

doi:10.6041/j.issn.1000-1298.2013.08.027

基于组合核函数的籼稻重度不宜存检测模型^{*}

石礼娟¹ 谢彪彪² 谢新港³ 吴 洋³

(1. 华中农业大学理学院, 武汉 430070; 2. 武汉科技大学机械自动化学院, 武汉 430081;
3. 华中农业大学工学院, 武汉 430070)

摘要: 为给籼稻储存品质的判定提供一种快速无损检测手段,对 80 份重度不宜存籼稻和 80 份非重度不宜存籼稻的近红外反射光谱进行了实验研究。根据训练样本非线性可分的特点,选择支持向量机方法建立定性模型。在对不同核函数的特性进行分析和研究的基础上,定义了一种新的核函数——组合核函数。该组合核函数是多项式核函数与径向基核函数的线性组合,将两者各自的特点融合在一起兼具内推和外推性能。实验结果表明,以这两种函数的线性组合作为核函数且调节因子为 0.7 时,所建立的模型综合性能最好。所建模型的训练集正确识别率为 97.21%,测试集正确识别率为 93.25%。

关键词: 籼稻 近红外光谱 组合核函数 定性模型 支持向量机
中图分类号: S123; TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2013)08-0165-04

Identification Model of Severely Unstorable Indica Paddy Based on Combined Kernel Function

Shi Lijuan¹ Xie Biaobiao² Xie Xin'gang³ Wu Yang³

(1. College of Science, Huazhong Agricultural University, Wuhan 430070, China
2. School of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China
3. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China)

Abstract: In order to explore a rapid and non-destructive method for identifying storable quality of indica paddy, experiments were conducted on 80 samples of severely unstorable indica paddy and 80 samples of other indica paddy. Support vector machine (SVM) method was selected to build qualitative model according to training samples' characteristics. A combined kernel function was defined after analyzing and studying different kernel function. The proposed combined kernel function was the linear combination of polynomial function and radial basis function, and it combined each advantage of these two functions to acquire both interpolation and extrapolation properties of kernel function. The experiments showed that the model had the best comprehensive performance when the linear combination polynomial function and radial basis function was taken as the kernel function of SVM, and the control factor was 0.7. The training identification rate was 97.21%, and the test identification rate was 93.25%.

Key words: Indica paddy NIR spectroscopy Combined kernel function Qualitative model Support vector machine

引言

特别是超过正常储存年限以后,稻谷的内部结构逐渐松弛,酶活性降低,呼吸能力衰退,生活力减弱,这是稻谷自身的生理、生化变化过程^[1]。稻谷陈化不

收稿日期: 2012-12-13 修回日期: 2012-12-31
^{*} 国家重大科技专项资助项目(4005-09500216)、中央高校基本科研业务费专项资助项目(2011PY038)和华中农业大学科技创新基金(SRF)资助项目(2012029)
作者简介: 石礼娟,副教授,主要从事机器视觉研究,E-mail: slj2002@mail.hzau.edu.cn
通讯作者: 谢新港,副教授,主要从事农产品自动检测研究,E-mail: xxg@mail.hzau.edu.cn

仅会影响蒸煮后的食用品质,而且如果长期积压,就转化为“陈化粮”,从而失去食用价值,甚至对食用者的身体健康造成损害。我国自2006年12月1日开始实施的GB/T 20569—2006《稻谷储存品质判定规则》国家标准中对籼稻和粳稻分别按储存品质分为宜存、轻度不宜存和重度不宜存3类。

目前,稻谷的储存品质检测方法大都是基于其化学成分组成发生变化的化学检测方法,有过氧化物酶测定法^[2]、酸性指示剂的分光光度法^[3]和脂肪酸值测定法^[4]。因为近红外光谱分析技术具有速度快、效率高、成本低、测试重现性好、测量方便等特点,不仅用于定量分析,也被广泛应用于品种识别^[5-6]、产地鉴别^[7]等定性分析中。本研究采集了重度不宜存籼稻和非重度不宜存籼稻的近红外反射光谱,根据光谱数据样本的特点和各核函数的优势,构建基于组合核函数的重度不宜存籼稻定性鉴别模型。

