Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 65 (2005) 118–128
Talanta
www.elsevier.com/locate/talanta

Artificial neural networks based on genetic input selection for quantification in overlapped capillary electrophoresis peaks

Yaxiong Zhang [a], Hua Li [a,*], Aixia Hou [a], Josef Havel [b]

[a] *Institute of Analytical Science, Northwest University, Xi'an 710069, China*
[b] *Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlárská 2, 611 37 Brno, Czech Republic*

Received 23 March 2004; received in revised form 12 May 2004; accepted 20 May 2004
Available online 28 July 2004

Abstract

The application of multilayer perceptron artificial neural networks (MLP ANN) based on genetic input selection for quantification of the unresolved peaks in micellar electrokinetic capillary chromatography (MECC) is reported. An optimization strategy for genetic input selection was also proposed. When the corresponding CE peaks cannot be resolved completely only by separation techniques, MLP ANN based on genetic input selection can be a suitable tool to resolve the problem. Both the spectra and the electrophoretograms of the unseparated analytes were used as the multivariate input data. The two kinds of the data were suitable for quantification of overlapped CE peaks by MLP ANN based on genetic input selection. The study also shows that the applying of genetic input selection in MLP ANN can improve the precision of quantification in both completely and partially overlapped CE peaks to some extent.

Keywords: Micellar electrokinetic capillary chromatography; Overlapped peaks; Artificial neural networks; Genetic input selection; Quantification

1. Introduction

Capillary electrophoresis (CE) has been a promising separation technique in the field of analytical chemistry [1]. The analytes studied in CE should be resolved completely for quantitative analysis. But sometimes the completely electrophoretic separation is difficult or impossible. In these cases, resolution based on chemometrics is necessary and suitable.

To improve the quantification of overlapped peaks in CE and other separation techniques, different chemometrics methods have been applied. Partial least squares regression (PLS) has been proven to be a powerful tool to improve the quantification in poorly resolved peaks in different separation techniques [2–4], multilinear regression (MLR) has also been used successfully in quantification in overlapped peaks [5]. Kalman filtering [6] and time domain derivative chromatograms [7] can also be employed to improve the resolution of the unresolved peaks. The chromatographic profiles recorded at given wavelength or the spectrochromatograms taken at the maximum of the peaks have been employed as multivariate data [8]. Artificial neural networks (ANN) have been increasingly applied in different separation techniques recently. ANN has shown its unique merits in capillary zone electrophoresis (CZE) [9–11], MECC [12] and high performance liquid chromatography (HPLC) [13,14]. Moreover, ANN can also be used to improve the precision of analysis in CE [15,16]. In the field of quantification in unresolved peaks [8] and chiral separation in CE [17], ANN can also give better calculated results. ANN is called "soft modeling" tool, without knowing or establishing an explicitly defined mathematical model [18], which makes it be applied widely in chemistry and separation science. In the authors' laboratory, other mathematical models based on chemometrics have been developed to resolve the overlapped peaks in CE [19,20].

Although ANN can be used for multicomponent analysis [21], its multivariate input data (corresponding spectra of the selected wavelengths or spectrochromatograms) may cause "over fitting" of the trained networks [22]. In order to avoid "over fitting" and improve the calculated results of ANN, it is necessary to select the suitable input parameters from

* Corresponding author. Fax: +86 29 8830 3448.
E-mail address: huali@nwu.edu.cn (H. Li).

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.05.050

all the input data available. Generally speaking, the performance of a network can be improved by reducing the number of inputs, even sometimes at the cost of losing some input information. The problem can be resolved by two possible approaches. One is to retain all the input parameters, but compact them into a smaller number in order to avoid losing any important input information. The other is to identify the input parameters that are not important to the performance of the networks, and then remove the less important input parameters from the input data. The second approach can be performed by genetic algorithm (GA), which is an optimization strategy to search for binary strings efficiently. GA is a chemometrical technique based on Darwin's evolution rule to simulate the evolution of a population. GA has been used in wavelength selection for multicomponent determination by PLS [23], in selecting variables for a PLS regression model [24], and to select some predictor variables for PLS model in differential pulse polarography [25]. GA was also applied in principal component selection for principal component regression (PCR) [26–28]. Moreover, the combination of GA and ANN to improve the calibration of the vapor concentrations of three analytes in ternary mixtures has been reported [29]. Despite the applications of GA to optimize the ANN models [30,31], as far as we know, genetic input selection for ANN applied in CE of overlapped peaks has not been reported. Therefore, in this paper, we have applied ANN approach with genetic input selection to achieve the quantitative analysis of overlapped CE peaks.

2. Theory

2.1. Artificial neural networks

ANN is a kind of information processing chemometrical technique. It simulates some properties of human brain. ANN is often applied in the field of regression or classification. The theory of ANN has been described thoroughly in several papers [32–34]. Although different algorithms of ANN of multilayer perceptron (MLP) have been developed, conjugate gradient descent (CGD) algorithm [35] is one of the most widely used. In this paper, ANN of MLP based on CGD algorithm was applied to model the relationship between the concentrations of the unresolved analytes and the corresponding input parameters. The theory of CGD ANN of MLP is given briefly here. ANN of MLP is composed of some logic units and the connection weights between the units. ANN of MLP is divided into three levels in order to understand the process of information processing. The three levels are called input layer, hidden layer and output layer, respectively. There are some logic units in each layer. The logic units are the basic information-processing unit in MLP ANN. Linear post-synaptic potential (PSP) function and logistic activation function were applied in MLP ANN in this paper. The sum-squared error function monitoring the training process of MLP ANN was used.

