

基于近红外光谱技术的鲢鱼营养成分的快速分析

徐文杰^{1,2}, 刘欢^{1,2}, 陈东清^{1,2}, 熊善柏^{1,2*}

(1. 华中农业大学食品科学技术学院, 武汉 430070; 2. 国家大宗淡水鱼加工技术研发分中心(武汉), 武汉 430070)

摘要: **目的** 通过采集鲢鱼的近红外光谱数据和测定鱼肉营养成分含量探索鲢鱼营养成分的快速分析方法。**方法** 采集 254 个鲢鱼鱼肉样品的近红外光谱数据, 经过多元散射校正、正交信号校正、数据标准化等 20 种方法预处理, 在 1000~1799 nm 光谱范围内, 结合化学实测值分别采用偏小二乘法、主成分分析和 BP 人工神经网络技术、偏小二乘法和 BP 人工神经网络技术建立鲢鱼营养成分近红外定量模型。**结果** 鲢鱼鱼肉粗蛋白含量为 12.05%~19.05%, 粗脂肪含量为 0.24%~5.27%, 水分含量为 72.62%~80.58%, 灰分含量为 0.46%~1.50%, 数据范围较大, 可满足建模要求。在 3 种建模方法中, 近红外光谱数据结合偏小二乘法建立的鲢鱼营养成分模型最优, 所得的粗蛋白、粗脂肪、水分和灰分的近红外定量模型的相关系数分别为 0.9969、0.9925、0.9831 和 0.9976。**结论** 采用近红外光谱数据和偏小二乘法建立的模型具有较好的预测能力, 能较为准确、快速地分析出鲢鱼鱼肉粗蛋白、粗脂肪、水分和灰分的含量。**关键词:** 近红外光谱; 鲢鱼; 营养成分; 偏小二乘; 主成分分析; 人工神经网络

Fast analysis on nutrient of silver carp based on near infrared spectroscopy

XU Wen-Jie^{1,2}, LIU Huan^{1,2}, CHEN Dong-Qing^{1,2}, XIONG Shan-Bai^{1,2*}

(1. College of Food Science and Technology, Huazhong Agricultural University, Wuhan 430070, China;
2. National R&D Branch Center for Conventional Freshwater Fish Processing (Wuhan), Wuhan 430070, China)

ABSTRACT: Objective To explore a rapid analysis method for nutrient of silver carp through collecting near infrared spectroscopy and determining nutrient. **Methods** The near infrared (NIR) spectra of 254 silver carp samples were collected. The diffuse reflectance spectra of samples were performed with different spectral pretreatments, such as multiplicative scatter correction (MSC), orthogonal signal correction (OSC), and standardization (S). The near infrared quantitative analysis models were obtained by partial least square (PLS) regression, principal component analysis (PCA) combined with back propagation artificial neural network (BP-ANN), partial least square combined with back propagation artificial neural network with 1000~1799 nm, respectively. **Results** The results showed that the protein content of silver carp ranged from 12.05% to 19.05%, the fat content from 0.24% to 5.27%, the moisture content from 72.62% to 80.58%, and the ash content from 0.46% to 1.50%. The nutrient measured values met the modeling requirements. The analysis models obtained by PLS were the best. The correlation coefficients of the models were 0.9969, 0.9925, 0.9831 and 0.9976 for protein, fat, moisture and ash content, respectively. **Conclusion** The results indicated that the

基金项目: 国家现代农业产业技术体系专项(CARS-46-23)、“十二五”国家科技支撑计划项目(2013BAD19B10)

Fund: Supported by China Agriculture Research System (CARS-46-23) and National Key Technology R&D Program in Twelfth-Five-Year Plan (2013BAD19B10)

*通讯作者: 熊善柏, 教授, 主要研究方向为水产品加工及贮藏工程。E-mail: xionsb@mail.hzau.edu.cn

*Corresponding author: XIONG Shan-Bai, Professor, College of Food Science and Technology, Huazhong Agricultural University, No. 1, Shizishan Street, Hongshan District, Wuhan 430070, China. E-mail: xionsb@mail.hzau.edu.cn

models exhibited an acceptable fitting accuracy and predictive ability of analysis of protein, fat, moisture and ash content of silver carp by NIRS.

KEY WORDS: near infrared spectroscopy; silver carp; nutrient; partial least square; principal component analysis; artificial neural network

鲢鱼作为四大家鱼之一, 是我国淡水鱼的主要品种。淡水鱼作为一种高蛋白、低脂肪、营养丰富的健康食品, 已成为我国居民优质动物蛋白的重要来源^[1]。随着人们生活水平的提高, 市场对水产品的需求量越来越大, 水产品的营养质量也越来越受到人们的关注^[2-3]。如何快速获得水产品的营养信息已成为人们关注的热点问题。目前, 水产品营养成分含量多采用化学方法测定, 费时费力且操作复杂。而近红外光谱(near infrared spectroscopy, NIRS)分析技术具有快速、无损的技术优势, 不仅可用于品种鉴别^[4-6], 还可用于农产品的品质评价^[7-9]。本实验室已经验证近红外光谱技术可以用于草鱼营养成分的快速检测分析^[10], 为得到准确性更高、稳定性更强的近红外定量分析模型, 本实验在测定 254 个鲢鱼样品的粗蛋白、粗脂肪、水分和灰分含量和采集各样品近红外光谱的基础上, 在 1000~1799 nm 范围内分别采用偏最小二乘法(partial least square, PLS)、主成分分析(principal component analysis, PCA)和 BP 人工神经网络技术(back propagation artificial neural network, BP-ANN)及偏最小二乘法和 BP 人工神经网络技术 3 种方法对鲢鱼原始光谱数据进行了建模分析, 以期得到具有较高准确性和预测能力的鲢鱼营养成分定量模型。

1 材料与方法

1.1 材料与试剂

鲢鱼, 分别于 2012 年 11~12 月和 2013 年 3~5 月, 从武汉市集贸市场采集到不同批次、不同规格的鲜活鲢鱼样品 202 个, 从荆州水产品养殖基地(循环水养殖和非循环水养殖)采集样品 16 个, 从石首水产品研究所采集样品 36 个, 共 254 个样品, 鱼体质量为 1000~3000 g/尾。

