

文章编号: 0253-2239(2008)02-0326-05

应用红外光谱技术快速检测月桂酸单甘油酯的品质指标

冯凤琴 张辉 王莉 何勇*

(浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029)

摘要 月桂酸单甘油酯是用途广泛的食物添加剂,在其制备过程中经分子蒸馏得到的制备品中会有月桂酸、甘油等杂质。用化学滴定或气相色谱等传统方法检测制备品中的月桂酸单甘油酯纯度及其杂质含量过程相当繁琐。为了对月桂酸单甘油酯制备品的品质进行快速定量,先利用气相色谱法确定不同工艺下的月桂酸单甘油酯产品中各成分的含量,再利用傅立叶红外光谱仪对月桂酸单甘油酯制备品进行分析,得到它们的光谱数据曲线,并结合主成分分析和反向传播神经网络建立回归模型。通过对实验结果的均方根误差预测值 P_{RMSE} 以及相关系数 r 辨析,预测月桂酸单甘油酯含量的结果为 $P_{RMSE} = 3.6376, r = 0.9950$, 预测甘油含量的结果为 $P_{RMSE} = 1.4764, r = 0.9795$, 预测月桂酸含量的结果为 $P_{RMSE} = 1.2859, r = 0.9247$ 。结果表明,应用光谱分析方法能够较好检测月桂酸单甘油酯、月桂酸和甘油含量。

关键词 医用光学与生物技术; 傅里叶变换红外光谱; 月桂酸单甘油酯品质检测; 主成分分析; 反向传播神经网络
中图分类号 TH744.1 文献标识码 A

Rapid Detection of the Quality Indicators in Glycerol Monolaurate Products Using Fourier Transform Infrared Spectroscopy

Feng Fengqin Zhang Hui Wang Li He Yong

(College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China)

Abstract Glycerol monolaurate (GML) is a food additive which is widely used. The products prepared by esterification of glycerol and lauric acid followed by molecular distillation have lauric acid, glycerol and other impurities. To detect the glycerol monolaurate, lauric acid and glycerol in the products is a very tedious process using traditional methods such as chemical titrimetry and gas chromatography. For fast quality detection of the products that has been analyzed on gas chromatography, Fourier transform infrared spectroscopy (FTIS) was used in this study. Principal component analysis (PCA) and back-propagation artificial neural networks (BP-PANN) were used to build the regression model. Then this model was used to detect the glycerol monolaurate, lauric acid and glycerol of samples in the validation set according to their spectra. Root mean square error prediction (RMSEP) and correlation coefficient (r) of prediction were used as the evaluation standards, the predicted results of GML was $P_{RMSE} = 3.6376, r = 0.9950$, the predicted results of glycerol was $P_{RMSE} = 1.4764, r = 0.9795$ and the predicted results of lauric acid was $P_{RMSE} = 1.2859, r = 0.9247$. The results indicated that it was feasible to detect glycerol monolaurate, lauric acid and glycerol content using spectroscopic technique.

Key words medical optics and biotechnology; Fourier transform infrared spectroscopy; quality detection of glycerol monolaurate; principal component analysis (PCA); back-propagation artificial neural networks (BP-ANN)

1 引 言

月桂酸单甘油酯(Glycerol monolaurate, GML)通称 Monolaurin, 国外商品名为 Lauricidin, 分子式为 $\text{CH}_3(\text{CH}_2)_{10}\text{COOCH}_2\text{CHOHCH}_2\text{OH}$, 其结构中

既有亲水性基团, 又有亲油性基团, 是一种能保持油水体体系稳定的性能优异的乳化剂, 同时又是一种国际公认的安全、高效、广谱的抑菌剂, 其抑菌效果不会因 pH 的改变而改变, 一般在 pH4~8 范围内均

