained the contour perimeter and area, approximate contour perimeter and area, approximate contour points and radius of the contour circle. According to the characteristics of rice flour block sample data, the SVM classification algorithm was used to analyze the sample set composed of multi feature data of rice flour block. **Results:** Compared with five algorithms, the average accuracy of GBDT classification algorithm was 89% with elapsed time of 1.10 s. The average accuracy of KNN classification algorithm was 88% with elapsed time of 0.23 s. The average accuracy of lo-

gistic regression classification algorithm was 88% with elapsed time of 0.68 s. The average accuracy of random forest classification algorithm was 87% with elapsed time of 0.47 s. The average accuracy of tree classification algorithm was 87% with elapsed time of 0.084 s. SVM classification algorithm had the highest average detection accuracy, up to 95%, and the shortest average elapsed time of 0.000 97 s. **Conclusion:** SVM classification algorithm has the characteristics of high accuracy and low elapsed time, which adapt the on-line detection of defective rice flour.

Keywords: machine learning; support vector machines; rice flour blocks; detection of defection

米粉块是可食用块状干米粉,口感好且方便保存和运输,在中国南方地区深受消费者欢迎^[1]。米粉块一般由自动生产流水线加工而成,在加工成型环节会产生缺损米粉块,如不及时发现并处理,会对米粉块批量生产带来质量下降的影响。目前,企业多采用人工检测,随着工作时间加长,工人疲劳度增加,检测效率和准确性大大下降^[2]。机器学习是基于数据集建立数理模型进行研究推理,并可以衍生独立的计算模式,被广泛用于解决工程应用和科学领域的复杂问题。如:张先洁等^[3]运用支持向量机(SVM)识别番茄果实成熟阶段准确率高达 94.27%;Zhu 等^[4]基于深度特征和支持向量机的胡萝卜外观质量识别准确率为 98.17%。Laxmi 等^[5]多类别直觉模糊双支持向量机在植物叶片识别中的应用有较好的泛化能力。通过上述文献的学习,结合缺损米粉块形状各异,其特征参数无规律^[6-7],研究拟提出运用机器学习对缺损米粉块进行检测,利用相机对传输带上的米粉块进行实时拍照,经图像处理提取米粉块相关特征数据作为机器学习的检测数据^[8-9],通过支持向量机分类算法,对数据进行分析后检测出缺损米粉块,以期实现缺损米粉块在线检测与分拣。

1 支持向量机分类

支持向量机(SVM)是基于统计学习理论中结构风险

基金项目:江西省教育厅科学技术研究项目(编号:GJJ181506)

作者简介:谭卢敏,女,江西理工大学应用科学学院讲师,硕士。

通信作者:冯新刚(1980—),男,江西理工大学应用科学学院讲师,硕士。E-mail:gztm2017@163.com

收稿日期:2022-01-10

最小化原则提出的,适用于有限数据集下的样本分类和回归处理,是机器学习一种有监督的学习模式^[10-11]。SVM 算法用于分类问题的基本思路是寻找两类线性样本中的一个最优分类面,使得该分类面到两类样本数据点的距离最大,对于线性可分样本数据,SVM 找到合适的参数 (ω, b) ,得到最优分类面函数,即为决策函数如式(1)所示。

$$f(x) = \text{sgn}\{\langle \omega \cdot x \rangle + b\} = \text{sgn}\left\{\sum_{i=1}^n a_i^* y_i (x_i \cdot x_j) + b^*\right\} \quad (1)$$

通过决策函数可以对线性未知样本进行分类判别。对于近似线性可分数据,以上最优分类面并不能把所有样本都正确分类,为此,引入松弛因子 ξ 和惩罚因子 C ,在经验风险和推广性能之间找到一个均衡点,让训练模型

有一定的容错率,同时对未知样本的分类正确率满足设计要求。

对于非线性可分样本,SVM 利用非线性变化核函数方法,用满足 Mercer 条件的核函数得到原始空间中非线性学习算法,通过该方法将原空间转换到某线性特征空间后进行处理。其中根据原样本数据特点选择合适的核函数对转换后特征空间的线性化程度有较好的帮助^[12-13]。

2 米粉块数据分析

米粉块加工属于食品加工,对加工条件要求较高,所以对流水线加工的米粉块通过非接触式的工业相机拍照获取原始图像,经图像处理获得米粉块数据。通过相机拍摄获得的米粉块图片共 160 张,作为样本用于模型训练,部分图片信息如图 1 所示。



