

蒋雪松,周宏平. 基于遗传算法的农产品品质无损检测研究进展[J]. 江苏农业科学,2013,41(12):17-20.

# 基于遗传算法的农产品品质无损检测研究进展

蒋雪松,周宏平

(南京林业大学机械电子工程学院,江苏南京 210037)

**摘要:**介绍了遗传算法(GA)最优化理论的基本原理和方法。论述了遗传算法的具体构造步骤,研究了基于遗传算法用于农产品质量评判的模型建立与实现,并对遗传算法在苹果、番茄、梨子、鱼、牛奶等农产品的外部感官品质、内部物理品质评判及农产品产地溯源鉴别中的具体应用作了分析。结果表明,遗传算法结合模式识别、神经网络等数据处理方法在农产品品质检测中将获得越来越广泛的应用。

**关键词:**遗传算法;农产品;无损检测

**中图分类号:** S11+6 **文献标志码:** A **文章编号:** 1002-1302(2013)12-0017-03

近年来,随着生活水平的不断提高和饮食条件的改善,农产品的品质要求越来越受到人们的重视。农产品的品质包括外部感官品质和内部特性品质,一般通过肉眼或理化分析仪器来判断,结果通常由于主观因素存在相当大的个体差异,难以坚持统一、客观的标准。目前,基于机器视觉、近红外光谱、电子鼻技术等快速检测技术在农产品的品质无损检测中得到了很好的应用。遗传算法(GA)作为一种全局优化搜索算法,模拟了自然界遗传选择和自然淘汰的生物进化模式,由美国密执安大学 Holland 教授首先提出,他的学生对该理论进行了发展<sup>[1]</sup>。GA 数据处理方法结合图像处理技术、近红外光谱技术和电子鼻技术等无损检测技术,可以对检测的原始数据进行优化,从而为开发在线无损的自动检测系统打下良好的基础。

## 1 遗传算法基本原理简介

作为一种搜索算法,GA 的搜索过程是从一组随机产生的初始解(群体)开始的,问题的每个解对应于群体中的一个个体(染色体),其好坏的判断依据是适应度函数。“遗传”就是指这些群体中的个体根据适应度,根据“选择、交叉、变异运算”等算法,依靠后代对上代的反复迭代,进行不断的进化从而得到较好的个体,获取问题的近似最优解的过程。其原理如图 1 所示。GA 的处理过程包含以下几个环节和步骤。

### 1.1 编码

GA 把问题空间的研究对象通过编码机制按一定顺序,由特定的符号排成遗传空间的字符串,而不对研究对象直接讨论。将问题的解从求解空间的模型表示或映射成遗传解空间字符串的形式称为编码。在遗传空间,对 2 个染色体进行交叉可生成新的染色体,再由遗传空间向问题空间进行映射,在问题空间中可生成解个体,这个过程叫做译码(图 2)。对

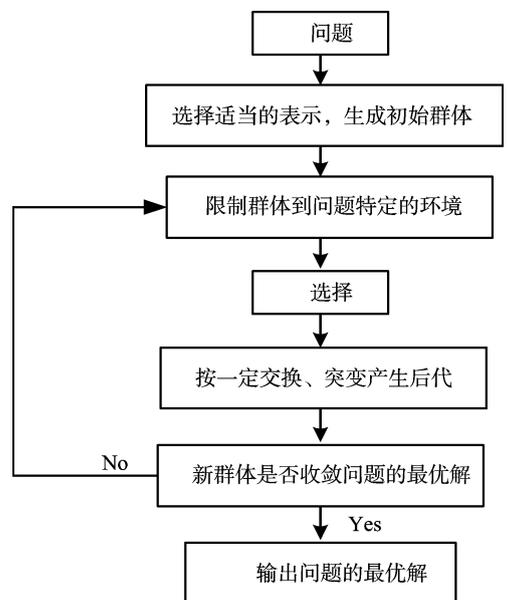


图1 遗传算法的基本原理

编码的基本要求是,编码需尽可能简明,且设计变量与染色体之间需一一对应。GA 中“码”的理解十分广泛。在优化问题上,一个码可以对应于一个可能的解;在分类问题,码可解释