收稿日期: 2007-06-22; 收到修改稿日期: 2007-08-22

基金项目: 国家自然科学基金(30671213)和教育部高等学校优秀青年教师教学科研奖励计划(02411)资助课题。

作者简介: 冯凤琴(1964—), 女, 河北石家庄人, 副教授, 博士, 主要从事食品化学方面的研究。E-mail: fengfq@zju.edu.cn

* 通信联系人。E-mail: yhe@zju.edu.cn

有效,且抗菌谱广,对食品中常见细菌、霉菌、酵母均有较强作用,同时还具有抑制多种病毒和原生动物的功效。

一般测定月桂酸单甘脂含量的方法有高碘酸氧化滴定法和高效液相色谱法等。冯凤琴等^[1]建立了气相色谱法测定月桂酸单甘酯含量的方法,该法比高碘酸氧化滴定法等耗时短、精密度高、测量成本低。

傅里叶变换红外光谱(Fourier transform infrared spectroscopy, FTIR)技术是一种快速、无损的检测手段,目前已广泛应用于化工、食品以及其它一些领域,例如在黑木耳产地鉴别^[2]、胆结石中蛋白质的近红外光谱特性^[3]等方面的应用。另外,田格等^[4]应用二维傅里叶变换红外(2DFTIR)光谱研究了细菌合成聚羟基丁酸酯在升温时的预熔行为以及降温后的重结晶过程;吴航等^[5]利用该方法研究了交联淀粉的退化过程,均取得了较好的效果。

红外光谱分析中的关键就是化学计量学方法,如何从庞大的光谱数据中最大限度地提取相关信息是化学计量学研究的核心。主成份分析(Principal component analysis, PCA)是多元统计中的一种数据挖掘技术,在不丢失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量,解决了由于谱带的重叠而无法分析的困难。反向传播神经网络模型是一个强有力的学习系统,能够实现输入与输出之间的高度非线性映射。目前使用最多的是多层结构的误差反向传播学习算法,并且已经证明此种模型可以逼近任何连续的非线性曲线。已经有学者将主成分分析和反向传播神经网络结合处理光谱数据^[6~8],结果表明这是一种有力地光谱建模方法。

2 材料与方法

2.1 样品与仪器

实验所用的月桂酸单甘油酯制备品由杭州康源食品科技有限公司提供,共 180 个样本,均呈白色蜡状粉末。因生产工艺不同,其中月桂酸单甘油酯、二酯及月桂酸、甘油等含量也不同。

实验所采用的气相色谱仪是山东鲁南瑞虹化工仪器有限公司 SP-6890 气相色谱仪,气相色谱分析软件为浙江大学智能信息工程研究所 N-2000 双通道色谱工作站。

实验所采用的红外光谱仪是日本 JASCO Model FT/IR-4100 傅里叶红外光谱仪,仪器测量范围 7800~350 cm^{-1} ,实验选定光谱检测范围为 4000~

400 cm^{-1} ,仪器分辨力 16 cm^{-1} 、8 cm^{-1} 、4 cm^{-1} 、1 cm^{-1} 四级,实验中设定的分辨力为 4 cm^{-1} 。整个实验过程保持室内温度在 25 $^{\circ}\text{C}$ 左右。为了增加透过率,将 95% 的溴化钾(作为光谱背景,不会影响待测样本)和 5% 的样本进行混合,然后通过压片机将混合物压成直径约 5 mm,厚度 2 mm 的圆柱体,放入光谱仪中进行光谱透过率测试。每个样本测试 16 次,并对测试值取平均。光谱数据采集软件为 FT/IR-4000 红外光谱仪附带的 Spectra Manger Version 2。光谱分析和数据处理使用的软件为 The Unscrambler 9.6 (CAMO Software AS, Norway) 和 Matlab R2006b (The MathWorks, Inc. USA)。

色谱柱为 OV-17 玻璃填充柱(长 1.5 m,内径 3 mm,岛津公司),光离子检测器(FID)检测。进样室温度 320 $^{\circ}\text{C}$,检测器温度为 330 $^{\circ}\text{C}$ 。柱子采用程序升温,分 3 个阶段:130~220 $^{\circ}\text{C}$,每分钟升温 10 $^{\circ}\text{C}$;220~280 $^{\circ}\text{C}$,每分钟升温 15 $^{\circ}\text{C}$;280 $^{\circ}\text{C}$ 恒温停留 10 min。载气为氮气,流速为 60 mL/min。样品衍生化方法以及采用十四酸作为内标制作标准曲线的方法参照^[1]。

