

虚拟组分-人工神经网络用于中药紫外光度法中多组分的同时测定

白立飞, 张海涛, 张寒琦, 王红霞, 王洪艳*

吉林大学化学学院, 吉林 长春 130012

摘要 采用虚拟组分自修正、自拟合的方法消除了干扰组分的影响, 实现了人工神经网络(ANN)-紫外分光光度法不经分离的中药多组分浓度的同时测定。建立了包含训练网络和拟合网络的双网络 ANN 算法, 提高了 ANN 算法的自学习、自判别能力, 使复杂中草药体系多组分浓度预报的准确度大大提高。对 21 种不同来源的秦皮中秦皮甲素和秦皮乙素的含量进行了预测, 预测结果与 HPLC 法相比较, 以相对误差小于 10% 计, 预测准确率大于 90%。该法对秦皮甲素和秦皮乙素的测定精密度分别为 0.37%, 1.5%。

关键词 人工神经网络; 虚拟组分; 秦皮甲素; 秦皮乙素

中图分类号: O657.3 **文献标识码:** A **文章编号:** 1000-0593(2007)01-0126-05

引言

中草药是一个非常复杂的混合物分析体系, 组分含量和背景干扰变化大且无规律可循。目前中药的分析手段多采用高效液相色谱(HPLC)法和薄层扫描法^[1, 2], 前者仪器成本高, 分析时间长, 后者重现性不好。紫外光度法虽然操作简单、快速, 但只能测一类组分的含量, 而不能得到某单一组分的含量^[3-5]。人工神经网络(ANN)是一种模拟人脑功能的新颖信息处理系统。它通过模拟人脑神经系统处理信息的方式, 以数学网络拓扑结构为理论基础, 对于处理非线性体系有其独到之处, 具有一定的自适应, 自组织, 自学习及自动建模功能。人工神经网络的发展为混合体系多组分同时测定提供了有效的工具^[6-8]。但是人工神经网络-紫外光度法目前还主要用于简单复方制剂的分析^[9-11], 在中草药多组分测定中的应用未见报道。本文提出了一种双网络 ANN 算法, 用其处理中草药紫外吸收光谱, 以此提高网络的自学习, 自判别和自拟合能力。利用 ANN 的双网络对虚拟组分(干扰组分的总和)自修正、自拟合, 消除了干扰组分的影响, 实现了中草药主要成分不经分离的同时测定。

1 双网络 ANN 原理

传统的人工神经网络是用一组已知结果的样本作为学习集, 利用网络学习及误差修正得到满足学习样本集的一组网络权值, 由此组权值连接网络各节点, 构成了神经网络。对

于未知样品, 只要输入测量信号, 便可得到一组输出结果, 结果是否准确再无判断。由于学习样本集的数据结构与待测样品中各组分的分布特征不一定相符, 预测浓度与实际浓度有时会产生很大的偏差。这是 ANN 算法用于中药多组分同时测定的瓶颈。

本文主要是在训练网络的基础上, 增加了一个拟合网络, 提高了对未知样本预测结果的自判别、自拟合功能, 为学习样本集的选择提供了依据。计算过程包括两个网络模块: I. 训练与预测网络模块(Training and prediction network module): 此模块的数学原理及计算过程与一般的 ANN 算法相同。选取任意一个训练集(包括样本的吸光度和 n 个组分浓度), 在已知的 n 个组分浓度中再任意加入一个或 m 个虚拟组分浓度, 与学习集的吸光度信息一起输入到网络 I 中进行训练, 此时训练集的组份数为 $n+m$ 个。利用网络学习及误差修正得到满足该训练集的一组网络权值。输入一个未知样品的吸光度值, 利用训练得到的网络参数得到 $n+m$ 个浓度输出值。II. 拟合与检验网络模块(Simulation and checking network module): 将模块 I 得到的 $n+m$ 个浓度输出值作为模块 II 的输入值, 重新构建一个网络(与一般的 ANN 算法相同)进行训练拟合, 以未知样品实测吸光度信号作为目标值, 用网络输出吸光度值与实测吸光度值之间的拟合误差来判断预测结果的准确度, 若拟合误差小于 12%, 则输出预测组分的浓度。若拟合误差无法满足要求, 则返回模块 I, 改变虚拟组分含量, 重新进行训练拟合。经过 k 次改变虚拟组分, 拟合误差仍达不到要求, 且网络输出吸光度与第 $k-1$ 次网络输出吸光度值之差小于 10^{-2} 时, 网络将调用