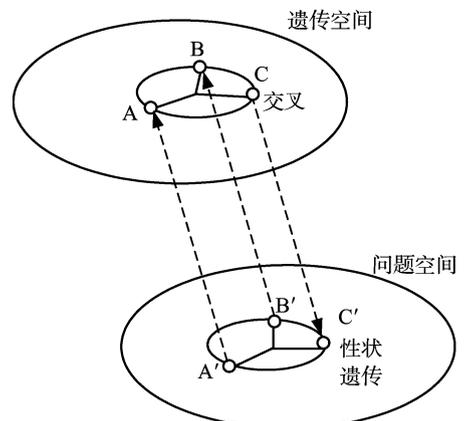


图2 编码与解码的空间映射

收稿日期:2013-10-17

基金项目:江苏省自然科学基金(编号:BK2011823);江苏省高校自然科学基金(编号:10KJB460005)。

作者简介:蒋雪松(1979—),男,江苏南通人,博士,讲师,研究方向为农产品无损检测技术。Tel:(025)85427149;E-mail:xsjiangzju@126.com。

为一个规则,也就是说码的前半部为输入或前件,后半部为输出或后件、结论等。

### 1.2 适应度函数

适应度函数用于评价个体的优劣程度。一般过程为:首先对遗传空间的基因型编码串进行译码,得到问题空间的解个体,然后将解个体值经过该函数转换为该个体的适应度。对于优化问题,可以把问题的模型函数的变种作为适应度函数。

### 1.3 遗传操作

通过编码设定好初始群体后,遗传操作的目的是根据问题域中个体的适应度大小选择个体,通过选择、交叉、变异等遗传操作,演化产生出越来越好的近似解。从优化搜索的

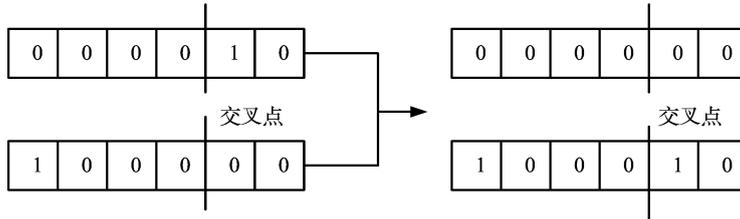


图3 单点交叉示意

### 1.4 控制因素

控制 GA 处理效果的主要参数包括个体编码串长度(串长)、每一代群体的大小(字符串的个数)、交叉率、变异率、终止代数、代沟等。串长的选取与问题所要求的求解精度或决策变量的个数有关。群体大小表示群体中所含个体的数量。交叉率是指染色体交叉的概率,一般设为 0.4 ~ 0.9;变异率指基因座的基因值做变动的概率,一般设为 0.01 ~ 0.1,代沟用于控制每代中种群被替换的比例;终止代数是表示 GA 运行结束条件的一个参数<sup>[3]</sup>。

## 2 遗传算法在农产品品质无损检测中的应用

传统上 GA 在规划设计、优化组合、信号分析、图像处理等领域有较多应用。近些年,伴随着计算机视觉、近红外光谱、电子鼻等技术在农产品品质检测中的应用,GA 显示了强大的数据搜索和优选能力。

### 2.1 农产品外部感官品质检测

农产品的外部感官品质包括外形、大小和表面缺陷等。近年来,计算机视觉及图像处理技术和 GA 相结合被用于对农副产品的检测与分选。龙满生等借助于 GA 对苹果的果形和颜色进行了识别训练,建立了对 2 个 3 层的前馈网络<sup>[4]</sup>,分别用于苹果形状、颜色特征和缺陷的识别和分级,改善了前馈神经网络的学习速度。每 1 个个体对应于 1 个神经网络,采取适应度比例法的选择算子,利用“小生境遗传法”防止基因漂移;采用浮点编码方式进行选择、交叉和变异。结果表明,该系统能够实现对苹果综合外观品质的分级与检测。此外,桃、西红柿等农产品的外部品质也可利用该系统进行评判。

沈明霞等将 GA 和 BP 算法相结合,用于苹果的计算机图像的识别中,将苹果分为 3 等<sup>[5]</sup>。他们用傅里叶描述子提取苹果的形状特征,使用 GA 在解空间中进行搜索定位,从而进行神经网络权值的训练和优化,使网络的实际输出向理想输出逼近。群体规模、交叉概率和变异概率分别设置为 50、0.8