首先利用气相色谱仪对所有样品进行分析,得到每个样品中月桂酸单甘油酯、月桂酸和甘油含量,并将此结果作为参考值。利用红外光谱仪得到所有样品的光谱曲线,将实验中所有样品随机分为建模集和预测集。利用建模集光谱数据优化反向传播神经网络模型。随后应用此模型对预测集样本的月桂酸单甘油酯、月桂酸和甘油含量进行预测,将预测结果与参考值进行比较,并以此来评价光谱模型性能的优劣。

2.2 主成分分析

设有 p 个原始指标: x_1, x_2, \dots, x_p , 用来评价 n 个单位,则共有 np 个数据,这 p 个原始指标之间往往存在着一定的相关性。主成分分析的目的是要将这些原始指标组合成新的不相关的指标: y_1, y_2, \dots, y_p , 以使各指标在整个经济过程中的作用容易解释。这些综合指标表现为原始指标的线性函数:

$$\begin{aligned} y_1 &= l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p, \\ y_2 &= l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p, \\ &\dots, \\ y_p &= l_{p1}x_1 + l_{p2}x_2 + \dots + l_{pp}x_p, \end{aligned} \quad (1)$$

由于所组合成的新指标 y_1, y_2, \dots, y_p 彼此不相关,有可能从中选择主要成分,通过对主要成分的重点分析,达到综合评价的目的。

通过数学计算可以将 p 个原始指标 x_1, x_2, \dots, x_p 的总方差分解为新的不相关指标 y_1, y_2, \dots, y_p 的方差之和 $\lambda_1 + \lambda_2 + \dots + \lambda_p$, 并使第一个综合指标 y_1 方差达到最大(贡献率最大), 第二个综合指标 y_2 方差达到次大, 依此类推。一般前面几个综合指标 $y_1, y_2, \dots, y_k (k < p)$ 即可包含总方差中的绝大部分。也就是说, 主成分分析可以使原始指标的大部分方差集中于少数几个主成分综合指标上。通过对这几个主成分的分析来实现对总体的综合评价。

主成分分析法的计算步骤, 可分为五个主要步骤:

- 1) 列出院士指标数值矩阵 \mathbf{X} ;
- 2) 计算 \mathbf{X} 的相关矩阵 \mathbf{R} ;
- 3) 计算相关矩阵 \mathbf{R} 的特征值 λ 和特征向量 \mathbf{L} (即指标 x 的系数);

4) 计算贡献率和累计贡献率, 据以确定主成分的个数, 并建立主成分方程; 每个主成分的贡献率等于它的特征值除以原始指标个数, 累计贡献率等于各主成分的贡献率顺序相加, 根据一定的选择标准, 如果前 k 个主成分的累计贡献率大于或等于所规定的门限值 ($k < p$), 则可选定这 k 个主成分, 根据特征向量 \mathbf{L} 建立这 k 个折成得分线性方程:

$$\begin{aligned} y_1 &= l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p, \\ y_2 &= l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p, \\ &\dots\dots, \\ y_k &= l_{k1}x_1 + l_{k2}x_2 + \dots + l_{kp}x_p. \end{aligned} \quad (2)$$

5) 解释各主成分的意义, 并将各单位的原始指标数值代入方程中计算综合评价进行分析比较。

3 结果与讨论

3.1 气相色谱图谱

把一典型样品衍生化后, 按气相色谱法的色谱条件进行分析, 得到的色谱图如图 1。把该色谱图与各组分标准品的色谱图相对应, 可以确定 4 个峰, 按先后顺序分别是: 甘油、月桂酸、十四酸(内标物)、月桂酸单甘油酯。月桂酸、甘油、月桂酸单甘油酯的标准曲线线性相关显著, 相关系数均大于 0.99。

