基于高光谱成像的生鲜猪肉细菌总数预测建模方法研究

王 伟[1], 彭彦昆[1], 张晓莉[2]

1. 中国农业大学工学院, 北京 100083
2. 佐治亚州立大学生物系, 亚特兰大, 美国 P.O. Box 4010

摘 要 生鲜猪肉中细菌总数(TVC)超标会直接危害大众, 为此研究验证高光谱成像技术结合相应的建模方法预测生鲜猪肉中TVC的可行性。针对非线性、小样本问题, 以及光谱维和空间维的大数据量问题, 在综合比较偏最小二乘回归(PLSR)、人工神经网络(ANNs)和最小二乘支持向量机(LS-SVM)3种建模方法的基础上, 最终选取了LS-SVM方法组建模型。3种建模方法综合比较的结果表明, LS-SVM同时兼顾了训练精度和泛化能力两方面的性能, 使其都能做到最优, 与标准平板菌落计数法所检测TVC的决定系数分别为0.987 2和0.942 6, 校正均方根误差和预测标准均方根误差分别为0.207 1和0.217 6, 其建模性能优于其他方法。研究结果表明, 高光谱成像技术结合LS-SVM预测建模方法可作为快速、非破坏预测生鲜猪肉TVC的有效手段。

关键词 生鲜猪肉; 细菌总数; 高光谱成像系统; 最小二乘支持向量机
中图分类号: O433.4 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2010)02-0412-05

引 言

我国猪肉产量占据世界猪肉总产量的半壁江山, 但我国猪肉生产的卫生状况还不容乐观。细菌总数(total viable bacteria count, TVC)是衡量猪肉卫生状况的重要指标, 当肉类中的细菌数超标, 就会变成致病菌危害大众, 因此, 肉类的微生物安全问题已引起人们的广泛关注。

针对生鲜猪肉TVC的检测, 传统的检测方法大多基于显微镜[1]、细菌分离[2]、免疫学[3,4]和基于核酸分析[5]等方法, 这些方法虽然有效, 但采样准备和检测时间长、效率低、操作繁琐、费时费力、对试样具有破坏性, 更重要的是这些方法难以实现自动化及计算机化、不适合在线检测, 不能充分满足国内消费和出口贸易的需要。一些最近的研究采用傅里叶变换红外光谱(Fourier transform infrared spectroscopy, FTIR)检测牛肉的腐败程度[6,7]; Duboisa等采用近红外化学成像检测食品中自含微生物[8], 研究表明可以通过观察细菌在一些特定近红外波长的光吸收/反射差异辨识细菌; 图像光谱技术集成了传统的图像和光谱技术, 可同时获得被测物体的空间和光谱信息, 相比传统的多光谱方法不能同时兼顾共线性消除和信息充分获取的矛盾, 高光谱图像技术因其高的光谱分辨率已成为农产品品质评估和安全检测的强有力工具[9-13]。Park等[14,15]开发了高光谱图像系统用于检测禽肉表面的粪便和消化物污染从而间接检测其中所含的大肠埃希氏菌E. coli O157:H7。然而, 高光谱图像技术反映样品光谱和空间特征的信息量大, 如何处理这一复杂信息, 建模方法的合理与否对最终预测能力的影响十分明显。本文在高光谱图像技术的基础上, 致力于探索生鲜猪肉TVC的合理预测建模方法。

1 仪器与试样准备

1.1 实验仪器

所采用的高光谱成像系统如图1所示。系统主要包括一个高性能的背后照明式CCD相机(Sencicam QE, Germany)及其控制单元; 一台波长范围在400~1 000 nm的图像光谱仪(ImSpector V10E, Spectral Imaging Ltd., Finland), 其光谱分辨率为2.8 nm, 波长间隔为0.74 nm; 由石英卤钨灯和稳压电源组成的光源系统(Oriel Instruments 66882, USA); 试样载物台及其调节机构及聚光镜等。在使用前按照文献[16]的方法进行空间和光谱维校正。试样载物台可沿高度方向进行调节, 光源为5 mm圆形光束其发散角小于20°。肉样放置在载物台上, 当光源照射在肉表面时, 肉组织的漫反射光经聚光镜、光谱仪, 然后通过CCD相机形成光谱图像, 经