硼酸、甲基红、次甲基蓝, 分析纯, 国药集团化学试剂有限公司; 浓盐酸、浓硫酸, 分析纯, 信阳市化学试剂厂。

1.2 仪器与设备

Supnir-2720 近红外光谱仪(杭州聚光科技股份有限公司), AC2105 型分析天平(德国赛多利斯公司), K600 食物调理机(德国博朗电器公司), 烘箱(上海精宏实验设备有限公司)。

1.3 试验方法

1.3.1 鲢鱼肉样本的制备

将鲜活样品于 10 °C 左右宰杀, 取鱼背脊肉, 绞碎, 分装, 随机编号后置于 4 °C 冰箱中保存备用, 放置时间不超过 1 d。

1.3.2 近红外光谱扫描

将制备好的鱼肉样品装入直径为 70 mm、高度为 20 mm 的圆形黑色样品盒内, 并用样品盒盖压平, 排除其底部气泡。以装满 1 个容器为 1 个试验样品, 进行近红外光谱采集。近红外光谱仪经 30 min 预热并进行白板参比和性能测试后开始测试。卤钨灯置于鱼肉样品的正上方 200 mm 处, 光线经鱼肉样品漫反射至光谱仪探头, 光谱仪对每个样品扫描 3 次, 样品的光谱曲线取 3 次扫描的平均值。光谱采集条件: 扫描波长为 1000~1799 nm; 仪器带宽为 1 nm; 扫描间隔为 1 nm; 光谱重复性优于 0.2 nm; 信噪比优于 2000:1; 工作温度范围为 15~20 °C(通过空调调节控制); 测量时间为 3 s。

1.3.3 粗蛋白的测定

取制备好的鲢鱼鱼肉样品 2~5 g, 依据 GB 5009.5-2010《食品安全国家标准: 食品中蛋白质的测定》^[11]中的凯氏定氮法测定($F = 6.25$)。取其平均值($n=3$)。

1.3.4 粗脂肪的测定

取制备好的鲢鱼鱼肉样品 10 g, 依据 GB 5009.6-2010《食品安全国家标准: 食品中脂肪的测定》^[12]中的索氏抽提法测定。取其平均值($n=3$)。

1.3.5 水分的测定

取制备好的鲢鱼鱼肉样品 2~5 g, 依据 GB 5009.3-2010《食品安全国家标准: 食品中水分的测定》^[13]中的直接干燥法(101~105 °C)测定。取其平均

值($n=3$)。

1.3.6 灰分的测定

取制备好的鲢鱼鱼肉样品 3~10 g, 依据 GB 5009.4-2010《食品安全国家标准: 食品中灰分的测定》^[14]中的方法测定。取其平均值($n=3$)。

1.3.7 建模方法与异常样本剔除

在建立模型时, 为了保持样本的多样性, 确定模型的预处理方法, 首先对所有样本进行了预建模分析, 然后再剔除样本, 进行正式建模分析, 将预建模的结果与剔除异常样本后的模型结果进行比较, 得出较优的定量模型。

1) 鲢鱼营养成分模型的建模方法

根据 254 个鱼肉样品的粗蛋白、粗脂肪、水分和灰分含量和近红外光谱数据, 选用不同方法建立鲢鱼鱼肉样品的营养成分模型。对模型进行优化, 寻找最佳光谱预处理方法和最佳波段范围。通过定标相关系数 (coefficient of determination of calibration, R_C)、验证相关系数 (correlation coefficient of validation, R_P)、定标标准偏差 (standard error of calibration, SEC)、验证标准偏差 (bias-corrected standard error of prediction, SEP)、交互验证标准误差 (standard error of cross validation, SECV) 等参数对模型进行内部验证, 最后通过外部验证考察模型的准确性和适应性^[15]。

对 254 个数据进行 K-S 分组, 80% 样本用于建立 NIRS 模型, 称为定标集, 20% 样本用于检验所建 NIRS 模型的准确性和可靠性, 称为验证集。根据鱼肉样品的近红外光谱特性, 选取 51 个样品为验证集, 其余为定标集。

2) 异常样本的剔除

在建立模型过程中, 发现异常样本时, 需要对其进行分析。剔除异常样本一般要遵循预测浓度残差、重构光谱残差、光谱 PLS 分解主成分得分的聚类分析以及杠杆值与学生化残差等标准^[16]。

1.4 光谱数据处理与分析

光谱数据采用近红外光谱仪自带的 RIMP 化学计量学软件进行处理。将鲢鱼营养成分模型分别采用偏最小二乘法、主成分分析和 BP 人工神经网络技术、偏最小二乘法和 BP 人工神经网络技术 3 种方法建立。

2 结果与分析

2.1 样品的光谱采集与化学分析结果

由表 1 可知, 254 个鲢鱼样品的粗蛋白含量为 12.05%~19.52%, 平均为(16.56±1.41)%, 粗脂肪含量为 0.24%~5.27%, 平均为(1.61±0.78)%, 水分含量为 72.62%~80.58%, 平均为(76.97±1.36)%, 而灰分含量为 0.46%~1.50%, 平均为(1.15±0.16)%, 数据范围较大, 可满足近红外光谱技术定量建模的要求。

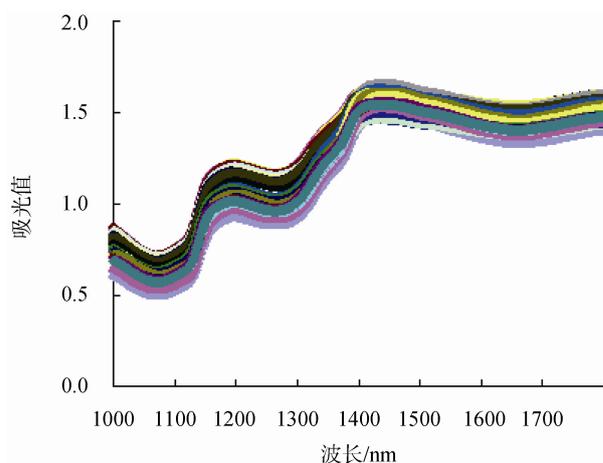


图 1 254 个鲢鱼样品的近红外光谱图

Fig. 1 NIR spectra of 254 silver carp samples

表 1 鲢鱼肉样本营养成分的化学测量值
Table 1 The chemical measured values of nutrient of silver carp

指标	样品数	最大值/%	最小值/%	平均值/%	方差
粗蛋白	254	19.52	12.05	16.56	1.41
粗脂肪	254	5.27	0.24	1.61	0.78
水分	254	80.58	72.62	76.97	1.36
灰分	254	1.50	0.46	1.15	0.16