The initial search direction of CGD is given by:

$$d_0 = -g_0 \tag{1}$$

Subsequently, the search direction is updated using the Polak–Rebiere formula [36]:

$$d_{j+1} = -g_{j+1} + \beta_j d_j \tag{2}$$

$$\beta_j = \frac{g_{j+1}^{\mathrm{T}}(g_{j+1} - g_j)}{g_j^{\mathrm{T}} g_j} \tag{3}$$

2.2. Genetic algorithm

Genetic algorithm is a global optimization algorithm. It searches for optimal binary strings by operating initially random population of the binary strings. The search process is composed of artificial mutation, crossover and selection. GA is a kind of simulation of the evolution of a population in nature based on Darwin's evolution law [37].

In the algorithm, each binary string represents a mask, which determines what kind of input parameters should be employed to construct the neural networks. In a binary string, "0" indicates that the input parameter should be abandoned, and "1" means that the parameter is necessary for the neural networks. For example, for nine original input parameters represented by a binary mask 100100001, the first, the fourth, and the ninth input parameter should be regarded as the suitable input data to corresponding neural networks, while the other parameters should be removed from the input data set.

GA randomly generates a population of such binary strings, and then simulates the natural selection process to search for the superior binary strings by mutation, crossover and selection. The superior strings are bred together and form a population of new generation. After the processing for several new generations, successively better strings are produced. Generally, the best binary mask of the last generation during the GA processing is regarded as the solution of the input parameters for the neural networks.

In the GA for input selection of neural networks, each mask represented by a binary string was used to construct a training set. Generalized regression neural networks (GRNN) were applied to test the training set. GRNN can be trained extremely quickly, which makes it possible to perform a large number of evaluations required by GA. None-linear functions can be modeled accurately using GRNN. Moreover, GRNN is comparatively sensitive to the inclusion of irrelevant input variables. Therefore, GRNN was served as the fitness function of GA in this paper. The standard Holland GA [38] was adopted in this work.

2.3. Optimization strategy

In this paper, ANN adopted a high number of input variables to predict the concentrations of the corresponding analytes. For this approach, problems of data analysis may be

caused. If the number of weights of MLP ANN exceeds the number of samples for the training of MLP ANN to some extent, "over fitting" may occur [22]. In the case of a high number of input variables, irrelevant, redundant, and noisy variables might be included in the input data set. On the other hand, meaningful variables could be hidden [39]. For high number of input variables, the probability of chance correlation increases [40]. At last, a high number of input variables may prevent MLP ANN from finding optimal models [41]. Therefore, genetic input selection is necessary in order to improve the predicted results of MLP ANN.

In this paper, the optimization strategy includes two steps. The first step is based on parallel GA runs for the same data set. Many variable selection strategies perform the variable selection procedure only by one single run of GA. For the single run approach, some drawbacks will be caused despite its shorter computation time. In GA, the binary strings of the initial population are generated randomly, so different runs of GA often result in similar but not identical combinations of the input variables [42]. GA can select irrelevant variables by chance correlation even validation procedure was employed [43]. In literature [43], it is also reported that different runs of GA often result in similar but not the same variables selected due to collinearity of data. That is to say, one single run of genetic input selection is not a robust approach. The input variables that appeared in the final generation of each GA run were collected in the process of all the GA runs. The variables then were ranked according to the frequency of their appearance in the last generations of all the GA runs. In the second step, the input variables were added to MLP ANN according to their frequencies selected by GA in a stepwise procedure, i.e. the input variable(s) with the highest frequency by GA selection procedure was (were) regarded as the most suitable input parameter(s) for MLP ANN model and was (were) added to the MLP ANN first, then the second highest, the third highest, and so on. The prediction error of MLP ANN model by the input data set was compared to that of a previous MLP ANN model. If the prediction error was descending, it was reasonable that the new adopted input variable(s) was (were) acceptable for the MLP ANN model. Then, the stepwise procedure was repeated to add the next important variable to MLP ANN input data set until the prediction error was not improved significantly. The prediction error of ANN was calculated by the following equation:

$$\text{Prediction error} = \frac{\sqrt{\sum_{p=1}^{P}(i_{\text{ptrue}} - i_{\text{ppred}})^2}}{\sqrt{\sum_{p=1}^{P}(i_{\text{ptrue}})^2}} \quad (4)$$

where "i_{ptrue}" and "i_{ppred}" are the target and predicted value of one sample of MLP ANN, respectively, and "P" is the number of all the samples to train MLP ANN model.

Finally, when all the suitable input variables were selected, the architecture of MLP ANN was experimentally determined by Trajan Automatic Network Designer based on simulated annealing method [44] and conjugate gradient descent approach [35].

3. Experimental

3.1. Reagents

Dibazolum, vitamin B1, promethazine hydrochloride, and chloroquine phosphate were purchased from Baoji Medicine Corporation (Shaanxi, China). Compound reserpine tablets were from Changzhou Medicine Corporation (Jiang Su, China). Sodium dodecyl sulphate (SDS) was from Hongyan (Tianjin, China). The mixture solution of dibazolum and vitamin B1 and that of promethazine hydrochloride and chloroquine phosphate were prepared using corresponding volumes of acetone as solvent, and then the two mixtures were prepared to be experimental samples by mixing appropriate volumes of twice distilled water. The background electrolytes (BGE) were composed of appropriate amount of KH_2PO_4, $Na_2B_4O_7{\cdot}10H_2O$, and SDS. The concentration of SDS in the BGE system was high enough to form pseudo stationary phase, which assured that the mechanism of the electrophoretic separation was based on MECC. The pH of the separation system was adjusted to 9.00 with appropriate volume of dilute hydrochloric acid. All the chemicals used in the experimental were analytical reagent grade purity.