角度而言,遗传操作可实现优胜劣汰的进化机制,逐步实现最优解。“选择”是用来确定如何按适应度的大小,从父代群体中选取一定的个体遗传到下一代群体中,以及这些个体将产生怎样的子代个体,是指淘汰原个体的操作。“交叉”是指从群体中按较大的概率选择 2 个个体,对 2 个个体的某个或某些位进行交换,从而形成 2 个新的个体,它包括交叉点位置的选取和怎样进行部分基因的交换。最简单的单点交叉示意图 3。“变异”是指用一定的等位基因来替换原来个体编码串中的某些基因,从而形成一个新的个体。GA 中的变异运算是一种随机算法,只是产生新个体的辅助算法,它决定了 GA 的局部搜索能力。交叉运算和变异运算的相互配合,共同完成对搜索空间的全局搜索和局部搜索<sup>[2]</sup>。

和 0.01,优化计算限制在 150 代,最佳个体的结果出现在 141 代,结果是收敛的。最后用红富士苹果对得到的网络进行了预测,准确率超过 80%。

番茄的缺素症状可以通过叶片表面的颜色进行反映。毛罕平等利用 GA 从提取的缺素叶片图像中选取 12 个颜色及纹理特征项,排列好顺序后采用二进制的编码方式,染色体长度为 12,每代的群体个数取 20,依据适应度函数选取适当的选择、交叉、变异算子对特征群体进行遗传操作<sup>[6]</sup>。结果验证了 GA 对特征项优化组合的可靠性要好于人工方式对特征项的选择。

为了实现对黄瓜植株中氮、镁元素亏缺的无损快速检测,石吉勇等以 K-最近邻域法(KNN)模型识别率作为 GA 中染色体的适应度值,采用二进制编码对近红外光谱的特征子区间进行了编码、优选,选择出 8 个特征子区间,主成分为 5,相邻个数为 5,最大遗传代为 80,交换率和变异率分别取 0.95、0.05 时建立了最优 GA-KNN 诊断模型<sup>[7]</sup>。结果表明,模型对应的训练集识别率为 98%,预测集识别率为 96%,近红外光谱结合 GA-KNN 算法能够有效地对特征波数点进行选测,实现黄瓜植株氮、镁元素亏缺的快速诊断。

Moscetti 等用 GA 和线性判别式分析(LDA)技术结合近红外光谱开发了板栗虫害的自动检测系统<sup>[8]</sup>。他们在 1 100 ~ 2 300 nm 的光谱范围中用 GA 选取了 2 ~ 6 个特征波长,判断的错误率为假阴性 16.81%,假阳性 0.00%,总的错误率为 8.41%。

### 2.2 农产品内部物理品质检测

农产品的内部品质包括成熟度、营养成分、水分、糖酸度、pH 值、可溶性固形物含量等指标。

为研究苹果等农产品的成熟度,赵杰文等提出一个基于 GA 优化神经网络识别模型的电子鼻系统,对苹果挥发出来的乙烯进行检测<sup>[9]</sup>。在获得的传感器阵列的 5 条响应曲线中分别提取 5 个特征参数作为输入向量,运用 GA 优化的神经网络把这 25 个特征参数所表达的信息融合起来进行模式识

别。GA 中他们采用浮点编码方式,适应度函数为均方根平方的倒数,种群取 50~100,选用稳态轮盘赌选择法,交叉概率和变异概率分别取 0.65~0.9 和 0.001~0.1。结果表明,该遗传神经网络对富士苹果训练集的判断正确率为 100%,对预测集的测试正确率达到 96.4%。

番茄等农产品的成熟度还与表面颜色特征有关,因此,也可以通过它们的表面颜色判断内部品质。张长利等通过对番茄表面颜色特征的分析,提取 HIS 模型的 H 值作为网络的输入,利用 GA 原理将多层前馈神经网络各层之间的权值同时编码为一条染色体,采用二进制编码方式,将神经网络输出值的均方误差产生目标函数,适应度函数取目标函数的倒数。采用单点交叉,群体规模、交叉概率和变异率分别为 50、0.40 和 0.003,终止条件为目标函数小于等于 0.05<sup>[10]</sup>。研究证明利用 GA 对番茄成熟度的自动检测准确率为 94%。

为了对孵化鸡蛋的可成活性进行自动检测,陈佳娟等将遗传神经网络与计算机视觉技术相结合,用色度直方图分析将孵化鸡蛋表面的颜色特征参数作为遗传神经网络的输入,对网络进行二进制编码,适应度函数为网络能量函数的倒数,采用一点交叉和两点交叉,群体规模和代间隙分别为 50、1.0,对多层前馈神经网络的拓扑结构与权值进行了分析,结果表明,对于孵化 3d 以上的鸡蛋,方法准确率较高。鲁棒性和速度较好<sup>[11]</sup>。

