

基于粒子群聚类的牛肉含水率光谱检测技术*

唐 鸣 徐 杨 彭彦昆 汤修映 田潇瑜 牛力钊

(中国农业大学工学院,北京 100083)

摘要:采用市场当日上架的生鲜牛肉外侧最长肌制作样本,在波长 900 ~ 2 300 nm 内进行光谱检测和分析。利用基于粒子群算法(PSO)的聚类分析方法,对光谱信息进行优化以减少计算量,提高回归模型精度。该算法以经过多元散射校正(MSC)、变量标准化(SNV)等方法预处理后的光谱信息作为目标矩阵,以波长为目标进行聚类,根据聚类结果对不同波段进行重新组合,并建立偏最小二乘回归(PLSR)模型。结果表明,利用 PSO 聚类分析方法在 900 ~ 1 400 nm 波段内获得的生鲜牛肉含水率预测模型最优, $R_c = 0.920 5$, $R_v = 0.919 1$ 。该方法能够有效减少光谱的数量,提升回归模型的预测结果。

关键词:牛肉 总含水率 标准粒子群算法 聚类分析 偏最小二乘回归

中图分类号: S123; TP301.6 **文献标识码:** A **文章编号:** 1000-1298(2014)10-0220-06

引言

基于光谱分析技术的生鲜牛肉无损检测方法,其中一个重要步骤是牛肉含水率的检测。目前,以回归分析为主的含水率预测模型应用较多,赵丽丽等对肉制品的含水率进行了预测,其预测集相关系数为 0.9^[1-5]。

波长 300 ~ 2 500 nm 范围内包含了 H₂O 及 -OH 的若干个较强吸收波长,因此可见/近红外波段的光谱常用于和水分相关的研究中^[6-11]。目前,基于近红外光谱分析技术的预测模型以回归模型为主,采用较多的是传统回归分析法,例如多元线性回归以及非线性的偏最小二乘回归(Partial least square regression, PLSR)等。近年来,结合智能算法的非线性回归方法也得到了广泛应用且效果良好,例如采用人工神经网络结合偏最小二乘回归的方法对生菜叶片含水率的预测^[12],采用支持向量机的偏最小二乘方法对初制绿茶的含水率进行了预测^[13],均取得了比单纯使用传统回归方法更高的预测集相关系数。就运算过程来看,单纯的非线性的回归方法算法复杂,收敛效率不高,因此国内外学者也探讨了采用优化算法与之相结合的方法以降低维度及其运算量,例如利用遗传算法,对即将进行回归分析的数据矩阵进行优化处理,明显提高了预测效果^[14-16]。Shi 等发现,粒子群算法

采用速度-位移的模型操作简单,避免了进化类算法中遗传的复杂操作,在运算速度和运算寻优能力上优势比较显著,在工程应用方面取得了比遗传算法更好的效果^[2]。

粒子群优化算法(Particle swarm optimization, PSO)是一种非监督的机器学习算法,也是基于群体迭代搜寻最优解的优化算法。国内对于粒子群算法及其改进算法的应用逐渐得到普及,但到目前为止,在光谱分析上的应用极少^[17],而以光谱数据作为目标矩阵进行聚类优化的应用尚未见报道。本文采用基于 PSO 的聚类分析和非线性回归相结合的方法对牛肉含水率进行检测。

1 材料和方法

1.1 实验系统

在自建的肉品品质及安全光谱无损检测系统中进行实验,主要仪器包括:NIRQuest 型光谱仪,分辨率约为 6 nm,标称波长范围 900 ~ 2 500 nm(实际范围约为 870 ~ 2 490 nm);Vivo 反射式面光源,波长范围 360 ~ 2 200 nm。其中光谱仪通过数据线连接计算机,由计算机发出相应的控制指令实施光谱信息的采集并保存数据。光源为独立调节式,位于计算机中的光谱仪处理系统,能够根据采集条件(光强及探头距离)自动调整积分时间;对原始光谱实施 5 次平均和 3 点平滑。

收稿日期:2013-09-06 修回日期:2013-10-10

*公益性行业(农业)科研经费资助项目(201003008)和“十二五”国家科技支撑计划资助项目(2012BAH04B00)

