Simultaneous Determination of Vitamin B Complex Using Wavelet Neural Network†

YIN, Chun-Sheng*,[a,b](印春生) GUO, Wei-Min[a](郭卫民) LIU, Shu-Shen[c](刘树深)
SHEN, Yang[a](沈阳) PAN, Zhong-Xiao[a](潘忠孝) WANG, Lian-Sheng[b](王连生)
[a]*Department of Applied Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China*
[b]*School of the Environment, Nanjing University, Nanjing, Jiangsu 210093, China*
[c]*College of Bioengineering, Chongqing University, Chongqing 400044, China*

A simultaneous determination of four components of B-group vitamin, using a novel wavelet-based neural network (WNN), combined with correlation coefficient and standard deviation approach for wavelength selection, was reported in this work. Eleven representative wavelength points were selected from each original UV spectrum, based on correlation coefficients and standard deviations of the observed data. A family of wavelet basic functions built from Morlet wavelet was adopted to improve the transfer quality of output data and solve the problems of training difficultly involved in neural networks. The predicted results, with fitting correlation coefficients (R = 0.9998—0.9999) and rooted mean squares errors (RMS = 0.0578—0.1478), are satisfactory.

Keywords Wavelet neural network, UV spectra, vitamin B, wavelength selectivity

Introduction

Chemometric approaches were adopted to carry out simultaneous determination problems of multi-components by more and more analytical chemists in recent years.[1-3] Among these methods, it is well known that artificial neural network (ANN) is a high performance nonlinear analyzer and obtained extensive applications.[4-7] ANN is capable of covering more relations between the input and output data than classical regression analysis such as least squares (LS), and multivariate linear regression (MLR) without prior knowledge of the relationship between variables involved in the research system.[8-10] These leading works mostly discussed the appropriate wavelength range selection, which is very important to the simultaneous determinations of multi-components.[6,7]

Wavelet theory, a novel field which interests not only mathematicians but also scientists studying acoustics, fluid mechanics, and chemistry, involves representing general signals in terms of simpler and fixed building blocks of constant scales and positions. Unlike Fourier transformation, the wavelet transformation has dual localization both in frequency and in time, which makes it an excellent nonlinear analysis tool.[11] Recently, wavelet neural networks (WNN) have been proposed based on the theories of feed-forward neural networks and wavelet decompositions.[12-16] These theoretical studies have proven the exciting superiority of WNN over conventional back-propagation (BP) neural networks,[17] and the same efficiency of functional approximation can be obtained with WNN model in a smaller size structure.[18] Especially, WNN inherits a series of merits of the wavelet transform. It has a good identification for the singular values (particularly the sharp peak values) and the minute structural differences between spectra, and is good at processing a mutated signal, and regulates adaptively the dilation coefficient in the light of the signal characteristics. It has been found that WNN has not only

* E-mail: csyin@nju.edu.cn
Received December 1, 2000; revised and accepted May 8, 2001.
Project supported by the National Natural Science Foundation of China (No. 20075024).
†Communication of this paper was published in *Computer and Applied Chemistry*, **1999**, *16*, 378.

a good capability of modeling for the internal data set (the training set), but also a powerful prediction ability to the external data set (the prediction set).[19]

Vitamins are essential to keep the normal physiological and biochemical functions of human being and animals.[20] VB_1, VB_2, VB_6 and NAA (Nicotinamide) are commonly encountered four B-group vitamins. Determination of B-group vitamins such as VB_1, VB_2, VB_6 and NAA (Nicotinamide), *etc.*, can be carried out using UV spectrophotometry. Their ultraviolet spectra, however, are often severely overlapped. In spectral quantitative analysis, the selection of an appropriate wavelength range is significantly important to the simultaneous determination of multicomponents.[4,5,21,22] It is very convenient to directly separate them in their mixture from each other with the aid of WNN.

In the present research, 46 UV spectra of four B-group vitamins including their mixtures were measured at the range of the wavelength from 210.5 nm through 330.0 nm. WNN, combined with the correlation coefficient and the standard deviation method,[23] is employed to implement simultaneous determination of four components of the B-group vitamins in the UV spectral analysis, and the satisfying results are obtained.