收稿日期: 2009-02-12, 修订日期: 2009-05-16
基金项目: 国家自然科学基金项目(30771244), (863计划)项目(2008AA10Z210)和北京市自然科学基金项目(6082017)资助
作者简介: 王 伟, 1975年生, 中国农业大学工学院讲师 e-mail: wxw@mail.buaa.edu.cn; playerwxw@cau.edu.cn
© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

数据采集卡生成 16 位图像数据文件。

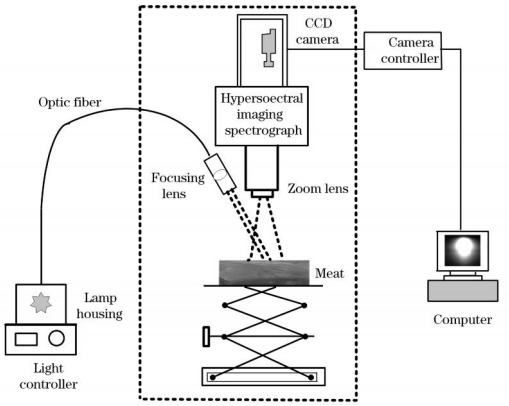

Fig. 1 Hyperspectral imaging system configuration

1.2 试样准备

2008 年 1 月从北京市顺义区某肉联厂购置宰后 24 小时出自同一头猪的新鲜猪里脊肉, 运至中国农业大学农畜产品无损检测实验室, 无须清洗、剔除脂肪或结缔组织、接种细菌等准备工作, 将肉样均匀分割成长@宽@厚为 4 cm@6 cm@4 cm 的肉块, 剔除不规则肉块后共获得 26 块试样。所有试样用密封袋包装后先在 4℃冰箱中保鲜存放 24 h。为加速肉样的腐败过程, 存放 24h 后每隔 2 h, 即在存放后的第 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 和 24 h, 分别取出 2 块试样放置在位于室温 20~ 22 ℃的工作台上。其中一块作为校正集试样, 另一块作为预测集试样, 直至所有试样均从冰箱中取出, 这样校正集和预测集均包含 13 个试样。

1.3 高光谱图像采集和细菌总数检测

当所有试样均从冰箱中取出后, 在室温中醒 20 min, 用高光谱成像系统获取所有试样的光谱图像, 为减小位置的影响, 采集图像时使每个试样在载物台水平面上每间隔 120°旋转一次, 分别采集相应位置的光谱图像, 并以其平均值作处理用。在扫描试样漫反射图像的过程中, 拉上窗帘并关闭所有照明灯以防止外部环境光的影响。设定 CCD 的曝光时间为 0.025 s, 调整光密度使其最大值为相机总动态范围的 1/3。在采集所有试样图像之前, 为校正相机暗电流和室内照明对图像的影响, 分别采用盖住镜头和采集标准白板图像的方法获取黑、白图像, 然后按公式(1)计算相对反射光谱

$$R = (R_s - R_d)/(R_r - R_d) \quad (1)$$

其中 R 为相对反射密度, R_s 为试样原始图像反射密度, R_r 为标准参考白板反射密度, R_d 为黑色图像。所有的光谱数据在各单个波长沿扫描线上取平均值。包括图像参数在内的图像数据均以 16 位二进制格式文件存放。

获取所有样品的光谱图像后在 2 h 内送至中国农业大学微生物检测实验室, 依据国标采用平板菌落计数法进行 TVC 的检测, 以此作为标准参考数据。表 1 为用国标方法测得的 13 个校正集样本的 TVC, log(TVC)与在某一单波长处的平均反射密度对照表。可见生鲜猪肉 TVC 随时间的变化规律及其与平均反射密度的关系均为显著非线性。而且, 基于试验的探索性目的和条件所限, 样本数量有限。因此, 为解决该类问题采用传统的多元线性回归(MLR), 偏最小二乘回归(PLSR), 甚至是基于经验风险最小化原则的人工神经网络(ANNs)建模方法均有可能难以取得良好的预测结果。