3.2 红外光谱图谱

样本的红外透射光谱曲线如图 2 所示(为了能够观察清楚光谱曲线, 随机选择十条曲线)。图 2 中横坐标为红外光谱波数, 纵坐标为光谱透射率。从图 2 中可以看出, 不同样品的光谱曲线交错重叠。光谱数据存在不同程度的平移, 还存在一些噪声, 所以在进行光谱数据分析之前首先要进行预处理。

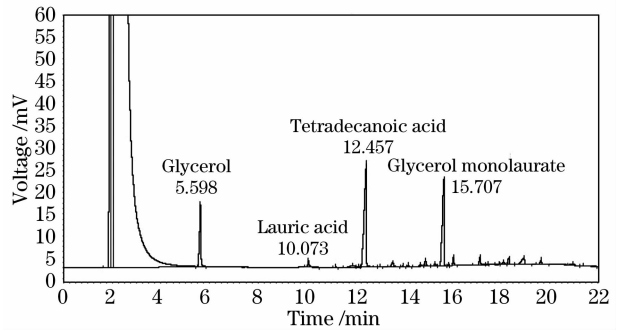


图 1 一个典型的气相色谱分析图

Fig. 1 Gas chromatograms of a typical sample

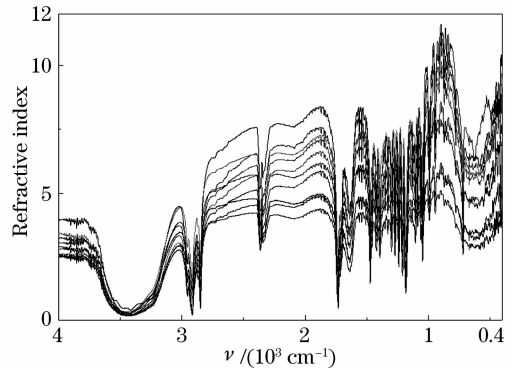


图 2 十个样品的红外透射光谱图

Fig. 2 IR transmittance spectra of ten samples

3.3 光谱数据预处理

为了去除来自高频随机噪声、基线漂移、样本不均匀、光散射等影响, 需要进行光谱预处理来消除噪声。首先采用移动平均平滑法 (Moving average smoothing) 平滑, 选用平滑点数为 3, 此时能很好滤除各种因素产生的高频噪声, 随后再进行多元散射校正 (MSC) 和标准正态变量变换 (SNV) 以消除基线漂移、光散射以及样本不均匀的影响^[9]。

3.4 主成分分析

红外光谱仪采集得到的每个样本的光谱曲线包含 3736 个透射率值, 如此庞大的数据集不能直接用于反向传播神经网络模型的建立。为了提高建模运算速度和精度, 减少运算量, 通过主成分分析对原始光谱输入变量进行降维压缩, 此步骤应用软件 The Unscrambler 9.6 进行主成分分析, 可以得到前 10 个主成分累计贡献率和剩余残差如表 1 所示。前 5 个主成分的累计贡献率 (Reliabilities) 已经达到了 90.787%, 而且交互验证的剩余残差 (Residual variances, RV) 最小, 通过对预测样本结果均方根误差 (RMSEP) 校验, 5 个主成分的均方根误差值最小, 所以选用前 5 个主成分作为优化输入特征子集, 进行反向传播神经网络建模。

表 1 主成分累计贡献率和剩余残差

Table 1 The reliabilities and residual variances (RV) of principal components (PCs)

Principal component	Reliabilities /%	RV /10 ⁻³
PC1	40.443	23.49
PC2	73.017	10.64
PC3	81.376	7.347
PC4	86.937	5.153
PC5	90.787	1.014
PC6	93.808	2.443
PC7	95.300	1.854
PC8	96.464	3.634
PC9	97.180	1.113
PC10	97.429	1.395

3.5 反向传播神经网络建模及预测

建模集样本的 5 个主成分作为神经网络的输入,建立三层反向传播神经网络模型。一般来说反向传播神经网络具有三层就能够满足大多数问题的要求,即对于任何在闭区间内的一个连续函数都可以用单隐层的反向传播网络逼近,因而一个三层反向传播网络就可以完成任意的 n 维到 m 维的映射。模型的输入节点数 5,输出节点数 3,通过对网络训练误差的循环寻优确立网络模型的隐含层节点数为 8。神经网络的训练函数选用 Trainlm,该函数的学习算法为 L-M 反传算法,具有更好的收敛效果;学习函数选择梯度下降动量学习函数(Learngdm),同

