

# 基于 Elman 网络的近红外光谱技术多组分定量分析研究

刘波平<sup>1,2</sup>, 秦华俊<sup>3</sup>, 罗 香<sup>2</sup>, 曹树稳<sup>3</sup>, 王俊德<sup>1\*</sup>

1. 南京理工大学现代光谱研究室, 江苏 南京 210014
2. 江西省分析测试中心, 江西 南昌 330029
3. 南昌大学食品科学教育部重点实验室, 江西 南昌 330047

**摘 要** 研究了 Elman 神经网络(反馈神经网络, Recurrent Network)在近红外光谱定量分析中的应用。以饲料样品为实验材料, 采用 Elman 网络建立了饲料中苯丙氨酸(Phe)、赖氨酸(Lys)、酪氨酸(Tyr)和胱氨酸(Cys)四种氨基酸含量的近红外光谱定量分析模型。用偏最小二乘法(partial least squares, PLS)将原始数据压缩为主成分, 取前 3 个主成分的 12 个吸收峰值输入 Elman 网络, 网络中间层神经元个数为 47。Elman 网络模型对样品 4 个氨基酸含量的预测决定系数( $r^2$ )分别为 0.960, 0.981, 0.979, 0.952。表明所建 Elman 网络预测模型通过近红外光谱能够较准确预测饲料中苯丙氨酸、赖氨酸、酪氨酸和胱氨酸四种氨基酸的含量, 为通过近红外光谱技术进行多组分定量分析提供了新思路。

**关键词** 近红外光谱; Elman 网络; 偏最小二乘法; 多组分定量分析

**中图分类号:** O657.3 **文献标识码:** A **文章编号:** 1000-0593(2007)12-2456-04

## 引 言

近红外光谱作为一种分析手段是根据样品化学成分在近红外光谱区域的光学吸收波长不同, 通过对已知化学成分含量的样品与其近红外反射光谱的回归分析, 建立定标方程, 从而对同一种类及其相似类型的未知样品进行估测。近红外光谱定量分析采用的建模方式有很多, 如主成分分析、多元线性回归、逐步线性回归、偏最小二乘、人工神经网络等<sup>[1-5]</sup>。其中 PLS 压缩的主成分所含信息能精确地表达输入与输出间的关系, 建立的模型具有较高的预测精度<sup>[6, 9]</sup>; 神经网络作为一种新兴的化学计量学方法, 有着其他经典方法无可比拟的特点, 它的测定结果准确, 受干扰影响小, 是多组分同时定量测定中的一种首选方法<sup>[8]</sup>, Elman 网络是一种动态神经网络, 它是在 BP 网络基本结构的基础上, 通过存储内部状态使其具备映射动态特征的功能, 从而使系统具有适应时变特性的能力, 比前向神经网络具有更强的计算能力<sup>[9]</sup>。目前, Elman 网络在组分预测主要应用在色谱和紫外<sup>[10-18]</sup>。本文将 Elman 网络应用于近红外光谱定量分析, 建立了实现饲料中四种氨基酸含量的同时预测模型, 取得了良好的效果。

## 1 Elman 网络基本原理

Elman 神经网络是 Elman 于 1990 年提出的, 该模型在前馈网络的隐含层中增加一个承接层, 作为一步延时算子, 达到记忆的目的, 从而使系统具有适应时变特性的能力, 能直接反映动态过程系统的特性<sup>[9]</sup>。Elman 神经网络一般分为 4 层: 输入层、中间层(隐含层)、承接层和输出层, 其输入层、隐含层和输出层的连接类似于前馈网络, 输入层的单元仅起信号传输作用, 输出层单元起线性加权作用。隐含层单元的传递函数可采用线性或非线性函数, 承接层又称为上下文层或状态层, 它用来记忆隐含层单元前一时刻的输出值, 可以认为是一个一次延时算子。