Table 1 TVCs, log(TVC) and reflectance intensity of pork meat for the calibration samples

样本	1	2	3	4	5	6	7	8	9	10	11	12	13
TVC(CFU/g, @10^7)	0.21	2.15	4.00	131.70	7.15	2.50	14.10	301.70	7.55	61.95	8.80	23.8	6.85
log(TVC)	6.32	7.33	7.60	8.14	7.85	7.40	8.15	8.49	7.88	7.84	7.94	8.38	7.84
反射密度(@10^3)	1.58	1.38	1.55	1.41	1.61	1.46	1.56	1.57	1.38	1.49	1.65	1.39	1.39

2 最小二乘支持向量机的回归原理

基于统计学习理论基础的支持向量机方法[17, 18](support vector machines, SVM), 通过核函数将低维非线性问题隐式的映射成高维的线性问题, 使问题简化, 同时在结构风险最小化原则基础上通过间隔最大化方法, 可同时兼顾训练精度和泛化能力两方面性能, 因此该方法在分类、回归和人脸识别等诸多领域均获得了广泛应用。同时, 为避免因复杂优化算法带来的计算速度问题, 这里拟采用一种改进的 SVM 回归方法即最小二乘支持向量机[19](least square support vector machines, LS-SVM)对 TVC 进行预测, 其估计函数的一般形式为

$$y(x) = \sum a_i K(x, x_i) + b \quad (i = 1, 2, \ldots, N) \quad (2)$$

针对核函数的选取问题, 依据 TVC 预测问题的实际, 本文选取径向基核函数

$$K(x_j, x_i) = \exp\{-\|x_j - x_i\|^2/2R^2\} \quad (3)$$

3 结果与讨论

3.1 光谱图像预处理

光谱分析采用自编软件包, 为克服肉样厚度和表面不平度差异, 首先对相对反射光谱图像采用 SNV 校正。在此基础上, 分别获取一阶和二阶微分光谱。试验分析表明, 采用二阶微分光谱作为原始数据, 联合 LS-SVM 建模方法进行预测能够取得良好的结果。

3.2 细菌总数的 PLSR 和 ANNs 预测建模方法

为比较各种建模方法预测 TVC 的能力, 分别采用 PLSR 和 ANNs 方法, 采用校正集样品对光谱 520 个波段的全波段数据进行建模, 然后依据所建模型对预测集样品进行验证, 性能以决定系数(R^2)、校正均方根误差(RMSEC)和预测均方根误差(RMSEP)来度量, PLSR 和 ANNs 模型的预测结果分别如图 2 和图 3 所示。

PLSR 方法, 因其在提取主成分时考虑到与待分析组分的相关性, 所建线性模型通常具有较高的预测精度, 因此本

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

处采用 PLSR 方法所建模型对预测集 13 个样本的 TVC 进行了预测，结果如图 2 所示，R^2 = 0.719 4 和 RMSEV = 0.457 9，但从预测曲线可以看出，其对部分样本的预测结果非常好，而对其他部分的预测结果却很差，总体来看呈现出一种两极分化的状态，预测效果欠佳。

