

# 基于支持向量机的新鲜与解冻许氏平鲷 (*Sebastes schlegeli*)近红外光谱鉴别技术\*

刘申申<sup>1,2</sup> 孙永<sup>1</sup> 周德庆<sup>1①</sup>

(1. 中国水产科学研究院黄海水产研究所 青岛 266071; 2. 上海海洋大学食品学院 上海 201306)

**摘要** 本研究建立了一种近红外光谱技术,用于鉴别鱼类是否经过解冻处理。首先测定了120个样品的近红外光谱,通过主成分分析对原始光谱数据进行降维处理,再结合支持向量机建模进行分类鉴别。对所有建模样品的主成分1和2按得分值绘制得分图,进行分析聚类,并将前10个主成分的得分值作为支持向量机的输入,优化惩罚参数 $c$ 和核函数参数 $g$ ,对90个样本训练;用未知的30个样本进行判别验证,建立鉴别鲜、冻许氏平鲷的支持向量机分类模型,预测准确率达100%。研究表明,近红外光谱技术结合主成分分析和支持向量机可以作为一种简便、快速、准确的方法用于判断鱼类是否经过解冻处理。

**关键词** 近红外光谱;许氏平鲷;解冻;主成分分析;支持向量机;鉴别

**中图分类号** TS254 **文献标识码** A **文章编号** 2095-9869(2015)06-0134-05

鱼类富含水分和营养物质,容易发生腐败变质,对保鲜保存的要求较高。冷冻作为鱼类等水产品长期贮藏的方法,可大大延长货架期,但易造成品质的下降。冷冻鱼市场价格比新鲜鱼低,一些商贩常常将冷冻鱼解冻后冰藏,冒充新鲜鱼出售,不仅损害了消费者的合法权益,更不利于市场的公平竞争。因此,为了保护消费者的利益,保障水产品贸易的有序进行,亟需建立一种快速鉴别鱼类是否经过冷冻处理的方法。

目前,已建立一些比较成熟的新鲜鱼鉴别方法(Bozzetta *et al.*, 2012; Duflos *et al.*, 2002; Baixas-Nogueras *et al.*, 2007),但这些方法存在费时、成本高、需要专业人员操作等缺点。近红外光谱法(Near-infrared spectroscopy, NIR)因其快速、简便、成本低、无损和多组分同时测定等优点受到人们的重视,已经越来越多地应用于鱼类及其制品的品质评价,包括化学组成测定(徐文杰等, 2013; Khodabux *et al.*, 2007)、微生物快速测定(Lin *et al.*, 2006; Sone *et al.*, 2011)、感官评价(Isaksson *et al.*, 2002)以及鉴别(Uddin *et al.*, 2004、

2005)等方面,而国内在判断鱼类是否经过冷冻处理的文献报道较少。

近红外光谱主要来源于分子振动的倍频与合频,含有大量数据信息,但很难从光谱图中直接得出物质组成信息,必须依靠化学计量学方法(梁逸曾等, 2012)。主成分分析(Principal component analysis, PCA)主要是对数据进行降维处理,将原变量转化为少数综合变量,且这些变量能表征原变量信息,相互正交(梁逸曾等, 2012; 张菊华等, 2011)。支持向量机(Support vector machine, SVM)是Vapnik于1995年新提出的学习算法(Vapnik, 1995),它采用结构风险最小化原理,兼顾训练误差和泛化能力,在解决小样品、非线性、高维数、局部极小值等模式识别问题中表现出多方面的优势(邓乃扬等, 2004; 李国正等, 2004)。

许氏平鲷(*Sebastes schlegeli*),俗称黑头鱼,属鲷形目、鲷科、平鲷属,一般体长20–30 cm,吻较尖,背及两侧灰褐色,为冷温性近海底层鱼类,广泛分布于中国黄渤海和东海,其肉质细嫩、低脂肪、软硬适