图 1 部分米粉块图片

Figure 1 Pictures of some rice noodles

对样本图片进行图像处理,获取米粉块轮廓的周长和面积、近似轮廓的周长和面积、近似轮廓点数、轮廓外接圆半径 6 个特征数据,并且给每个样本图片定义了分类标签,“0”表示合格米粉块,“1”表示缺损米粉块。

米粉块图像处理流程如图 2 所示。图像经过灰度化处理、二值化处理和图像形态学处理后滤除原图中的干扰信息,提高了米粉块特征信息提取的准确度。

经过图像形态学处理后得到清晰的米粉块二值轮廓图,用数字化二值图像轮廓扫描算法提取米粉块轮廓信息,根据轮廓信息进一步计算得到轮廓外接圆信息和轮廓近似点,轮廓近似点是

根据 Douglas-Peucker 算法逼近原轮廓,得到更少的顶点数,再根据轮廓近似点获得近似轮廓信息,近似轮廓信息利用多边形逼近原始轮廓,进一步规范米粉块轮廓信息。

部分米粉块以各流程处理后的图片如图 3 所示。其中:轮廓信息图片的红色线是根据轮廓信息在原图上画出的米粉块轮廓线;外接圆信息图片的棕色线是根据外接圆信息在原图上画出的米粉块外接圆线;近似轮廓信息图片的紫色线是根据近似轮廓信息在原图上画出的米粉块近似轮廓线。

根据轮廓信息、外接圆信息和近似轮廓信息计算米粉块的周长、面积、外接圆半径等特征数据,对特征数据进行分析发现,单一特征数据对缺损检测不能提供准确的依据,比如由于加工的原因合格米粉块的轮廓面积会有一定差别,当缺损面积较小时其轮廓面积可能会大于合格米粉块的,因此会降低缺损米粉块检测的准确度。经过试验分析,采用米粉块的多特征数据进行检测,有利于提高检测准确度。