Methodology

The structure of WNN employed in this study is shown in Fig. 1. The computer program is written in Visual C ++6.0 and run on a PC-II personal computer.

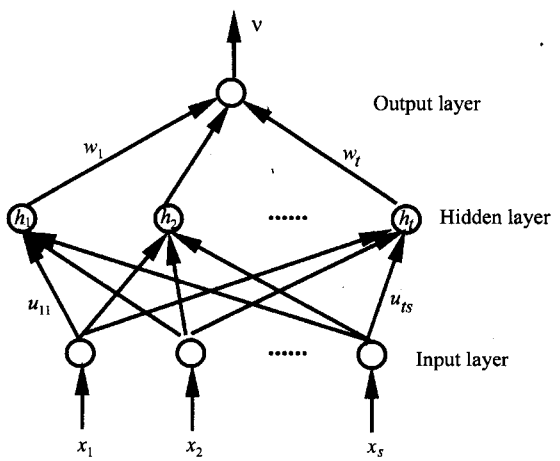

Fig. 1 Architecture of the wavelet neural network.

where x_s denotes input values, u_{ts} and w_t node connection weights in the input and hidden layers respectively and were initialized as some random values. The input signal was propagated forward through the network using the following function:

$$\nu_n = \sum_{t=1}^{N} w_t h\left(\frac{\sum_{i=1}^{s} u_{ti}x_n(i) - b_t}{a_t}\right) \quad (1)$$

where v is output values, s the number of variables, $1 \leqslant t$, $i \leqslant s$, a_t dilation parameter, b_t translation parameter, and h (h_t) is taken as a Morelet wavelet

$$h(t) = \cos(1.75t)\exp(-\frac{t^2}{2}) \quad (2)$$

Experimental

Shimadzu UV-2100 (Shimadzu, Kyoto, Japan), Beckman DU-7HS (Beckman, America) and UV-752 (Shanghai, China) systems were employed for spectral measurements. PHS-3C (Shanghai, China) acidimeter was used for regulating the pH value of the studied solution. Sample solutions of four vitamins, VB1, VB2, VB6, and NAA (Nicotinamide), which meet the demands of the pharmacopoeia of People's Republic of China,[24] were prepared by using routine methods to dilute into the reserve solution. Hydrochloric acid (HCl) and other reagents used are of analytical standard. And the distilled water used in this experiment was prepared.

According to the pharmacopoeia of People's Republic of China, 0.1 g of medicinal vitamin B_s was exactly weighed and dissolved in HOAc-NaOAc buffer solution (pH = 4.0) to prepare stock solutions. Proper quantities of stock solutions were taken to compose 46 samples of different concentration covering 0—0.7412 mol/L, 0—0.3985 mol/L, 0—1.2157 mol/L, and 0—2.0473 mol/L for VB_1, VB_2, VB_6, and NAA respectively (refer to Table 1). All the samples were measured at 240 wavelength points from 210.5 nm to 330.0 nm every 0.5 nm.

Working data set

Data selection

In this work, a correlation coefficient and standard

deviation method[23] is utilized to select 11 wavelength points from each spectrum of 240 wavelength points to process for WNN. The correlation coefficient $Corr(x, y)$ between wavelength points x and y and the standard deviation of x, $Std(x)$, are denoted respectively as follows:

$$Corr(x,y) = \frac{\sum xy - \frac{1}{n}\sum x \cdot \sum y}{[(\sum x^2 - \frac{1}{n}(\sum x^2))(\sum y^2 - \frac{1}{n}(\sum y^2))]^{1/2}} \quad (3)$$

and

$$Std(x) = sqrt(Var(x)) = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} \quad (4)$$

Here, $Std(x)$ denotes the relative contributions of x to the whole signal. If $Corr(x, y)$ is too big (approximate to 1.0), the functions of x and y are considered as iterative, and the two data need to be compressed. If $Std(x)$ is too small (approximate to 0.0), the effect of x may be negligible, and the datum can be deleted. In this work, as $Corr(x, y) > 0.98$, or $Std(x) < 0.08$, the wavelength points x and y are merged into one, and the wavelength point x is deleted. Finally, 11 wavelength points are selected from 240 wavelength points of each spectral signal. The signals retrieved with the 11 wavelength points are shown in Fig. 2, taking six representative spectra from the whole sample set as examples, and these six original spectra are shown in Fig. 3. Fig. 2 is reconstructed from the 11 wavelength points, which well represents the features of the original signals in Fig. 3.