\* 中央级公益性科研院所基本科研业务费专项资金(20603022013018)资助。刘申申, E-mail: shen616@126.com

① 通讯作者:周德庆,研究员, E-mail: zhoudeqing@ysfri.ac.cn

收稿日期:2014-12-01, 收修改稿日期:2015-03-10

口, 深受人们的喜爱(张德胜等, 2012)。本研究选取许氏平鲷作为样品, 采用 PCA 和 SVM 相结合的方法建立新鲜与解冻许氏平鲷的近红外光谱鉴别技术。

## 1 材料与方法

### 1.1 实验材料

从青岛某市场购买不同规格的鲜活许氏平鲷 60 尾。样品重量为 269.00–720.20 g/尾, 体长为 21.00–29.90 cm/尾。

### 1.2 主要仪器

傅里叶变换红外光谱仪(Bruker Tensor 27, 德国布鲁克光谱仪器公司), 近红外积分球附件(美国 Pike 公司)。

数据处理软件: OPUS 7.0 光谱分析软件; MATLAB R2014a 软件。

### 1.3 实验方法

**1.3.1 样品制备** 鲜活鱼买回后立即宰杀、去鳞、去内脏、清洗干净。每条鱼沿脊背一分为二, 一侧作为 1 个样品, 共 120 个样品。将其样品分为两组, F 组为新鲜组, T 组为解冻组, 每组 60 个样品。对 F 组鱼去除鱼皮以后沿背部采肉, 用绞肉机混匀绞碎, 装到样品袋中, 尽快用于近红外光谱分析; T 组放到样品袋中, 密封, 储存在  $-20^{\circ}\text{C}$  冰箱中, 10 d 后取出,  $4^{\circ}\text{C}$  冰箱过夜解冻, 之后处理同上, 用于近红外光谱分析。

**1.3.2 光谱采集** 取按照 1.3.1 制备的鱼肉样品装入培养皿中, 轻轻压实以排除培养皿底部的气泡, 且保持培养皿表面平整光洁。培养皿置于积分球光谱采集系统之上。每份样品重复扫描 3 次, 取平均光谱作为该样品的原始光谱。光谱采集条件为扫描波数:  $10000\text{--}4000\text{ cm}^{-1}$ ; 扫描次数: 64; 分辨率:  $8\text{ cm}^{-1}$ ; 测量方式: 漫反射测量; 近红外光谱扫描采集样品光谱时保持稳定的实验室温度和湿度。工作温度范围为  $23\text{--}25^{\circ}\text{C}$ ; 相对湿度范围为 40%–45%。

## 2 结果与分析

### 2.1 近红外光谱图

120 个样品的近红外原始光谱曲线如图 1 所示, F 组样品与 T 组样品的光谱曲线相互交错, 肉眼很难观察到二者的区别。

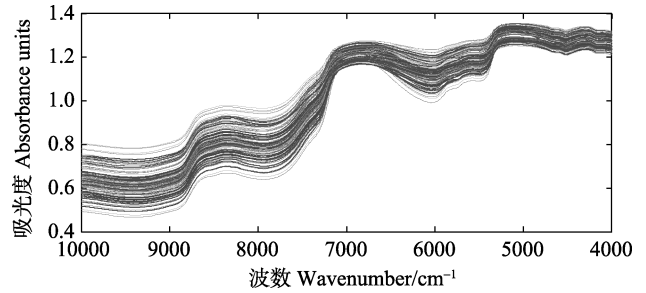


图 1 新鲜与解冻许氏平鲷的原始光谱  
Fig.1 The original NIR spectra of fresh and frozen-thawed *S. schlegeli*

分别对 F 组和 T 组样品的标准近红外光谱图取其平均光谱曲线(图 2)。从图 2 中可以看出, F 组和 T 组鱼肉的光谱有明显区别, 并具有一定的特征性和指纹性, 这一差异是鉴定新鲜和解冻许氏平鲷的基础。F 组样品吸光度大于 T 组, 相差约 0.02–0.06 个单位。肌肉组织间隙的大小导致了光散射的差异, 这是 F 组与 T 组样品能够有效区别的依据(张晓敏, 2012)<sup>1)</sup>。一般认为, 新鲜鱼类的细胞完整, 当光线进入新鲜鱼体内之后, 鱼体细胞不仅吸收光线, 还通过散射作用改变光线传入方向, 使其传到下一个细胞, 增加了光线传输距离, 导致吸光度的增加。而当鱼类经过解冻处理以后, 冰晶的生长致使细胞破裂, 内容物外流, 光散射的细胞数量减少, 导致吸光度的减少(Uddin *et al.*, 2005)。

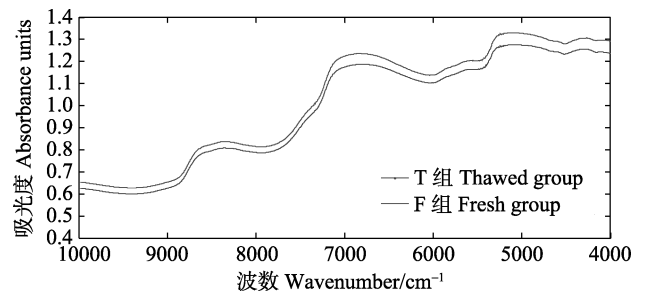


图 2 新鲜与解冻许氏平鲷的平均红外光谱  
Fig.2 The average NIR spectra of fresh and frozen-thawed *S. schlegeli*

