

利用傅立叶变换简化苹果糖度的预测模型的研究

屠振华, 籍保平, 史波林, 朱大洲
(中国农业大学食品科学与营养工程学院 北京 100083)

摘要: 本文通过离散傅立叶变化, 将苹果光谱信号转换为一系列的傅立叶系数。实验表明运用较少的傅立叶系数(15个)建立苹果糖度的PLS和PCR模型的效果与原始光谱(88个变量)建立的PLS模型的效果基本一致。这可以表明选择前若干个傅立叶系数建模, 可以包含原光谱图的绝大部分信息。这样既可以较好的压缩图谱, 减少计算量, 同时也可以达到定量分析的要求。

关键词: 傅立叶系数; 压缩图谱; 苹果糖度; 近红外光谱

Simplification of Prediction Model for Apple Sugar Content Using Fourier Analysis

TU Zhen-hua, JI Bao-ping, SHI Bo-lin, ZHU Da-zhou
(College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China)

Abstract: A Fourier transform (FT) was used as a tool to reduce the number of variables in pattern recognition. It is carried out in the frequency domain starting from the first 15 Fourier coefficients (FFT coefficients) as the input to the PLS (partial least square) method and the PCR (principal component regression) method. Comparison between the PLS FFT method and other calibration methods applied in the original domain (wavelength domain WD) and in the FFT domain. The results demonstrate that FT is a fast and powerful reduction method. This also shows that the method can be used to investigate the structure of the data.

Key words: Fourier coefficients; feature selection; apple sugar content; NIR

中图分类号: TS207.3

文献标识码: A

文章编号: 1002-6630(2006)10-0135-04

在农产品品质的近红外光谱无损检测中, 偏最小二乘法(partial least square, PLS)和主成分回归(principal component regression, PCR)等多元校正方法常被用于建立预测模型。但是由于农产品是非常复杂的生物体, 其组分多, 相应的衍生物复杂, 在近红外谱区内, 由于吸收相互干涉, 造成重叠多谱带复杂。因此在农产品光谱分析中与待测组分不相关的信息必定会对近红外光谱预测产生影响, 使预测模型变得复杂。傅立叶变换是一种在信号分析中广泛运用的方法, 将原始采样的时间序列转换成易于分析和处理的谱图。通过离散傅立叶变换, 可以将光谱信号转换为一系列的傅立叶系数。McClure 由 Beer 定律出发, 证明了傅立叶系数与被测样品浓度之间存在线性关系^[1]。这样我们就可以将多元统计的方法和傅立叶系数结合起来, 建立线性模型。由于原光谱主要信息集中在傅立叶系数的前若干个值上^[2], 因此利用傅立叶回归建立模型可以显著降低计算量, 达

到图谱压缩的目的。

本研究运用傅立叶系数和偏最小二乘法结合建立了苹果糖度的预测模型, 比较原始光谱图的预测模型, 表明利用傅立叶变换可以很好的简化苹果糖度的预测模型。

1 傅立叶回归的原理

傅立叶变换(FT)是一种十分重要的信号处理技术, 它能够实现频域函数和时域函数之间的转换, 实质是把原光谱分解成许多不同频率的正弦波的叠加和^[3]。根据需要可通过 FT 对原始光谱数据进行平滑、插值、滤波、拟合及提高分辨率等运算, 或用 FT 频率谱即权系数(傅立叶系数)直接参与模型的建立。在 NIR 光谱分析中, 傅立叶变换可用来对光谱进行平滑去噪、数据压缩以及信息的提取。

通常检测器采集到的 $x(t)$ 是有限长度, 等间隔的数据点 $x(t_n)$ 构成的序列, 即数字信号, 有:

