

# 基于光谱技术的支持向量机判别牛肉新鲜度

马世榜<sup>1,2</sup>, 徐杨<sup>1\*</sup>, 彭彦昆<sup>1</sup>, 汤修映<sup>1</sup>

(1. 中国农业大学工学院, 北京 100083;  
2. 南阳理工学院, 南阳 473004)

**摘要:** **目的** 实现生鲜牛肉新鲜度等级的无损快速判别。**方法** 用可见/近红外光谱检测系统, 获取储存1~18 d的36块牛肉样品的400~1600 nm范围的光谱信息, 以挥发性盐基氮理化值为分类依据。用多元散射校正(MSC)、变量标准化(SNV)、SG平滑预处理方法处理光谱数据, 分别建立牛肉新鲜度的支持向量机分类模型。**结果** MSC+SG预处理后所建立的分类模型预测能力最好, 训练集和测试集的回判识别率和预测识别率分别为96.30%、100%, 验证集识别率为88.89%。**结论** 可见/近红外光谱结合支持向量机, 对牛肉新鲜度进行无损快速判别是可行的。

**关键词:** 可见/近红外光谱; 支持向量机; 牛肉新鲜度; 无损分类

## Assessment of beef freshness based on spectral technology with support vector machine

MA Shi-Bang<sup>1,2</sup>, XU Yang<sup>1\*</sup>, PENG Yan-Kun<sup>1</sup>, TANG Xiu-Ying<sup>1</sup>

(1. College of Engineering, China Agricultural University, Beijing 10083, China;  
2. Nanyang Institute of Technology, Nanyang 473004, China)

**ABSTRACT: Objective** To establish a rapid and non-destructive discrimination method of freshness level of fresh beef. **Methods** A total of 36 beef samples' reflectance spectra in 400~1600 nm were collected by a laboratory visible/near-infrared spectroscopy system. These samples were stored for 1~18 d. The total volatile basic nitrogen (TVB-N) values of sample were taken as the reference of classification. The multiplicative scatter correction (MSC), standard normalized variate (SNV), and savitzky-golay (SG) smoothing methods were used as the pretreatment method for reflectance spectra processing, respectively. The classifier models of support vector machine (SVM) were built for identification of beef freshness. **Results** The MSC combined with SG smoothing was the best pretreatment, and the predictive ability of classification model built was the best. The identification rates of training set and prediction set were 96.30% and 100%, respectively. The identification rate of validation set was 88.89%. **Conclusion** It is feasible to use visible/near-infrared spectrum combined SVM for rapid and non-destructive identification of freshness level of fresh beef.

**KEY WORDS:** visible/near-infrared spectrum; support vector machine; beef freshness; non-destructive classification

基金项目: 公益性行业(农业)科研专项(201003008)

\*通讯作者: 徐杨, 教授, 博士生导师, 主要研究方向: 农畜产品品质安全无损检测技术与装备。Email: xuyang@cau.edu.cn

## 1 引言

肉品新鲜度是指肉的新鲜程度,是衡量肉品是否符合食用要求的客观标准,是评价肉品品质的重要指标,也是消费者选购肉类食品的主要依据。牛肉作为我国第二大消费肉类食品,其新鲜度的检测判别越来越受到重视。目前用于检测判别新鲜度的方法主要是感官检查和检测挥发性盐基氮(total volatile basic nitrogen, TVB-N)、菌落总数、pH 值等理化指标进行判断的方法<sup>[1]</sup>,其中检测 TVB-N 指标是 GB2707-2005 规定的检测判别肉类食品新鲜度的方法,规定 TVB-N 15 mg/100 g 为新鲜肉,可以安全食用,TVB-N>15 mg/100 g 为不新鲜肉,不可以食用。感官检查方便快捷,但易受主观因素的影响;检测理化指标方法准确,但操作专业性要求高、周期长、费用高,并常常需要对被检测肉品进行破坏处理。因此需要对肉品新鲜度无损快速检测判别方法进行研究。