3.2. Apparatus and conditions

All the electrophoretic separations were performed on a Beckman Coulter P/ACE 5500 CE instrument equipped with a photo diode array detector (DAD). The range of the scanning wavelength was from 190 to 600 nm. The uncoated fused-silica capillary of Yongnian optic fiber plant (Hebei, China) was used. The total length of the capillary is 57 cm, and its length to the detector is 50 cm. The inner diameter of the capillary is 75 μm. The MECC experiments were performed under a constant voltage of 20 kV in normal polarity mode, i.e. the detector was at the negative electrode (outlet) side at the end of the capillary. During all the separation performance, the temperature in the capillary was kept constantly at 20 °C. The sample solutions were injected to the capillary column by high-pressure approach within 6 s. The electrophoretograms were detected at 195 nm. The spectra were recorded from 195 to 365 nm every 5 nm.

3.3. Software and data processing

All the calculations of MLP ANN and the simulations of GA input selections were carried out using Trajan software version 3.0 (Durham, UK). P/ACE workstation software version 1.2 was applied to collect and evaluate the electrophoretic data on a Pentium III personal computer. The original electrophoretic data of the selected wavelength were converted to ASCII files in order to perform further mathematical processing by the P/ACE workstation software. The

data points of maximum absorbance in electrophoretograms were searched for by a library function of MATLAB 6.5.

4. Results and discussion

Promethazine hydrochloride, chloroquine phosphate, vitamin B1 and dibazolum are important components in compound reserpine tablets. Under the separation conditions in this paper, the CE peaks of the other components with ultra violet ray absorbability in compound reserpine tablets can be resolved completely. However, vitamin B1 and dibazolum can only be partially separated, while promethazine hydrochloride and chloroquine phosphate cannot obtain their separation at all. So quantification based on chemometrics is necessary for the four components. The electrophoretogram of compound reserpine tablets is shown in Fig. 1.

4.1. Analysis of mixture of promethazine hydrochloride and chloroquine phosphate in completely overlapped CE peaks

The electrophoretogram and the spectrum of promethazine hydrochloride and chloroquine phosphate of one sample under the MECC separation conditions mentioned above are shown in Fig. 2. It is shown that the CE peaks of the two compounds are completely overlapped.

4.1.1. The concentrations of the mixed solution series diluted from one original concentration

In this case, all the concentrations of the mixed solutions were obtained by diluting one mixture. The spectra and electrophoretograms of the binary mixtures were applied to

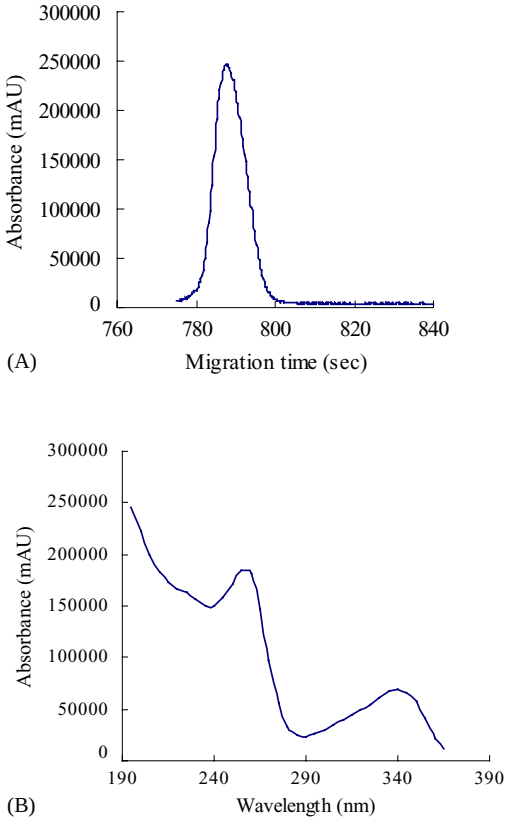

Fig. 2. (A) Electrophoretogram of the mixture of promethazine hydrochloride and chloroquine phosphate at 195 nm. (B) Spectrum of the mixture of promethazine hydrochloride and chloroquine phosphate from 195 to 365 nm.

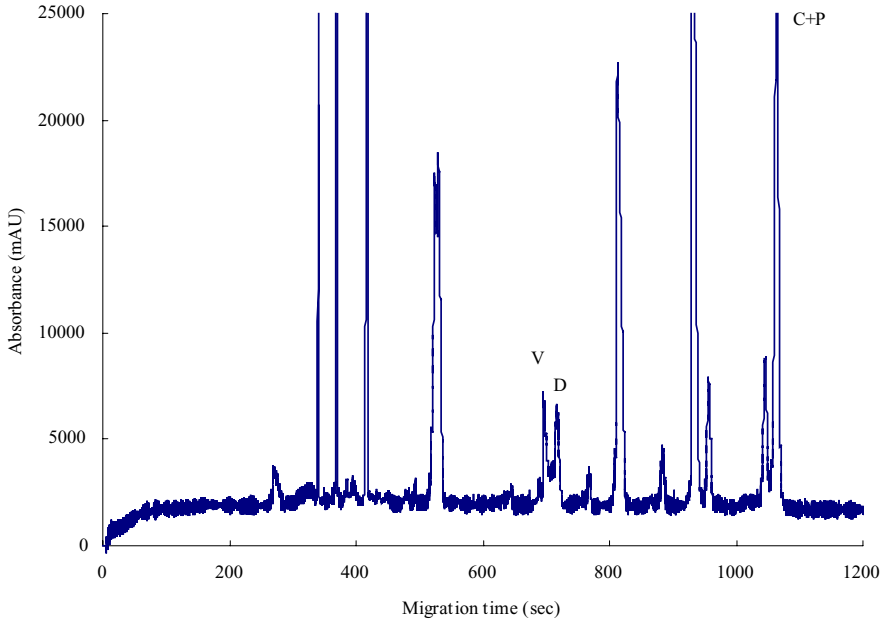

Fig. 1. The electrophoretogram of compound reserpine tablets at 195 nm. V: vitamin B1; D: dibazolum; C: chloroquine phosphate; P: promethazine hydrochloride.

predict the concentrations of the two components in the mixed solutions by genetic input selection and MLP ANN.