1 材料与方法

1.1 实验材料与仪器

实验材料来自湖北省某粮库,包括已经重度不宜存籼稻80份和非重度不宜存籼稻(宜存或不宜存类)80份。

实验设备采用Thermo Scientific公司的ANTARIS II型傅里叶近红外光谱仪。该光谱仪配备了镀金积分球漫反射体、PbS检测器、石英玻璃样品杯,可产生波数3 800~12 000 cm⁻¹内的近红外光谱。光谱采集时,扫描次数为64次,分辨率为8 cm⁻¹。分析软件为TQ Analyst 8.0和Matlab 7.6。

1.2 光谱采集

每次装样时尽量保证石英杯内样品的高度均匀。每隔波数3.86 cm⁻¹采集反射强度并转换成1 g(1/R),取平均值后以文件形式保存在计算机中。扫描前将样品按顺序编号,扫描时光谱文件以该样品的编号命名。共得到80个重度不宜存籼稻的光谱,随机选择40个作为训练集,40个作为测试集;80个非重度不宜存籼稻样品,随机选择40个作为训练集,40个作为测试集。

1.3 建模方法

支持向量机(Support vector machine, SVM)是在统计学习理论的VC维理论和结构风险最小化原则基础上发展出来的一种机器学习方法^[8]。它主要是针对二类模式识别问题提出的,样本线性可分时,在原始输入空间构造最优分类超平面,使得训练样本不仅可以被无误差地划分为不同类,而且每一类数据支持向量与超平面之间的距离最大。对于非线性情况,引入核映像将输入空间中的训练集样本

通过函数映射到高维特征空间,然后再在高维空间中构造最优超平面,并得到分类器的决策函数^[9]。

目前,常用的非线性核函数有

非齐次多项式核函数

$$K(x, y) = (\langle x, y \rangle + 1)^q \quad (1)$$

径向基核函数(RBF)

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (2)$$

Sigmoid 核函数

$$K(x, y) = \tanh(v \langle x, y \rangle + c) \quad (3)$$

式中 q, γ, v, c ——参数

2 结果与讨论

2.1 基于单一核函数建模

计算40个重度不宜存和40个非重度不宜存籼稻样品的光谱每个波数下吸光度的平均值,得到2条平均光谱(图1)。再计算两组光谱在每个波数下吸光度的组内标准差,得到两类标准偏差光谱图(图2)。由图1可知,重度不宜存籼稻的平均吸光度与非重度不宜存籼稻的平均吸光度在近红外短波区域有明显差别。标准偏差光谱图(图2)显示出两类样本光谱的标准偏差均小于0.02,说明两类样品的类内差距很小。这一特性充分表明通过近红外光谱鉴别重度不宜存籼稻和非重度不宜存籼稻的可行性。

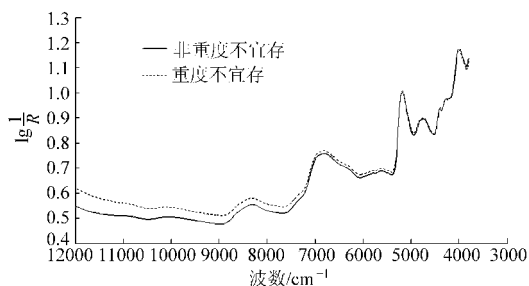


图1 平均光谱

Fig. 1 Average spectra

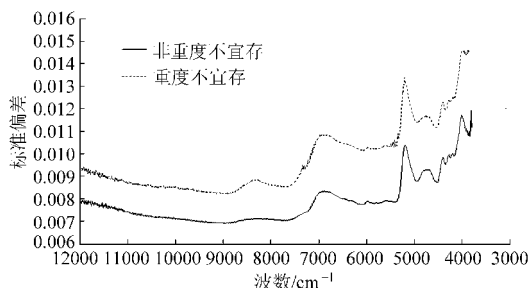


图2 标准偏差光谱

Fig. 2 Standard deviation spectra

一般来说,原光谱中存在由于基线漂移、仪器状态变化而产生的噪声信息,为消除这些信息对建模的影响,采用不同预处理方法即多元散射校正(MSC)、标准归一化(SNV)、一阶导数与Norris平滑