Data pretreatment

According to literature,[25] normalizing input variables tends to make the training process better behaved by improving the numerical condition of the optimization problem and ensuring that various default values involved in initialization and termination are appropriate. Indeed, the input data in the input layer and the output data in the output layer were preprocessed by:

$$Out = \frac{In - Min}{Max - Min} \times (0.8 - 0.2) + 0.2 \quad (5)$$

where *In* is the original data, *Out* is the transformed data, *Max* and *Min* are respectively maximum and minimum absorbances in the original data. The transformed data are used as the node inputs of the input layer. Here the purpose of data preprocess is to represent all training samples in the same range, and makes the WNN training faster and reduce the chances of getting stuck in local optima. Also, weight decay can be done more conveniently with standardized inputs.[25]

Results and discussion

Six representative UV spectra of the four B-group vitamins and their mixture in the range from 210.5 through 330.0 nm are shown in Fig. 3. As seen from Fig. 3, the UV spectra overlap severely. It is difficult to determine them simultaneously without the aid of some suitable chemometrics or mathematical methods.

The calculation may be carried out according to the following steps, which are inter-verified.

Optimization of neural network parameters

A global search method, combined with partly trial-and-error procedure, is used in the optimization of WNN topological structure. Herein a gradient descent procedure has been used during training, so the learning rate (η) should be a small value (say < 0.02 in this paper). The WNN simulator will automatically search the most proper network structural parameters, when the predicted error denoted as rooted mean squares error (*RMS*) for the prediction set is the least. The change ranges of momentum factor (α), learning rate (η), node points in the hidden layer, and training number are set at 0.90—0.99, 0.001—0.02, 4—15, and 1000—50000, respectively. At last the α, η, node points in the hidden layer, and training number are determined at 0.96, 0.004, 6, and 19000 to train WNN. To optimize the neural network parameters, a split-sample validation (SSV) procedure[25] is used in the WNN

modeling, *i.e.*, 9 representative samples are randomly selected from the working data set as the test set (the external set) and the remaining 37 samples are used as the training set (the internal set). They are never used any way during the training but used to verify the reliability of the established WNN model lately. Once the WNN parameters are optimized and determined, the training for WNN is easy to do. After 19000 iterations a good predictive model is generated with the correlation coefficient $R = 0.9999$ and $RMS = 0.0668$ for the training set of 37 samples, and $R = 0.9999$ and $RMS = 0.1191$ for the prediction set of 9 samples. The calculated concentration values are given in Table 1.

Leave-one-out cross validation

The above validation method, however, reduces the amount of data available for both training and validation. And because not all the data in the working set are validated, the predicting capability of WNN for all the data in the working set can not be determined. To overcome this disadvantage, a "leave-one-out" cross validation method[25] is used in this work to verify the reliability of the established WNN model. The "leave-one-out" cross validation often works well for continuous error function such as the root mean squared error. It is generally considered as an excellent validation approach for the reliability of the obtained neural network (NN) model. Here the cross-validated values have been calculated by the WNN technique, with $R = 0.9993$ and $RMS = 0.2148$.

The calculated correlation coefficients R and rooted mean squares error RMS between observed and calculated concentrations of four B-group vitamins by WNN are listed in Table 2. According to Table 2, the similar results obtained by SSV and CV procedures indicate that WNN is a powerful tool for processing the multi-component system in this work. The results obtained by the above validation methods for 4-component system sufficiently demonstrate that the application of WNN in training and prediction are both very successful for the mixed system. This conclusion could also be resulted from the relative standard deviation (RSD), listed in Table 1. As shown in Table 1, the data of RSD for all four vitamin B_s are much low, $-3.96\% \leq RSD \leq 3.68\%$, which meet the requirement of multi-component analysis.