烟叶作为一种特殊的农产品,含水量对品质有很大的影响。段史江等利用图像处理技术分析了鲜烟叶及烘烤过程中烟叶图像的颜色特征及纹理特征,以其为输入指标,利用 Matlab 程序确定遗传因子及其他参数,建立了基于 GA 的最小二乘支持向量机(LS-SVM)预测模型和 BP 神经网络模型,用于烟叶含水量的判断<sup>[12]</sup>。结果表明,前者的预测精度更高。

牛奶中脂肪、蛋白质和乳糖等营养成分是牛奶品质的重要因素。李庆波等采用 GA 对近红外光谱建模过程中的波段进行了优选<sup>[13]</sup>。采取二进制编码方式,每个波长对应一个基因,适应度函数取,经过复制、交叉、变异后,测量结果得到显著改善,波长个数从原始谱图的 100 个减少为 16 个,可以满足牛奶成分含量的检测精度要求。结果表明,GA 不仅能够减少波长个数,提高测量速度,节省资源,而且能够有效提高测量精度。

虞佳佳等通过 GA 优化神经网络权阈值,避免了神经网络过程中出现的局部最优的缺点,建立了芒果的近红外光谱糖度、酸度预测模型,提高了预测精度<sup>[14]</sup>。选择 PRESS 值最小作为 GA 的优化标准来优化信息值的权阈值,变异率、群落代分别为 0.9、100;适应度函数为误差倒数的平方。应义斌等使用改进的 GA,在建立近红外光谱的多元校正模型中对波长进行自动选取,用于对梨的糖度、酸度和 pH 值以及西瓜的可溶性固形物进行快速无损检测<sup>[15-16]</sup>。

GA 还被用于鱼类等农产品的新鲜度检测中<sup>[17]</sup>。

### 2.3 农产品的产地溯源检测

农产品的产地对品质、价格和质量有一定的影响。陈永明等利用基于 GA 的近红外光谱技术快速无损地鉴别橄榄油产地<sup>[18]</sup>。他们利用全局搜索-GA 提取了 3 种不同产地的橄榄油的光谱漫反射特征曲线的特征波长,作为主成分分析的输入量,采取的适应度函数的独立变量数为 3,结果表明,GA

抽取特征波长方法正确,能有效压缩光谱数据,该方法能快速无损对橄榄油产地进行分类,同时对其他油类产地的鉴别有一定的借鉴意义。

Tewari 等使用也基于 GA 的人工神经网络和偏最小二乘法结合近红外光谱技术<sup>[19]</sup>,实现了 22 种不同种类的柑橘的产地鉴别和含糖量检测,结果准确、可靠。

### 3 结论及展望

近年来,基于图像处理、近红外光谱和电子鼻技术的农产品品质快速无损检测技术发展迅速。这些新技术和传统的仪器方法不同,一般不能直接得到需要的分析结果,往往需要结合某些数据处理方法来进行建模,然后通过模型对农产品的品质指标进行预测和分类。GA 这种全局搜索方法不需要其他任何先决条件或辅助信息,本质上是一种不依赖具体问题的直接搜索方法。GA 与模式识别、神经网络等结合,可以对模型的输入值进行特征值的筛选,实现模型的优化,从而提高模型的预测精确性和简洁性。随着 GA 的日趋完善和成熟,它在农产品检测中将有巨大的潜力和广阔的应用空间。