作者简介:唐鸣,博士生,主要从事电子测量及无损检测研究,E-mail: tang_ming@foxmail.com

通讯作者:徐杨,教授,博士生导师,主要从事无损检测技术及新材料研究,E-mail: xuyang@cau.edu.cn

1.2 材料及光谱采集

制备实验样品的牛肉属于同一牛肉品种,为超市当日上架冷鲜肉,取背部最长肌,在 4℃ 恒温条件下运送至实验室。根据光源特点制备样品规格为 6 cm × 6 cm × 7 cm。在室温条件下,快速制作样本,尽量减小样品在恒定环境中因存放时间不同而造成的误差。制成的牛肉样本在采集光谱之前,需要在空气中曝露 30 min 左右^[18],目的在于使得样品表面水分自然挥发,以避免因表面水分反射率过高所带来的影响。采集光谱时将样本平置于光源载物面上,对每个样品在其表面 3 个不同位置共计实施 3 次采集,并以所有光谱均值作为该样本的最终光谱信息。

所获取到的反射光谱采用主成分分析等方法剔除其中的异常值^[19-20],最终得到用于建模的样本数量共计 96 个,即 96 条有效光谱曲线。

肉品含水率的测量全过程依据 GB/T 9695.15—2008《肉与肉制品 水分含量测定》进行操作。

1.3 粒子群优化算法

粒子群算法中 z_i ($i = 1, 2, \dots, d$, 下同) 为粒子的位置矢量,根据要解决问题的度量标准,计算位置矢量当前的适度值,并根据度量标准评价位置优劣;粒子的飞行速度为 v_i , p_i 为第 i 个粒子目前已计算得到的最优位置, p_g 为所有粒子目前已计算得到的最优位置。因此,粒子飞行速度和位置关系为

$$v_{id}^{k+1} = v_{id}^k + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \quad (1)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (2)$$

式中 d ——目标矩阵的种类维度

k ——迭代次数

r_1, r_2 —— $[0, 1]$ 之间的随机数,用来保持群体的多样性

c_1, c_2 ——学习因子,也可称之为加速因子

所谓标准粒子群算法,便是在上述算法中增加了权重因子之后的粒子群算法。标准粒子群算法中,权重因子的作用体现在影响收敛速度、权衡局部最优和全局最优的能力上^[21],权重因子的增加使得粒子飞行速度重新构建为

$$v_{id}^{k+1} = w v_{id}^k + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \quad (3)$$

权重因子是由 Shi 等提出并以 w 表示^[2]。实验表明, w 较小时(小于 0.8),如果能够在初始空间内找到全局最优解,则此时所用时间将是该算法的最短运算时间,否则将难以找到最优解。 w 较大(小于 1.2)时,将增加寻找最优解所耗费的时间。因此,在粒子群算法工作过程中,通常采用动态的调整权重,即

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{I_{\max}} k \quad (4)$$

式中 w_{\max} ——初始权重 w_{\min} ——最终权重

I_{\max} ——最大迭代次数

除 k 之外,前 3 个参数加上学习因子 c_1 和 c_2 、类别数目 K ,这 6 个参数是粒子群算法开启之前所需要设定的参数。

1.4 光谱的聚类 and 建模

基于 PSO 算法,对光谱矩阵以波长为对象实施聚类分析,为建立回归分析模型准备数据。

1.4.1 聚类分析

聚类分析(Cluster analysis)是指将一大类中的所有个体按照相互之间的相似程度归入几个子类当中,使得类内的相似性、类间的差异性达到最大,即将 n 个个体的集合分为 N 个类组($N < n$)^[22]。将基于 PSO 的聚类分析方法用于光谱矩阵的优化,能够发现并掌握各波长之间相互的作用和调控的关系,聚类的结果以相似度准则进行考察。光谱表达矩阵为 M_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)。算法中粒子更新所依据的适应度函数为均方差函数(MSE)。

$$\tilde{D} = \frac{1}{M} \sum_{i=1}^M (d_{\min}(x_i))^2 \quad (5)$$

$$\text{其中 } d_{\min}(x_i) = \min_{y_j \in Y} d(x_i, y_j) \quad (6)$$

式中 $d(x_i, y_j)$ ——欧氏距离

$$d(x, y) = \sqrt{\sum_{j=1}^m (M_{xj} - M_{yj})^2} \quad (x, y = 1, 2, \dots, n) \quad (7)$$