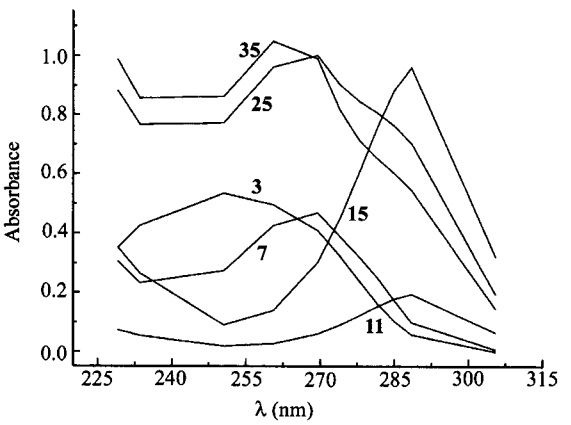

Fig. 2 Spectra retrieved from 11 wavelength points from samples **3,7,11,15,25** and **35**.

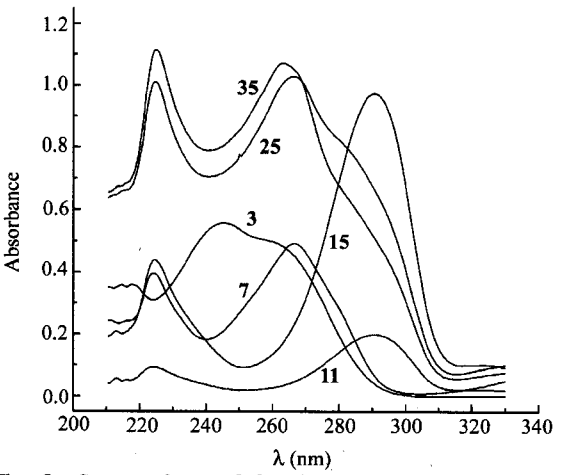

Fig. 3 Spectra of original data from **3,7,11,15,25** and **35**.

Conclusion

The WNN inherits a series of merits of the wavelet transform, such as a good recognized ability to singular values (specially, the sharp peak values) and the fine structural destinations between the observed spectra, so it is good at handling the mutated signals. All these make WNN easily handle a large number of date.

In general, using a neural network technique is time-consuming and difficult to be carried out with the data of an entire spectral signal for the simultaneous multi-component determination. In this work, the correlation coefficient and the standard deviation methods are used as indicators to select only 11 wavelength points from 240 wavelength points of each original UV spectral signal, without loss of information in the original experimental data. Correspondingly, the topological structure of WNN diminishes. The results indicate that accurate estimation and prediction can be obtained by WNN.

Table 1 Synthetic and calculated concentrations values for four B-group vitamins by WNN (Unit: mol/L × 10^{-4})