光谱分析技术具有非接触、无需对被测样品进行预处理、测试重新性好等无损、快速检测的优点,已经广泛应用于农畜产品的无损检测研究<sup>[2-6]</sup>。侯瑞峰等<sup>[7]</sup>用近红外光谱结合聚类分析方法对猪肉的新鲜度进行了初步研究,蔡健荣等<sup>[8]</sup>用近红外光谱和联合区间偏最小二乘法预测猪肉的挥发性盐基氮,预测相关系数和标准差分别为 0.8238 和 4.17。支持向量机(support vector machine, SVM)是基于统计学习理论的结构风险最小化原理的机器学习方法,能够较好的解决非线性、小样本和高维数等学习问题,具有较好的泛化能力,在分类判别、预测优化等方面已有研究应用<sup>[9]</sup>。孙晓荣等<sup>[10]</sup>利用近红外技术结合支持向量机对马铃薯和玉米淀粉类别进行了判别,赵杰文等<sup>[11]</sup>利用支持向量机建立分类模型对不同产地苹果进行判别,回判识别率和预测识别率分别为 87%和 100%。本研究利用可见/近红外光谱技术结合支持向量机判别分类方法,以牛肉的 TVB-N 指标为依据,建立生鲜牛肉的新鲜度支持向量机判别模型,与前期的定量预测 TVB-N 研究不同,直接对牛肉的新鲜度等级进行定性判别,实现牛肉新鲜度的无损、快速判别。

## 2 材料与方法

### 2.1 试验材料、仪器与试剂

从北京美廉美超市购取经解僵、排酸后刚上市的

福成牌内蒙古小黄牛背最长肌,用无菌刀分割为尺寸大小约为 6 cm×5 cm×2 cm 的 36 块,用自封保鲜袋包装后放置在 4 °C 左右的保鲜箱内,运至实验室后,整齐无挤压地保存在 4 °C 冰箱中储藏。为获取牛肉从新鲜到腐败整个周期的数据,整个试验周期设为 18 d,每天从冰箱中随机取出 2 块样品进行检测试验。试验系统采用文献<sup>[12]</sup>中搭建的可见/近红外光谱检测系统。系统主要由光谱仪、计算机、光源、载物台和封闭舱等组成。其中光谱仪(荷兰 AVANTES 公司)为双通道,两通道光谱波长测量总范围为 200~1750 nm,分辨率分别为 0.5 nm 和 2.4 nm,光源为 2 个卤钨灯(12 V, 20 W),呈 45°对称安放在样品上方。封闭舱的主要作用是构成一暗箱环境,避免外界光线对检测结果的干扰。此外还包括肉搅拌机(广东美的电器制造有限公司)、电子天平(上海超平科学仪器有限公司)、KDY-9820 凯氏定氮仪(北京市通润源机电技术有限公司)、PE 保鲜膜等仪器设备材料。

主要试剂: 0.01 mol/mL 盐酸标准溶液(北京益利精细化学品有限公司), 1.0%氧化镁(青海铁源钾镁有限责任公司)混悬液, 2.0%硼酸(北京化工厂)吸收液, 1 g/L 的甲基红乙醇指示剂(西陇化工股份有限公司), 1 g/L 的溴甲酚绿乙醇指示剂(天津市津科精细化工研究所)和去离子水等。

### 2.2 试验方法

#### 2.2.1 光谱数据采集及理化值测量

每次试验前,把牛肉样品随机从冰箱中取出,去除保鲜袋,在空气中自然暴露 20 min,使其表面水自然挥发;用可见/近红外光谱检测系统,采集牛肉样品表面 400~1600 nm 范围的光谱反射强度信息。每个样品采集前需调节载物台的高度,保证每块样品上表面距离光纤探头间的距离为 13 cm;每个样品采集不同 6 点的光谱信息,取 6 点光谱信息的平均值作为该样品的最终光谱信息;测量点避开结缔组织。

测量反映样品新鲜度的 TVB-N 理化指标,评价样品的新鲜度。TVB-N 的测量按照 GB/T5009.44-2003 中的半微量定氮法,用 KDY-9820 凯氏定氮仪进行测量。样品绞碎前去除结缔组织。每次测量前,做 3 个样品空白试验,取平均值作为样品空白试验值,每个样品做 3 次样液试验,取平均值作为该样品的最终样液试验值。

#### 2.2.2 支持向量机基本原理<sup>[13,14]</sup>

设训练集样本为 $(x_1, y_1), \dots, (x_i, y_i), i=1, 2, 3, \dots, n, x_i$

为第  $i$  块牛肉的光谱数据信息,  $y_i$  为第  $i$  块牛肉所属的新鲜度类别。根据国标牛肉新鲜度等级分为新鲜和不新鲜两类, 对两类样本进行分类, 取  $y_i = [-1, +1]$ 。支持向量机遵循结构风险最小化准则构造决策超平面使负样本和正样本之间的分类间隔最大。假设所有能够分开这两类不同样本平面的集合为

$$y_i [w \cdot x_i + b] > 1 - \xi_i \quad (i=1, 2, 3 \dots n) \quad (1)$$

式中  $\xi_i$  为松弛因子, 且  $\xi_i \geq 0$ 。式(1)的最优求解问题可以变换为对下式(2)的最小化, 求得  $\alpha_i$