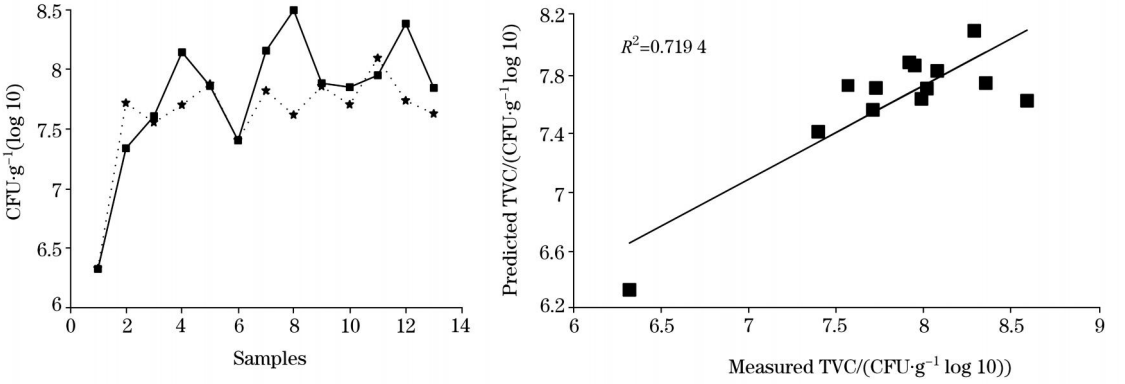

Fig. 2 Prediction result of PLSR

1: Measured value; 2: Predicted value

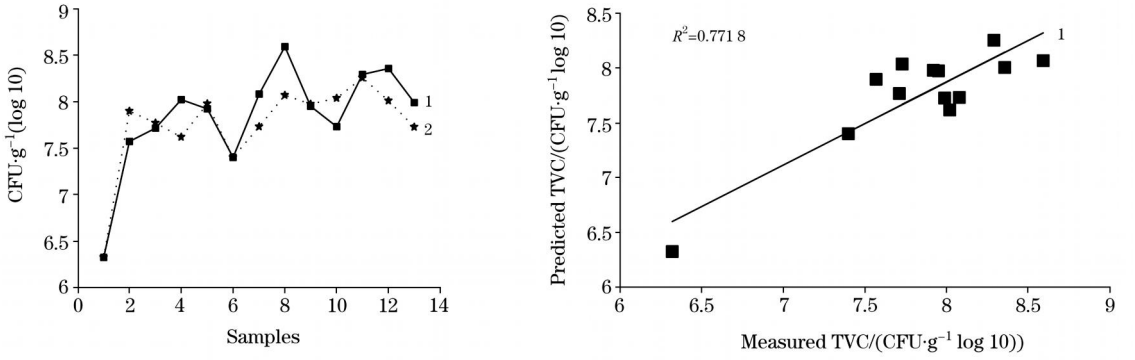

Fig. 3　Prediction result of ANNs

1: Measured value; 2: Predicted value

在众多的非线性建模方法中，ANNs 因其结构简单易于实现已成为强有力的建模工具，因此，这里采用它所建模型的预测结果如图 3 所示，相比较 PLSR 方法其性能有所增强，R^2 = 0.771 8，RMSEC = 0.421 8，预测结果基本反映了 TVC 的变化趋势，但相对于其校正过程的结果，反映出模型的稳定性不好，泛化能力一般。这是因为 ANNs 方法过多地依赖于训练集数据的质量，对噪声等无用信号往往一同进行训练，容易导致过拟合使得泛化性能变差。因此，若要对生鲜猪肉 TVC 进行更好的预测，需要更为有效的方法。

3.3　细菌总数的 LS-SVM 预测建模方法研究

采用 LS-SVM 方法所得到的模型对校正集 13 样本的 TVC 预测值和采用国标法获得的标准值之间的关系分别为 R^2 = 0.987 2 和 RMSEC = 0.169 1。而对预测集 13 个样本进行预测得到 R^2 = 0.942 6 和 RMSEP = 0.217 6。如图 4 所示