收稿日期: 2005-11-22, 修订日期: 2006-02-22

基金项目: 国家自然科学基金项目(30371757)和吉林省科技厅项目(20030551-7)资助

作者简介: 白立飞, 女, 1978 年生, 吉林大学化学学院博士研究生

* 通讯联系人

e-mail: wang_hy@jlu.edu.cn

另一个训练样本集,重复以上操作。双网络人工神经网络算法由于增加了拟合与检验模块,有效的选择了学习样本和虚拟组分,大大提高了未知样品浓度预测的准确度,实现了中草药复杂体系的多组分同时测定。

2 实验部分

2.1 仪器与试剂

UV-1200 型紫外-可见分光光度计(北京瑞利公司), Agilent 1100 液相色谱仪, WR-C 微波消解器(北京盈安美诚科学仪器有限公司), 秦皮甲素, 秦皮乙素等对照品均购自中国药品生物制品检定所(北京), 所用乙醇为分析纯, 水为二次重蒸水, 所用秦皮为市售药材。

标准储备液的配制: 分别准确称干燥的秦皮甲素 7.8 mg 和秦皮乙素 4.9 mg, 溶于无水乙醇, 并用无水乙醇定容至 25 mL, 作为标准储备液。

吸收光谱: 取适量标准储备液, 用无水乙醇稀释至 10 mL, 在 200~500 nm 波长范围内, 间隔 0.5 nm, 扫描吸收光谱。

2.2 样品处理

将秦皮生药在 80 °C 烘干 3 h, 粉碎, 过四十目筛, 准确称取 0.5 g 放入微波消解罐中, 加入 80% 的乙醇 20 mL, 在 300 kPa 下萃取 5 min, 冷却至室温过滤, 滤液置于 50 mL 容量瓶中, 用乙醇稀释至刻度。取适量样品提取液于 10 mL 容量瓶中, 用无水乙醇稀释至刻度, 在选定的波长下测定吸光度。

3 结果与讨论

3.1 吸收光谱

从图 1 可以看出秦皮甲素和秦皮乙素的吸收光谱严重重叠, 根据吸收光谱特征, 并考虑 200 nm 附近干扰因素较多, 选取秦皮的测定波长为 320, 325, 330, 335, 340, 345, 350, 355, 360, 365 nm。

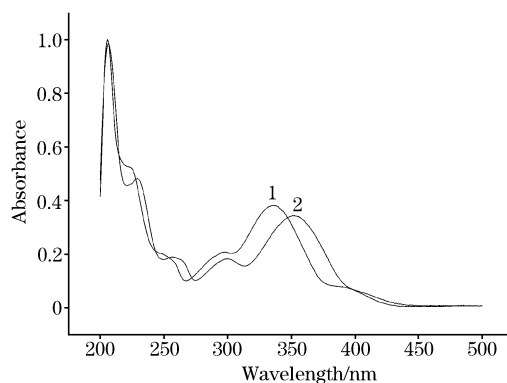


Fig. 1 Absorption spectra of aesculetin and aesculin

1: Aesculin($3.12 \mu\text{g} \cdot \text{mL}^{-1}$); 2: Aesculetin($2.94 \mu\text{g} \cdot \text{mL}^{-1}$)

3.2 虚拟组分个数对测定结果的影响

中药是一个复杂的混合物分析体系, 其化学成分复杂, 且成分含量还受生长环境、采集时间、生长年数、储藏时间和加工方法等因素的影响。从文献[12]液相色谱图可以看出, 除已知的秦皮甲素和秦皮乙素外, 其他成分的含量并不低, 且无法在测定过程中加以消除或掩蔽。因此采用紫外光度法会对测定结果带来很大影响。本文利用虚拟组分排除法, 可有效消除未知组分的干扰, 对不同来源的秦皮中秦皮甲素和秦皮乙素的浓度进行预测, 收到了较好的效果。所谓虚拟组分排除法就是将不可测的未知组分设为一个或几个虚拟的组分, 在一定范围内任意赋予它们一定的浓度值, 加入到学习样本中参与训练网络的训练和学习, 并在检验模块中利用模拟吸光度值与实测吸光度值之间的拟合误差进行检验, 以此判断虚拟组分的合理性。由于有了检验模块, 虚拟组分含量可以动态变化。为了便于考察虚拟组分的影响, 暂且将虚拟组分 R 的取值定为已知有效组分的倍数。10 个已知样本加入不同个数虚拟组分时, 拟合误差及计算结果的变化见表 1。随着虚拟组分个数的增加, 拟合误差先减小后变大, 且拟合误差较小时计算结果较准确, 因此适合的虚拟组分个数为 1~2 个, 为计算方便, 我们选择加入一个虚拟组分。