2.2 鲢鱼营养成分模型的建立

为了建立最优的鲢鱼营养成分模型, 本文采用 Savitzky-Golay 平滑、多元散射校正 (multiplicative scatter correction, MSC)、数据标准化 (standardization, S) 等 20 种方法对光谱预处理, 并分别用偏最小二乘法、主成分分析和 BP 人工神经网络技术、偏最小二乘法和 BP 人工神经网络技术建立 3 种统计模型, 并分析对比, 以选择出预测能力最强的鲢鱼营养成分模型。

利用 RIMP 化学计量学软件, 计算出鱼肉样本粗蛋白、粗脂肪、水分和灰分的马氏阈值和学生残差, 将异常样本剔除, 蛋白质模型剔除 7 个, 脂肪模型剔除 6 个, 水分模型剔除 5 个, 灰分模型剔除 7 个。其中, 四种营养成分剔除的样本有重叠, 共剔除异常样本 11 个。由此而建立的鲢鱼营养成分模型比剔除异常样本之前的模型准确性有一定的提高。

2.2.1 偏最小二乘回归模型

偏最小二乘法是 NIRS 定量分析中应用最多的多元校正方法^[17]。偏最小二乘法不仅可以对光谱矩阵进行分解, 同时也可以对浓度矩阵进行分解^[18-19], 因此偏最小二乘法可以过滤掉原始数据矩阵中的无用信息, 以保证保留的主成分一定与组分浓度相关, 此

外, 偏最小二乘法适用于多变量模型的建立, 可有效消除噪声、减少误差、提高模型稳定性和模型的预测精度^[20]。

分别比较 20 种光谱预处理方法, 得到偏最小二乘模型的最佳光谱预处理方法。以交互验证标准误差 SECV、定标标准偏差 SEC、验证标准偏差 SEP、定标相关系数 R_C 和验证相关系数 R_P 等为评价指标对模型的最优波段进行选择, SECV、SEC 和 SEP 越小, R_C 和 R_P 越高表明模型精确度越高, 预测性能越好。采用留一交互验证法, 根据交互验证的 PRESS 值来确定模型的最佳主因子数, PRESS 值越小, 说明模型的预测能力越强。采用偏最小二乘法建立模型, 20 种光谱预处理方法下粗蛋白模型的统计参数见表 2。

表 2 20 种光谱预处理条件下偏最小二乘法建立的粗蛋白模型的统计参数
Table 2 Statistical parameters of crude protein model under 20 different spectral pretreatments by PLS

预处理方法	定标标准偏差 SEC	定标相关 系数 R_C	交互验证标准 误差 SECV	验证标准 偏差 SEP	验证相关 系数 R_P	主因子 PF
Savitzky-Golay 平滑、Savitzky-Golay 导数、多元散射校正(MSC)、均值中心化	0.0067	0.8839	0.0072	0.0075	0.8493	6
Savitzky-Golay 平滑、Savitzky-Golay 导数、标准正态变量变换(SNV)、均值中心化	0.0059	0.9101	0.0072	0.0068	0.8842	9
Savitzky-Golay 平滑、Savitzky-Golay 导数、去趋势校正(DT)、均值中心化	0.0058	0.9150	0.0069	0.0070	0.8797	9
Savitzky-Golay 平滑、差分求导、多元散射校正(MSC)、均值中心化	0.0064	0.8946	0.0070	0.0074	0.8559	6
不进行预处理(N)	0.0012	0.9969	0.0057	0.0046	0.9459	20
数据标准化(S)	0.0012	0.9965	0.0054	0.0048	0.9401	20
多元散射校正(MSC)、Savitzky-Golay 平滑	0.0045	0.9500	0.0070	0.0066	0.9008	14
Savitzky-Golay 平滑	0.0022	0.9891	0.0065	0.0057	0.9223	20
多元散射校正(MSC)、标准化(S)	0.0025	0.9852	0.0073	0.0071	0.8795	15
多元散射校正(MSC)、Savitzky-Golay 导数	0.0044	0.9513	0.0093	0.0072	0.8519	7
多元散射校正(MSC)	0.0012	0.9965	0.0066	0.0066	0.9051	20
净分析信号(NAS)	0.0011	0.9969	0.0049	0.0046	0.9458	14
正交信号校正(OSC)	0.0008	0.9984	0.0051	0.0047	0.9434	16
去趋势校正(DT)	0.0013	0.9962	0.0058	0.0053	0.9320	17
Savitzky-Golay 导数	0.0038	0.9640	0.0082	0.0062	0.8878	8
基线校正、标准正态变量变换(SNV)	0.0014	0.9955	0.0065	0.0065	0.9068	19
多元散射校正(MSC)、基线校正	0.0016	0.9941	0.0065	0.0065	0.9057	18
基线校正、Savitzky-Golay 导数	0.0038	0.9640	0.0082	0.0062	0.8878	8
Savitzky-Golay 平滑、Savitzky-Golay 导数、基线校正	0.0042	0.9553	0.0081	0.0068	0.8709	9
Savitzky-Golay 平滑、Savitzky-Golay 导数	0.0056	0.9218	0.0073	0.0053	0.9239	10

由表 2 可知, 原始光谱数据经正交信号校正处理后的模型定标集相关系数最高, 为 0.9984, 验证集相关系数为 0.9434, 而原始光谱数据经净分析信号处理后的定标集相关系数为 0.9969, 验证集相关系数为 0.9458, 且交互验证标准误差和验证标准差均低于正交信号校正处理后的模型。所以经正交信号校正处理后的粗蛋白模型还有待于进一步研究, 偏最小二乘法建立的粗蛋白模型的最佳光谱预处理方法为净分析信号。在最优的光谱预处理方法下, 不同波段的模型统计参数见表 3。

由表 3 可知, 波段范围为 1000~1350、1450~1799 nm 时, 定标集相关系数最高, 为 0.9986, 但验证集相

关系数较低, 为 0.9331, 且交互验证标准误差 SECV 较高, 可能是出现了过拟合的现象。而波段范围为 1000~1799 nm 时, 模型有较高的验证相关系数、较低的验证标准偏差和交互验证标准误差。因此, 偏最小二乘法建立的粗蛋白模型最佳光谱波段范围为 1000~1799 nm, 最佳主因子数为 14。在最优条件下, 模型定标相关系数为 0.9969, 标准差为 0.0011, 验证相关系数为 0.9458, 标准差为 0.0046。