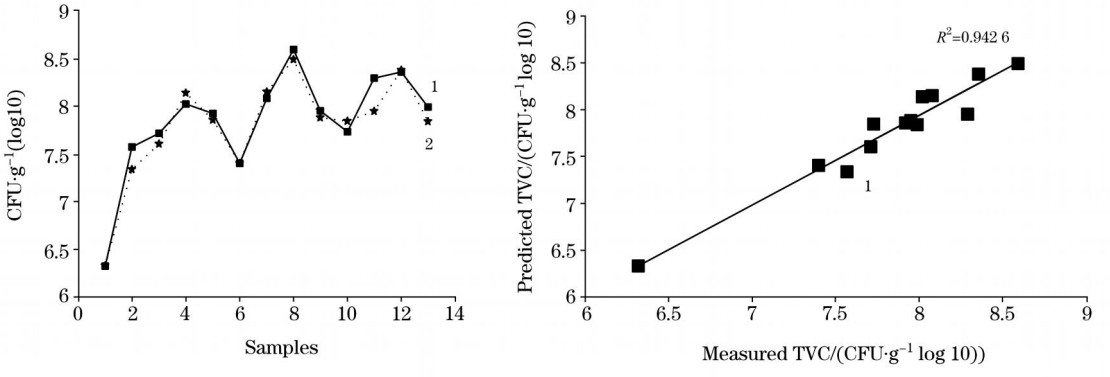

Fig. 4　Prediction result of LS-SVM

1: Measured value; 2: Predicted value

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

预测结果已比较理想。可见, LS2SVM 建模方法能够同时兼顾训练精度和泛化能力两方面的性能。但 LS2SVM 参数选择的是否合理直接决定着其性能的优劣, 针对该问题截至目前尚无固定可循的理论和方法, 这里采用贝叶斯证据框架对其参数进行推断优化[21], 可以避免像交叉验证方法一样事先需要准备交叉验证集; 而且贝叶斯框架还能对不同网络结构的性能进行比较, 从而选取最优核函数或核参数[21, 22]。

3.4 4 种建模方法建模性能的比较

表 2 分别以 R^2 和 RMSEP 为指标对 3 种建模方法的性能进行了比较。可见, 虽然神经网络方法的训练效果最佳, 但其预测泛化性能一般; 而 LS2SVM 同时兼顾了训练精度和泛化能力两方面的性能, 在两者之间做到了最佳折衷使得两方面都能做到最优。综合比较的结果表明, 其 LS2SVM 的建模性能优于其他 3 种建模方法, 因此我们采用它作为 TVC 的预测建模方法。

Table 2 Comparative result of the three modeling methods

指标	PLSR	ANNs	LS2SVM
R^2	0.719 4	0.771 8	0.942 6
RMSEC	0.198 2	0.051 3	0.169 1
RMSEV	0.457 9	0.421 8	0.217 6

4 结 论

相比于传统的实验室化学分析等方法的费时费力、对试样具有破坏性的缺憾, 光谱学方法可以有效地实现非破坏、快速检测肉中 TVC 的目的, 尤其是高光谱图像方法, 具有高的光谱分辨率, 它集成了传统的图像和光谱技术, 可同时获得被测物体的空间和光谱信息, 从而有效克服了传统多光谱方法不能同时兼顾共线性消除和信息充分获取的矛盾, 因此本文采用高光谱图像技术对生鲜猪肉的 TVC 进行了预测。与此同时, 高光谱图像技术反映样品光谱和空间特征的信息量大而且复杂, 建模方法的确定对最终预测能力的影响十分明显, 本文在高光谱图像技术的基础上, 针对小样本非线性问题, 通过比较四建模方法的性能, 探索出针对生鲜猪肉 TVC 合理有效的预测建模方法。预测结果表明, LS2SVM 性能优于其他建模方法, 结合高光谱成像系统, 可以作为预测生鲜猪肉 TVC 的有效手段。

应该看到, 尽管本文的研究初步取得了良好的结果, 但采用高光谱 520 个全波段数据进行建模, 一方面数据量巨大, 存储需求高, 运算时间长; 另一方面, 全波长数据的各个波段之间存在着多重共线性, 不利于充分提取最为重要的信息。因此, 为了研制出实用的便携式 TVC 检测仪器, 在本研究的基础上, 进一步的研究应着重致力于 TVC 的高光谱特征波长的优化选取问题。同时, 为避免在空间位置上遗漏重要信息, 在图像的采集方式上, 有必要采取以面扫描方式取代线扫描方法, 以便充分利用肉腐败过程中的光扩散信息。