Elman 型神经网络的特点是隐含层的输出通过承接层的延迟与存储, 自联到隐含层的输入, 这种自联方式使其对历史状态的数据具有敏感性, 内部反馈网络的加入增加了网络本身处理动态信息的能力, 从而达到了动态建模的目的。Elman 网络的非线性状态空间表达式为

$$\begin{aligned} \mathbf{y}(k) &= g(\omega^3 \mathbf{x}(k)) \\ \mathbf{x}(k) &= f(\omega^1 \mathbf{x}_c(k) + \omega^2 (\mathbf{u}(k) - 1)) \\ \mathbf{x}_c(k) &= \mathbf{x}(k-1) \end{aligned}$$

其中,  $\mathbf{y}$ ,  $\mathbf{x}$ ,  $\mathbf{u}$ ,  $\mathbf{x}_c$  分别表示  $m$  维输出结点向量,  $n$  维中间层结点单元向量,  $r$  维输入向量和  $n$  维反馈状态向量。 $\omega^3$ ,

收稿日期: 2006-08-25, 修订日期: 2006-11-28

基金项目: 教育部南昌大学食品科学重点实验室开放基金项目(NCU200404)和江西省星火计划项目(2005年)资助

作者简介: 刘波平, 1963年生, 南京理工大学化工学院博士研究生 \* 通讯联系人 e-mail: jdwang@mail.njust.edu.cn

$\omega^2$ ,  $\omega^1$  分别表示中间层到输出层、输入层到中间层、承接层到中间层的连接权值。 $g(\cdot)$  为输出神经元的传递函数, 是中间层输出的线性组合。 $f(\cdot)$  为中间层神经元的传递函数, 常采用 S 函数。Elman 网络也采用 BP 算法进行权值修正, 学习指标函数采用误差平方和函数

$$E(\omega) = \sum_{k=1}^n [\tilde{y}_k(\omega) - \hat{y}_k(\omega)]^2$$

其中  $\tilde{y}_k(\omega)$  为目标输出向量。

## 2 仪器与实验材料

### 2.1 仪器

试验使用美国 Nicolet 公司的 460 型傅里叶拓展近红外光谱仪及近红外光纤探头来采集饲料样本光谱, 扫描范围 4 200~10 000  $\text{cm}^{-1}$ , 扫描次数为 80 次, 分辨率为 8  $\text{cm}^{-1}$ 。相关软件: OMNIC7.0, TQ7.0 及 Matlab7.0。饲料中苯丙氨酸(Phe)、赖氨酸(Lys)、酪氨酸(Tyr)和胱氨酸(Cys)四种氨基酸含量测定使用美国 Waters 公司的 2487 型液相色谱(HPLC)。

### 2.2 实验材料

试验样本采自 70 个不同厂家不同批次饲料样品, 用粉碎机粉碎至颗粒为 40 目, 全部样本分训练集 45 个, 预测集 25 个。用石英杯装约占容积 1/4 的饲料粉碎样品, 置于光纤架上分直立、侧放, 不同部位扫描样品近红外光谱, 每个样品采集 8 个谱图(图 1), 取平均值建模。

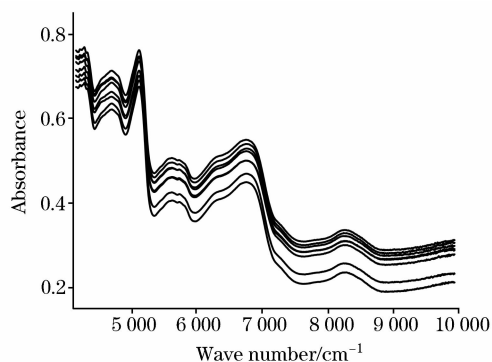


Fig. 1 The spectra of one sample scanned for 8 times

## 3 结果与讨论

### 3.1 光谱数据预处理

对谱图数据范围进行选择的时候可以采用全谱, 但是全谱的计算工作量是非常大的。而且, 并非所有谱图数据对建模都有用, 应该尽量剔除信息较弱、与样品缺乏相关关联的光谱区域数据以减少校正模型的计算工作量。剔除无用数据的方法有方差处理、一阶、二阶微分等, 择优原理都是将处理后光谱变化最明显的区域视为最有效的光谱范围。原始谱图经一阶微分处理后如图 2 所示, 从中可以选定原始谱图中 4 290~6 220  $\text{cm}^{-1}$ (1), 6 785~7 585  $\text{cm}^{-1}$ (2)两个波段光谱数据作为 PLS 压缩提取主成分的基础数据。