据此可知,采用聚类分析的方法对目标矩阵 M 进行优化处理的目的是:每个波长都对应了 96 个条件,理想状态下每个波长依据条件数的顺序变化而变化的规律应当相同,而实际情况是,被检测对象在物理及化学性质上或多或少存在着不同,对应着不同的波长或不同的波段,将具有不同的吸收特点和变化趋势,因此聚类的结果便是根据这种特点及趋势,对相同变化特点或趋势的波长实现聚集。

1.4.2 偏最小二乘回归

聚类分析的结果为光谱矩阵的优化和组合提供了依据,新的波段组合采用回归分析的方法为参考值建立回归模型,较为广泛应用的方法为 PLSR。

PLSR 是传统的回归分析方法,能够较好地消除多重共线性等问题,取得较好的回归精度。PLSR 采用了主元素提取的方法,在抽取主成分的过程中能够较为完整地保留光谱中的数据,采用迭代的方法进行运算,运算过程中累积误差较小,它是回归分析方法中精度较高的一种。

2 结果与分析

2.1 牛肉含水率测量结果

采用 GB/T 9695.15—2008 的方法测得用于实验的牛肉样品含水率,96 个有效样本中,依据样本编号顺序,每取 4 个样本中,前 3 个作为校正集数据,后 1 个作为验证集数据。如表 1 所示牛肉样品含水率,校正集和预测集所选用的样本,含水率范围类似,分布规划较为合理。

表 1 牛肉样品含水率

Tab. 1 Statistics data of beef water content %

样本集	含水率	均值	均方差
校正集	71.09 ~ 79.22	75.59	2.14
预测集	72.01 ~ 79.01	75.05	1.49
总样本	71.09 ~ 79.22	75.39	1.89

2.2 光谱的预处理

经过光谱仪简单处理过的光谱矩阵,用于定量分析之前需要进行预处理。常用的方法包括多元散射校正 (Multiplication scatter correction, MSC), 变量标准化 (Standard normal variate, SNV) 等。实践证明,经过 MSC 处理后得到的光谱数据可以有效地消除散射影响,增强与成分含量相关的光谱吸收信息^[23]。另外,光谱仪在截止波长附近,信噪比很低,通常要去掉这一段信息,因此本实验实际用于建模的光谱其波长范围为 896 ~ 2 283 nm。

图 1 为截取后经过 MSC 预处理的反射率光谱。截取后的光谱矩阵中,最后变量作为聚类样本数共 230 个,牛肉的样本数量作为聚类条件数共 96 个。

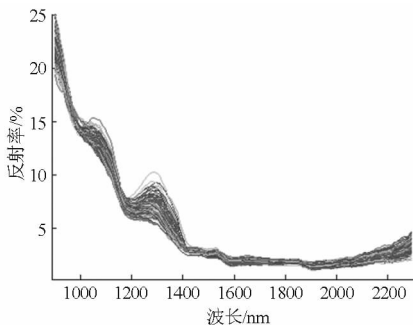


图 1 样品的近红外反射率光谱

Fig. 1 Reflectance of NIR spectra of beef samples

2.3 聚类结果及回归模型

针对目标矩阵 $M_{230 \times 96}$, 采用 Matlab2009 编写程序,根据文献[2, 24-25]及经验,第 1 次设定粒子群算法工作参数,其中学习因子 $c_1 = c_2 = 1.2$, 权重因子 $w_{\max} = 0.9$, $w_{\min} = 0.6$, 设定迭代次数为 100 次,初步设定分为 10 类,得到聚类结果如图 2 所示。

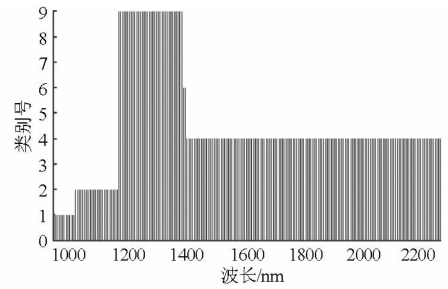


图 2 反射率光谱第 1 次聚类结果

Fig. 2 Result of first clustering

由图 2 可知,聚类分析结果以 10 组为参考值,最终根据光谱自身的特点构成了 5 个聚类中心,即纵坐标类别号为 1、2、4、6、9 的 5 个波段。

根据上述结果,可尝试将聚类总数预先设置为 5 类,随着类别维数的降低,本次聚类可适当调整最终权重 w_{\min} ,这个变化使得寻优过程中动态变化的权重值与上次聚类过程相比有所改变,聚类结果的改变可能会带来建模效果的提升。因此,第 2 次设定参数 $c_1 = c_2 = 1.2$, 权重因子 $w_{\max} = 0.9$, $w_{\min} = 0.4$, 分 5 类进行聚类,结果如图 3 所示。本次聚类结果实际构成了 5 个聚类中心,即 5 个波段组,由图 3 纵坐标可知,编号依次为 1 ~ 5 号。