收稿日期: 2006-08-14

作者简介: 屠振华(1983-), 男, 硕士研究生, 研究方向为食品加工新技术。

$$x(t) \in \{x(t_n)\}, \quad (n=0, L, N-1) \quad (1)$$

故可将表示 $x(t_n)$ 为有限的数字信号序列：

$$x[n], \quad (n=0, 1, L, N-1)$$

离散傅立叶变换(DFS)可写成：

$$X(k) = x[n]e^{-j2\pi kn/N} \quad (k=0, 1, L, N-1) \quad (2)$$

通过离散傅立叶变换, 可以将近红外光谱信号转化为一系列的傅立叶系数^[4]：

$$F(k) = \sum_{n=0}^{N-1} x[n] \exp\left(\frac{-j2\pi kn}{N}\right), \quad (k=0, 1, L, N-1) \quad (3)$$

其中：序列 $x[n]$ 为每个样本的在 N 个波长点下的吸光度值, $F(k)$ 为经过离散傅立叶变换(DFS)得到的 N 个在频率域下的傅立叶系数值。

McClure 由 Beer 定律出发^[1], 经过简单的数学推导, 证明了傅立叶系数与被测样品浓度之间存在线性关系, 这样我们就可以将多元统计的方法和傅立叶系数结合起来, 建立回归模型。本试验中运用傅立叶系数和偏最小二乘法来建立苹果糖度的预测模型。

2 材料与方法

2.1 仪器及其参数选择

试验仪器为天津大学研制生产的声光可调谐滤光器(AOTF)近红外光谱仪, 仪器可采集的光谱范围 1000 ~ 1850nm, 光谱分辨率 < 1nm, 光源 50W/220V 卤钨灯, 铟镓砷(InGaAs)探测器, 信噪比(SNR) > 2500; 实验中光谱的采集范围 1065 ~ 1625nm, 波长个数为 88 个, 扫描次数 8 次/s。每次测试前都必须先使仪器预热 30min 以上。本实验采用漫反射测量方法, 使用光纤技术, 光通过入射光纤透入到苹果内部, 经内部组织漫反射后携带相应的内部品质信息的光返回苹果表皮, 再通过接收光纤进入光谱仪的探测器。

可溶性固形物的真值由北京万成北增精密仪器有限公司生产的 WZ-113 型便携式折光糖度仪(经过计量部门检测)测量, 测量单位是 Brix/ATC, 范围是 0 ~ 32%。

2.2 材料

苹果样品品种为陕西红富士, 共 80 个样品。购回后置于 4℃ 冰柜中贮藏。光谱检测试验在环境温度可控的实验室(本研究中环境温度控制为 26℃)内进行。

2.3 方法

实验前, 先取出冷藏的红富士苹果 80 个, 将它们置于实验室温度下 12h。以使苹果整体温度达到与环境温度一致。然后从中随机的选取 60 个作为校正集, 剩下的 20 个作为预测集, 并分别编号。

2.3.1 光谱测量

以直径为 30mm, 厚度为 2mm 的 BaSO₄ 标准块作为参比。实验时, 在带皮苹果的最大横径上等距离地选

取 4 个点进行光谱扫描, 取这 4 个点的平均光谱作为该带皮苹果的原始光谱。这样, 校正集的样本数为 60 个, 预测集的样本数为 20 个。

在光谱采集时, 综合考虑随机误差和系统误差, 采样次数选为 5 次, 即对于苹果的某一采集点在一次检测过程中光谱仪器采样的光谱图为 5 张, 然后对这 5 张光谱图进行平均, 则所得光谱为该点的光谱。因此, 代表整个苹果的原始光谱为 5 × 4 = 20 张光谱的平均光谱。图 1 为实验采集的 80 个样品的光谱图。