4.1.1.1. Case I: Analysis of the binary mixtures by spectra. Using DAD, three-dimensional electrophoretograms can be recorded in CE. Therefore, quantification can be achieved using either the electrophoretograms at a given wavelength or spectra at some selected wavelengths. All the spectra data were collected at the maximum of the CE peaks, and the absorbencies of 35 wavelengths (from 195 to 365 nm) formed one input vector for each sample. In order to train MLP ANN, 18 samples were included in the data set. Three of all the samples selected randomly were used as test set, and the others were training samples. So the training process of MLP ANN can be monitored to avoid "over training". Under the training conditions, the Trajan software used to perform MLP ANN training can search for the best iterative times automatically. So "over training" can be avoided conveniently. For more than 30 input parameters of MLP ANN should be investigated, genetic input selection strategy was performed. The corresponding performance parameters for the genetic input selection strategy are given in Table 1. The frequency of the variables not being selected in the last generations during 10 parallel GA runs is shown in Fig. 3. It is clear that the input variables suggested by GA are nearly the same. So in the second step of optimization strategy, the input variables can be selected easily. The prediction errors of MLP ANN versus the number of input variables not being used are shown in Fig. 4. It is reasonable that 28 spectra data points were the selected input variables to MLP ANN. According to Trajan automatic network designer in 1000 iteration times with unit penalty 0.01, 28:2:2 MLP ANN was constructed. The average prediction errors of promethazine hydrochloride and chloroquine phosphate and the standard deviations (S.D.) of the prediction errors during 10 parallel runs of the designed MLP ANN were calculated, respectively. Moreover, the performance of the automatically designed 35:1:2 MLP ANN under the same operating conditions as above was also investigated without the optimization strategy. This time, the average prediction errors of promethazine hydrochloride and chloroquine phosphate in 10 MLP ANN parallel runs were also investigated, so were done the S.D. of the prediction errors. All the calculated results of the two MLP ANN, the corresponding performance parameters and the structures of the designed MLP ANN are shown in Table 1. From the results in the table, we can conclude that the genetic input selection improved the predictive ability of neural networks under the investigated conditions.

4.1.1.2. Case II: Analysis of the binary mixtures by electrophoretograms. The data points of electrophoretogram measured at 195 nm were collected from corresponding ASCII files. At first, the maximum absorbance data point of the electrophoretogram was searched for by a library function of MATLAB (version 6.5). Then, 22 data points were collected symmetrically from the two sides of the maximum data point to form the input data of electrophoretogram. Twenty experimental samples were constructed to train the MLP ANN, in which four samples selected randomly were served as test set, and the others were training samples. The performance parameters for the genetic input selection process are given in Table 1. But this time, all of the 22 input variables should be selected to train MLP ANN according to the optimization strategy. The 22:1:2 architecture of MLP ANN used to calculate the predicted results was designed automatically in 100 iteration times with the unit penalty 0.01. The average prediction errors of promethazine

Table 1
The average prediction errors of promethazine hydrochloride and chloroquine phosphate in Section 4.1.1

Compounds	Case I (with GA)	Case I (without GA)	Case II (with GA)	Case II (without GA)
Promethazine hydrochloride[a] (%)	10.32	25.88	12.73	12.73
n[b]	10	10	10	10
S.D.[c] (%)	0.59	1.87	0.21	0.21
Chloroquine phosphate[a] (%)	10.34	25.84	12.72	12.72
n[b]	10	10	10	10
S.D.[c] (%)	0.55	1.93	0.16	0.16
MLP ANN	28:2:2	35:1:2	22:1:2	22:1:2
Generation of GA	1000	–	1000	–
Population of GA	1000	–	1000	–
Mutation rate	1	–	1	–
Crossover rate	0.3	–	0.3	–
Unit penalty factor of GA	0.01	–	0.01	–
Smoothing factor	0.3	–	0.3	–
Iteration times of design	1000	1000	100	100
Unit penalty of design	0.01	0.01	0.01	0.01

[a] The average prediction error of the corresponding compound.
[b] The number of the parallel runs of the designed MLP ANN.
[c] The standard deviation.

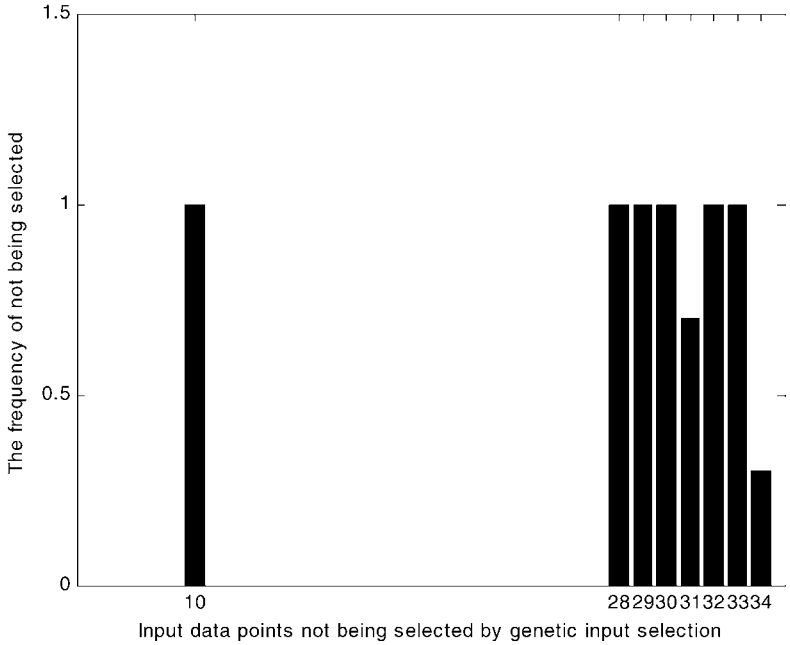

Fig. 3. The frequency of the data points not being selected by genetic input selection in Section 4.1.1.1.

hydrochloride and chloroquine phosphate in 10 parallel MLP ANN runs, the corresponding performance parameters, the S.D. of the prediction errors and the structures of the designed MLP ANN are also shown in Table 1.