3 SVM 机器学习检测缺损米粉块

使用 SVM 机器学习方法对米粉块的多特征数据组成的样本集进行分析,实现缺损检测。米粉块样本集如表 1 所示。

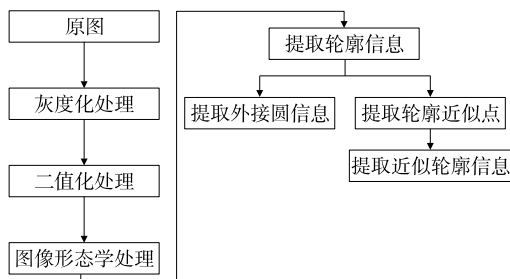


图 2 米粉块图像处理流程图

Figure 2 Flow chart of rice flour block image processing

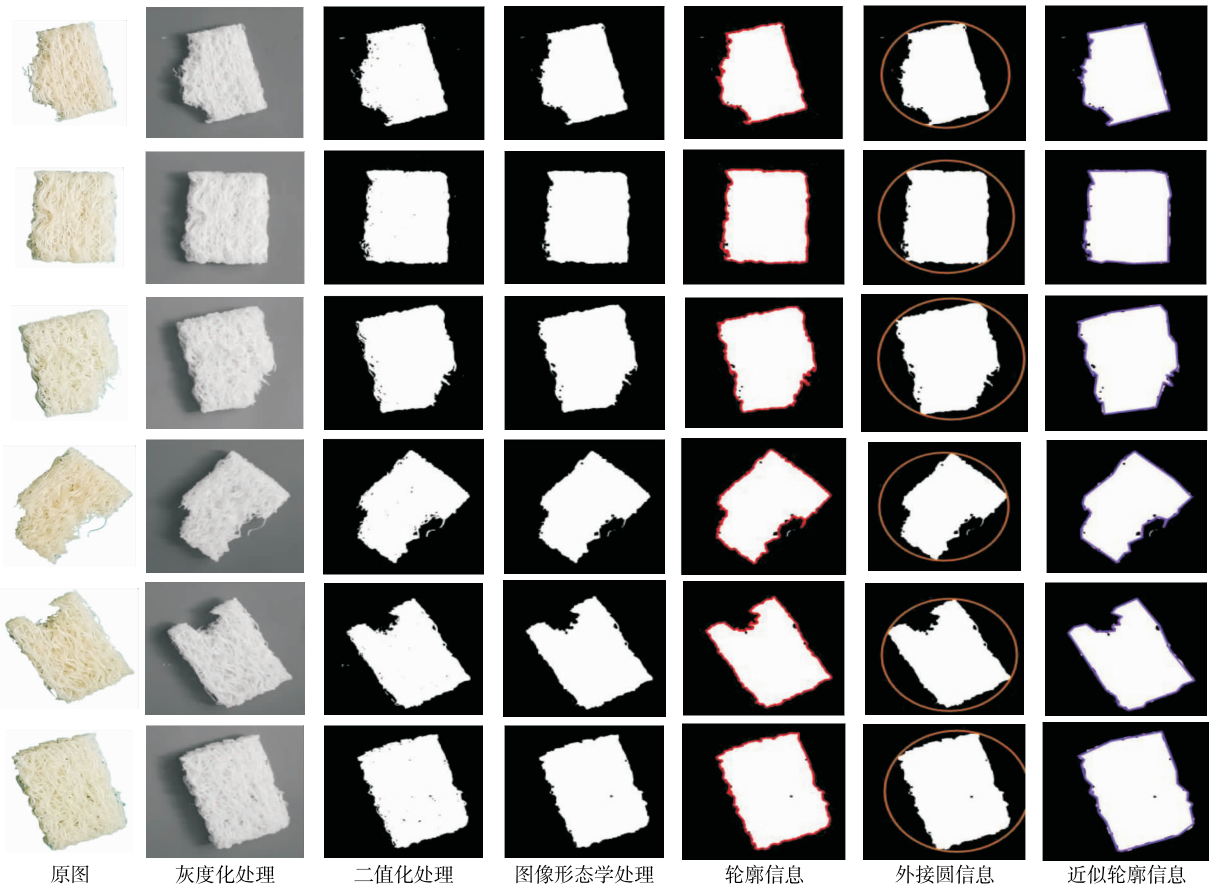


图 3 部分米粉块图像处理流程各步骤处理结果

Figure 3 Processing results of each step of image processing flow of some rice flour blocks

表 1 部分米粉块特征数据集							
Table 1 Characteristic data set of some rice flour blocks							
轮廓面积	轮廓周长	近似轮廓点数	近似轮廓面积	近似轮廓周长	外接圆半径	分类标签	
9 381.484	442.359	12	9 102.387	394.689	70.570	1	
12 119.641	466.719	8	11 975.183	413.186	74.942	0	
10 215.282	488.599	9	10 091.118	425.706	71.194	1	
9 119.845	417.499	11	8 971.458	373.831	66.198	1	
10 784.386	428.903	6	10 183.422	389.728	70.570	0	
11 258.348	457.557	8	11 014.820	407.984	73.692	1	
10 341.812	428.430	6	9 740.865	384.941	71.710	1	
10 936.223	433.184	8	10 720.810	389.069	68.537	1	
11 954.248	447.400	6	11 607.020	400.855	72.345	0	
8 197.154	368.416	9	8 095.756	325.483	55.845	1	

每个米粉块由 6 个特征数据和一个分类标签数据组成,通过对数据进行 SVM 机器学习,得到训练模型,从而实现米粉块的在线检测,其检测流程如图 4 所示。

从表 1 可以看出,米粉块的特征数据大小不一,在进行分析之前需要进行数据清洗,采用 min-max 标准化对

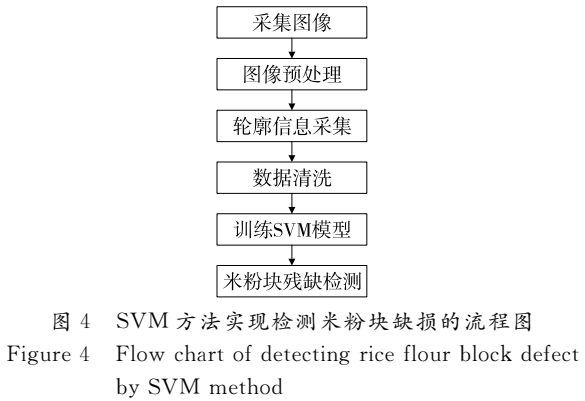


图 4 SVM 方法实现检测米粉块缺损的流程

Figure 4 Flow chart of detecting rice flour block defect by SVM method

样本数据进行归一化处理,如表 2 所示。

处理后的样本数据随机分为训练数据集和测试数据集,两个数据集中的样本数按照 7 : 3 的比例进行分配,代入 SVM 模型中进行训练和测试。

其中为了确定合适的核函数,分别选择 rbf 核函数、linear 核函数、poly 核函数,对训练集进行训练,用交叉验证法寻找最优模型参数,得到各自的优化模型,然后把测试集分别代入这些模型进行分类测试并记录分类准确率和测试用时,如表 3 所示。