No.	Obs.[a]				Calcd.[a]				*RSD* (%)			
	VB_1	VB_2	VB_6	NAA	VB_1	VB_2	VB_6	NAA	VB_1	VB_2	VB_6	NAA
1	0.1482	0	0	0	0.1495	-0.0008	0.0007	-0.0008	0.84			
2	0.2965	0	0	0	0.2959	-0.0007	0.0001	0.0017	-0.21			
3	0.4447	0	0	0	0.4480	-0.0004	0.0017	0.0026	0.73			
4	0.7412	0	0	0	0.7401	0.0013	0.0012	-0.0038	-0.16			
5	0	0.0797	0	0	-0.0041	0.0793	-0.0010	-0.0064		-0.47		
6	0	0.1594	0	0	-0.0009	0.1596	0.0008	0.0038		0.08		
7	0	0.2391	0	0	-0.0013	0.2409	-0.0009	-0.0001		0.72		
8	0	0.3985	0	0	0.0020	0.3969	-0.0006	0.0019		-0.40		
9	0	0	0.2431	0	0.0008	-0.0007	0.2427	0.0073			-0.20	
10	0	0	0.4863	0	-0.0011	-0.0001	0.4830	0.0002			-0.67	
11	0	0	0.7294	0	0.0013	-0.0010	0.7348	0.0064			0.73	
12	0	0	1.2157	0	-0.0001	0.0005	1.2133	-0.0015			-0.20	
13	0	0	0	0.4095	0.0011	0.0007	0.0019	0.4245				3.68
14	0	0	0	0.8189	0.0032	-0.0004	0.0015	0.8067				-1.49
15	0	0	0	1.2284	0.0004	0.0013	0.0012	1.2227				-0.47
16	0	0	0	2.0473	0.0002	0.0022	-0.0023	2.0359				-0.56
17	0.1482	0.0797	0.2431	0.4095	0.1437	0.0786	0.2377	0.3976	-3.08	-1.33	-2.24	-0.290
18	0.1482	0.1594	0.4863	0.8189	0.1457	0.1591	0.4873	0.8161	-1.74	-0.22	0.20	-0.35
19	0.1482	0.2391	0.7294	1.2284	0.1484	0.2375	0.7319	1.2268	0.08	-0.67	0.33	-0.13
20	0.2965	0.1594	0.7294	0.4095	0.2952	0.1618	0.7304	0.4030	-0.44	1.50	0.13	-1.58
21	0.2965	0.2391	0.2431	0.8189	0.2922	0.2404	0.2403	0.8160	-1.46	0.54	-1.16	-0.36
22	0.4447	0.0797	0.7294	0.8189	0.4447	0.0800	0.7255	0.8096	0.00	0.40	-0.53	-1.14
23	0.4447	0.1594	0.2431	1.2284	0.4427	0.1580	0.2394	1.2358	-0.47	-0.90	-1.54	0.60
24	0.1482	0.0983	0.4231	1.0237	0.1488	0.0981	0.4202	1.0204	0.34	-0.16	-0.67	-0.32
25	0.1838	0.1382	0.6662	0.8189	0.1852	0.1388	0.6662	0.8111	0.74	0.48	0.00	-0.96
26	0.2224	0.1780	0.3647	0.6142	0.2232	0.1798	0.3617	0.6061	0.36	1.00	-0.84	-1.32
27	0.2580	0.2179	0.6079	0.4095	0.2604	0.2189	0.6142	0.4105	0.93	0.49	1.04	0.26
28	0.3321	0.1196	0.5446	0.9172	0.3330	0.1199	0.5446	0.9180	0.27	0.27	0.00	0.09
29	0.3706	0.1594	0.2431	0.7125	0.3745	0.1603	0.2433	0.7052	1.04	0.58	0.08	-1.02
30	0.4062	0.1993	0.4863	0.5077	0.4101	0.2006	0.4887	0.5083	0.95	0.67	0.50	0.11
31	0.4447	0.2391	0.7294	1.2284	0.4439	0.2380	0.7328	1.2309	-0.20	-0.48	0.47	0.20
32	0.1482	0.2391	0.2431	1.2284	0.1532	0.2394	0.2409	1.2309	3.34	0.13	-0.92	0.20
33	0.4447	0.2391	0.2431	0.4095	0.4433	0.2375	0.2424	0.4079	-0.33	-0.69	-0.32	-0.38
34	0.1482	0.0797	0.7294	1.2284	0.1471	0.0790	0.7294	1.2382	-0.78	-0.90	0.00	0.80
35	0.4447	0.0797	0.7294	0.4095	0.4453	0.0785	0.7270	0.4136	0.13	-1.57	-0.33	1.02
36	0.2965	0.1594	0.4863	0.8189	0.2931	0.1592	0.4857	0.8353	-1.15	-0.15	-0.13	2.00
37	0.1482	0.0664	0.0160	1.3922	0.1468	0.0638	0.0232	1.4127	-0.96	-3.96	44.5	1.47
38	0.5930	0	0	0	0.5927	0.0007	0.0008	0.0054	-0.05			
39	0	0.3188	0	0	0.0039	0.3199	0.0015	-0.0011		0.33		
40	0	0	0.9726	0	-0.0006	0.0001	0.9784	0.0029			0.60	
41	0	0	0	1.6379	-0.0091	0.0025	-0.0171	1.6158				-1.35
42	0.2965	0.0797	0.4863	1.2284	0.2943	0.0795	0.4820	1.2309	-0.73	-0.30	-0.89	0.20
43	0.4447	0.2391	0.4863	0.4095	0.4450	0.2387	0.4897	0.4062	0.07	-0.17	0.70	-0.80
44	0.2965	0.0797	0.3015	1.1219	0.3027	0.0788	0.2974	1.1236	2.10	-1.10	-1.35	0.15
45	0.4447	0.0797	0.2431	1.2284	0.4436	0.0785	0.2403	1.2464	-0.27	-1.67	-1.18	1.47
46	0.1482	0.2391	0.7294	0.4095	0.1535	0.2358	0.7372	0.4193	3.56	-1.38	1.07	2.40

[a] Experimental and predicted values; Note: Samples from 38 through 46 make up the prediction set.