4.1.2. The concentrations of the mixed solution series with a constant total concentration

In this case, the summation of the concentrations of the two analytes in each mixed solution was constant. But the concentrations of the same component in different mixed solutions were different. The spectra and electrophoretograms of the binary mixtures were applied to predict the concentrations of the two analytes in the mixed solutions by genetic input selection and MLP ANN.

4.1.2.1. Case I: Analysis of the binary mixtures by spectra. As in previous case of analysis of the samples diluted from one original solution, 35 spectra data at the maximum of the overlapped CE peaks of promethazine hydrochloride and chloroquine phosphate were employed to predict the concentrations of the corresponding components. Twenty samples were constructed to train the MLP neural networks,

Table 2
The average prediction errors of promethazine hydrochloride and chloroquine phosphate in Section 4.1.2

Compounds	Case I (with GA)	Case I (without GA)	Case II (with GA)	Case II (without GA)
Promethazine hydrochloride[a] (%)	4.95	5.26	7.84	7.84
n[b]	10	10	10	10
S.D.[c] (%)	0.20	0.41	0.29	0.29
Chloroquine phosphate[a] (%)	4.93	5.24	7.82	7.82
n[b]	10	10	10	10
S.D.[c] (%)	0.21	0.40	0.30	0.30
MLP ANN	32:2:2	35:2:2	22:1:2	22:1:2
Generation of GA	100	–	1000	–
Population of GA	100	–	1000	–
Mutation rate	1	–	1	–
Crossover rate	0.3	–	0.3	–
Unit penalty factor of GA	0	–	0.01	–
Smoothing factor	0.3	–	0.3	–
Iteration times of design	1000	1000	100	100
Unit penalty of design	0.01	0.01	0.01	0.01

[a] The average prediction error of the corresponding compound.
[b] The number of the parallel runs of the designed MLP ANN.
[c] The standard deviation.

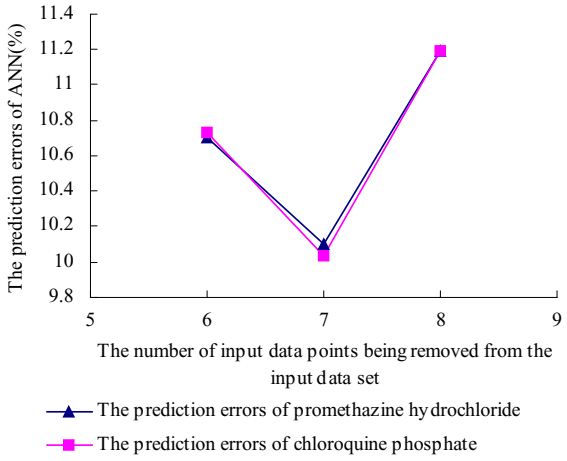

Fig. 4. The optimization of the input variables in Section 4.1.1.1.

in which four samples selected randomly were served as test set, and the others were training samples. The performance parameters for the genetic input selection are listed in Table 2. The genetic input selection procedure suggested that three spectra data points should be abandoned in constructing the training data set of MLP ANN. Moreover, 20 parallel GA runs gave the same suggestions. Then, it was reasonable that the second step of the optimization strategy can be omitted. For the 32 input variables of spectra data, a 32:2:2 MLP ANN was designed by Trajan automatic network designer with 1000 iteration times and the unit penalty 0.01. However, when no optimization strategy for input variables was applied, a 35:2:2 architecture of MLP ANN was designed automatically using the same parameters as above. All the calculated results of the two automatically designed MLP ANN in 10 parallel runs, the corresponding performance parameters, and the structures of the designed MLP ANN are included in Table 2. The calculated results of this part also indicate that MLP ANN with the optimization strategy can give better-predicted results by fewer input variables.

4.1.2.2. Case II: Analysis of the binary mixtures by electrophoretograms. When the suitable data points in the ASCII file of the electrophoretogram were collected to form an input data set, 22 channels of a data window were used as inputs to MLP ANN. Twenty samples were applied to train the MLP ANN, in which four samples selected randomly were served as test set, and the others were training samples. The optimization strategy was also applied to the input variable selection with the identical performing parameters as those being used in Section 4.1.1.2. But the optimization strategy still suggested that no input variables should be removed from the input data set. During 100 iteration times, a 22:1:2 MLP ANN was designed with the unit penalty 0.01. The calculated results by the MLP ANN of this section, the corresponding performance parameters, and the structures of the designed MLP ANN were also shown in Table 2.