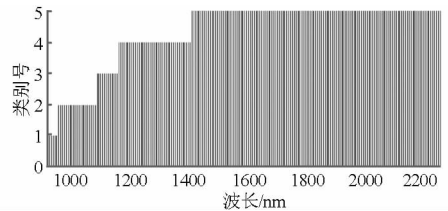


图 3 反射率光谱第 2 次聚类结果

Fig. 3 Result of second clustering

由图 2、3 可得,纵坐标代表的类别号以及各个波段起止波长的对应关系如表 2 所示。采用图 2 及图 3 所展示的聚类结果,以 PLSR 方法建立回归模型,全波段的 PLSR 模型预测结果: R_c 为 0.905 7, SEC 为 0.008 0, R_v 为 0.862 9, SEV 为 0.012 3。对聚类所得的各个波段组进行不同的组合并建立 PLSR 回归模型,最终结果如表 3 所示,其中主成分个数选取的标准能够描述原始信息 98% 以上。

表 2 反射率光谱聚类结果

Tab. 2 Clustering results and corresponding wavelength

图 2		图 3	
类别号	起止波长/nm	类别号	起止波长/nm
1	896 ~ 972	1	896 ~ 928
2	979 ~ 1 140	2	934 ~ 1 082
9	1 146 ~ 1 371	3	1 088 ~ 1 160
6	1 377 ~ 1 384	4	1 166 ~ 1 416
4	1 390 ~ 2 283	5	1 422 ~ 2 283

表 3 基于聚类结果的模型分析
Tab. 3 Analysis of model based on clustering results

序号	图 2 所示数据处理结果					图 3 所示数据处理结果				
	组合	R_c	SEC	R_v	SEV	组合	R_c	SEC	R_v	SEV
1	1-2	0.890 3	0.008 6	0.135 1	0.012 8	1-2	0.866 1	0.009 5	0.314 4	0.026 2
2	1-2-9	0.889 1	0.008 7	0.919 8	0.013 5	1-2-3	0.893 1	0.008 5	0.201 6	0.016 7
3	1-2-9-6	0.887 9	0.008 7	0.931 1	0.013 5	1-2-3-4	0.895 9	0.008 4	0.891 6	0.013 9
4	1-2-9-4	0.904 9	0.008 1	0.866 7	0.012 2	1-2-3-5	0.960 0	0.005 3	0.712 9	0.017 9
5	1-2-6	0.887 1	0.008 7	0.5080	0.016 4	1-2-4	0.917 9	0.008 1	0.908 9	0.009 0
6	1-2-6-4	0.951 4	0.005 8	0.848 7	0.016 1	1-2-4-5	0.925 6	0.007 2	0.847 2	0.014 4
7	1-2-4	0.949 6	0.005 9	0.798 5	0.016 7	1-2-5	0.948 0	0.006 0	0.722 4	0.015 0
8	1-9	0.869 9	0.009 3	0.825 0	0.010 4	1-3	0.792 1	0.011 6	0.682 7	0.015 0
9	1-9-6	0.870 8	0.009 3	0.836 9	0.010 7	1-3-4	0.887 3	0.008 7	0.892 2	0.017 4
10	1-9-6-4	0.918 9	0.007 5	0.846 6	0.016 1	1-3-4-5	0.921 0	0.007 4	0.857 8	0.015 8
11	1-6	0.838 5	0.010 3	0.731 3	0.013 0	1-4	0.848 4	0.010 0	0.773 5	0.013 5
12	1-6-4	0.941 0	0.006 4	0.835 1	0.015 8	1-4-5	0.935 3	0.006 7	0.759 5	0.018 0
13	1-4	0.943 1	0.006 3	0.790 6	0.016 4	1-5	0.937 6	0.006 6	0.554 9	0.020 2
14	2-9	0.935 4	0.006 7	0.901 4	0.007 5	2-3	0.895 4	0.008 4	0.867 9	0.017 6
15	2-9-6	0.947 5	0.006 1	0.773 8	0.010 5	2-3-4	0.953 9	0.005 7	0.891 0	0.014 3
16	2-9-6-4	0.918 5	0.007 5	0.789 5	0.014 8	2-3-4-5	0.919 7	0.007 4	0.861 4	0.012 6
17	2-6	0.906 1	0.008 0	0.460 5	0.021 3	2-4	0.920 5	0.007 4	0.919 1	0.011 1
18	2-6-4	0.932 3	0.006 9	0.751 6	0.014 8	2-4-5	0.942 5	0.006 3	0.849 0	0.014 5
19	2-4	0.931 1	0.006 9	0.647 2	0.016 5	2-5	0.968 5	0.004 7	0.697 3	0.016 6
20	9-6	0.930 9	0.006 9	0.798 5	0.012 3	3-4	0.903 0	0.008 1	0.838 7	0.014 4
21	9-6-4	0.918 4	0.007 5	0.647 7	0.018 2	3-4-5	0.911 9	0.007 8	0.738 4	0.015 8
22	9-4	0.916 6	0.007 6	0.636 9	0.018 2	3-5	0.936 7	0.006 6	0.565 1	0.020 1
23	6-4	0.904 0	0.008 1	0.599 7	0.026 0	4-5	0.920 0	0.008 2	0.570 8	0.018 8