采用同样的方法得到偏最小二乘法建立的粗脂肪、水分和灰分模型的最佳光谱数据预处理方法、最佳光谱波段和最佳主因子数, 结果见表 4。所建最优模型的内部验证结果见表 5。

表 3 不同光谱波段条件下偏最小二乘法建立的粗蛋白模型的统计参数

Table 3 Statistical parameters of crude protein model under different spectral bands by PLS

波段/nm	定标准偏差 SEC	定标相关系数 R_c	交互验证标准误差 SECV	验证标准偏差 SEP	验证相关系数 R_p	主因子 PF
1000~1799	0.0011	0.9969	0.0049	0.0046	0.9458	14
1000~1350、1450~1799	0.0008	0.9986	0.0061	0.0049	0.9331	17
1000~1300、1700~1799	0.0065	0.9065	0.0083	0.0060	0.9063	7
1160~1350、1500~1799	0.0024	0.9861	0.0062	0.0049	0.9372	10
1000~1200、1300~1450	0.0050	0.9419	0.0074	0.0070	0.8930	9
1000~1200、1300~1650	0.0011	0.9973	0.0050	0.0068	0.8742	16
1000~1450	0.0027	0.9833	0.0067	0.0079	0.8636	12
1000~1650	0.0011	0.9972	0.0050	0.0067	0.8779	16
1000~1700	0.0013	0.9960	0.0052	0.0063	0.8924	14
1000~1760	0.0008	0.9987	0.0050	0.0052	0.9276	17

表 4 鲢鱼营养成分偏最小二乘模型的最佳条件

Table 4 The optimum parameters of PLS models of silver carp

组分	波段范围/nm	预处理方法	最佳主因子数
粗蛋白	1000~1799	净分析信号(NAS)	14
粗脂肪	1000~1799	正交信号校正(OSC)	10
水分	1000~1799	数据标准化(S)	16
灰分	1000~1799	正交信号校正(OSC)	14

表 5 鲢鱼营养成分偏最小二乘模型的内部验证结果

Table 5 Internal validation results of PLS models of silver carp

组分	定标准偏差 SEC	定标相关系数 R_c	交互验证标准误差 SECV	验证标准偏差 SEP	验证相关系数 R_p
粗蛋白	0.0011	0.9969	0.0049	0.0046	0.9458
粗脂肪	0.0010	0.9925	0.0022	0.0043	0.7579
水分	0.0025	0.9831	0.0058	0.0072	0.8045
灰分	0.0001	0.9976	0.0005	0.0006	0.9175

由表 4 和表 5 可知, 采用偏最小二乘法建立的近红外定量模型, 模型的最佳光谱波段均为 1000~1799 nm, 粗蛋白模型的最佳光谱预处理方法为净分析信号, 最佳主因子数为 14, 定标相关系数为 0.9969; 粗脂肪模型的最佳光谱预处理方法为正交信号校正, 最佳主因子数为 10, 定标相关系数为 0.9925; 水分模型的最佳光谱预处理方法为数据标准化, 最佳主因子数为 16, 定标相关系数为 0.9831, 灰分模型的最佳光谱预处理方法为正交信号校正, 最佳主因子数为 14, 定标相关系数为 0.9976, 说明粗蛋白、粗脂肪、水分和灰分数学模型的预测值和真实值之间具有较高的相关性。

2.2.2 主成分分析和 BP 人工神经网络模型

主成分分析是将多项指标转化为少数几项综合指标, 用综合指标来解释多变量的方差-协方差结构, 它是利用降维技术用少数几个综合变量来代替原始多个变量, 这些变量尽量的保留了原始变量的信息, 并且变量间不相关^[21]。人工神经网络是一种大规模并行的非线性动力系统^[22], 能够实现输入与输出之

间的高度非线性映射, 在信号和模式识别中的应用非常广泛^[23]。

人工神经网络的参数设置为: 粗蛋白模型的输入层节点数为 14, 粗脂肪模型的输入层节点数为 10, 水分模型的输入层节点数为 16, 灰分模型的输入层节点数为 14, 4 个模型的隐含层的节点数均为 10, 输入层到隐含层的初始权重、隐含层到输出层的初始权重为 -0.5~0.5 的随机数, 隐含层的转化函数为线形函数, 输出层的转化函数为对数函数, 初始学习速率为 0.1, 动量项为 0.9。采用主成分分析和 BP 人工神经网络技术建立模型, 20 种光谱预处理方法下粗蛋白模型的统计参数见表 6。

由表 6 可知, 原始光谱数据经标准化处理后的模型定标集相关系数和验证集相关系数均最高, 为 0.8307 和 0.8677, 且定标标准偏差和验证标准偏差均较低。因此, 主成分分析和 BP 人工神经网络建立的粗蛋白模型的最佳光谱预处理方法为标准化。在最优的光谱预处理方法下, 不同波段的模型统计参数见表 7。

表 6 20 种光谱预处理条件下主成分分析和人工神经网络建立的粗蛋白模型的统计参数
Table 6 Statistical parameters of crude protein model under 20 different spectral pretreatments by PCA-ANN

预处理方法	定标标准 偏差 SEC	定标相关 系数 R_c	验证标准 偏差 SEP	验证相关 系数 R_p
Savitzky-Golay 平滑、Savitzky-Golay 导数、多元散射校正(MSC)、均值中心化	0.0088	0.8099	0.0069	0.8465
Savitzky-Golay 平滑、Savitzky-Golay 导数、标准正态变量变换(SNV)、均值中心化	0.0086	0.8178	0.0072	0.8395
Savitzky-Golay 平滑、Savitzky-Golay 导数、去趋势校正(DT)、均值中心化	0.0083	0.8288	0.0070	0.8495
Savitzky-Golay 平滑、差分求导、多元散射校正(MSC)、均值中心化	0.0084	0.8267	0.0071	0.8427
不进行预处理(N)	0.0093	0.7788	0.0081	0.7837
数据标准化(S)	0.0083	0.8307	0.0065	0.8677
多元散射校正(MSC)、Savitzky-Golay 平滑	0.0087	0.8089	0.0074	0.8318
Savitzky-Golay 平滑	0.0093	0.7778	0.0081	0.7835
多元散射校正(MSC)、标准化(S)	0.0097	0.7779	0.0089	0.7763
多元散射校正(MSC)、Savitzky-Golay 导数	0.0130	0.5237	0.0091	0.7634
多元散射校正(MSC)	0.0086	0.8129	0.0074	0.8342
净分析信号(NAS)	0.0127	0.5426	0.0100	0.6399
正交信号校正(OSC)	0.0104	0.7107	0.0084	0.7601
去趋势校正(DT)	0.0086	0.8182	0.0072	0.8367
Savitzky-Golay 导数	0.0129	0.5648	0.0096	0.7389
基线校正、标准正态变量变换(SNV)	0.0086	0.8126	0.0074	0.8338
多元散射校正(MSC)、基线校正	0.0089	0.8029	0.0072	0.8415
基线校正、Savitzky-Golay 导数	0.0129	0.5648	0.0096	0.7389
Savitzky-Golay 平滑、Savitzky-Golay 导数、基线校正	0.0083	0.8301	0.0074	0.8368
Savitzky-Golay 平滑、Savitzky-Golay 导数	0.0082	0.8339	0.0073	0.8411