(FD + Norris)、二阶导数与 Norris 平滑(SD + Norris)对原始光谱进行处理,再对处理后的光谱数据进行主成分分析,然后取前 10 个主成分计算建模集中每个样品光谱到两类的类中心(平均光谱)的马氏距离,具体统计结果如表 1 所示。

表 1 不同预处理下建模集样本的马氏距离统计结果

Tab.1 Results of Mahalanobis distance with different preprocessing

预处理方法	至本类中心马氏距离			至另类中心马氏距离		
	平均值	最大值	标准差	平均值	最小值	标准差
MSC	0.96	1.62	0.25	6.47	5.74	0.31
SNV	0.96	1.61	0.25	6.54	5.81	0.31
FD + Norris	1.36	1.52	0.21	3.87	2.94	0.32
SD + Norris	1.36	1.46	0.21	3.18	2.22	0.31

表 1 中,至本类中心马氏距离的 3 个统计值(平均值、最大值、标准差)描述了类内距离,各指标值越小,类内聚合度越好;至另类中心马氏距离 3 个统计值(平均值、最大值、标准差)描述了类间距离,各指标值越大,类间距离越大。比较 4 种不同预处理方法至本类中心马氏距离的平均值和至另类中心马氏距离的平均值可知,前两种预处理方法相差不大,但第 3、4 种预处理方法与前两种预处理方法有显著差异。第 3、4 种预处理方法下的类内距离大而类间距离小,综合考虑类内距离小,类间距离大的聚类原则,淘汰第 3、4 两种预处理方法。第 1、2 种预处理方法下,类内距离的各指标值几乎没有差异,只是第 2 种方法的类间距离要稍大一点,因此,本文选择 SNV 方法作为预处理方法。

由于 PCA 中前 2 个主成分贡献率最高,计算建模集中 80 个样本所对应第 1 主成分和第 2 主成分的得分。如图 3 所示,图中的每个点对应一个样本。主成分的得分可以反映所有样本之间的相似性和独特性,因此基于样本的主成分得分图能够揭示样本的内部特征和聚类情况。

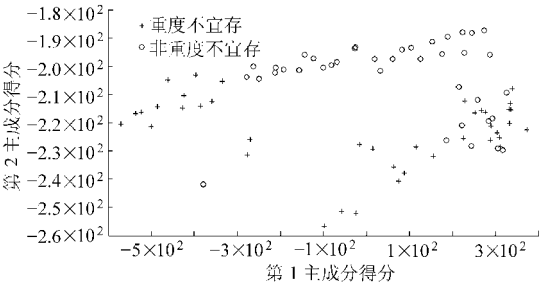


图 3 PCA 前两个主成分得分分布

Fig.3 PCA scores plots of the first two PCs

由图 3 可知,两类样本线性不可分,因此本研究以支持向量机理论为基础,用非线性核函数将两类样本映射到高维空间,然后再构造最优分类平面。

分别用多项式和径向基作为核函数构造了 2 种 SVM 模型,并分别用训练集和测试集对其进行检验,实验在 Matlab 7.0 中进行。

(1)非齐次多项式核函数

非齐次多项式核函数中只有一个参数 q ,它表征多项式的阶数,取不同 q 值进行训练与测试,实验结果如表 2 所示。表中, R_c 为训练集正确识别率, R_t 为测试集正确识别率, R_p 为相对误差, $R_p = (R_c - R_t)/R_c$ 。

表 2 不同 q 值下的 SVM 模型识别率

Tab.2 Identification accuracy by selecting different values of q for SVM model

q	$R_c/\%$	$R_t/\%$	R_p
1	68.37	67.54	0.012 1
2	73.45	71.28	0.029 5
3	80.65	78.19	0.030 5
4	86.35	82.51	0.044 5
5	82.35	78.13	0.051 2
6	73.65	68.34	0.072 1
7	70.35	64.32	0.085 7