2.4 分析与讨论

基于 PSO 的聚类方法,粒子结构的设计是基于聚类中心的,即聚类的结果是根据聚类中心而得到的,算法中各个参数的调整,是对初始聚类中心的微调,同时也是对寻优路线的调整。由于优化的度量标准是一定的,因此参数调整实际效果便是在不同聚类水平上对目标进行聚类。通常,本身结构较好,类间分离度较为紧致、类间差异性较大的目标矩阵,算法中参数的微小改变,不会造成聚类结果之间较大的差别^[21]。本文实验结果也验证了这一点。

如上所述,图 2 和图 3 分别显示了光谱在变换前后的聚类结果,光谱所代表的信息保持了较好的一致性,受到权重因子变化的影响有限,对比而言,大约 1 000 ~ 1 100 nm、1 200 ~ 1 400 nm 和 1 400 ~ 2 300 nm 3 个波段,始终被聚合到了同一个类别中,该结果也反映出目标矩阵 M 是较为可靠的数据集,即光谱信息能够较为可靠的反映出样本的实际情况。

图 2 设置初始聚类中心 10 个,最终形成了 5 个聚类中心。从目标矩阵本身的意义讲,聚类中心代表的是:96 个牛肉样品的光谱信息共计表现出 5 个特征。根据表 3 结果可知,编号为 2 和 9 的 2 类

对应波段相邻,2-9 组合的波段回归模型预测效果最好, $R_v = 0.901 4$,验证集误差为 0.007 5。

图 3 设置初始聚类中心为 5 个,并改变了粒子群算法运行的条件,最终也构成 5 个聚类中心。聚类结果与图 2 相比更加紧致和直观。其中采用编号为 2 和 4 的 2 类进行建模,具有较好的预测效果, $R_v = 0.919 1$,验证集误差为 0.011 1。

根据实际建模结果可知,较好的回归结果均出现在小于 1 400 nm 的波段内。

由此可见,虽然 PSO 参数的调整带来了运行效果的差异,但是目标矩阵本身较为稳定的聚集为 5 个类组,这种结果与光谱矩阵 M 的稳定性和可靠性是分不开的。

3 结论

(1)牛肉总含水率的无损检测技术,以聚类分析为具体应用形式,基于标准粒子群优化算法对光谱数据进行优化处理,再采用处理后不同波段的组合结果建立回归模型。

(2)以 PSO 聚类为先导的处理方法,起到了降低目标矩阵的维度、提高运算速度的效果,与传统 PLSR 下的全波段模型相比,有效的提高回归模型的

预测精度,相对于全波段回归结果的 $R_v = 0.8629$ 、预测集均方差 0.0123,该方法能够取得的最好结果为 $R_v = 0.9191$ 、预测集均方差 0.0111。