Table 1 Effects of virtual component number

Samples No.	Virtual component number 0			Virtual component number 1 ($R=2.4$)			Virtual component number 2 ($R_1=2.0, R_2=0.4$)			Virtual component number 3 ($R_1=1.5, R_2=0.5, R_3=0.4$)		
	Prediction value/%	Simulation error/%		Prediction value/%	Simulation error/%		Prediction value/%	Simulation error/%		Prediction value/%	Simulation error/%	
1 Shiyt	0.697	0.713	<9.9	0.704	0.711	<1.81	0.699	0.716	<0.11	0.705	0.714	<0.79
2 Henan	0.711	0.547	<3.36	0.672	0.562	<11.4	0.676	0.553	<0.06	0.689	0.544	<0.82
3 Baict	0.699	0.181	<1.28	0.658	0.160	<2.45	0.66	0.184	<2.0	0.663	0.185	<2.2
4 Hebei	1.01	0.31	<3.28	1.11	0.282	0.6	1.1	0.291	1.6	1.10	0.289	1.29
5 Jilin	1.22	0.221	<14.4	1.22	0.221	<6.5	1.22	0.220	<4.2	1.23	0.221	<10.4
6 Duob	0.563	0.362	<8.0	0.538	0.153	<10.5	0.54	0.308	<1.93	0.597	0.286	<2.5
7 Jilin	1.37	0.20	<4.01	1.36	0.194	<0.8	1.36	0.2	<1.44	1.37	0.198	2.35
8 Jilin	1.21	0.216	<15.0	1.19	0.216	<8.2	1.205	0.216	<8.5	1.21	0.22	<9.8
9 Rende	1.19	0.219	<11.1	1.19	0.219	<2.0	1.20	0.219	<5.3	1.19	0.22	<9.5
10 Jilin	1.23	0.217	<6.3	1.23	0.215	<3.6	1.24	0.217	<5.9	1.23	0.216	<4.0

3.4 虚拟组分含量的动态调节

从表 2 可以看出随着虚拟组分值 R 的变化, 拟合误差有着明显的改变, 计算结果亦有一定的变化, 且拟合误差较小时, 计算结果较准确, 说明利用网络拟合误差可以控制虚拟组分的量, 以获得准确的分析结果。例如样本 1, 2, 6, 8 是在 $R=2.6$ 时, 获得了小于 10% 网络拟合误差, 而此时, 2 个组分的预测误差也小于 10%, 满足分析要求。样本 5 在 $R=0$ 时即满足条件, 且随着虚拟组分的调节, 网络拟合误差

无明显变化, 说明无需加入虚拟组分。样本 3, 4 在 $R=0$ 就满足要求, 调节后反而使误差变大, 但在未知样品计算中若 $R=0$ 时满足要求, 程序会自动跳出循环, 报出结果。由表 2 可以看出, 通过虚拟组分动态调节, 以网络拟和误差小于 12%, 浓度预测结果的相对误差小于 10% 为标准, 未知组分浓度预报准确率不小于 90%。这说明我们用网络拟合误差是否满足给定的误差限可以动态调节虚拟组分的量, 从而消除干扰组分的影响。

Table 2 Effects of virtual component value on simulation error and results

Samples	1 Baicaotang			2 Tongrentang			3 Duobang			4 Jilin			
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	
Error/%													
0	21.2	26.4	7.7	-22.2	-24.1	7.4	7.24	7.8	-2.9	-9.41	3.0	8.6	
2.0	22	18.6	4.5	19	-20	5.7	16	8.9	-3.2	10	3.4	11.8	
虚拟组分 R	2.2	10.5	13.7	-1.9	23	-21	5.7	4.9	5.5	-3.2	12.8	-0.1	10.1
	2.3	25	11.6	-8.3	23	-20.1	5.1	5.1	6.0	-4.0	7.1	1.1	10.1
	2.6	9.6	8.2	-1.9	1.52	9.7	3.3	19	9.0	-4.7	6.7	3.2	13.9
Samples	5 Siyitang			6 Henan			7 Jilin			8 Anguo			
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	
Error/%													
0	1.18	0.43	1.4	28.8	43.2	-23.4	-18	33.6	-5.7	-34.5	18.5	-0.7	
2.0	-2.12	2.2	0.7	14.3	9.4	-1.4	25	36.5	-4.8	-17	6.1	-3.2	
虚拟组分 R	2.2	-1.81	1.4	1.1	11.4	8.5	-0.5	25	29.1	-1.3	-11.7	6.1	
	2.3	12.1	2.6	-4.0	11.3	8.3	-0.5	16.5	25.6	-1.2	-10.8	5.3	
	2.6	0.5	1.6	2.1	0.9	6.7	-1.1	30	25.9	-5.7	0.8	5.4	