(2)径向基核函数

径向基核函数只有一个参数 γ ,它是正则化系数,表示核函数的宽度。取不同 γ 值进行训练与检验,取值均为 2 的整数次幂,实验结果如表 3 所示。

表 3 不同 γ 值下的 SVM 模型识别率

Tab.3 Identification accuracy by selecting different values of γ for SVM model

γ	$R_c/\%$	$R_t/\%$	R_p
2	76.25	71.27	0.065 3
4	80.65	75.64	0.062 1
8	87.35	80.28	0.080 9
16	98.45	87.56	0.110 6
32	98.45	85.89	0.127 6
64	98.45	85.16	0.135 0

由表 2 和表 3 所示的训练集正确识别率和测试集正确识别率可以看出单一的核函数所建立的 SVM 模型效果并不理想,但不同的核函数有不同的优势。对比表 1 和表 2 的实验结果发现,用多项式核函数设计的 SVM 模型,训练集正确识别率与测试集正确识别率虽然不高,但两者之间的差距较小,而且多项式的阶数越小,训练集正确识别率与测试集正确识别率之间的差距越小,这充分说明多项式核函数有着良好的全局性质,具有很强的外推能力,而且阶数越低,外推能力越强。用径向基核函数设计的 SVM 模型,训练集正确识别率与测试集正确识别率较高,但两者之间的差距也比较大,训练集正确识

别率随着参数值的增大而增大,这充分说明径向基核函数局部性很强,其优势在于内推能力强,其内推能力随着参数的增大而增强。

2.2 基于组合核函数建模

为了提高 SVM 模型的综合性能,采用组合核函数来实现 SVM 算法,这样就可以充分利用两种核函数的优势,做到取长补短。不同的核函数的非负线性组合仍然满足 Mercer 条件^[8],因此多种核函数的非负线性组合可以作为一种新的核函数来构造 SVM 模型,使得新的核函数既具有良好的学习能力又具有较好的推广能力。定义以下带调节因子的组合核函数

$$K_{RP}(x,y) = \beta K_R(x,y) + (1-\beta)K_P(x,y) \quad (0<\beta<1) \quad (4)$$

式中 K_{RP} ——由径向基核函数与多项式核函数线性组合的核函数

β ——调节因子,用于调节组合核函数的内推能力与外推能力

K_R ——径向基核函数

K_P ——多项式核函数

径向基核函数的参数 γ 选择 16,此时训练集正确识别率和测试集正确识别率最高。多项式核函数的阶数 q 选择 1,此时训练集正确识别率和测试集正确识别率之间的差距最小。为得到合适的调节因子,分别取以下 9 组不同的值进行训练和检验,结果如表 4 所示。

由表 4 可知,调节因子 β 越大,训练集正确识别率和测试集正确识别率越高,但两者的差距也越大,

也就是说,随着调节因子增大,所构建的模型的外推能力减弱而内推能力增强。综合模型的内推与外推两方面的需求,最终选择 $\beta = 0.7$ 。所建模型的训练集正确识别率为 97.21%,测试集正确识别率为 93.25%。

表 4 基于组合核函数的模型识别结果
Tab.4 Recognition results of model based on combined kernel function

β	$R_c/\%$	$R_t/\%$	R_p
0.1	70.82	69.72	0.015 5
0.2	78.24	76.78	0.018 7
0.3	84.38	82.42	0.023 2
0.4	89.76	87.56	0.024 5
0.5	92.82	89.35	0.037 4
0.6	95.78	91.93	0.040 2
0.7	97.21	93.25	0.040 7
0.8	97.89	91.39	0.066 4
0.9	98.28	88.95	0.094 9

3 结束语

为了建立籼稻重度不宜存的定性模型,分别用多项式核函数和径向基核函数实现支持向量机算法。研究了两种核函数的特点并定义了一种新的核函数——组合核函数,使得新的核函数既具有良好的学习能力又具有较好的推广能力。研究结果表明,多项式核函数具有较强的外推能力,而径向基核函数的优势在于内推能力强。以两种函数的线性组合作为核函数且调节因子为 0.7 时,所建立的模型综合性能最好。

参 考 文 献

1 王展,梁秋慧,刘英.大米新鲜度快速检测方法研究[J].粮食与饲料工业,2007(3):5~6.
Wang Zhan, Liang Qiuhui, Liu Ying. A study on a method for detecting freshness of milled rice[J]. Cereal & Feed Industry, 2007(3):5~6. (in Chinese)

2 Matsukura U, Kaneko S, Momma M. Method for measuring the freshness of individual rice grains by means of a color reaction of catalase activity[J]. Journal of the Japanese Society for Food Science and Technology, 2000, 47(7): 523~528.