Matlab 工具箱中的 learngd 函数相比它能提供更快的收敛速度,动量允许网络不但根据当前梯度而且还能根据误差曲面最近的趋势响应。就像一个低通滤波器一样,动量允许网络忽略误差曲面的小特性。没有动量,网络又可能在一个局部最小中被卡住。有了动量网络就能够平滑这样的最小。动量能够通过把权重变得与上次权重变化对动量的影响由一个动量常量来决定,它能够设定为 0 到 1 之间的任意值。当动量常量为 0 时,权重变化根据梯度得到。当动量常量为 1 时新的权重变化等于上次的权重变化,梯度值被忽略了。神经网络各层之间的传递函数选用 Logsig(S 函数的对数函数),函数 Logsig 可将神经元的输入(范围为整个实数集)映射到区间(0, 1)中;性能函数选择均方误差规范化函数(Msereg),可以避免反向传播网络中由于学习速率因子取值不当造成的误差,能很好地抑制过拟合问题的发生,因而能减少网络训练的迭代次数;设定网络的训练步数为 1000,训练目标为 0.0001,迭代次数的设定是经验值,只要能充分达到或接近训练目标即可,步数不宜选定过大,否则容易引起过拟合。

用优化的神经网络对预测集样本进行仿真。图 3 即为预测结果,横坐标各为三种成分含量的参考值(真实值),纵坐标为神经网络仿真后的预测值。

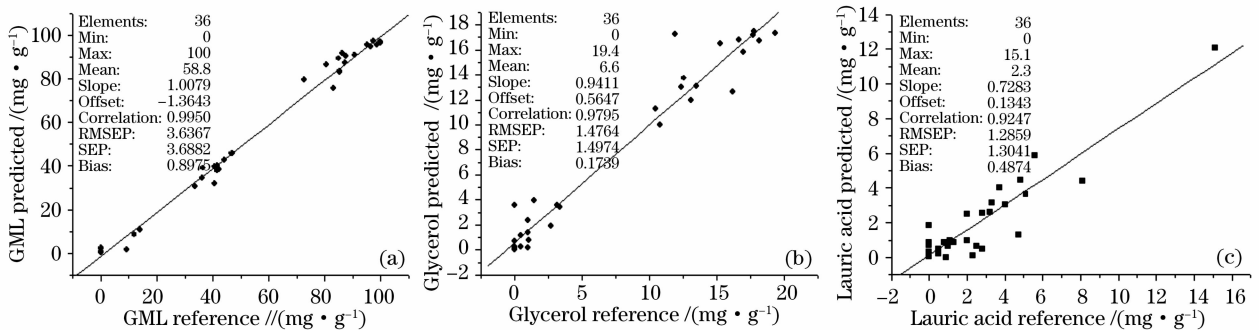


图 3 月桂酸单甘油酯(a),甘油(b),月桂酸(c)的预测结果

Fig. 3 Predicted results of GML (a), glycerol (b), lauric acid (c)

图 3(a)是对预测样本的月桂酸单甘油酯含量的预测结果,其 $P_{RMSE}=3.6376$, $r=0.9950$;图 3(b)是对预测样本甘油(Glycerol)含量的预测结果,其 $P_{RMSE}=1.4764$, $r=0.9795$;图 3(c)是对预测样本月桂酸(Lauric acid)含量的预测结果,其 $P_{RMSE}=1.2859$, $r=0.9247$ 。通过对预测结果的比较发现,利用红外光谱对月桂酸单甘油酯的预测结果相关系数达到了 0.9950,取得了比较好的效果。对于产品中的杂质甘油以及月桂酸的预测结果相对差一些。