表 7 不同光谱波段条件下主成分分析和神经网络建立的粗蛋白模型的统计参数
Table 7 Statistical parameters of crude protein model under different spectral bands by PCA-ANN

波段/nm	定标标准偏差 SEC	定标相关系数 R_C	验证标准偏差 SEP	验证相关系数 R_P
1000~1799	0.0083	0.8307	0.0065	0.8677
1000~1350、1450~1799	0.0083	0.8278	0.0065	0.8648
1000~1300、1700~1799	0.0087	0.8097	0.0078	0.8046
1160~1350、1500~1799	0.0087	0.8097	0.0078	0.8027
1000~1200、1300~1450	0.0085	0.8190	0.0067	0.8635
1000~1200、1300~1650	0.0084	0.8248	0.0069	0.8506
1000~1450	0.0083	0.8289	0.0068	0.8605
1000~1650	0.0082	0.8427	0.0069	0.8653
1000~1700	0.0083	0.8488	0.0063	0.8777
1000~1760	0.0081	0.8533	0.0069	0.8781

由表 7 可知, 波段范围为 1000~1760 nm 时, 模型具有最高的定标集和验证集相关系数, 为 0.8533 和 0.8781, 且定标标准偏差较低。因此, 主成分分析和神经网络建立的粗蛋白模型的最佳光谱波段范围为 1000~1760 nm, 在最优条件下, 模型定标相关系数为 0.8533, 标准差为 0.0081, 验证相关系数为 0.8781, 标准差为 0.0069。

采用同样的方法得到主成分分析和神经网络建立的粗脂肪、水分和灰分模型的最佳光谱预处理方法和最佳光谱波段结果见表 8。所建最优模型的内部验证结果见表 9。

由表 8 和表 9 可知, 采用主成分分析和 BP 神经网络技术建立的近红外定量模型, 模型的最佳光谱预处理方法均为数据标准化, 粗蛋白模型的最佳波段范围为 1000~1760 nm, 定标相关系数为 0.8533; 粗脂肪模型的最佳波段范围为 1000~1799 nm, 定标相关系数为 0.8732; 水分模型的最佳波段范围为 1000~1760 nm, 定标相关系数为 0.8520, 灰

分模型的最佳波段范围为 1000~1350、1450~1799 nm, 定标相关系数为 0.8848。

2.2.3 偏最小二乘和 BP 神经网络模型

采用偏最小二乘和 BP 神经网络技术建立鲢鱼营养成分模型, 比较 20 种光谱预处理方法, 得到偏最小二乘和 BP 神经网络模型的最佳光谱预处理方法, 结果见表 10。

表 8 鲢鱼营养成分主成分分析和神经网络模型的最佳条件

Table 8 The optimum parameters of PCA-ANN models of silver carp

组分	波段范围/nm	预处理方法
粗蛋白	1000~1760	数据标准化(S)
粗脂肪	1000~1799	数据标准化(S)
水分	1000~1760	数据标准化(S)
灰分	1000~1350、1450~1799	数据标准化(S)

表 9 鲢鱼营养成分主成分分析和神经网络模型的内部验证结果
Table 9 Internal validation results of PCA-ANN models of silver carp

组分	定标标准偏差 SEC	定标相关系数 R_C	验证标准偏差 SEP	验证相关系数 R_P
粗蛋白	0.0081	0.8533	0.0069	0.8781
粗脂肪	0.0044	0.8732	0.0056	0.6693
水分	0.0076	0.8520	0.0076	0.7789
灰分	0.0008	0.8848	0.0008	0.8658

表 10 20 种光谱预处理条件下偏最小二乘和人工神经网络建立的粗蛋白模型的统计参数
Table 10 Statistical parameters of crude protein model under 20 different spectral pretreatments by PLS-ANN

预处理方法	定标标准 偏差 SEC	定标相关 系数 R_C	验证标准 偏差 SEP	验证相关 系数 R_P
Savitzky-Golay 平滑、Savitzky-Golay 导数、多元散射校正(MSC)、均值中心化	0.0066	0.8992	0.0065	0.8740
Savitzky-Golay 平滑、Savitzky-Golay 导数、标准正态变量变换(SNV)、均值中心化	0.0066	0.8979	0.0063	0.8831
Savitzky-Golay 平滑、Savitzky-Golay 导数、去趋势校正(DT)、均值中心化	0.0065	0.9027	0.0064	0.8815
Savitzky-Golay 平滑、差分求导、多元散射校正(MSC)、均值中心化	0.0063	0.9140	0.0070	0.8519
不进行预处理(N)	0.0076	0.8856	0.0072	0.8990
数据标准化(S)	0.0046	0.9500	0.0063	0.9188
多元散射校正(MSC)、Savitzky-Golay 平滑	0.0066	0.9059	0.0070	0.8517
Savitzky-Golay 平滑	0.0076	0.8758	0.0071	0.8925
多元散射校正(MSC)、标准化(S)	0.0052	0.9430	0.0063	0.8772
多元散射校正(MSC)、Savitzky-Golay 导数	0.0051	0.9528	0.0073	0.8356
多元散射校正(MSC)	0.0049	0.9451	0.0066	0.8793
净分析信号(NAS)	0.0145	0.5077	0.0140	0.5944
正交信号校正(OSC)	0.0087	0.8106	0.0070	0.8468
去趋势校正(DT)	0.0071	0.8796	0.0074	0.8321
Savitzky-Golay 导数	0.0085	0.8231	0.0075	0.8358
基线校正、标准正态变量变换(SNV)	0.0049	0.9453	0.0066	0.8799
多元散射校正(MSC)、基线校正	0.0040	0.9634	0.0065	0.9022
基线校正、Savitzky-Golay 导数	0.0085	0.8231	0.0075	0.8358
Savitzky-Golay 平滑、Savitzky-Golay 导数、基线校正	0.0101	0.8095	0.0100	0.8298
Savitzky-Golay 平滑、Savitzky-Golay 导数	0.0087	0.8112	0.0079	0.8495