3 韩占江,王伟华,李帅.几种大米新陈度检验方法的比较研究[J].湖北农业科学,2009,48(7):1 736~1 742.
Han Zhanjiang, Wang Weihua, Li Shuai. Comparative study of several methods for determination freshness of rice[J]. Hubei Agricultural Sciences, 2009, 48(7):1 736~1 742. (in Chinese)

4 Hachiya M, Asanome N, Goto T, et al. Fluorescence imaging with UV-excitation for evaluating freshness of rice[J]. JARQ—Japan Agricultural Research Quarterly, 2009, 43(3): 193~198.

5 刘雪梅,章海亮.基于 DPLS 和 LS-SVM 的梨品种近红外光谱识别[J].农业机械学报,2012,43(9):160~164.
Liu Xuemei, Zhang Hailiang. Identification of varieties of pear using near infrared spectra based on DPLS and LS-SVM model [J]. Transactions of the Chinese Society for Agricultural Machinery, 2012,43(9):160~164. (in Chinese)

6 Cozzolino D, Fassio A, Restaino E, et al. Verification of silage type using near-infrared spectroscopy combined with multivariate analysis[J]. Journal of Agricultural and Food Chemistry, 2008, 56(1): 79~83.

7 赵杰文,蒋培,陈全胜.雪莲花产地鉴别的近红外光谱分析方法[J].农业机械学报,2010,41(8):111~114.
Zhao Jiewen, Jiang Pei, Chen Quansheng. Discrimination of snow lotus from different geographical origins by near infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(8):111~114. (in Chinese)