R was the content ratio of virtual component and aesculin, E_1 was relative error of simulation absorbance and measure absorbance, E_2 and E_3 were relative error of aesculin and aesculetin respectively by ANN and HPLC method.

3.5 双网络用于训练集的选择

本方法在对未知样本进行浓度预测时, 利用检验模块(模块 2)可随机选取训练集。当通过调整虚拟组分的含量, 无法满足吸光度拟合误差(网络拟合误差)的要求时, 程序会自动返回训练模块(模块 1), 更换另一个训练集。在表 2 中样本 7 无论怎样调节虚拟组分, 计算误差均不能满足要求,

且变化值趋于稳定, 此时应更换训练集。在表 3 中 4 号样本(等同表 2 中样本 7), 通过 3 次更换学习样本, 得到了满意的结果。表 3 中所示的 6 个样本选择不同训练集时, 结果都有着明显的差异, 而且通过选择不同训练集都得到了满意的结果。因此可以根据网络拟合误差值选择合适的学习样本, 提高了计算的准确度。

Table 3 Effects of different training set on prediction results

Samples No.	Relative error/%					
	Training set 1		Training set 2		Training set 3	
	Aesculin	Aesculetin	Aesculin	Aesculetin	Aesculin	Aesculetin
1 Tongrt	9.7	-9.1	1.2	-11	1.2	-10.4
2 Baict	24.6	-1.9	10	-3.1	0.9	11.8
3 Rende	11.7	8.1	8.5	8.5	-10.8	-4.4
4 Jilin	16.5	1.8	25.6	-1.2	-0.4	3.8
5 Duobang	9.4	3.2	-0.9	9.3	-16.1	3.5
6 Henan	-16.3	-8.1	-4.4	-2.8	-11.3	4.3

3.6 实际样品预测及不同计算方法比较

购买了不同地区不同药店的 21 个秦皮药材, 按着实验方法进行处理, 并测得紫外光谱吸光度值。将 40 个训练样本组合成 8 个学习训练的子集, 每个子集有 15~20 个学习样品。实际样品预测时, 将测得的样品吸光度值输入网络, 选取任意一个学习训练子集, 通过调整虚拟组分的含量(%), 使拟合误差满足要求。若多次调整后, 误差仍无法满足要

求, 且误差变化不大, 则需更换另一个训练子集。由于不知道实际样品组分的含量, 预测时, 虚拟组分的含量可以在 1%~5% 之间任意取值。

为了考察双网络 ANN 算法的优越性, 将其与偏最小二乘(PLS)方法进行了比较。用与双网络 ANN 算法相同的训练集作为 PLS 的校准模型, 对 21 个样品进行浓度预测, 预测结果如表 4 所示。与 HPLC 测量结果相比较, 以结果相对