(3)基于 PSO 的聚类分析对目标矩阵的优化效果显著,在提高预测精度的前提下,优化了有效光谱波段,显著降低了运算量,提高了检测效率。

参 考 文 献

- 1 赵丽丽,张录达,宋忠祥,等.近红外光谱定量检测研制腊肉制品品质的研究[J].光谱学与光谱分析,2007,27(1):46-49. Zhao Lili, Zhang Luda, Song Zhongxiang, et al. Quality analysis of Chinese bacon with near infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2007, 27(1):46-49. (in Chinese)
- 2 Shi Y, Eberhart R C. A modified particle swarm optimizer[C]//Proceedings of IEEE International Congress on Evolutionary Computation, 1998:69-73.
- 3 高海兵,高亮,周驰,等.基于粒子群优化的神经网络训练算法研究[J].电子学报,2004,32(9):1572-1574. Gao Haibing, Gao Liang, Zhou Chi, et al. Particle swarm optimization based on algorithm for neural network learning[J]. Acta Electronica Sinica, 2004, 32(9):1572-1574. (in Chinese)
- 4 刘锦萍,郁金祥.基于改进的粒子群算法的多元线性回归模型参数估计[J].计算机工程与科学,2010,32(4):101-105. Liu Jinping, Yu Jinxiang. Parameter estimation of multiple linear regression models based on the improved particle swarm optimization algorithm[J]. Computer Engineering & Science, 2010, 32(4):101-105. (in Chinese)
- 5 Narsaiah K, Jha S N. Nondestructive methods for quality evaluation of livestock products[J]. Journal of Food Science and Technology, 2012, 49(3):342-348.
- 6 Naganathan G K, Grimes L M, Subbiah J, et al. Visible/near-infrared hyperspectral imaging for beef tenderness prediction[J]. Computer and Electronics in Agriculture, 2008,64(2):225-233.
- 7 Gamal E, Sun Dawen, Paul A. Near-infrared hyperspectral imaging for predicting color, pH and tenderness of fresh beef[J]. Journal of Food Engineering, 2012, 110(1):127-140.
- 8 Liu Y, Chen Y R, Ozaki Y. Two-dimensional visible/near infrared correlation spectroscopy study of thermal treatment of chicken meats[J]. Journal of Agricultural and Food Chemistry, 2000, 48(3):901-908.
- 9 杨建松,孟庆祥,任丽萍,等.近红外光谱法快速评定牛肉品质[J].光谱学与光谱分析,2010,30(3):685-687. Yang Jiansong, Meng Qingxiang, Ren Liping, et al. Rapid evaluation of beef quality by NIRS technology[J]. Spectroscopy and Spectral Analysis, 2010,30(3):658-687. (in Chinese)
- 10 孙晓明,卢凌,张佳程,等.牛肉化学成分的近红外光谱检测方法的研究[J].光谱学与光谱分析,2011,31(2):379-383. Sun Xiaoming, Lu Ling, Zhang Jiacheng, et al. Research on prediction chemical composition of beef by near infrared reflectance spectroscopy[J]. Spectroscopy and Spectral Analysis, 2011,31(2):379-383. (in Chinese)
- 11 Prieto N, Roehe R, Lavin P. Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review[J]. Meat Science, 2009, 83(2):175-186.
- 12 毛罕平,高洪燕,张晓东.生菜叶片含水率光谱特征模型研究[J].农业机械学报,2011,42(5):166-170. Mao Hanping, Gao Hongyan, Zhang Xiaodong. Spectral characteristic model of lettuce leaves' water content[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(5):166-170. (in Chinese)
- 13 李晓丽,程术希,何勇.基于漫反射光谱的初制绿茶含水率无损检测方法[J].农业工程学报,2010,26(5):195-201. Li Xiaoli, Cheng Shuxi, He Yong. Nondestructive measurement of moisture content of green tea in primary processing based on diffuse reflectance spectroscopy[J]. Transactions of the CSAE, 2010, 26(5):195-201. (in Chinese)
- 14 Bangalore A S, Schaffer R E. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least squares regression: application to near-infrared spectroscopy[J]. Analytical Chemistry, 1996, 68(23):4200-4212.
- 15 孙俊,毛罕平,羊一清,等.基于冠层光谱特性的水稻叶片含水率模型[J].农业工程学报,2009,25(9):133-136. Sun Jun, Mao Hanping, Yang Yiqing, et al. Model of moisture content of paddy rice leaf based on canopy spectral reflectance[J]. Transactions of the CSAE, 2009, 25(9):133-136. (in Chinese)
- 16 Li L, Cheng Y B, Ustin S, et al. Retrieval of vegetation equivalent water thickness from reflectance using genetic algorithm (GA)-partial least squares (PLS) regression[J]. Advances in Space Research, 2008,41(11):1755-1763.
- 17 何建平,金萍.离散粒子群优化算法在硅钢涂层近红外光谱厚度检测中的应用研究[J].光谱学与光谱分析,2011,31(9):2416-2419. He Jianping, Jin Ping. Study on the application of DBPSO algorithm to thickness measurement of surface insulation coating of silicon steel by NIR spectrometry[J]. Spectroscopy and Spectral Analysis, 2011,31(9):2416-2419. (in Chinese)
- 18 Ripoll G, Alberti P, Panea B, et al. Near-infrared reflectance spectroscopy for predicting chemical, instrumental and sensor quality of beef[J]. Meat Science, 2008, 80(3):697-702.
- 19 Osborne S D, Jordan R B, Kunnemeyer R. Method of wavelength selection for partial least squares [J]. Analyst, 1997, 122(12):1531-1537.
- 20 陈斌,邹贤勇,朱文静.PCA结合马氏距离法剔除近红外异常样品[J].江苏大学学报:自然科学版,2008,29(4):277-279.