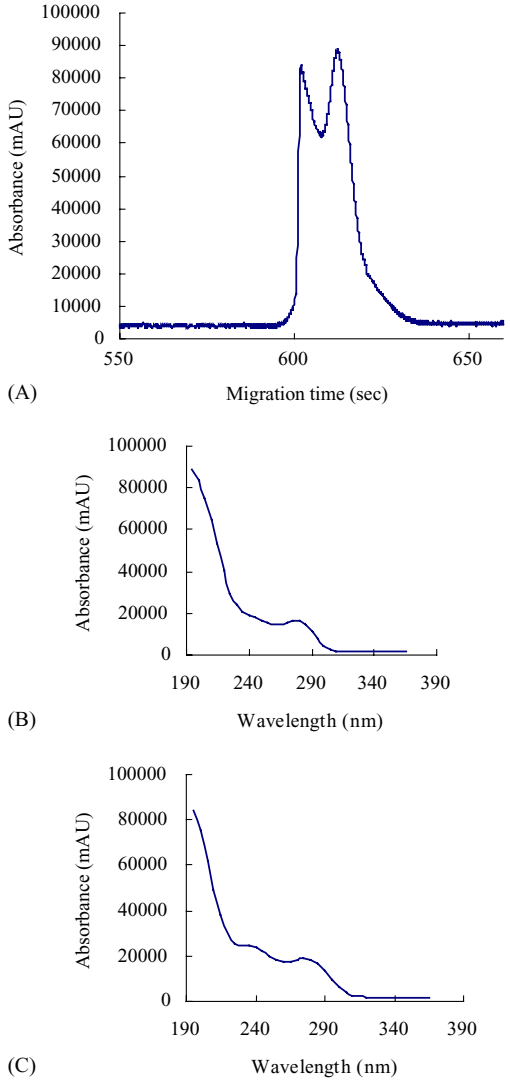

Fig. 5. (A) Electrophoretogram of the mixture of vitamin B1 and dibazolum at 195 nm. (B) The spectrum of dibazolum from 195 to 365 nm. (C) The spectrum of vitamin B1 from 195 to 365 nm.

4.2. Analysis of mixture of vitamin B1 and dibazolum in partially overlapped CE peaks

Vitamin B1 and dibazolum were partially separated under the performance conditions of MECC. The electrophoretogram and the spectrum of the two compounds being measured in one sample by DAD are shown in Fig. 5. Therefore, both the electrophoretograms and the spectra can be used in quantification of the two compounds in partially overlapped CE peaks by MLP ANN approach. Seventy data points at the maximum of CE peaks were applied as the input spectra data to MLP ANN. The first 35 data was from vitamin B1, and the other 35 belongs to dibazolum. For partially resolved CE peaks of binary mixtures, two local maximum absorbance data points appear in the ASCII file of the corresponding electrophoretogram. The two data points can be searched for using a library function of MATLAB (version 6.5). The data

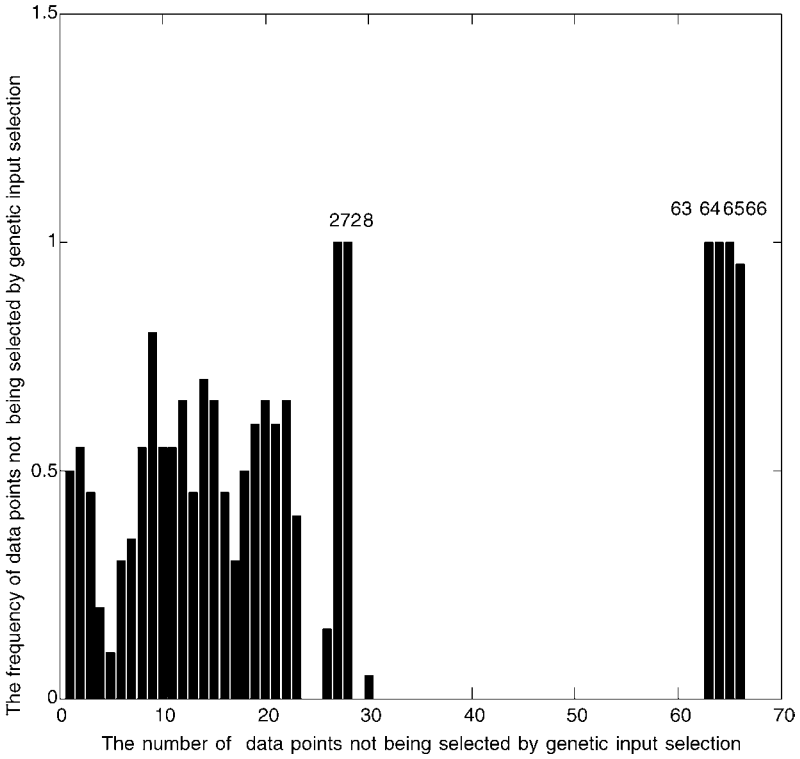

Fig. 6. The frequency of input variables not being selected by genetic input selection in Section 4.2.1.1.

points of the two local maximum values and those between and outside the two local maximum peaks were collected to form an input vector of electrophoretogram. The absorbance data points outside the two maximum points were selected symmetrically. Like the previous part of the paper dealing with the quantification of promethazine hydrochloride and chloroquine phosphate in their completely overlapped CE peaks, the quantification of vitamin B1 and dibazolum in their partially overlapped CE peaks was also investigated under the two conditions, i.e. one series of the concentrations of the two compounds being studied in their mixed solutions were obtained by diluting one original mixed solution, and the other series of the concentrations of the mixed solutions of the two components being investigated had a constant total concentration in each solution although the concentrations of each component in different solutions were varied.

Table 3
The average prediction errors of vitamin B1 and dibazolum in Section 4.2.1

Compounds	Case I (with GA)	Case I (without GA)	Case II (with GA)	Case II (without GA)
Vitamin B1[a] (%)	4.28	12.34	5.21	6.58
n[b]	10	10	10	10
S.D.[c] (%)	0.75	0.87	0.53	1.71
Dibazolum[a] (%)	4.28	12.47	5.20	6.63
n[b]	10	10	10	10
S.D.[c] (%)	0.72	0.84	0.49	1.72
MLP ANN	64:1:2	70:1:2	29:1:2	40:1:2
Generation of GA	1000	–	1000	–
Population of GA	1000	–	1000	–
Mutation rate	1	–	1	–
Crossover rate	0.3	–	0.3	–
Unit penalty factor of GA	0.01	–	0.01	–
Smoothing factor	0.3	–	0.3	–
Iteration times of design	100	100	100	100
Unit penalty of design	0.01	0.01	0.01	0.01

[a] The average prediction error of the corresponding compound.
[b] The number of the parallel runs of the designed MLP ANN.
[c] The standard deviation.