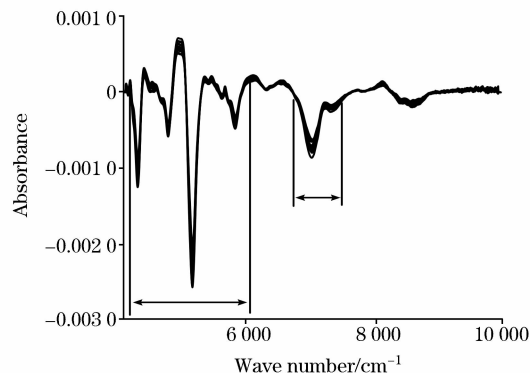


Fig. 2 The spectra processed by first derivative

### 3.2 PLS 法压缩提取主成分

用 PLS 法将 45 个训练集样品的光谱数据变换压缩为主成分, 前三个主成分得分及累计得分达 99.169%, 如表 1 所示。基本能涵盖样品信息。

Table 1 The result of spectra data processed by PLS

Principal component	Variance/%	Cumulative variance/%
1	90.123	90.123
2	7.946	98.070
3	1.099	99.169

### 3.3 网络的设计与训练

确定 Elman 网络的结构是预测性能的基础, 实际上机构的确定尤其是中间层神经元个数的确定是经验性的问题, 需要大量的实验。由样本数据可知, 网络输入层神经元数目应为 45, 输出层应为 4 个, 根据经验, 将中间层神经元个数设置为 43, 45, 47, Matlab 命令如下。

`net=newelm(threshold,[a(i),4],{'tansig','purelin'})`

其中  $a(i) = 43, 45, 47$ ; 'tansig','purelin' 分别表示中间层和输出层传递函数。训练之前设定训练步数为 1 000, 训练目标误差为 0.001。图 3 为网络取不同中间层神经元个数的预测误差曲线, 从中可以看出中间层取 47 个神经元时, 误差曲线相对较为平稳, 从而确定网络中间层神经元个数为 47。

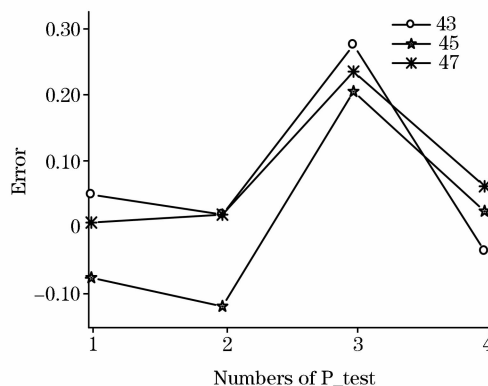


Fig. 3 The error curve of different number of hidden node

## 3.4 预测集样品对 Elman 网络模型的检验

表 2 是 Elman 网络模型对 25 个预测集样品的苯丙氨酸、

赖氨酸、酪氨酸和胱氨酸四种氨基酸含量预测结果。从预测决定系数( $R^2$ )可以看出模型的预测能力很好。

Table 2 The percentage content of four amino acids calibrated by HPLC and forecasted by Elman model in feedstuff powder