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

式(2)中  $K(x_i, x_j)$  是核函数, 常用的核函数有线性核函数、多项式核函数、径向基(RBF)核函数和 Sigmoid 核函数。选择合适的核函数和惩罚系数  $C$ , 由下式分类函数(3)得到最优分类平面。

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

### 2.2.3 数据处理方法

由于肉品样品表面纹理及样品背景的影响, 会导致光谱产生偏移或漂移现象, 同时仪器噪声及外界高频随机噪声干扰和光谱的多重共线性的存在, 均会影响校正模型建立的质量和未知样品预测的准确性, 为提高判别模型的判别精度和稳定性, 提高预测精度, 采用多元散射校正(multiplication scatter correction, MSC)、变量标准化(standard normalized variate, SNV)、SG(Savitzky-Golay)平滑方法对光谱数据进行预处理, 由预处理后的光谱数据建立支持向

量机分类判别模型, 根据所建模型判别率的高低, 确定最佳预处理方法和最佳模型。所建模型的预测判别率越高, 且训练集和测试集的判别率越接近, 说明所建模型预测判别能力和稳定性越好。方法的具体实现采用 Matlab 7.6.0(美国 Mathworks 公司)软件编程进行。

## 3 结果与分析

### 3.1 样品光谱及 TVB-N 值

图 1(a)所示是 36 个牛肉样品原始反射光谱曲线。从图上可以看出, 不同牛肉样品光谱曲线在 1~18 d 的储存期内整体变化趋势基本一致, 只是强度有所差异, 但部分光谱曲线存在有漂移现象, 局部存在有干扰噪声, 需要进行预处理。图 1(b)是 36 个样品原始光谱曲线经过 MSC 和 SG 共同预处理后的光谱图, 其中 SG 平滑处理, 采用的阶次为 2, 窗口大小为 9。可以看出经 MSC 和 SG 共同处理后, 光谱漂移现象基本消除, 局部干扰噪声也明显减少。

图 2 是 1~18 d 的储存期内 36 个牛肉样品, 每天检测 2 个样品的 TVB-N 值, 取其平均值绘制的变化趋势图, 实折线代表每天两个样品的 TVB-N 的平均值, 短竖直线代表两个样本 TVB-N 检测值与其平均值之间的误差。从图上可以看出, 牛肉在 4 环境下, 1~18 d 的储存期内 TVB-N 值是呈不断增加的变化趋势, 即越来越不新鲜, 其中大约在第 7 d 以后 TVB-N 的值超出国标中规定的 15 mg/100 g, 即为不新鲜肉, 不可以食用。变化趋势图中局部有下降趋势, 主要可能是由于样品间存有差异及试验误差造成的。