### 2.2 主成分分析结果

对 F 组和 T 组, 共 120 个样品进行主成分分析聚类, 前 10 个主成分的特征值即累计可信度如表 1 所示。

由于前两个主成分累计可信度已达到 98.000% 以上, 所以仅用前两个主成分就可以表示原始近红外光谱的信息(戴素贤等, 1999)。图 3 表示 120 个样品的主成分 1、2 得分, 其中 X 轴表示样品的第 1 主

1) 张晓敏. 近红外光谱技术快速评估鲈鱼新鲜度的方法研究. 浙江工商大学硕士研究生学位论文, 2012, 41–62

表 1 主成分与累计可信度

Tab.1 Reliabilities of principal components (%)

主成分 Principal components	累计可信度 Cumulative reliability	主成分 Principal components	累计可信度 Cumulative reliability
PC1	90.770	PC6	99.982
PC2	99.456	PC7	99.990
PC3	99.877	PC8	99.993
PC4	99.928	PC9	99.996
PC5	99.968	PC10	99.997

成分得分(PC1), Y 轴表示样品的第 2 主成分得分(PC2)。图 3 中, F 组和 T 组许氏平鲈明显分为两类, 说明主成分 1、2 对新鲜和解冻许氏平鲈有较好的聚类作用。从图 3 可以看出, F 组的样品大多位于 X 轴上方, T 组样品大多位于 X 轴下方, 且分布较新鲜组比较集中, 但 F 组样品分布分散。为了提高预测精确度, 本研究采用 SVM 结合 PCA 的方法建立新鲜和解冻许氏平鲈的鉴别模型。

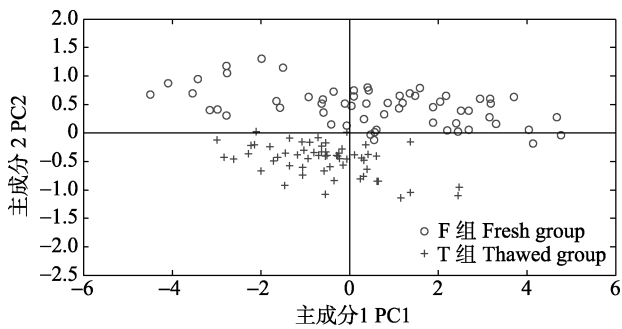


图 3 主成分 1 和 2 的得分

Fig.3 Principal component score plot of fresh and frozen-thawed *S. schlegeli*

2.3 支持向量机分析结果

由于全波段从 10000–4000  $\text{cm}^{-1}$  共有 1556 个数据点, 如果采用全光谱计算, 不仅计算量大, 而且有些区域样品的光谱信息很弱, 与样品的组成或性质间缺乏相关性。由表 1 可知, 前 10 个主成分的累计可信度为 99.997%, 这 10 个综合变量可以代表原变量包含的绝大部分信息。因此, 将 90 个样本(随机分别从 F 组和 T 组中抽取 45 个样品)的前 10 个主成分的得分作为训练集, 将剩余 30 个样本(F 组和 T 组各 15 个样品)的前 10 个主成分的得分作为测试集。另外, 设定 F 组样品标签为 0, T 组样品标签为 1。

SVM 主要是建立一个分类超平面作为决策曲面, 使正例和反例之间的隔离边缘最大化(史峰等, 2010)。本研究选择了 RBF 径向基核函数, 其中, 惩罚参数  $c$  ( $c$  为回归误差的权重)和径向基核参数  $g$  是

两个需要确定的重要参数。因为当  $c$  值太大时, 会使样品数据出现过学习现象, 而  $c$  值太小, 会使样品数据出现欠学习现象(刘星等, 2014)。因此, 常采用交互验证(Cross validation, CV)意义下的网格搜寻(Grid search)、遗传算法(Genetic algorithm, GA)和粒子群优化算法(Particle Swarm Optimization, PSO)参数寻优, 然后利用最佳参数  $c$  和  $g$  进行 SVM 网络训练, 构建模型。