No.	Phe/%			Lys/%			Tyr/%			Cys/%		
	Pre.	HPLC	relative error	Pre.	HPLC	relative error	Pre.	HPLC	relative error	Pre.	HPLC	relative error
1	1.22	1.26	-3.17	0.72	0.72	0.00	0.56	0.55	1.82	0.24	0.31	-22.58
2	1.06	1.04	1.92	0.64	0.62	3.23	0.27	0.23	17.39	0.22	0.23	-4.35
3	1.21	1.17	3.42	0.57	0.52	9.62	0.55	0.54	1.85	0.33	0.30	10.00
4	1.13	1.14	-0.88	0.51	0.50	2.00	0.33	0.29	13.79	0.17	0.20	-15.00
5	1.36	1.33	2.26	0.47	0.55	-14.55	0.60	0.55	9.09	0.35	0.32	9.37
6	0.74	0.85	-12.94	0.81	0.80	1.25	0.41	0.47	-12.77	0.00	0.01	-
7	0.96	1.08	-11.11	0.60	0.66	-9.09	0.57	0.60	-5.00	0.14	0.18	-22.22
8	0.91	1.04	-12.50	0.33	0.33	0.00	0.38	0.33	15.15	0.21	0.17	23.53
9	1.09	1.12	-2.68	0.66	0.68	-2.94	0.49	0.55	-10.91	0.32	0.30	6.67
10	1.09	1.06	2.83	0.66	0.68	-2.94	0.56	0.62	-9.68	0.31	0.35	-11.43
11	1.26	1.28	-1.56	0.75	0.73	2.74	0.67	0.60	11.67	0.38	0.36	5.56
12	0.96	0.87	10.34	0.62	0.59	5.08	0.60	0.56	7.14	0.30	0.25	20.00
13	1.21	1.11	9.01	0.54	0.54	0.00	0.51	0.53	-3.77	0.13	0.18	-27.78
14	1.11	1.15	-3.48	0.59	0.58	1.72	0.52	0.57	-8.77	0.26	0.23	13.04
15	2.62	2.69	-2.60	1.68	1.68	0.00	1.99	1.92	3.65	1.01	0.96	5.21
16	1.17	1.24	-5.65	0.54	0.52	3.85	0.49	0.54	-9.26	0.24	0.27	-11.11
17	1.19	1.15	3.48	0.74	0.68	8.82	0.74	0.63	17.46	0.29	0.25	16.00
18	1.30	1.19	9.24	0.68	0.66	3.03	0.62	0.59	5.08	0.33	0.31	6.45
19	0.82	0.88	-6.82	0.50	0.53	-5.66	0.38	0.41	-7.32	0.26	0.32	-18.75
20	1.09	1.04	4.81	0.93	0.85	9.41	0.87	0.82	6.10	0.38	0.31	22.58
21	0.80	0.89	-10.11	0.55	0.57	-3.51	0.44	0.47	-6.38	0.18	0.22	-18.18
22	1.52	1.42	7.04	1.18	1.17	0.85	1.18	1.20	-1.67	0.38	0.40	-5.00
23	0.84	0.90	-6.67	0.51	0.56	-8.93	0.58	0.62	-6.45	0.25	0.27	-7.41
24	1.08	1.15	-6.09	0.83	0.88	-5.68	0.90	0.94	-4.26	0.36	0.32	12.50
25	1.28	1.27	0.79	0.93	0.89	4.49	0.89	0.81	9.88	0.30	0.32	-6.25
$r^2$	0.960			0.981			0.979			0.952		

## 4 结 论

由于目前 Elman 网络应用在组分预测方面并不十分广泛,在工程学方面应用较多。本文采用 Elman 网络建立了近红外光谱法同时检测苯丙氨酸、赖氨酸、酪氨酸和胱氨酸四

种氨基酸含量的定量模型,结果表明,除近红外光谱技术本身对低含量组分的检测尚存有局限性导致出现胱氨酸含量预测误差较大的情况外,其他三种氨基酸含量的预测结果良好。为近红外光谱定量分析实现多组分预测提供了一种新的方法。