Table 2 *R* and *RMS* between observed and calculated concentrations of four B-group vitamins

Sort		VB_1	VB_2	VB_6	NNA
SSV*	*R*	0.9998	0.9999	0.9999	0.9999
	RMS	0.1478	0.0578	0.1444	0.1264
CV*	*R*	0.9997	0.9991	0.9992	0.9994
	RMS	0.1739	0.1671	0.2723	0.2459

* SSV, CV: split-sample validation and cross validation.

References

1 Pan, Z. X.; Pan, D. J.; Sun, P. Y.; Zhang, M. S.; Zuberbuhler, A. D.; Jung, B. *Spectrochim. Acta Part A* **1997**, *53*, 1629.

2 Brown, C. W.; Lynch, P. F.; Obremskl, R. J.; Lavery, D. S. *Anal. Chem.* **1982**, *54*, 1472.

3 Mark, H. *Appl. Spectrosc.* **1988**, *42*, 1427.

4 Lucasius, C. B.; Beckers, M. L. M.; Kateman, G. *Anal. Chim. Acta.* **1994**, *286*, 135.
5 Lucasius, C. B.; Dane, A. D.; Kateman, G. *Anal. Chim. Acta.* **1993**, *82*, 647.
6 Garrido-French, A.; Jouan-Rimbauf, D.; Massart, D. L.; Kuttatharmmakul, S.; Martinez-Galera, M.; Martinez-Vidal, J. L. *Analyst* **1995**, *120*, 2787.
7 Costadinnova, L.; Nedeltcheva, T. *Analyst* **1995**, *120*, 2217.
8 Gasteiger, J.; Zupan. J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 503.
9 Burns, S. A.; Whitesides, G. M. *Chem. Rev.* **1993**, *93*, 2583.
10 Sumpter, B.; Getino, C.; Noid, D. *Annu. Rev. Phys. Chem.* **1994**, *45*, 439.
11 Mallat, S. G. *IEEE T. Pattern Anal.* **1989**, *11*, 674.
12 Zhang, Q. H.; Benveniste, A. *IEEE T. Neural Net.* **1992**, *23*, 889.
13 Pati, Y. C.; Krishnaprasad, P. S. *IEEE T. Neural Net.* **1993**, *4*, 73.
14 Zhang, J.; Walter, G. G.; Miao, Y.; Lee W. N. *IEEE T. Signal Proces.* **1995**, *43*, 1485.
15 Qian, W.; Clarke, L. P. *P. IEEE.* **1996**, *84*, 1458.
16 Zhang, Q. H. *IEEE T. Neural Net.* **1997**, *8*, 227.
17 Liu, W.; Li, J. P.; Xiong, J. H.; Pan, Z. X.; Zhang, M. S. *Chin. Sci. Bull.* **1997**, *42*, 822.
18 Pati, Y. C.; Krishnaprasad, P. S. *IEEE T. Neural Net.* **1993**, *4*, 73.
19 Liu, L.; Guo, Q. X. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 133.
20 Zhou, Q. Z. *Acta Pharm. Sin.* **1987**, *22*, 896.
21 Brown, C.W.; Lynch, P. F.; Obremskl, R. J.; Lavery, D. S. *Anal. Chem.* **1982**, *54*, 1472.
22 Mark, H. *Appl. Spectrosc.* **1988**, *42*, 1427.
23 Chen, X. L. *Probability Theory and Statistics*, Press of University of Science and Technology of China, Hefei, **1996** (in Chinese).
24 Pharmacopoeia Committee of Health Ministry of the People's Republic of China, *The Pharmacopoeia of the People's Republic of China*, 2nd section, Chemical Industry Press, Beijing, **1995** (in Chinese).
25 http://www.faqs.org/faqs/ai-faq/neural-nets

(E200012262 SONG, J.P.; LING, J.)