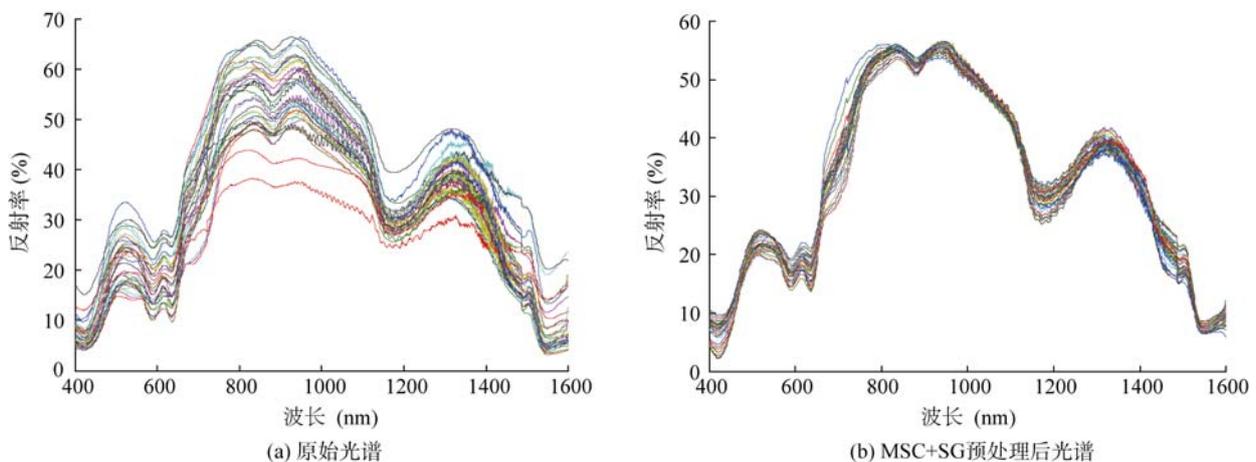


图 1 牛肉样品光谱

Fig. 1 Spectrograms of beef samples

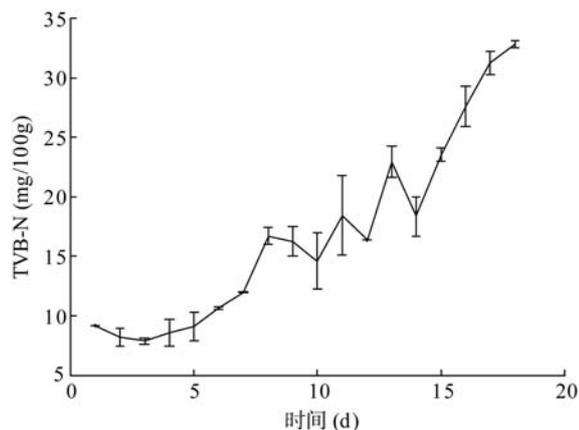


图2 储存期挥发性盐基氮变化趋势

Fig. 2 Changing trend of TVB-N during the storage time

按照 3.1 的原则把 36 个样品随机分为训练集和测试集, 训练集 27 个样品用于训练建立支持向量机的分类模型, 测试集 9 个样品用于测试训练集所建分类模型的预测能力和稳定性。训练集和测试集样品的 TVB-N 值分布统计情况如表 1 所示。训练集样品的 TVB-N 值变化范围较大, 且包含测试集的变化范围。

表 1 训练集与测试集 TVB-N 值统计结果  
Table 1 Statistical results of TVB-N value in training and test set

样本集	样本数	最大值	最小值	平均值	标准差
训练集	27	33.10	7.46	17.09	8.05
测试集	9	30.26	7.91	16.35	8.01

### 3.2 支持向量机最优参数的选取

支持向量机分类模型建立时, 选取不同的核函数对所建模型的性能影响较大, 如前文所述主要的核函数有线性核函数、多项式核函数、径向基核函数和 Sigmoid 核函数, 径向基核函数对分类问题有较好的拟合效果<sup>[15,16]</sup>, 径向基核函数可以把非线性样本数据映射到高维特征空间, 可以处理具有非线性关系的样本数据, 因此本文取径向基函数作为核函数。惩罚因子  $C$  和核函数参数  $\sigma$  对模型的判别精度影响较大, 其值的确定过程等于模型的建立过程。初始化设定  $C$  的取值范围为  $2^{-8} \sim 2^8$ ,  $\sigma$  的取值范围为  $2^8 \sim 2^{-8}$ , 采用网格搜索和留一交叉验证法进行优化确定, 其中设定  $C$  参数的搜索步长为 1,  $\sigma$  参数的搜索步长为 0.1。

### 3.3 支持向量机建模、判别及模型验证

按照 GB2707-2005 规定依据理化指标 TVB-N 的

值, 把 TVB-N  $15 \text{ mg}/100 \text{ g}$  的样品标识为“+1”标签代表新鲜, TVB-N  $>15 \text{ mg}/100 \text{ g}$  的样品标识为“-1”标签代表不新鲜。训练集 27 个样品中新鲜样品为 10 个, 不新鲜样品为 17 个, 测试集 9 个样品中新鲜样品为 4 个, 不新鲜样品为 5 个。用前述 MSC、SNV 和 SG 不同预处理方法及组合方法对样品光谱数据进行处理后, 用训练集样品的光谱数据信息和对应的类别标签进行训练, 建立支持向量机分类模型, 用测试集样品光谱数据及对应类别标签对所建模型进行验证, 结果如表 2 所示。从表中可以看出, 用不同预处理方法对光谱进行预处理后所建分类模型, 对训练集的回判识别率和测试集的预测识别率不同。SNV 预处理后所建分类模型的预测能力比较差, 回判识别率和预测识别率分别只有 59.26% 和 55.56%; MSC 预处理后所建分类模型存在过训练的现象, 回判识别率达到 100%, 但预测识别率仅达到 66.67%; SG 预处理后所建模型预测能力较好, 回判识别率和预测识别率分别为 96.30% 和 88.89%; MSC 和 SG 共同预处理后所建模型的预测能力最好, 回判识别率和预测识别率分别达到 96.30% 和 100%, 能够较好的对牛肉新鲜度进行等级判别。