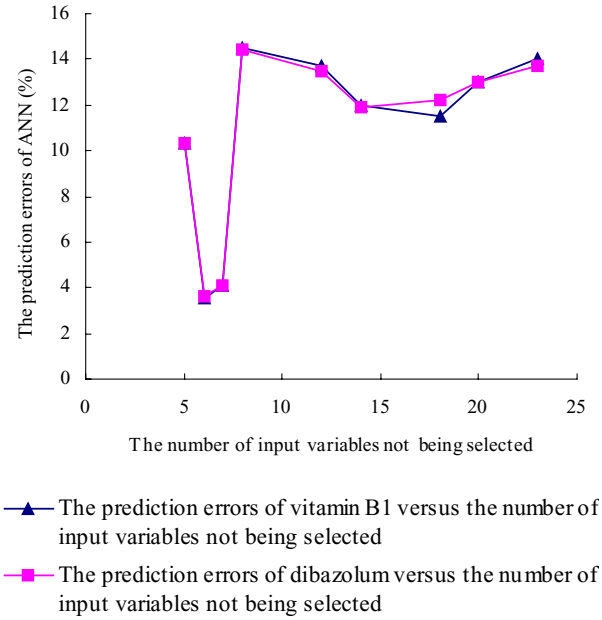

Fig. 7. The optimization process of input variables in Section 4.2.1.1.

In Section 4.2.1, the concentrations of the two components in their mixed solutions under the first experimental conditions were predicted by the proposed MLP ANN approach. The frequencies of the input variables not being selected during 20 parallel GA runs are shown in Fig. 6. In the second step of the optimization strategy, the process of input variable optimization is shown in Fig. 7. All the predicted results, the structures of the automatically designed MLP ANN and the corresponding performance parameters are listed in Table 3. In Section 4.2.2, the concentrations of the two compounds in their mixed solutions prepared according to the second experimental conditions were investigated. The corresponding results are given in Table 4.

In the two tables, corresponding spectra data were used as input variables to MLP ANN in Case I, and the electrophoretogram data were input to MLP ANN in Case II. Under each of the investigated conditions, 20 samples were used to train the corresponding MLP ANN. Four of the 20 samples selected randomly were served as test set, and the others were training samples. From the calculated results in the two tables, we can also conclude that the predictive ability of MLP ANN was improved to some extent applying the proposed input selection strategy based on GA in the experimental conditions investigated in this paper.

5. Prediction of the concentrations in experimental samples

The proposed approach was applied to the quantification of vitamin B1 and dibazolum, promethazine hydrochloride and chloroquine phosphate in their mixed solutions under the experimental conditions investigated above. The input variables of spectra or electrophoretograms were selected according to the proposed optimization strategy. The samples, of which concentrations were to be predicted, were selected randomly. Moreover, the samples to be investigated were not included in the training set of corresponding networks. The performance was based on leave-one-out cross validation strategy [45], i.e. only the sample to be predicted was removed from the training set of MLP ANN. The comparisons of the target and predicted concentrations of all samples being studied are shown in Fig. 8.

The results show that the predicted concentrations are in good consistent with the target values for both the

Table 4
The average prediction errors of vitamin B1 and dibazolum in Section 4.2.2

Compounds	Case I (with GA)	Case I (without GA)	Case II (with GA)	Case II (without GA)
Vitamin B1[a] (%)	6.13	6.51	6.70	7.45
n[b]	10	10	10	10
S.D.[c] (%)	0.29	1.27	0.33	0.74
Dibazolum[a] (%)	6.11	6.52	6.70	7.45
n[b]	10	10	10	10
S.D.[c] (%)	0.29	1.25	0.33	0.74
MLP ANN	63:1:2	70:1:2	50:1:2	55:1:2
Generation of GA	1000	–	1000	–
Population of GA	1000	–	1000	–
Mutation rate	1	–	1	–
Crossover rate	0.3	–	0.3	–
Unit penalty factor of GA	0.01	–	0.01	–
Smoothing factor	0.3	–	0.3	–
Iteration times of design	100	100	100	100
Unit penalty of design	0.01	0.01	0.01	0.01

[a] The average prediction error of the corresponding compound.
[b] The number of the parallel runs of the designed MLP ANN.
[c] The standard deviation.

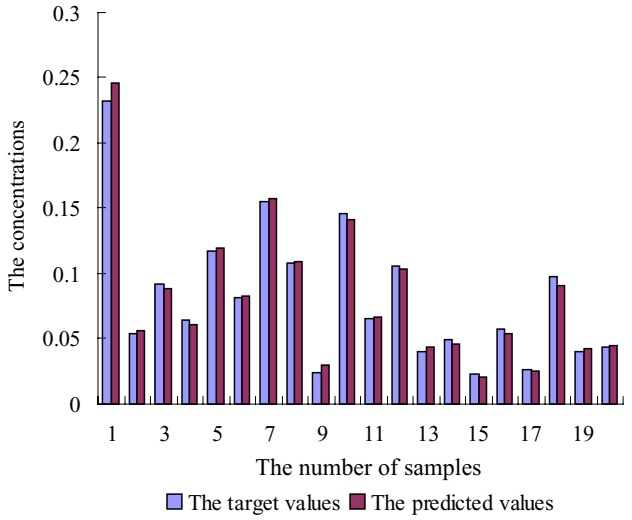

Fig. 8. The predicted concentrations vs. the target concentrations.

spectrum and electrophoretogram data as input variables. Further more, the results also proved that the four analytes could be quantified in overlapped MECC peaks by the proposed MLP ANN approach based on genetic input selection.

6. Conclusions

Quantification of multicomponent in MECC is possible even in the case of overlapped peaks by means of MLP ANN approach. The data from spectra or electrophoretograms was suitable for MLP ANN prediction. Moreover, the predicted results of MLP ANN were improved to some extent applying the input optimization strategy based on genetic input selection. At the same time, the structures of MLP ANN were simplified by removing the unnecessary input variables from the input data sets. The proposed approach was also tested using a leave-one-out cross validation procedure. The calculated results indicate that GA–MLP ANN approach was a promising tool to resolve overlapped CE peaks even applying a simpler network. Then, the quantification of corresponding analytes can be achieved in the case of unresolved CE peaks. That is to say, the time consumption in searching for optimal experimental conditions can be shortened when complete separation of some components is difficult.