如表 2 所示, 任意给定两组  $c$  和  $g$  值时, 测试集分类准确率分别为 86.67% 和 96.67%, 因此, 随机选择的参数无法保证测试集分类准确率达到最理想的效果。于是, 采用 3 种方法对  $c$  和  $g$  两个参数寻优。在 3 种参数寻优方法中, 用 GA 和 POS 方法选择出的最佳参数  $c$  和  $g$  构建模型时, 所建模型对测试集的预测错 1 个, 即 1 个冷冻样品被误判为新鲜样品, 此时预测集分类准确率为 96.67%; 而采用 CV 方法优化的参数  $c$  和  $g$  构建模型, 该模型的测试集分类准确率最高, 为 100%。由此可见, CV 参数寻优方法可以提高 SVM 模型的稳健性和泛化性能。

利用 CV 方法优化的最佳的参数进行 SVM 网络训练, 建立预测模型, 图 4 为测试集的预测结果。图 4 中样本 1–15 为 F 组样品, 样本 16–30 为 T 组样品, 即样本 1–15 对应的样本标签为 0, 样本 16–30 对应

表 2 SVM 参数优化与模型结果

Tab.2 The kernel function parameters and model results of support vector machine

参数优化方法 Parameter optimization methods	Best $c$	Best $g$	测试集分类准确率 Classification accuracy of test sets /%
任意给定 1 (Random set 1)	10.000	3.000	86.670(26/30)
任意给定 2 (Random set 2)	5.000	2.000	96.670(29/30)
CV	0.707	0.500	100.000(30/30)
GA	9.165	1.162	96.670(29/30)
POS	1.500	1.700	96.670(29/30)

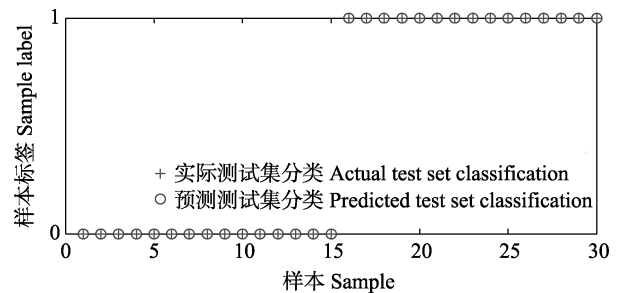


图 4 采用 CV 优化参数的测试集分类结果

Fig.4 Classification result of SVM with CV optimal parameters

的样本标签为 1。由图 4 可以看出, 预测测试集分类与实际预测集分类相同, 即模型对测试集的判断结果均正确, 预测准确率为 100%。由此说明, 用 SVM 方法鉴别新鲜鱼和解冻鱼是可行的。

### 3 讨论

本研究应用 NIR 在 10000–4000  $\text{cm}^{-1}$  波数范围内对 120 个新鲜和解冻许氏平鲷样品采集近红外光谱, 结合 PCA 和 SVM, 建立新鲜和冷冻–解冻许氏平鲷的鉴别模型。该模型的预测效果较好, 预测准确率达到 100%, 说明近红外光谱技术可以简便、快速、准确地鉴别许氏平鲷是否经过解冻处理。

由于新鲜鱼类经过解冻处理, 冰晶的生长使细胞破裂, 内容物外流, 严重影响肉的品质, 导致肌肉蛋白质结构、颜色、脱水、质构、脂肪氧化、pH 的变化。因此, 在以后的研究中, 可以考虑将近红外光谱数据与理化指标数据结合起来, 并且从增加样品数量和分析比较其他判别方法(如 BP 人工神经网络等)角度提高判别的精度和稳定性, 为近红外光谱技术应用用于鉴别新鲜和解冻鱼类奠定基础。

### 参 考 文 献

邓乃扬, 田英杰. 数据挖掘中的新方法—支持向量机. 北京: 科学出版社, 2004, 1–223

史峰, 王小川, 郁磊, 等. MATLAB 神经网络 30 个案例分析. 北京: 北京航空航天大学出版社, 2010, 112–121

刘星, 毛丹卓, 王正武, 等. 薏仁种类的近红外光谱技术快速鉴别. 光谱学与光谱分析, 2014, 34(5): 1259–1263

徐文杰, 李俊杰, 贾丹, 等. 近红外光谱技术分析草鱼营养成分. 食品安全, 2013, 34(20): 161–164

李国正, 王猛, 曾华军, 译. 支持向量机导论. 北京: 电子工业出版社, 2004, 1–82

张菊华, 朱向荣, 李高阳, 等. 近红外光谱法结合化学计量学方法用于茶油真伪鉴别分析. 分析化学, 2011, 39(5): 748–752

张德胜, 陆波, 汪晓慧. 深水抗风浪网箱养殖许氏平鲷技术研究. 水产科技情报, 2012, 39(2): 92–95

梁逸曾, 许青松. 复杂体系仪器分析: 白、灰、黑分析体系及其变量解析方法. 北京: 化学工业出版社, 2012, 480–482

戴素贤, 谢赤军, 陈栋, 等. 七种高香型乌龙茶香气成分的主成分分析. 华南农业大学学报, 1999, 20(1): 113–117

Bozzetta E, Pezzolato M, Cencetti E, *et al.* Histology as a valid and reliable tool to differentiate fresh from frozen-thawed fish. J Food Protect, 2012, 75(8): 1536–1541