### 参考文献:

- [1] Holland J H. Adaptation in natural and artificial systems; an introductory analysis with applications to biology, control, and artificial intelligence[M]. East Lansing: University of Michigan Press, 1975.
- [2] 吉林林. 遗传算法研究综述[J]. 计算机应用与软件, 2004, 21(2): 69-73.
- [3] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999: 87-89.
- [4] 满清生, 何东健, 宁纪锋. 基于遗传神经网络的苹果综合分级系统[J]. 西北农林科技大学学报: 自然科学版, 2001, 29(6): 108-111.
- [5] 沈明霞, 姬长英, 李秀智. 基于遗传 BP 神经网络的苹果形状识别[J]. 粮油加工与食品机械, 2003(12): 64-66.
- [6] 毛罕平, 徐贵力, 李萍萍. 基于遗传算法的蔬菜缺失叶片图像特征选择研究[J]. 江苏大学学报: 自然科学版, 2003, 24(2): 1-5.
- [7] 石吉勇, 邹小波, 赵杰文, 等. 近红外光谱技术快速无损诊断黄瓜植株氮、镁元素亏缺[J]. 农业工程学报, 2011, 27(8): 283-287.
- [8] Moschetti R, Haff R P, Saranwong S, et al. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy[J]. Postharvest Biology and Technology, 2014, 87: 88-94.
- [9] 赵杰文, 邹小波, 潘胤飞, 等. 基于遗传神经网络的苹果气味识别方法研究[J]. 江苏大学学报: 自然科学版, 2004, 25(1): 1-4.
- [10] 张长利, 房俊龙, 潘伟. 用遗传算法训练的多层前馈神经网络对番茄成熟度进行自动检测的研究[J]. 农业工程学报, 2001, 17(3): 153-156.
- [11] 陈佳娟, 陈晓光, 纪寿文. 采用计算机视觉进行孵化鸡蛋成活可能性的自动检测[J]. 计算机应用与软件, 2001, 18(6): 5-10.
- [12] 段史江, 宋朝鹏, 马力, 等. 基于图像处理的烘烤过程中烟叶含水量检测[J]. 西北农林科技大学学报: 自然科学版, 2012, 40(5): 74-80.
- [13] 李庆波, 汪曛, 徐可欣, 等. 牛奶主要成分含量近红外光谱快速测量法[J]. 食品科学, 2002, 23(6): 121-124.
- [14] 虞佳佳, 何勇, 鲍一丹. 基于光谱技术的芒果糖度酸度无损检测方法研究[J]. 光谱学与光谱分析, 2008, 28(12): 2839-2842.

华 春,王仁雷,陈全战,等. 光合作用 C<sub>4</sub> 途径的进化[J]. 江苏农业科学,2013,41(12):20-21.

# 光合作用 C<sub>4</sub> 途径的进化

华 春<sup>1</sup>, 王仁雷<sup>2</sup>, 陈全战<sup>1</sup>, 周 峰<sup>1</sup>

(1. 南京晓庄学院生物化工与环境工程学院, 江苏南京 211171; 2. 江苏教育学院, 江苏南京 210013)

**摘要:**光合作用被称为地球上最重要的化学反应,研究光合作用碳同化途径多样性及其进化具有重要意义。光合生物中至少存在 3 种碳固定途径,介绍了 C<sub>4</sub> 途径的起源路线、起源时间、环境因素,重点介绍了 C<sub>4</sub> 途径的几个进化阶段。

**关键词:**C<sub>4</sub> 途径;进化;Kranz 解剖结构

**中图分类号:** Q945.11 **文献标志码:** A **文章编号:** 1002-1302(2013)12-0020-02

CO<sub>2</sub>、温度、湿度、水分直接影响植物的光合作用,全球气候变化导致植物产生了新的光合作用碳固定途径。新的碳固定途径包括 C<sub>4</sub> 途径、景天酸代谢 (Crassulacean acid metabolism, CAM) 途径,其中,光合作用产量最高的是 C<sub>4</sub> 途径。地球上 25 万种植物中,只有约 7 500 种植物是 C<sub>4</sub> 植物,但它们的初级生产力约占植物总生产力的 23%。在过去的 3 500 万年间,C<sub>4</sub> 途径经过了 66 次独立进化<sup>[1]</sup>。

## 1 C<sub>4</sub> 途径的起源路线

人们很早就知道 C<sub>4</sub> 途径的生理特征是通过众多进化途径积累起来的,近年来分子系统学、碳同位素比研究发现,C<sub>4</sub> 途径的进化具有高度重复性。此外,蛋白质进化的基因组分析也被应用于研究 C<sub>4</sub> 途径的起源。研究发现,禾本科 (Poaceae) 有 22~24 个 C<sub>4</sub> 途径进化分支,共约 4 500 种 C<sub>4</sub> 植物;莎草科 (Cyperaceae) 是第二大 C<sub>4</sub> 植物科目,有 6 个 C<sub>4</sub> 途径进化分支,共约 1 500 种 C<sub>4</sub> 植物;藜科 (Chenopodiaceae) 有 10 个进化分支,共约 500 种 C<sub>4</sub> 植物。这些植物中,约有 1 500 种是真双子叶植物<sup>[2]</sup>。莎草科荸荠属 (*Eleocharis*) 有 2 个独立 C<sub>4</sub> 途径进化分支;真双子叶植物醉蝶花属 (*Cleome*) 有 3 个 C<sub>4</sub> 途径进化分支,其中最典型的 C<sub>4</sub> 模式植物是西洋白花菜 (*C. gynandra*);粟米草属 (*Mollugo*) 有 2 个 C<sub>4</sub> 途径进化分支,可能都起源于同一种植物线叶粟米草 (*Mollugo cervi-*