由表 10 可知, 原始光谱数据经多元散射校正和基线校正处理后的模型, 定标集和验证集相关系数均较高, 为 0.9634 和 0.9022, 且定标标准偏差最低, 验证标准偏差相对较低, 因此偏最小二乘法和 BP 人工神经网络建立的粗蛋白模型的最佳光谱预处理方法为多元散射校正和基线校正。在最优的光谱预处理方法下, 不同波段的模型统计参数见表 11。

由表 11 可知, 波段范围为 1000~1760 nm 时, 定标集相关系数最高, 为 0.9641, 但验证集相关系数为 0.8926, 可能是出现了过拟合的现象, 而波段范围为 1000~1799 nm 时, 定标集和验证集相关系数均较高, 定标标准偏差和验证标准偏差均较低。因此, 偏最小二乘和人工神经网络建立的粗蛋白模型的最佳光谱波段范围为 1000~1799 nm, 在最优条件下, 模型定

标相关系数为 0.9634, 标准差为 0.0040, 验证相关系数为 0.9022, 标准差为 0.0065。

采用同样的方法得到偏最小二乘和人工神经网络建立的粗脂肪、水分和灰分模型的最佳光谱预处理方法和最佳光谱波段结果见表 12。所建最优模型的内部验证结果见表 13。

由表 12 和表 13 可知, 采用偏最小二乘和 BP 人工神经网络技术建立的模型, 粗蛋白模型的最佳波段范围为 1000~1799 nm, 最佳光谱预处理方法为多元散射校正和基线校正, 定标相关系数为 0.9634; 粗脂肪模型的最佳波段范围为 1000~1799 nm, 最佳光谱预处理方法为多元散射校正和 Savitzky-Golay 导数, 定标相关系数为 0.9784; 水分模型的最佳波段范围为 1000~1760 nm, 最佳光谱预处理方法为数据标准

表 11 不同光谱波段条件下偏最小二乘和人工神经网络建立的粗蛋白模型的统计参数
Table 11 Statistical parameters of crude protein model under different spectral bands by PLS-ANN

波段/nm	定标准偏差 SEC	定标相关系数 R_C	验证标准偏差 SEP	验证相关系数 R_P
1000~1799	0.0040	0.9634	0.0065	0.9022
1000~1350、1450~1799	0.0052	0.9373	0.0061	0.8903
1000~1300、1700~1799	0.0061	0.9340	0.0066	0.8618
1160~1350、1500~1799	0.0049	0.9518	0.0065	0.8720
1000~1200、1300~1450	0.0071	0.9078	0.0070	0.8464
1000~1200、1300~1650	0.0052	0.9418	0.0071	0.8453
1000~1450	0.0063	0.9245	0.0068	0.8585
1000~1650	0.0034	0.9736	0.0072	0.8663
1000~1700	0.0052	0.9381	0.0072	0.8508
1000~1760	0.0040	0.9641	0.0067	0.8929

表 12 鲢鱼营养成分偏最小二乘和人工神经网络模型的最佳条件
Table 12 The optimum parameters of PLS-ANN models of silver carp

组分	波段范围/nm	预处理方法
粗蛋白	1000~1799	多元散射校正(MSC)、基线校正
粗脂肪	1000~1799	多元散射校正(MSC)、Savitzky-Golay 导数
水分	1000~1760	数据标准化 (S)
灰分	1000~1799	基线校正、标准正态变量变换(SNV)

表 13 鲢鱼营养成分偏最小二乘和人工神经网络模型的内部验证结果
Table 13 Internal validation results of PLS-ANN models of silver carp

组分	定标准偏差 SEC	定标相关系数 R_C	验证标准偏差 SEP	验证相关系数 R_P
粗蛋白	0.0040	0.9634	0.0065	0.9022
粗脂肪	0.0018	0.9784	0.0051	0.6547
水分	0.0033	0.9762	0.0070	0.8169
灰分	0.0003	0.9777	0.0007	0.9017

化, 定标相关系数为 0.9762, 灰分模型的最佳波段范围为 1000~1799 nm, 最佳光谱预处理方法为基线校正和标准正态变量变换, 模型的定标相关系数为 0.9777。

2.2.4 最佳鲢鱼营养成分模型的确定

根据 3 种化学计量学方法建立的模型结果可知, 采用偏最小二乘法建立的鲢鱼营养成分最优。所建的粗蛋白模型定标相关系数为 0.9969, 验证相关系数为 0.9458; 粗脂肪模型的定标相关系数为 0.9925, 验证相关系数为 0.7579; 水分模型的定标相关系数为

0.9831, 验证相关系数为 0.8045; 灰分模型的定标相关系数为 0.9976, 验证相关系数为 0.9175, 说明粗蛋白、粗脂肪、水分和灰分数学模型对未知样品有很好的预测能力。拟合出的鲢鱼肉粗蛋白、粗脂肪、水分和灰分模型, 见图 2。其中鱼肉粗蛋白模型预测值和实测值的线性方程为 $Y_1=0.9989X_1 + 0.00017506$, 粗脂肪模型的线性方程为 $Y_2=0.98429X_2 + 0.00025904$, 水分模型的线性方程为 $Y_3=0.96651X_3 + 0.025784$, 灰分模型的线性方程为 $Y_4=0.99633X_4 + 0.000038918$ 。