Baixas-Nogueras S, Bover-Cid S, Veciana-Nogué MT, *et al.* Effects of previous frozen storage on chemical, microbiological and sensory changes during chilled storage of Mediterranean hake (*Merluccius merluccius*) after thawing. Eur Food Res Technol, 2007, 226(1): 287–293

Duflos G, Le Fur B, Mulak V, *et al.* Comparison of methods of differentiating between fresh and frozen-thawed fish or fillets. J Sci Food Agric, 2002, 82: 1341–1345

Isaksson T, Swensen LP, Taylor RG, *et al.* Non-destructive texture analysis of farmed Atlantic salmon using visual/near-infrared reflectance spectroscopy. J Sci Food Agric, 2002, 82(1): 53–60

Khodabux K, L'Omelette MSS, Jhaumeer-Laulloo S *et al.* Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. Food Chem, 2007, 102(3): 669–675

Lin MS, Mousavi M, Al-Holy M, *et al.* Rapid near infrared spectroscopic method for the detection of spoilage in rainbow trout (*Oncorhynchus mykiss*) fillet. J Food Sci, 2006, 71(1): S18–S23

Sone I, Olsen RL, Dahl R, *et al.* Visible/near-infrared spectroscopy detects autolytic changes during storage of Atlantic salmon (*Salmo salar* L.). J Food Sci, 2011, 76(3): 203–209

Uddin M, Okazaki E. Classification of fresh and frozen-thawed fish by near-infrared spectroscopy. J Food Sci, 2004, 69(8): 665–668

Uddin M, Okazaki E, Torza S, *et al.* Non-destructive visible/NIR spectroscopy for differentiation of fresh and frozen-thawed fish. J Food Sci, 2005, 70(8): 506–510

Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995

## Distinguishing the Fresh from the Frozen-Thawed *Sebastes schlegeli* Using the NIR Spectroscopy Identification Technology Based on Support Vector Machine

LIU Shenshen<sup>1,2</sup>, SUN Yong<sup>1</sup>, ZHOU Deqing<sup>1①</sup>

(1. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071;

2. College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306)

**Abstract** Frozen fish are usually less desirable than the fresh counterparts on the market. As a result, it has become a common issue that the frozen-thawed fish are disguised as fresh for a higher price. In this study the near infrared (NIR) spectroscopy was employed to separate the frozen-thawed fish from the fresh. One hundred and twenty prepared *Sebastes schlegeli* samples including 60 fresh and 60 frozen-thawed were scanned using a near-infrared spectroscopy system between 10000–4000 cm<sup>-1</sup> wavenumbers. The working model was based on the fact that the average NIR spectra of fresh fish were distinctive from the frozen-thawed and possessed certain characteristics and fingerprint resistance. Principal component analysis (PCA) was used for the dimension reduction of the spectra data. The first two principal components (PCs) explained over 98% of variances in all the spectral bands. Clustering was performed and analyzed based on the first two PCs of all samples. The principal component score plot demonstrated that the fresh (above the X axis) and the frozen-thawed samples (below the X axis) were well separated, and that the distribution of fresh samples was dispersal. In order to improve the accuracy of prediction, the support vector machine (SVM) classification model was developed to differentiate the fresh fish from the frozen-thawed, based on principal component analysis scores. The score values of the first ten PCAs were used as the input variables of the SVM, and the penalty parameter *c* and kernel function parameter *g* were optimized. Ninety samples were used for building the SVM model. This model was then applied to predict the rest 30 unknown samples, and the prediction rate was 100%. These results suggested that the near infrared spectroscopy combined with principal component analysis and support vector machine could be used as a rapid, simple and reliable method to identify the fresh and frozen-thawed fish.

**Key words** NIR; *Sebastes schlegeli*; Frozen-thawed; PCA; SVM; Identification

① Corresponding author: ZHOU Deqing, E-mail: zhoudq@ysfri.ac.cn