误差小于 10% 计, ANN 和 PLS 的预测准确率分别为 90%, 76%。PLS 拟合结果部分误差较大, 因为它不能自动选择合

适的校准集。当样本与校准集之间数据结构(待测组分及干扰组分含量分布)不相匹配时, 就会出现较大的拟合误差。

Table 4 Compare result of double-ANN with PLS

Sample No.	HPLC 法 content/%		PLS 法 Relative error/%		ANN 法 Relative error/%	
	Aesculin	Aesculetin	Aesculin	Aesculetin	Aesculin	Aesculetin
1	0.463	0.757	2.8	-1.4	2.4	0.8
2	0.375	0.598	9.2	-5.4	0.3	0.2
3	0.473	0.156	39.3	40.3	-2.9	-7.1
4	0.975	0.351	48.8	39.7	-0.82	-3.1
5	0.925	0.338	-79.7	-82.6	0	0.3
6	0.873	0.278	-8.3	-3.7	0.1	0.36
7	0.467	0.145	-6.8	2.3	-0.2	3.4
8	0.694	0.703	45.2	55.5	-6.8	9.4
9	0.619	0.565	4.29	-4.2	7.3	10.0
10	0.658	0.160	-12.6	12.8	11	-3.1
11	1.244	0.295	15.3	1.23	5.2	8.5
12	1.176	0.312	4.1	0.05	1.2	-11
13	1.291	0.2015	2.1	0.14	-5.5	-4.5
14	0.695	0.148	-0.06	-0.84	0.7	-0.9
15	1.244	0.253	-4.8	-4.5	0.8	-6.7
16	1.230	0.214	-7.4	-0.97	1.6	8.4
17	1.150	0.195	0.56	-7.1	4.3	0
18	1.35	0.248	-0.48	7.5	-4.4	-2.8
19	1.172	0.248	-1.4	-1.4	-0.9	9.3
20	1.290	0.196	0.30	-0.48	-5.0	-0.5
21	1.152	0.183	-0.13	1.6	6.1	5.5

4 结 论

从计算过程及结果可以看出, 改进的人工神经网络由于增加了一个模拟网络, 因此可通过对输入信息的反向模拟来

动态的选择学习样本集, 而且虚拟组分的动态调节有效地消除了干扰组分的影响, 从而克服了其他方法分离困难、操作时间长的缺点, 实现了不经分离对复杂中草药体系多组分浓度的同时快速测定。

参 考 文 献

- [1] Li K, Wang H J. Biomed. Chromatogr. , 2004, 18; 178.
- [2] Wang X, Li F W, Sun Q L, et al. J. Chromatogr. A, 2005, 1063; 247.
- [3] PU Xu-feng(蒲旭峰). West China Journal of Pharmaceutical Sciences(华西药科学杂志), 2001, 17(2): 4.
- [4] HUANG Qi-hua(黄启华). Chinese Traditional Patent Medicine(中成药), 2000, 22(12): 836.
- [5] LI Min-jing, YOU Jing-yan, LIU Zhong-ying, et al(李敏晶, 游景艳, 刘中英, 等). Chemical Journal of Chinese Universities(高等学校化学学报), 2004, 25(5): 850.
- [6] LIN Sheng-ling, XIE Chun-sheng, WANG Jun-de, et al(林生岭, 谢春生, 王俊德, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(6): 1135.
- [7] HE Chi-yang, SUN Yi-min, WU Gen-hua, et al(何池洋, 孙益民, 吴根华, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(5): 719.
- [8] CAO Yong-sheng, CHEN Yi-wei, ZU Jin-feng, et al(曹永生, 陈奕卫, 祖金凤, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(4): 751.
- [9] CHEN Zhen-ning(陈振宁). Chinese Journal of Analytical Chemistry(分析化学), 2001, 29(11): 1322.
- [10] YU Yu-mian, LI Zhi, LI Rong(余煜棉, 李 智, 李 荣). Journal of Instrumental Analysis(分析测试学报), 1996, 15(3): 7.
- [11] YAN Zheng-yu, JIANG Xin-min, KANG Ji-hong(严拯宇, 姜新民, 康继红). Journal of Science Analysis(分析科学学报), 1998, 14(1): 27.
- [12] WANG Xu, ZHANG Xian, et al(王 旭, 张 宪, 等). Chinese Journal of Experimental Traditional Medical Formulae(中国实验方剂学杂志), 2004, 10(2): 6.

Simultaneous Determination of Multi-Components in Chinese Herbal Medicine with UV Spectrometry by Virtual Components-Artificial Neural Network

BAI Li-fei, ZHANG Hai-tao, ZHANG Han-qi, WANG Hong-xia, WANG Hong-yan*
College of Chemistry, Jilin University, Changchun 130012, China

Abstract In the present paper, the simultaneous determination of multi-components in Chinese herbal medicine was performed by artificial neural network-UV spectrometry. The interference of other components was eliminated by self-revising and self-simulation of the virtual component. The double ANN including training and simulation network was established, and the capabilities of self-recognition and self-studying were improved. Therefore, the prediction accuracy of multicomponent content was improved greatly in the complicated Chinese herbal medicine system. The contents of aesculin and aesculetin, which were extracted from 21 Cortex Fraxinis, were predicted. Comparing the results with those of HPLC, the prediction accuracy was more than 90% within the relative errors less than 10%. The measurement precisions of aesculin and aesculetin were 0.37% and 1.5% respectively.

Keywords Artificial neural network; Virtual components; Aesculin; Aesculetin

(Received Nov. 22, 2005; accepted Feb. 22, 2006)

* Corresponding author