分析其原因是由于甘油和月桂酸在产品中含量较少,其光谱特征吸收峰容易收到主要成分月桂酸单甘油酯特征吸收峰的影响,导致分析结果误差稍大一些。

4 结 论

探索了利用傅里叶变换红外光谱技术检测月桂酸单甘油酯制备过程产品中月桂酸单甘油酯以及甘油、月桂酸两种杂质的含量。预测结果的相关系数

和均方根误差都达到了较好的效果,尤其是对制备品中的有效成分(月桂酸单甘油酯)的预测,效果更加理想。同传统方法相比,利用红外光谱技术进行检测,过程更加简洁,所需时间更短,所以该项技术更加容易实现实时在线检测。今后的研究方向,主要是优化预测模型,使其能够对产品中含量较少的杂质进行更加准确的预测,同时需研究制备过程中各种成分的敏感波段,以减少数据处理过程中的计算量和重叠波谱对预测结果的影响。

参 考 文 献

- 1 Feng Fengqin, Yang Hongjie. Study on the method for the determination of glycerol monolaurate (GML) by GC[J]. *J. Chinese Institute of Food Science and Technology*, 2005, **1**(5): 59~63
冯凤琴,杨宏杰. 气相色谱法测定月桂酸单甘酯含量的方法研究[J]. *中国食品学报*, 2005, **1**(5): 59~63
- 2 Shi Youming, Liu Gang, Liu Jianhong *et al.*. Identification of auricularia auricula from different regions by Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(1): 129~132
时有明,刘 刚,刘剑虹等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. *光学学报*, 2007, **27**(1): 129~132
- 3 Liu Gang, Xing Da, Wang Haimin *et al.*. Study of protein in human gallstones by Fourier transform infrared spectroscopy and surface-enhanced Raman spectroscopy[J]. *Acta Optica Sinica*, 2002, **22**(4): 441~446
刘 刚,邢 达,王海珉等. 胆结石中蛋白质的傅里叶变换红外光谱和表面增强拉曼光谱研究[J]. *光学学报*, 2002, **22**(4): 441~446
- 4 Tiang Ge, Wu Qiong, Sun Suqin *et al.*. Studies on pre-melting and crystallization process of biosynthesized poly(3-hydroxybutyrate) using two-dimensional Fourier-transform infrared spectroscopy[J]. *Chemical J. Chinese Universities*, 2002, **23**(8): 1627~1631
田 格,吴 琼,孙素琴等. 二维傅里叶变换红外(2DFTIR)相关光谱技术研究聚羟基丁酸酯(PHB)的熔融与结晶[J]. *高等学校化学学报*, 2002, **23**(8): 1627~1631
- 5 Huang Mingzhi, Chen Weimin, Huang Kelong *et al.*. Monitoring of solid-phase synthesis of 7-fluoro-2H-1,4-benzoxazine-3(4H)-ones via FTIR spectroscopy [J]. *Chemical J. Chinese Universities*, 2006, **27**(11): 2117~2119
黄明智,谌伟民,黄可龙等. FTIR 监测 7-氟-2H-1,4-苯并噁嗪-3(4H)-酮的固相合成[J]. *高等学校化学学报*, 2006, **27**(11): 2117~2119
- 6 He Y, Feng S J, Deng X F *et al.*. Study on lossless discrimination of varieties of yogurt using the Visible NIR-spectroscopy[J]. *Food Res. Int.*, 2006, **39**(6): 645~650
- 7 He Y, Li X L, Deng X F. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model[J]. *J. Food Engineering*, 2007, **79**(4): 1238~1242
- 8 He Yong, Feng Shuijuan, Li Xiaoli *et al.*. Study on fast discrimination of varieties of acidophilous milk using near infrared spectra[J]. *Spectroscopy and Spectral Analysis*, 2006, **26**(11): 2021~2023
何 勇,冯水娟,李晓丽等. 应用近红外光谱快速鉴别酸奶品种的研究[J]. *光谱学与光谱分析*, 2006, **26**(11): 2021~2023
- 9 Chu Xiaoli, Yuang Hongfu, Lu Wanzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, **16**(4): 528~539
褚小立,袁洪福,陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, **16**(4): 528~539