为验证所建分类模型的通用性和稳健性, 用第二批次试验样品对模型进行验证。第二批次试验样品取自北京市御香苑畜牧有限公司, 肉样品种为东北改良牛, 采取与第一批次试验相同的试验方案。按照试验时间顺序, 从 1~18 d 所有试验样品中, 随机取出 18 个样品进行验证。18 个样品的光谱数据经 MSC+SG 预处理后, 代入前述所建的支持向量机分类模型里进行分类判别, 判别识别率为 88.89% (16/18)。与第一批次试验的测试集识别率相比, 精度有所下降, 这可能主要是由于肉样品种差异造成的影响。但识别精度可以达到应用要求。

表 2 不同预处理方法建立 SVM 模型的结果  
Table 2 Results of SVM model with different pretreatment methods

预处理方法	训练集	测试集
	回判识别率/%	预测识别率/%
MSC	100(27/27)	66.67(6/9)
SNV	59.26(16/27)	55.56(5/9)
SG	96.30(26/27)	88.89(8/9)
MSC+SG	96.30(26/27)	100(9/9)

## 4 结 论

利用可见/近红外光谱检测系统结合支持向量机分类方法, 对牛肉新鲜度进行判别分类。通过对牛肉样品光谱不同预处理后所建支持向量机的分类模型的结果对比, 得出 MSC+SG 为最佳预处理方式, 所建支持向量机分类模型对训练集和测试集的回判识别率和测试识别率分别为 96.30% 和 100%, 验证集的回判识别率为 88.89%。结果表明, 可见/近红外光谱技术结合支持向量机分类方法, 能够较好地对牛肉新鲜度进行区分判别, 直接对牛肉新鲜度等级进行快速、无损判别是可行的。

### 参考文献

- [1] 王长远, 马万龙, 姜昱男. 猪肉新鲜度的检测及肉质综合评定[J]. 农产品加工, 2007, (10): 75-77.
- [2] 彭彦昆, 张雷蕾. 光谱技术在生鲜肉品质安全快速检测的研究进展[J]. 食品安全质量检测学报, 2010, 27(2): 62-67.
- [3] 马世榜, 徐杨, 田潇瑜. 生鲜肉类品质快速检测技术的研究进展[C]. 农产品质量安全与现代农业发展专家论坛论文集, 2011.
- [4] 许冠男, 郭培源, 袁芳. 猪肉新鲜度无损检测技术现状及发展方向[J]. 北京工商大学学报(自然科学版), 2010, 28(1): 14-17.
- [5] 刘魁武, 成芳, 应义斌. 猪肉品质检测的研究进展[J]. 农业工程学报, 2008, 24(S1): 239-242.
- [6] 徐霞, 成芳, 应义斌. 近红外光谱技术在肉品检测中的应用和研究进展[J]. 光谱学与光谱分析, 2009, 29(7): 1876-1880.
- [7] 侯瑞峰, 黄岚, 王忠义, 等. 用近红外漫反射光谱检测肉品新鲜度的初步研究[J]. 光谱学与光谱分析, 2006, 26(12): 193-2196.
- [8] 蔡健荣, 万新民, 陈全胜. 近红外光谱法快速检测猪肉中挥发性盐基氮的含量[J]. 光学学报, 2009, 29(10): 2808-2812.
- [9] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
- [10] 孙晓荣, 刘翠玲, 吴静珠, 等. SVM 方法在淀粉分类问题中的应用[J]. 食品工业科技, 2011, 32(11): 431-433.
- [11] 赵杰文, 呼怀平, 邹小波. 支持向量机在苹果分类的近红外光谱模型中的应用[J]. 农业工程学报, 2007, 23(4): 149-152.
- [12] 马世榜, 汤修映, 徐杨, 等. 可见/近红外光谱结合遗传算法无损检测牛肉 pH 值[J]. 农业工程学报, 2012, 28(18): 263-268.
- [13] Vapnik V. 统计学习理论的本质[M]. 张学工译. 北京: 清华大学出版社, 2000.
- [14] 吴达科, 马承伟, 杜尚丰. 支持向量机在斑潜蝇虫害叶片光谱分析中的应用[J]. 农业机械学报, 2007, 38(10): 87-90.
- [15] Park B, Abbott JA, Lee KJ, *et al.* Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and Gala apples [J]. Trans ASAE, 2003, 46(6): 1721-1731.
- [16] 杨磊, 王化增, 陈子凌. 基于支持向量机的油气储量价值等级评价[J]. 中国石油大学学报(自然科学版), 2012, 36(3): 192-196.

(责任编辑: 张宏梁)

### 作者简介



马世榜, 博士生, 讲师, 主要研究方向: 农畜产品品质安全无损检测技术与装备。  
E-mail: mshibang@126.com



徐杨, 教授, 博士生导师, 主要研究方向: 农畜产品品质安全无损检测技术与装备。  
E-mail: xuyang@cau.edu.cn