由表 3 可知,rbf 核函数准确率最高,用时较少;poly 核函数准确率在 3 种核函数中最低同时用时较长;linear 核函数虽然用时最短,但是准确率没有 rbf 核函数高,用时与 rbf 核函数相差不大;通过数据比较,选择 rbf 核函数的 SVM 模型。

为了进一步检验该模型的优势,以相同米粉块样本数据集在相同条件下用其他分类算法进行训练,并对测试集进行分类分析,结果如表 4 所示。通过对比,SVM 分类算法的平均准确率最高,平均用时最短。

4 结论

研究结果表明,用SVM分类算法进行米粉块缺损检

测相比 GBDT、KNN、Logistic Regression、Random Forest 和 tree 5 种分类算法准确率高,用时短,有利于实现缺损米粉块的在线快速检测。但该研究对样本数据的先验信息特征研究不够深入,仅用 3 种常用核函数进行试验分析,后续可以利用隐含在数据中的先验信息选择更合适的核函数,进一步提高米粉块缺损检测的准确率。

参考文献

[1] 黎东, 张周莉, 张福, 等. 2018~2020 年南充市米粉质量安全状况分析[J]. 食品安全质量检测学报, 2020, 11(15): 5 184-5 189.
LI Dong, ZHANG Zhou-li, ZHANG Fu, et al. Quality and safety analysis of rice flour in Nanchong from 2018 to 2020[J]. Journal of Food Safety and Quality Inspection, 2020, 11(15): 5 184-5 189.
[2] 黄维安, 蒙毅, 陈磊, 等. 干米粉加工中常见问题的原因分析与对策[J]. 粮油食品科技, 2020, 28(6): 158-165.
HUANG Wei-an, MENG Yi, CHEN Lei, et al. Cause analysis and countermeasures of common problems in dry rice flour processing[J]. Grain, Oil and Food Science and Technology, 2020, 28(6): 158-165.
[3] 张先洁, 汪小昂, 孙国祥, 等. 基于彩色点云图像的不同成熟阶段番茄果实数量的测定方法[J]. 华南农业大学学报, 2022, 43(2): 123-128.
ZHANG Xian-jie, WANG Xiao-chan, SUN Guo-xiang, et al. A method for measuring the number of tomato fruits at different maturity stages based on color point cloud image[J]. Journal of South China Agricultural University, 2022, 43(2): 123-128.
[4] ZHU Hong-fei, YANG Lian-he, FEI Jian-wu. Recognition of carrot appearance quality based on deep feature and support vector machine[J]. Computers and Electronics in Agriculture, 2021, 186(7): 1-8.
[5] LAXMI Scindhiya, GUPTA S K. Multi-category intuitionistic fuzzy twin support vector machines with an application to plant leaf recognition[J]. Engineering Applications of Artificial Intelligence, 2022, 110(4): 1-27.
[6] 程子华. 基于机器视觉的残缺饼干分拣系统开发[J]. 现代食品科技, 2022, 38(2): 1-6.
CHENG Zi-hua. Development of incomplete biscuit sorting system based on machine vision[J]. Modern Food Science and Technology, 2022, 38(2): 1-6.
[7] 李成吉, 张淑娟, 孙海霞, 等. 基于计算机视觉的核桃外观缺陷检测[J]. 现代食品科技, 2019, 35(8): 247-253, 246.
LI Cheng-ji, ZHANG Shu-juan, SUN Hai-xia, et al. Walnut appearance defect detection based on computer vision[J]. Modern Food Technology, 2019, 35(8): 247-253, 246.
[8] 李文秀, 栾秋平. 基于机器视觉的预包装食品检测[J]. 食品与机械, 2020, 36(9): 155-157, 176.
LI Wen-xiu, LUAN Qiu-ping. Detection of prepackaged food based on machine vision[J]. Food & Machinery, 2020, 36(9): 155-157, 176.