## 参 考 文 献

- [1] CHENG Zhong, CHEN De-zhao(成忠,陈德钊). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(6): 1046.
- [2] HE Yong, LI Xiao-li, SHAO Yong-ni(何勇,李晓丽,邵咏妮). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(5): 850.
- [3] ZHANG Lin, ZHANG Li-ming, LI Yan, et al(张琳,张黎明,李燕,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(4): 620.
- [4] Yande Liu, Yibin Ying. Postharvest Biology and Technology, 2005, 37: 65.
- [5] Meher L C, Sagar D V, Naik S N. Renewable and Sustainable Energy Reviews, 2006, 10(3): 248.
- [6] McClure W F, Crowell B, Stanfield D L, et al. Journal of Near Infrared Spectrosc. 2002, 10(3): 177.
- [7] Inmaculada G M, Noelia A G, José Miguel G C. Talanta, 2006, 69(3): 706.
- [8] LI Yan, SUN Xiu-yun, WANG Jun-de(李燕,孙秀云,王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(6): 773.
- [9] YANG Jian-gang(杨建刚). Applied Tutorial of Artificial Neural Networks(神经网络实用教程). Hangzhou: Zhejiang University

Press(杭州: 浙江大学出版社), 2001.

- [10] Rasit Köker. *Information Sciences*, 2006, 176: 1781.
- [11] JIA Chun-yu, LIU Hong-ming. *Journal of Iron and Steel Research, International*, 2006, 13(1): 31.
- [12] Lin Chi-ming, Huang Jih-jeng, Mitsuo Gen, et al. *Applied Mathematics and Computation*, 2006, 175(2): 1139.
- [13] Rasit Köker. *Engineering Applications of Artificial Intelligence*, 2005, 18(6): 685.
- [14] CAO Yue, ZHANG Cheng-liang(曹 跃, 张成良). *Journal of Huhan University of Technology(武汉理工大学学报)*, 2006, 28(2): 49.
- [15] ZHANG Ya-ou, LÜ Tian-sheng, DU Jian-fu(张亚欧, 吕恬生, 杜建福). *Journal of System Simulation(系统仿真学报)*, 2006, 18(2): 309.
- [16] GAO Ling, SHI Jun-xian, REN Shou-xin(高 玲, 石俊仙, 任守信). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2006, 26(1): 117.
- [17] LI Chun-yan, ZHU Dong-hua(林春燕, 朱东华). *Computer Application(计算机应用)*, 2006, 26(2): 476.
- [18] CHEN Feng, HU Shang-xu, YU Meng-huai(陈 锋, 胡上序, 俞蒙槐). *Chemical Journal of Chinese Universities(高等学校化学学报)*, 2000, 21(2): 193.

## Study on the Multicomponent Quantitative Analysis Using Near Infrared Spectroscopy Based on Building Elman Model

LIU Bo-ping<sup>1, 2</sup>, QIN Hua-jun<sup>3</sup>, LUO Xiang<sup>2</sup>, CAO Shu-wen<sup>3</sup>, WANG Jun-de<sup>1\*</sup>

1. College of Chemical Engineering, Nanjing University of Science and Technology, Nanjing 200014, China

2. Analytical and Testing Center of Jiangxi Province, Nanchang 330029, China

3. Key Laboratory of Food Science of MOE, Nanchang University, Nanchang 330047, China

**Abstract** The present paper introduces an application of near infrared spectroscopy(NIRS) multi-component quantitative analysis by building a kind of recurrent network(Elman)model. Elman prediction model for phenylalanine(Phe), lysine(Lys), tyrosine(Tyr) and cystine(Cys) in 45 feedstuff samples was established with good veracity. Twelve peak value data from 3 principal components straight forward compressed from the original data by PLS were taken as inputs of Elman, while 4 predictive targets as outputs. Forty seven nerve cells were taken as hidden nodes with the lowest error compared with taking 43 and 45 nerve cells. Its training iteration times was supposed to be 1 000. Predictive correlation coefficients by the model are 0.960, 0.981, 0.979 and 0.952. The results show that Elman using in NIRS is a rapid, effective means for measuring Phe, Lys, Tyr and Cys in feedstuff powder, and can also be used in quantitative analysis of other samples.

**Keywords** Near infrared spectroscopy (NIRS); Elman; PLS; Multi-component quantitative analysis

(Received Aug. 25, 2006; accepted Nov. 28, 2006)

\* Corresponding author