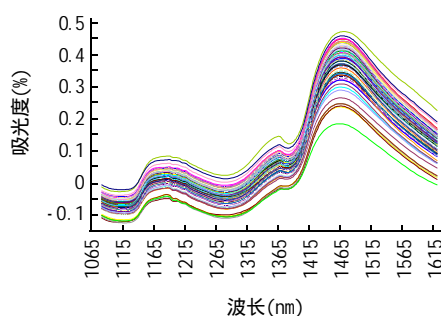


图 1 80 个苹果的近红外光谱图

Fig.1 NIR spectrum of 80 samples of apples

由图 1 所能看出的最明显特征就是一个波峰, 其位置在 1450 ~ 1490nm 左右, 根据近红外光谱的吸收特性, 在 1470nm 波长处, 是糖类的吸收波长, 因此可以初步判断近红外光谱对苹果的糖度分析能起到一定的作用。

2.3.2 苹果中可溶性固形物含量的测定

在 20℃ 的环境下, 将光谱采集后的苹果可食用部分榨汁、离心后, 用手持式糖度计(WYT0-32 型, 泉州韦达计量仪器厂)测定其可溶性固形物含量(按国标 GB10651-89 执行)作为整个苹果的可溶性固形物含量(糖度), 单位以白利度(°Brix)表示。具体可溶性固形物的测定结果如表 1 所示。校正集和预测集的样品的成分分布图见图 2。

表 1 苹果可溶性固形物化学值统计结果
Table 1 Statistic data of SSC in samples

样品项	校正集	预测集
No	60	20
Mean(°Brix)	13.06	13.295
Range(°Brix)	9.5 ~ 17.7	11.7 ~ 15.8
Std	2.02	1.25

2.4 数据处理

本实验的傅立叶系数计算采用 Origin 6.0 的程序计算得到。采用自编的程序使用 Matlab(7.04)建立了模型, 并通过相关系数 r 评价其相关性, 校正集样品的标准误差(RMSEC)和预测集样品的标准误差(RMSEP)来评价模型对未知样品的预测效果。

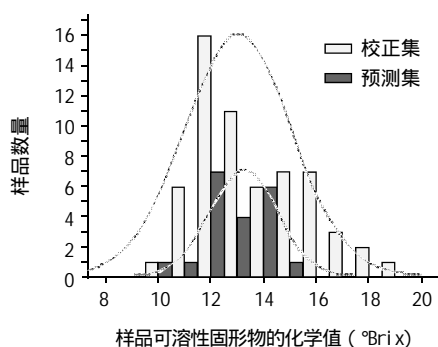


图2 苹果可溶性固形物化学值统计结果图

Fig.2 Statistic data of SSC in samples

相关计算公式如下^[5]：

$$r = \frac{\sum_{i=1}^m (z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} \quad (4)$$

$$RMSEC = \sqrt{\frac{1}{N-P-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

$$RMSEP = \sqrt{\frac{1}{NN} \sum_{i=1}^{NN} (y_{Ni} - \hat{y}_{Ni})^2} \quad (6)$$

式中， n 为样本数； N 为校正集样本数； NN 为预测集样本数； \hat{y}_i 为第 i 个样品的预测值； y_i 为第 i 个样品的测量值； y_m 为样品测量值的平均数； P 为使用的主成分数。

3 结果与分析

3.1 傅立叶回归和其参数选择

本研究对校正集的 60 个样本和预测集 20 个样本的光谱图进行离散傅立叶变换，得到这 80 个样本的傅立叶系数。图 3 为 5 号样本的原始光谱图。图 4 为该样本其变换后的傅立叶系数图。