(下转第 86 页)

表 2 部分米粉块 min-max 标准化后特征数据集
Table 2 Characteristic data set of some rice flour blocks after min max standardization

轮廓面积	轮廓周长	近似轮廓点数	近似轮廓面积	近似轮廓周长	外接圆半径	分类标签
0.398 29	0.623 44	0.800 00	0.378 33	0.688 79	0.702 81	1
0.787 45	0.821 81	0.400 00	0.821 77	0.872 87	0.911 47	0
0.516 79	1.000 00	0.500 00	0.530 95	0.997 49	0.732 62	1
0.361 11	0.420 99	0.700 00	0.358 12	0.481 19	0.494 15	1
0.597 68	0.513 85	0.200 00	0.545 20	0.639 41	0.702 81	0
0.665 04	0.747 20	0.400 00	0.673 53	0.821 11	0.851 85	1
0.534 78	0.510 00	0.200 00	0.476 88	0.591 76	0.757 24	1
0.619 26	0.548 71	0.400 00	0.628 15	0.632 85	0.605 79	1
0.763 94	0.664 49	0.200 00	0.764 94	0.750 15	0.787 53	0
0.229 97	0.021 27	0.500 00	0.222 94	0.000 00	0.000 00	1

表 3 不同核函数下 SVM 模型的平均准确率和平均用时
Table 3 Average accuracy and average time of SVM model under different kernel functions

核函数	平均准确率/%	平均用时/s
rbf	95	0.000 97
linear	90	0.000 88
poly	88	0.017 00

表 4 不同分类算法的平均准确率和平均用时
Table 4 Average accuracy and average time of different classification algorithms

算法	平均准确率/%	平均用时/s
GBDT 分类	89	1.10
KNN 分类	88	0.23
Logistic Regression 分类	88	0.68
Random Forest 分类	87	0.47
tree 分类	87	8.40×10^{-2}
SVM 分类	95	9.70×10^{-4}

- Engineering, 2021, 22(2): 141-154.
- [8] 史亚贝. 基于 DSP 的三自由度采摘机械手控制系统研究[J]. 农机化研究, 2022, 12(2): 34-38.
- SHI Ya-bei. Research on control system of 3-DOF picking manipulator based on DSP[J]. Agricultural Mechanization Research, 2022, 12(2): 34-38.
- [9] 李光, 章晓峰, 杨加超, 等. 基于残差 BP 神经网络的 6 自由度机器人视觉标定[J]. 农业机械学报, 2021, 52(4): 366-374.
- LI Guang, ZHANG Xiao-feng, YANG Jia-chao, et al. Vision calibration of 6-DOF robot based on residual BP neural network[J]. Journal of Agricultural Machinery, 2021, 52(4): 366-374.
- [10] MA Wei-ping, LI Wen-xin, SUN Jin-chuan, et al. Binocular vision target location method based on coarse fine stereo matching[J]. Computer Applications, 2020, 40(1): 227-232.
- [11] 宋海涛, 何文浩, 原魁. 一种基于 SIFT 特征的机器人环境感知双目立体视觉系统[J]. 控制与决策, 2019, 34(7): 1 545-1 552.
- SONG Hai-tao, HE Wen-hao, YUAN Kui. A robot environment perception binocular stereo vision system based on SIFT feature[J]. Control and Decision, 2019, 34(7): 1 545-1 552.
- [12] AHANDANI M A, ALAVIRAD H. Opposition-based learning in the shuffled differential evolution algorithm[J]. Soft Computing, 2016, 26: 64-85.
- [13] GAUTAM J V, PRAJAPATI H B, DABHI V K, et al. Empirical study of job scheduling algorithms in hadoop map reduce[J]. Cybernetics and Information Technologies, 2017, 21(1): 146-163.
- [14] 赵利平, 吴德刚. 基于小波与模糊相融合的苹果分级算法[J]. 食品与机械, 2020, 36(4): 142-145.
- ZHAO Li-ping, WU De-gang. Apple grading algorithm based on Wavelet and fuzzy fusion[J]. Food & Machinery, 2020, 36(4): 142-145.
- [15] AZUMAYA C M, DAYS E L, VINSON P N, et al. Screening for AMPA receptor auxiliary subunit specific modulators[J]. PLoS One, 2017, 12(3): 1 523-1 538.
- [16] 贺禹强, 刘故帅, 肖昇瑶, 等. 基于改进 GA-PSO 混合算法的变电站选址优化[J]. 电力系统保护与控制, 2017, 45(23): 143-150.
- HE Yu-qiang, LIU Gu-shai, XIAO Yi-yao, et al. Substation location optimization based on improved GA-PSO hybrid algorithm[J]. Power System Protection and Control, 2017, 45(23): 143-150.
- [17] 朱光耀. 基于无标定视觉伺服的全向移动机械臂跟踪控制[J]. 电子测量技术, 2020, 43(23): 23-29.
- ZHU Guang-yao. Tracking control of omnidirectional mobile manipulator based on uncalibrated visual servo[J]. Electronic Measurement Technology, 2020, 43(23): 23-29.
- [18] 王志中. 基于改进蚁群算法的移动机器人路径规划研究[J]. 机械设计与制造, 2018, 12(1): 242-244.
- WANG Zhi-zhong. Research on mobile robot path planning based on improved ant colony algorithm[J]. Mechanical Design and Manufacturing, 2018, 12(1): 242-244.
- [19] 于文妍, 杨坤林. 四旋翼无人机串级模糊自适应 PID 控制系统设计[J]. 机械设计与制造, 2019, 12(1): 227-231.
- YU Wen-yan, YANG Kun-lin. Design of cascade fuzzy adaptive PID control system for quadrotor UAV[J]. Mechanical Design and Manufacturing, 2019, 12(1): 227-231.