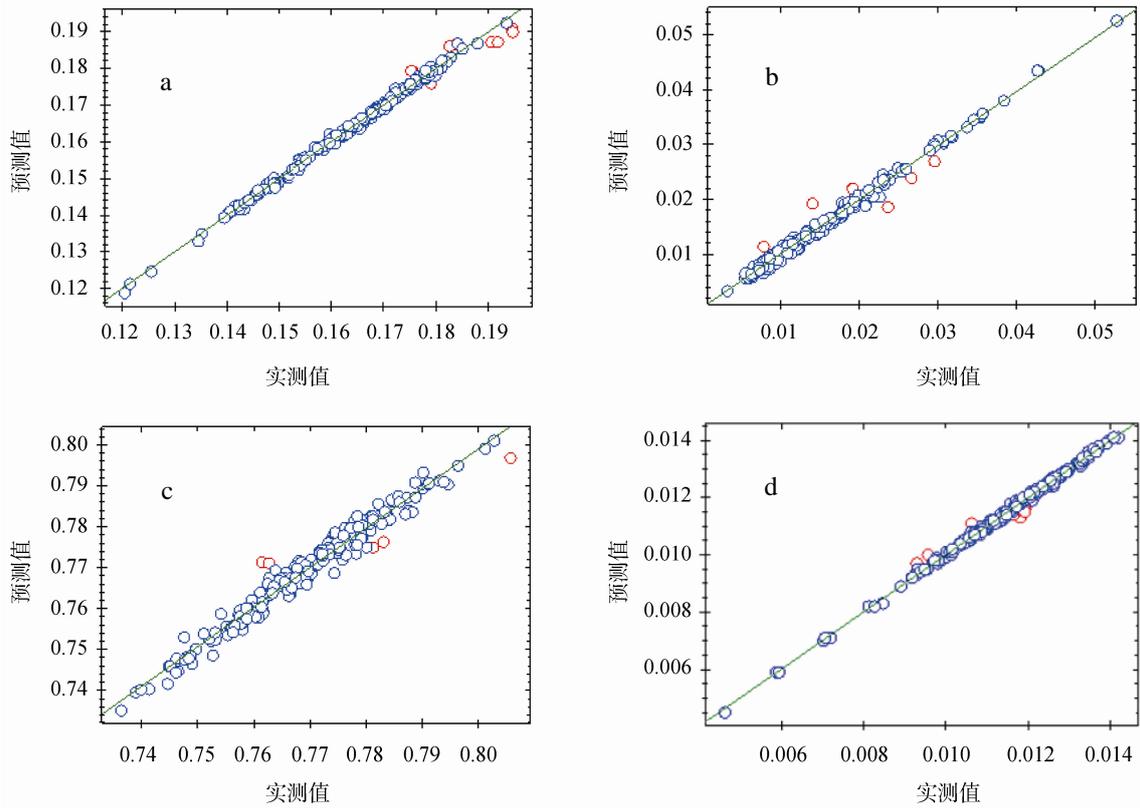


图 2 基于近红外光谱参数的鲢鱼营养成分预测模型

Fig. 2 Nutrient prediction models of silver carp based on the near infrared spectral parameters

a. 粗蛋白模型; b. 粗脂肪模型; c. 水分模型; d. 灰分模型

a. crude protein model; b. crude fat model; c. moisture model; d. ash model

2.3 鲢鱼营养成分模型的检验

随机选取一些模型之外的样品, 对采用偏最小二乘法建立的鲢鱼营养成分模型进行外部验证。采集验证集中 51 个样品的近红外光谱, 用所建立的营养成分模型分别对各样品的粗蛋白、粗脂肪、水分和灰分进行计算得到预测值, 结果见表 14, 并对其进行 t 检验, 4 者的 T 值分别为 0.5886、0.9197、0.7731 和 0.7500, 均小于 $f_{(0.05,50)}=1.9840$, 表明近红外光谱预测值与实测值之间不存在显著差异。由此可见, 原始光谱数据结合偏最小二乘法实现鲢鱼营养成分的快速分析是可行的。

3 结论

采集了 254 个鲢鱼样品的近红外漫反射光谱, 经过净分析信号、正交信号校正、多元散射校正等 20 种方法预处理, 得出了最优的光谱预处理方法,

从而滤掉原始光谱的无用信息, 有效消除噪声、减少误差。

偏最小二乘和 BP 人工神经网络模型更能体现多组分分析优势, 提高抗噪声能力和预测精度, 但在本文分析中采用偏最小二乘法建立的鲢鱼营养成分模型优于偏最小二乘和 BP 人工神经网络建立的模型, 可能是偏最小二乘法主要是线性回归方法, 而 BP 人工神经网络是一种非线性回归方法, 鲢鱼营养成分模型可能更趋近于线性回归的原因。所建立的粗蛋白最优模型定标相关系数为 0.9969, 验证相关系数为 0.9458; 粗脂肪模型的定标相关系数为 0.9925, 验证相关系数为 0.7579; 水分模型的定标相关系数为 0.9831, 验证相关系数为 0.8045; 灰分模型的定标相关系数为 0.9976, 验证相关系数为 0.9175, 说明粗蛋白、粗脂肪、水分和灰分数学模型对未知样品有较好的预测能力。

对所建模型进行外部验证, 结果表明, 采用偏

表 14 预测模型对鲢鱼肉样本营养成分含量的预测结果
Table 14 Predicted results of nutrient models of silver carp

	粗蛋白含量			粗脂肪含量			水分含量			灰分含量		
	真值	预测值	偏差	真值	预测值	偏差	真值	预测值	偏差	真值	预测值	偏差
最大值/%	19.24	19.09	0.59	3.97	2.34	1.77	80.27	79.29	1.69	1.38	1.40	0.19
最小值/%	13.81	13.37	-1.45	0.24	0.61	-0.92	74.78	74.92	-1.98	0.63	0.64	-0.11
平均值/%	16.70	16.84	-0.14	1.53	1.51	0.01	77.10	77.16	-0.06	1.16	1.15	0.01
方差/%	1.31	1.35	0.44	0.66	0.43	0.44	1.23	0.98	0.73	0.15	0.13	0.06

最小二乘法建立的鲢鱼营养成分模型对未知鲢鱼样品的营养成分有很好的预测能力,该方法具有准确、快速、可多组分同时测定等优点,建模后可同时大批量分析样品,在很大程度上简化了测试操作,可以作为鲢鱼营养成分的快速测定方法进行推广应用,具有一定的应用价值。