对得到的傅立叶系数分别利用前 5 个、10 个、15 个、20 个、25 个，用偏最小二乘法建立模型并进行预测。实验结果如表 2 所示。

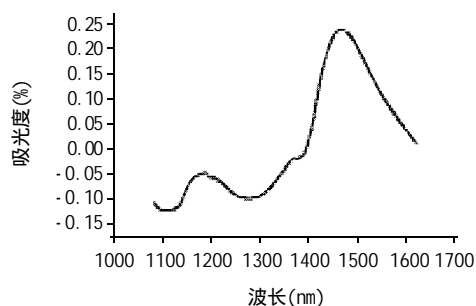


图3 5号样本的原始光谱图

Fig.3 The NIR spectra of the fifth sample of apples

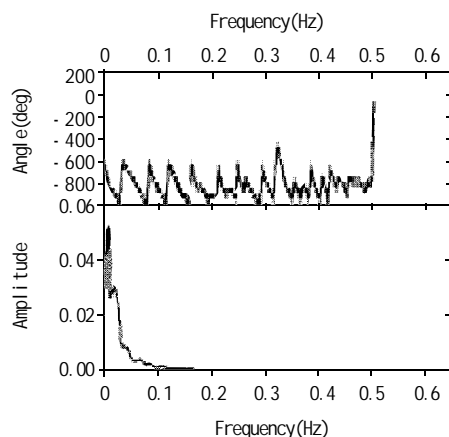


图4 5号样本的傅立叶系数图

Fig.4 The FFT coefficients of the fifth sample of apples

表2 各傅立叶系数 PLS 建模效果

Table 2 Results obtained for the selection of FFT coefficients with PLS

模型	变量数	主成分数	校正集		预测集
			决定系数	RMSEC(°Brix)	RMSEP(°Brix)
5	5	1	0.2274	1.9552	1.3723
10	10	8	0.8853	0.9338	0.8984
15	15	9	0.9056	0.8517	0.8236
20	20	9	0.9092	0.8359	0.8495
25	25	5	0.7504	1.3269	1.4124

结果表明：当选取前 15 个傅立叶系数用偏最小二乘法建立模型进行预测比用其它个数傅立叶系数用偏最小二乘法建立模型的效果更好（相关系数基本一致，预测标准偏差最小）。同时选取前 15 个可以最大的达到数据压缩的目的。用 15 个傅立叶系数用偏最小二乘法建立模型效果如图 5。

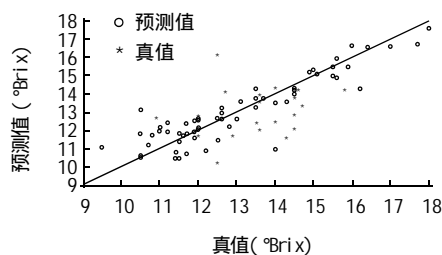


图5 用 15 个傅立叶系数用偏最小二乘法建立模型效果图

Fig.5 The 15 FFT coefficients -variables model: y predicted vs y observed

3.2 运用选取后的傅立叶系数各种建模方法结果比较

本实验分别对校正集的 60 个样本和预测集 20 个样本的选取后的傅立叶系数值运用偏最小二乘校正(PLS)，主成分回归(PCR)，建立模型，和用原始光谱运用偏最小二乘校正(PLS)建立的模型进行比较，比较结果如表 3。用原始光谱运用偏最小二乘校正(PLS)建立的模型效果如图 6。

表3 傅立叶系数 PLS、原始光谱 PRC、PLS 建模效果比较
Table 3 Comparison between the PLS FFT method and other calibration methods applied in the original domain(wavelength domain WD) and in the FFT domain

模型	变量数	主成分数	决定系数	RMSEP(°Brix)
PLS(FFT)	15	9	0.9056	0.8236
PCR(FFT)	15	11	0.9027	0.7792
PLS(WD)	88	6	0.8643	0.7680