(上接第 42 页)

- [22] 嵇林娜, 刘石雪, 汪凤玲, 等. 酸水解法测定脂肪的研究及方法优化[J]. 现代食品, 2020(21): 180-183.
- ZHUO Lin-na, LIU Shi-xue, WANG Feng-ling, et al. Study on the determination of fat by acid hydrolysis method and its optimization[J]. Modern Food, 2020(21): 180-183.
- [23] 何曼君, 陈维孝, 董西侠. 高分子物理[M]. 上海: 复旦大学出版社, 2001: 137-139.
- HE Man-jun, CHEN Wei-xiao, DONG Xi-xiang. Polymer physics[M]. Shanghai: Fudan University Press, 2001: 137-139.
- [24] HEMAR Y, HORNE D S. Electrostatic interactions in adsorbed protein layers probed by a sedimentation technique[J]. Journal of Colloid and Interface Science, 1998, 206(1): 138-145.
- [25] ISABELLE B, SERGE P. Molecular basis of Ca^{2+} -induced gelation in alginates and pectins: The egg-box model revisited[J]. Biomacromolecules, 2001, 2: 1 089-1 096.
- [26] 杜柏桥. 羧甲基纤维素钠(CMC)与酪蛋白的相互作用及其稳定性乳体系机理的研究[D]. 上海: 上海交通大学, 2008: 47-52.
- DU Bo-qiao. The interaction between carboxymethylcellulose (CMC) and casein micelle and the stabilization mechanism of acidified milk drinks induced by CMC[D]. Shanghai: Shanghai Jiao Tong University, 2008: 47-52.

(上接第 81 页)

- [9] 张红岩, 王永志, 刘庆红. 图像识别技术在食品包装缺陷检测中的应用[J]. 食品与机械, 2020, 36(8): 225-228.
- ZHANG Hong-yan, WANG Yong-zhi, LIU Qing-hong. Application of image recognition technology in food packaging defect detection[J]. Food & Machinery, 2020, 36(8): 225-228.
- [10] CORTES C, VAPNIK V N. Support-vector networks[J]. Machine Learnig, 1995, 20(3): 273-297.
- [11] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer, 1995: 133-135.
- [12] 梁礼明, 冯新刚, 陈云嫩, 等. 基于样本分布特征的核函数选择方法研究[J]. 计算机仿真, 2013, 30(1): 323-328.
- LIANG Li-ming, FENG Xin-gang, CHEN Yun-nen, et al. Research on kernel function selection method based on sample distribution characteristics[J]. Computer Simulation, 2013, 30(1): 323-328.
- [13] 王文剑, 田萌. 核选择研究进展[J]. 山西大学学报(自然科学版), 2017, 40(3): 460-471.
- WANG Wen-jian, TIAN Meng. Advances in kernel selection research[J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(3): 460-471.