*ana*)<sup>[3]</sup>。目前发现的所有 C<sub>4</sub> 植物都是被子植物,其中有 26 个单子叶植物进化分支,36 个真双子叶植物进化分支。2011 年,Sage 等又发现 4~6 个 C<sub>4</sub> 植物进化分支<sup>[1]</sup>。

## 2 C<sub>4</sub> 途径的起源时间

化石、同位素研究表明,C<sub>4</sub> 植物生态系统的扩展大约发生在 1 000 万年前。600 万~1 000 万年前中新世晚期 C<sub>4</sub> 植物快速膨胀,这可能与干旱、寒冷、火灾、CO<sub>2</sub> 浓度进一步降低有关。距今 500 万~700 万年前,热带稀树草原在非洲、亚洲、美洲广泛存在<sup>[4]</sup>。C<sub>4</sub> 植物的出现对中新世以来全球环境变化、现代环境格局的形成具有重要意义。研究表明,C<sub>4</sub> 植物的出现可以追溯到 1 200 万~1 500 万年前。C<sub>4</sub> 植物最早可能出现在 2 300 万年前的北美地区。对保存完好的植物化石硅酸体研究发现,具有 C<sub>4</sub> 植物的草原 1 900 万年前出现在北美地区。对花粉化石的同位素研究表明,具有 C<sub>4</sub> 植物的草原 3 300 万年前出现在欧洲西南部地区<sup>[4]</sup>。然而,化石、同位素研究都存在一定局限性。系统发生学分子钟研究表明,最早的 C<sub>4</sub> 植物起源于渐新世中期 (约 3 000 万年前) 的虎尾草亚科 (Chloridoideae),最古老的双子叶 C<sub>4</sub> 植物出现在 2 200 万~2 500 万年间。

## 3 C<sub>4</sub> 途径进化的环境因素

CO<sub>2</sub> 缺乏是 C<sub>4</sub> 途径进化的前提条件,分子钟研究表明,C<sub>4</sub> 植物在渐新世 CO<sub>2</sub> 浓度下降后才出现。促进 C<sub>4</sub> 途径进化的外部动力、环境因素包括高热、干旱、强光、盐渍、生态失调等。通过对 C<sub>3</sub>、C<sub>4</sub> 植物生存环境对比研究发现,强光、高温是促使 C<sub>4</sub> 途径进化的首要因素,高温会直接促进 C<sub>3</sub> 植物的光呼吸、暗呼吸,任何增强光呼吸的环境因子都会诱导 C<sub>4</sub> 途径

收稿日期:2013-05-04

基金项目:国家高技术研究发展计划 (编号:2012AA021701);江苏省自然科学基金青年基金 (编号:BK2012073);江苏省生态学重点学科建设项目。

作者简介:华 春(1963—),女,江苏江阴人,教授,从事植物生理学研究。E-mail:zfibca@163.com。

[15] Ying Y, Liu Y. Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy[J]. Journal of Food Engineering, 2008, 84(2): 206-213.

[16] Jie D, Xie L, Fu X, et al. Variable selection for partial least squares analysis of soluble solids content in watermelon using near-infrared diffuse transmission technique[J]. Journal of Food Engineering, 2013, 118(4): 387-392.

[17] 程 旒, 李小昱, 赵思明, 等. 鱼体新鲜度近红外光谱检测方法的比较研究[J]. 食品安全质量检测学报, 2013, 4(2): 427-

432.

[18] 陈永明, 林 萍, 何 勇. 基于遗传算法的近红外光谱橄榄油产地鉴别方法研究[J]. 光谱学与光谱分析, 2009, 29(3): 671-674.

[19] Tewari J C, Dixit V, Cho B K, 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]. Spectrochimica Acta A: Mol Biomol Spectrosc, 2008, 71(3): 1119-1127.