参考文献

- [1] 赵永峰, 胡海彦, 蒋高中, 等. 我国大宗淡水鱼的发展现状及趋势研究[J]. 中国渔业经济, 2012, 30(5): 91-99.
Zhao YF, Hu HB, Jiang GZ, *et al.* Current status and development trend on national conventional freshwater fishery industry [J]. Chin Fish Econom, 2012, 30(5): 91-99.
- [2] 张国琛, 毛志怀. 水产品干燥技术的研究进展[J]. 农业工程学报, 2004, 20(4): 297-300.
Zhang GC, Mao ZH. Research advances of aquatic product drying technologies [J]. Trans CSAE, 2004, 20(4): 297-300.
- [3] 贾晋, 周迪. 中国城乡居民粮食消费预测与结构优化—基于均衡营养目标的视角[J]. 农业经济与管理, 2013, 17(1): 55-64.
Jia J, Zhou D. Forecast of Grain Consumption of China's Urban and Rural Citizens and Structural Optimization: Based on Balanced Nutrition Angle [J]. Agric Econom Manag, 2013, 17(1): 55-64.
- [4] He Y, Li XL, Deng XF. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model [J]. J Food Eng, 2007, 79(4): 1238-1242.
- [5] Xie LJ, Ying YB, Ying TJ, *et al.* Discrimination of transgenic tomatoes based on visible/near-infrared spectra [J]. Anal Chim Acta, 2007, 584(2): 379-384.
- [6] 王磊, 郭中华, 金灵, 等. 基于近红外光谱技术乳制品品种快速无损鉴别[J]. 激光与红外, 2013, 43(10): 1133-1137.
Wang L, Guo ZH, Jin L, *et al.* Rapid and non-destructive test of dairy varieties based on near infrared spectroscopy [J]. Laser Infrared, 2013, 43(10): 1133-1137.
- [7] Prieto N, Ross DW, Navajas EA, *et al.* On-line application of visible and near infrared reflectance spectroscopy to predict chemical-physical and sensory characteristics of beef quality [J]. Meat Sci, 2009, 83(1): 96-103.
- [8] Prevolnik M, Skrlep M, Janes L, *et al.* Accuracy of near infrared spectroscopy for prediction of chemical composition, salt content and free amino acids in dry-cured ham [J]. Meat Sci, 2011, 88(2): 299-304.
- [9] Prieto N, Dugan M E R, Lopez-Campos O, *et al.* Near infrared reflectance spectroscopy predicts the content of polyunsaturated fatty acids and biohydrogenation products in the subcutaneous fat of beef cows fed flaxseed [J]. Meat Sci, 2012, 90(1): 43-51.
- [10] 徐文杰, 李俊杰, 贾丹, 等. 近红外光谱技术分析草鱼营养成分[J]. 食品科学, 2013, 34(20): 161-164.
Xu WJ, Li JJ, Jia D, *et al.* Nutrient analysis of grass carp (*Ctenopharyngodon idellus*) by near infrared spectroscopy [J]. Food Sci, 2013, 34(20): 161-164.
- [11] GB 5009.5-2010 食品中蛋白质的测定[S].
GB/T 5009.5-2010 Method for analysis of protein [S].
- [12] GB 5009.6-2010 食品中脂肪的测定[S].
GB/T 5009.6-2010 Method for analysis of fat [S].
- [13] GB 5009.3-2010 食品中水分的测定[S].
GB/T 5009.3-2010 Method for analysis of moisture [S].
- [14] GB 5009.4-2010 食品中灰分的测定[S].
GB/T 5009.4-2010 Method for analysis of ash [S].
- [15] 陆焯, 王锡昌, 刘源. 近红外光谱技术在鱼糜定性和定量上的应用[J]. 水产学报, 2011, 35(8): 1273-1279.
Lu Y, Wang XC, Liu Y. Qualitative and quantitative analyses of surimi with near infrared reflectance spectroscopy [J]. J Fish China, 2011, 35(8): 1273-1279.
- [16] 祝诗平. 近红外光谱品质检测方法研究[D]. 北京: 中国农业大学, 2003.
Zhu SP. Study Quality Detection Method on Near-infrared

- Spectroscopy [J]. Beijing: China Agricultural University, 2003.
- [17] 杨曙明, 张瑜. 近红外光谱分析技术及其在饲料质量分析中的应用[J]. 分析测试仪器通讯, 1997, 7(4): 215–220.
Yang SM, Zhang Y. Application of near infrared spectroscopy to quality control in feed products [J]. *Analyt Instrum Newslett*, 1997, 7(4): 215–220.
- [18] Schimann H, Joffre R, Roggy JC, *et al*. Evaluation of the recovery of microbial functions during soil restoration using near-infrared spectroscopy [J]. *App Soil Ecol*, 2007, 37(3): 223–232.
- [19] Tewari JC, Dixit V, Cho BK, *et al*. Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy [J]. *Spectrochim Acta Part A: Mol Biomol Spectrosc*, 2008, 71(3): 1119–1127.
- [20] 王欣, 叶华俊, 黎庆涛. 近红外光谱结合人工神经网络分析蔗汁的锤度和旋光度[J]. 光谱学与光谱分析, 2010, 30(7): 1759–1762.
Wang X, Ye HJ, Li QT. Determination of brix and POL in sugar cane juice by using near infrared spectroscopy coupled with BP-ANN [J]. *Spectroscopy Spectral Analys*, 2010, 30(7): 1759–1762.
- [21] 汤修映, 牛力钊, 徐杨, 等. 基于可见/近红外光谱技术的牛肉含水率无损检测[J]. 农业工程学报, 2013, 29(11): 248–254.
Tang XY, Niu LZ, Xu Y, *et al*. Nondestructive determination of water content in beef using visible/ near-infrared spectroscopy [J]. *Trans CSAE*, 2013, 29(11): 248–254.
- [22] 秦刚, 宋海燕, 陆辉山. 应用可见/近红外光谱技术快速鉴别山西陈醋品种[J]. 山西农业大学学报: 自然科学版, 2010, 30(4): 309–311.
Qin G, Song HY, Lu HS. Discrimination of mature vinegars using vis/ near infrared spectra [J]. *J Shanxi Agric Univ: Nature Science Edition*, 2010, 30(4): 309–311.
- [23] 吴春胤, 张文昭, 欧阳庆, 等. 基于 BP 神经网络模型的荔枝树叶面积测定方法[J]. 农业工程学报, 2007, 23(7): 166–169.
Wu CY, Zhang WZ, Ouyang Q, *et al*. BP neural network model for the measurement of the leaf area of litchi [J]. *Trans CSAE*, 2007, 23(7): 166–169.

(责任编辑: 赵静)

作者简介



徐文杰, 硕士, 主要研究方向为水产品加工。
E-mail: 406083958@qq.com



熊善柏, 教授, 硕士, 主要研究方向为水产品加工及贮藏工程。
E-mail: xiongsb@mail.hzau.edu.cn