- Chen Bin, Zou Xianyong, Zhu Wenjing. Eliminating outlier samples in near-infrared model by method of PCA-Mahalanobis distance[J]. *Journal of Jiangsu University: Natural Science Edition*, 2008, 29(4):277-279. (in Chinese)
- 21 纪震, 廖慧连, 吴青华. 粒子群算法及应用[M]. 北京: 科学出版社, 2009.
- 22 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. *软件学报*, 2008, 19(1):48-61.
Sun Jigui, Liu Jie, Zhao Lianyu. Clustering algorithm research[J]. *Journal of Software*, 2008, 19(1):48-61. (in Chinese)
- 23 芦永军, 曲艳玲, 宋敏. 近红外相关光谱的多元散射校正处理研究[J]. *光谱学与光谱分析*, 2007, 27(5):877-880.
Lu Yongjun, Qu Yanling, Song Min. Research on the correlation chart of near infrared spectra by using multiple scatter correction technique[J]. *Spectroscopy and Spectral Analysis*, 2007, 27(5):877-880. (in Chinese)
- 24 蒲蓬勃, 王鸽, 刘太安. 基于粒子群优化的模糊 C-均值聚类改进算法[J]. *计算机工程与设计*, 2008, 29(16):4277-4279.
Pu Pengbo, Wang Ge, Liu Taian. Research of improved fuzzy C-means algorithm based on particle swarm optimization[J]. *Computer Engineering and Design*, 2008, 29(16):4277-4279. (in Chinese)
- 25 张武, 陈剑, 高煜. 基于粒子群算法的发动机悬置系统稳健优化设计[J]. *农业机械学报*, 2010, 41(5):31-35.
Zhang Wu, Chen Jian, Gao Yu. Robust optimal design of an engine mounting system based on particle swarm optimization [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(5):31-35. (in Chinese)

Water Content Detecting of Beef Based on Spectral Analysis and Clustering Analysis of PSO Algorithm

Tang Ming Xu Yang Peng Yankun Tang Xiuying Tian Xiaoyu Niu Lizhao
(*College of Engineering, China Agricultural University, Beijing 100083, China*)

Abstract: Total water content is an important quality attribute for consumer satisfaction, and a more accurate pre-detecting method is necessary. The conventional method of partial least squares regression (PLSR) has been widely used in meat water content forecasting. In this study, the cluster analysis of particle swarm optimization algorithm was carried out and calibrated as one of the optimization methods of PLSR with the goal of reducing computation complexity and enhancing the prediction precision. Based on the novel method above, a predicting model of beef water content was developed in wavelength range of 900 ~ 2 300 nm, and the best predicting result with $R_c = 0.9205$ and $R_v = 0.9191$ was obtained in wavebands of 900 ~ 1 400 nm. The samples used in the experiment were beef longissimus collected from the supermarket in that day, and the water contents of samples were detected according to the national standard. Spectra of samples were acquired in reflectance spectral detection system and pre-treated procedure was carried out by means of multiplication scatter correlation method before model construction.

Key words: Beef Total water content Standard particle swarm optimization Clustering analysis
Partial least square regression