- Chen Li, Tu Kang, Zhao Yize, et al. Effects of 1-MCP and hot-air treatments on postharvest physiology and quality of Red Fuji apples[J]. Journal of Fruit Science, 2006,23(1):59~64. (in Chinese)
- 3 孙希生,王文辉,李志强,等. 1-MCP 对苹果采后生理的影响[J]. 果树学报,2003,20(1):12~17.
Sun Xisheng, Wang Wenhui, Li Zhiqiang, et al. Effects of 1-MCP treatment on physiology of apples after harvest[J]. Journal of Fruit Science, 2003,20(1):12~17. (in Chinese)
- 4 Fan X T. 1-methylcyclopropene inhibits apple ripening[J]. Journal of American Society for Horticultural Science, 1999,124(6):690~695.
- 5 Jennifer R, Dennis P, Murray D, et al. Influence of temperature and duration of 1-methylcyclopropene (1-MCP) treatment on apple quality[J]. Postharvest Biology and Technology,2002,24(3):349~353.
- 6 李学伟,赵晨霞,冯社章,等. 不同浓度的 1-MCP 处理对西洋梨常温后熟的影响[J]. 中国农学通报,2010,26(9):106~109.
Li Xuwei, Zhao Chenxia, Feng Shezhang, et al. Effects of 1-MCP treatment on pear storage at ambient temperature[J]. Chinese Agricultural Science Bulletin, 2010,26(9):106~109. (in Chinese)
- 7 孙希生,王志华,辛广,等. 不同处理条件下 1-MCP 对金冠苹果呼吸强度和品质的影响[J]. 果树学报,2004,21(2):141~144.
Sun Xisheng, Wang Zhihua, Xin Guang, et al. Responses of golden delicious apples to postharvest application of 1-MCP under different treatment conditions[J]. Journal of Fruit Science,2004,21(2):141~144. (in Chinese)
- 8 孙希生,王文辉,李志强,等. 1-MCP 对砀山酥梨保鲜效果的影响[J]. 保鲜与加工,2001,1(6):14~17.
Sun Xisheng, Wang Wenhui, Li Zhiqiang, et al. Effects of 1-MCP on cold storage of Dangshansuli pears[J]. Storage and Process, 2001,1(6):14~17. (in Chinese)
- 9 郝再彬. 植物生理实验[M]. 哈尔滨:哈尔滨工业大学出版社,2004.
- 10 Camps C, Guilermin P, Mauget J C, et al. Data analysis of penetrometric force/displacement curves for the characterization of whole apple fruits[J]. Journal of Texture Studies, 2005,36(4):387~401.
- 11 陈文烜,郅海燕,陈杭君,等. 1-MCP 结合减压贮藏对翠冠梨采后生理和品质的影响[J]. 中国食品学报,2010,10(4):228~232.
Chen Wenxuan, Gao Haiyan, Chen Hangjun, et al. Effects of 1-MCP treatment combined with hypobaric storage on postharvest physiology and quality of Cuiguang pear[J]. Journal of Chinese Institute of Food Science and Technology, 2010,10(4):228~232. (in Chinese)
- 12 程顺昌,冷俊颖,任小林,等. 不同环丙烯类乙烯抑制剂对苹果常温贮藏保鲜效果的影响[J]. 农业工程学报,2012,28(6):269~273.
Cheng Shunchang, Leng Junying, Ren Xiaolin, et al. Effect of different 1-substituted cyclopropenes as ethylene inhibitors on postharvest physiology of apple stored at ambient temperature[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012,28(6):269~273. (in Chinese)
- 13 程春梅,郑永华,郅海燕,等. 翠冠梨冷藏后用 1-MCP 处理对货架品质的影响[J]. 农业机械学报,2007,38(12):100~104.
Cheng Chunmei, Zheng Yonghua, Gao Haiyan, et al. Effects of 1-MCP treatment after cold storage on fruit quality in Cuiguang pears during shelf life[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007,38(12):100~104. (in Chinese)
- 14 Jamila C, Marie-Francoise D, Marie-Ghislaine G, et al. Physiological relationships among physical, sensory, and morphological attributes of texture in tomato fruits[J]. Journal of Experimental Botany,2007,58(8):1915~1925.
- 15 Infante R, Meneses C, Crisosto C H. Preconditioning treatment maintains taste characteristic perception of ripe 'September Sun' peach following cold storage[J]. International Journal of Food Science and Technology, 2009,44(5):1011~1016.
- 16 马庆华,王贵禧,梁丽松. 质构仪穿刺试验检测冬枣质地品质方法的建立[J]. 中国农业科学,2011,44(6):1210~1217.
Ma Qinghua, Wang Guixi, Liang Lisong. Establishment of the detecting method on the fruit texture of dongzao by puncture test[J]. Scientia Agricultura Sinica,2011,44(6):1210~1217. (in Chinese)
- 17 张鹏,李江阔,陈绍慧,等. 1-MCP 结合冰温贮藏磨盘柿的防褐保鲜效果[J]. 农业机械学报,2012,43(5):108~113.
Zhang Peng, Li Jiangkuo, Chen Shaohui, et al. Effect of 1-MCP combined with controlled freezing point storage on browning-preventing and freshness-keeping of mopan persimmon[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012,43(5):108~113. (in Chinese)

(上接第 168 页)

- 8 Vapnik V, Lerner A. Pattern recognition using generalized portrait[J]. Automation and Remote Control, 1963,24(6):774~780.
- 9 萧嵘,王继成,张福炎. 支持向量机理论综述[J]. 计算机科学,2000(3):1~3.
Xiao Rong, Wang Jicheng, Zhang Fuyan. Survey of support vector machine theory[J]. Computer Science, 2000(3):1~3. (in Chinese)
- 10 陈玲. 基于支持向量机的多类文本分类研究[D]. 重庆:重庆大学,2010.
Chen Ling. Study on multi-class text classification based on support vector machines[D]. Chongqing: Chongqing University, 2010. (in Chinese)