Acknowledgements

National Natural Science Foundation of China financially supported the work. It is gratefully acknowledged.

References

[1] N.M. Maier, P. Franco, W. Lindner, J. Chromatogr. A 906 (2001) 3.
[2] M. Martínez Galera, J.L. Martínez Vidal, A. Garrido Frenich, M.D. Gil García, J. Chromatogr. A 778 (1997) 139.
[3] A. Garrido French, J.L. Martínez Vidal, P. Parrilla, M. Martínez Galera, J. Chromatogr. A 778 (1997) 183.
[4] S. Sentellas, J. Saurina, S. Hernández-cassou, M.T. Galceran, L. Puignou, J. Chromatogr. A 909 (2001) 259.
[5] A. Cladera, E. Gómez, J.M. Estela, V. Cerda, J. Chromatogr. Sci. 30 (1992) 453.
[6] T.L. Cecil, R.B. Poe, S.C. Rutan, Anal. Chim. Acta 250 (1991) 37.
[7] J.A. Gimena Garcia, J. Giménez Plaza, J.M. Cano Pavón, J. Liq. Chromatogr. 17 (1994) 277.
[8] G. Bocaz-Beneventi, R. Latorre, M. Farková, J. Havel, Anal. Chim. Acta 452 (2002) 47.
[9] J. Havel, E.M. Peòa-Méndez, A. Rojas-Hernández, J.-P. Doucet, A. Panaye, J. Chromatogr. A 793 (1998) 317.
[10] M. Farková, E.M. Peòa-Méndez, J. Havel, J. Chromatogr. A 848 (1999) 365.
[11] V. Dohnal, M. Farková, J. Havel, Chirality 11 (1999) 616.
[12] J. Havel, J.E. Madden, P.R. Haddad, Chromatographia 49 (1999) 481.
[13] J. Havliš, J.E. Madden, A.L. Revilla, J. Havel, J. Chromatogr. B 755 (2001) 185.
[14] R.M. Latorre, S. Hernandez-Cassou, J. Saurina, J. Sep. Sci. 24 (2001) 427.
[15] P. Polášková, G. Bocaz, H. Li, J. Havel, J. Chromatogr. A 979 (2002) 59.
[16] H. Li, Y.X. Zhang, P. Pavla, J. Havel, Acta Chim. Sinca 60 (2002) 1264.
[17] V. Dohnal, H. Li, M. Farková, J. Havel, Chirality 14 (2002) 509.
[18] D.L. Massart, B.G.M. Vandeginste, S.M. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988.
[19] H. Li, F. Zhang, J. Havel, Electrophoresis 24 (2003) 3107.
[20] V. Dohnal, F. Zhang, H. Li, J. Havel, Electrophoresis 24 (2003) 2462.
[21] M. Bos, A. Bos, W.E. van der Linden, Analyst 118 (1993) 323.
[22] I.V. Teko, G.I. Poda, J. Med. Chem. 36 (1993) 811.
[23] J. Ghasemi, A. Niazi, R. Leardi, Talanta 59 (2003) 311.
[24] R. Leardi, A. Lupiáñez González, Chemometr. Intell. Lab. Syst. 41 (1998) 195.
[25] M.A. Alonso Lomillo, O. Dominguez Renedo, M.J. Arcos Martinez, Anal. Chim. Acta 449 (2001) 167.
[26] M.K. Hartnett, G. Lightbody, G.W. Irwin, Chemometr. Intell. Lab. Syst. 40 (1998) 215.
[27] A.S. Barros, D.N. Rutledge, Chemometr. Intell. Lab. Syst. 40 (1998) 65.
[28] U. Depczynski, V.J. Frost, K. Molt, Anal. Chim. Acta 420 (2000) 217.
[29] F. Dieterle, B. Kieser, G. Gauglitz, Chemometr. Intell. Lab. Syst. 65 (2003) 67.
[30] B.W. Kim, S.J. Park, Chemometr. Intell. Lab. Syst. 56 (2001) 39.
[31] E. Richards, C. Bessant, S. Saini, Chemometr. Intell. Lab. Syst. 61 (2002) 35.
[32] J. Zupan, J. Gasteiger, Anal. Chim. Acta 248 (1991) 1.
[33] B.G. Sumpter, C. Gettino, D.W. Noid, Annu. Rev. Phys. Chem. 45 (1994) 439.
[34] B.G. Sumpter, D.W. Noid, Annu. Rev. Mater. Sci. 26 (1996) 223.
[35] J.A. Kinsella, Network 3 (1992) 27.
[36] E. Polak, G. Ribiere, Operationelle 13 (1969) 35.
[37] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[38] J.H. Holland, Adaption in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975.
[39] M.B. Seasholtz, B. Kowalski, Anal. Chim. Acta 277 (1993) 165.

[40] D.J. Livingstone, D.T. Manallack, J. Med. Chem. 36 (1993) 65.
[41] D. Broadhurst, R. Goodacre, A. Jones, J.J. Rowland, D.B. Kell, Anal. Chim. Acta 348 (1997) 71.
[42] M.J. Arcos, M.C. Ortiz, B. Villahoz, L.A. Sarabia, Anal. Chim. Acta 339 (1997) 63.
[43] D. Jouan-Rim baud, D.L. Massart, O.E. de Noord, Chemometr. Intell. Lab. Syst. 35 (1996) 213.
[44] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Science 220 (1983) 671.
[45] S. Courtois, R. Phan-Tan-Luu, Analusis 26 (1998) 304.