参 考 文 献

[1] Champiat D, Matas N, Monfort B, et al. Luminescence, 2001, 16: 193.
[2] Hoon H S, Welson W W, Jeong S S. Immunology Letters, 2006, 106(2): 191.
[3] Durant J A, Young C R, Nisbet D J. International Journal of Food Microbiology, 1997, 38(2): 181.
[4] Nathalie Y F, Mulchandani A, Chen W. Analytical Biochemistry, 2001, 289: 281.
[5] Sylvie P, Francoise D, Burkhard M. Molecular and Cellular Probes, 2004, 18(6): 409.
[6] Ellis D I, Broadhurst D, Goodacre R. Analytica Chimica Acta, 2004, 514: 193.
[7] Ellis D I, Broadhurst D, Kell D B. Appl. Environ. Microbiol, 2002, 68: 2822.
[8] Duboisa J, Lewisb E N. Calvey E M. Food Microbiology, 2005, 22: 577.
[9] Hon H, Peng Y, Lu R. Transactions of the ASAE, 2007, 50(3): 963.
[10] Qiao J, Wang N, Ngadi M O, et al. Meat Science, 2007, 76: 1.
[11] Kim M S, Chen Y R, Mehl P M. Transactions of the ASAE, 2001, 44(3): 721.
[12] Lu R, Peng Y. Laser Focus World, 2005, 41: 99.
[13] Peng Y, Lu R. Postharvest Biology and Technology, 2008, 48(1): 52.
[14] Park B, Windham W R, Lawrence K C, et al. Proceedings of SPIE, 2002a, 4816: 308.
[15] Park B, Lawrence K C, Windham W R, et al. ASAE Paper, 2002b, (2): 313.
[16] Lu R, Chen Y R, Park B, et al. ASAE Paper, 1999, (99): 3120.
[17] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer2Verlag, 2005.
[18] Vapnik V N. IEEE Trans Neural Network (S1045-9227). 1999, 10(5): 988.
[19] Suykens J A K, Vandewalle J. Neural Network Letters (S1057-7122), 1999, 19(3): 293.
[20] Gestel T V, Suykens J A K. IEEE Transactions on Neural Network, 2001, 12: 809.
[21] Kowk J T. IEEE Transactions on Neural Network, 2000, 11: 1162.
[22] Law M H, Kowk J T. Proceedings of the 12th European Conference on Machine Learning. Freiburg, Germanny, 2001. 312.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Study on Modeling Method of Total Viable Count of Fresh Pork Meat Based on Hyperspectral Imaging System

WANG Wei[1], PENG Yan2kun[1], ZHANG Xiao2li[2]

1. College of Engineering, China Agricultural University, Beijing 100083, China
2. Department of Biology, Georgia State University, P. O. Box 4010, Atlanta, Georgia, USA

Abstract Once the total viable count (TVC) of bacteria in fresh pork meat exceeds a certain number, it will become pathogenic bacteria. The present paper is to explore the feasibility of hyperspectral imaging technology combined with relevant modeling method for the prediction of TVC in fresh pork meat. For the certain kind of problem that has remarkable nonlinear characteristic and contains few samples, as well as the problem that has large amount of data used to express the information of spectrum and space dimension, it is crucial to choose a logical modeling method in order to achieve good prediction result. Based on the comparative result of partial least2squares regression (PLSR), artificial neural networks (ANNs) and least square support vector machines (LS2SVM), the authors found that the PLSR method was helpless for nonlinear regression problem, and the ANNs method couldn. t get approving prediction result for few samples problem, however the prediction models based on LS2SVM can give attention to the little training error and the favorable generalization ability as soon as possible, and can make them well synchronously. Therefore LS2SVM was adopted as the modeling method to predict the TVC of pork meat. Then the TVC prediction model was constructed using all the 512 wavelength data acquired by the hyperspectral imaging system. The determination coefficient between the TVC obtained with the standard plate count for bacterial colonies method and the LS2SVM prediction result was 0.987 2 and 0.942 6 for the samples of calibration set and prediction set respectively, also the root mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP) was 0.207 1 and 0.217 6 individually, and the result was considerably better than that of MLR, PLSR and ANNs method. This research demonstrates that using the hyperspectral imaging system coupled with the LS2SVM modeling method is a valid means for quick and nondestructive determination of TVC of pork meat.

Keywords Fresh pork meat; Total viable count of bacteria; Hyperspectral imaging system; Least square support vector machines

(Received Feb. 12, 2009; accepted May 16, 2009)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net