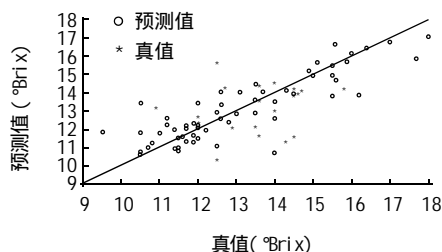


图6 原始光谱用偏最小二乘法建立模型效果图
Fig.6 The NIR spectra model: y predicted vs y observed

3.3 分析与讨论

本实验对校正集的 60 个样本和预测集 20 个样本的光谱图进行离散傅立叶变换,通过比较运用前 5 个、10 个、15 个、20 个、25 个傅立叶系数运用偏最小二乘法建立模型的效果表明,选择前 15 个傅立叶系数建立的 PLS 模型(相关系数 $r=0.9056$, 预设标准偏差 $RMSEP(^{\circ}Brix)=0.8236$)和运用前 15 个傅立叶系数建立的 PCR 模型(相关系数 $r=0.9027$, 预设标准偏差 $RMSEP(^{\circ}Brix)=0.7792$)的效果较好。同时,比较运用原始光谱建立的 PLS 模型(相关系数 $r=0.8643$, 预设标准偏差 $RMSEP(^{\circ}Brix)=0.7680$)的模型效果。我们可以看到用选用一些傅立叶系数建立的模型在压缩图谱的条件下,模型效果基本一致,原光谱图的绝大部分信息得到了保存。这表明针对本试验,选择少量傅立叶系数建模,完全可以包含原光谱图的绝大部分信息。这样既很好的压缩图谱,减小计算量的目的,同时也完全达到定量分析的要求,

起到一定的简化模型的效果。

同时,在比较选择前 15 个傅立叶系数建立的 PLS 模型(相关系数 $r=0.9056$, 预设标准偏差 $RMSEP(^{\circ}Brix)=0.8236$)和运用前 15 个傅立叶系数建立的 PCR 模型(相关系数 $r=0.9027$, 预设标准偏差 $RMSEP(^{\circ}Brix)=0.7792$)的模型效果。我们可以看到在压缩图谱后,由于变量较少,在保证模型效果的同时,可以采用计算时间更短的 PCR 方法来建立模型,这样也起到了一定简化的作用。

4 结论

本实验采用一种在信号分析中广泛运用傅立叶变化方法,将原始采样的时间序列转换成易于分析和处理的图谱。通过离散傅立叶变化,可以将光谱信号转换为一系列的傅立叶系数。之后运用偏最小二乘法建立傅立叶回归模型。通过比较不同个数傅立叶系数建模的效果,证明傅立叶回归的方法可以显著的减少计算量和提高计算速度。同时比较了原始光谱的模型效果,表明运用傅立叶系数建模是完全可行的。运用傅立叶回归的方法,对于数据量较大的近红外图谱,如用 CCD(2048 个波长点)型光谱仪采集的光谱,进行图谱压缩,提高计算速度是十分有意义的。

参考文献:

- [1] McClure W F, Hamid A, Giesbrecht F G, et al. Fourier analysis enhances NIR diffuse reflectance spectroscopy[J]. Applied Spectroscopy, 1984, 38: 322.
- [2] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
- [3] 胡广书. 数字信号处理[M]. 北京: 清华大学出版社, 2003.
- [4] Pasti L, Rimbaud D J, Massart D L, et al. Application of fourier transform to multivariate calibration of near-infrared data[J]. Analytica Chimica Acta, 1998, 365: 253.
- [5] 许禄. 化学计量学方法[M]. 北京: 科学出版社, 2004.

欢迎订阅

2007 年《食品与发酵工业》

入选“中国期刊方阵”期刊



全面报道食品和发酵工业两大领域国内外全新研究成果、学科前沿动态、产业发展趋势、生产经验,紧密跟踪企业经营状况及市场动态。既反映了行业的学术水平,又具有实用价值。

2007 年《食品与发酵工业》大 16 开本、200 页,单价 18 元,全年 216 元。全国各地邮局订阅,直接汇款至本刊编辑部订阅(免邮费)全年优惠价 180 元。

国际刊号:ISSN 0253-990X

国外代号:BM350

国内统一刊号:CN11-1802/TS

地址:北京市朝阳区霄云路32号

邮编:100027

电话:010-64645559

传真:010-64647111

E-mail: ffeo@vip.sina.com, ffeo@tom.com

http://spfx.chinajournal.net.cn

