Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 65 (2005) 853–860
Talanta
www.elsevier.com/locate/talanta

Modeling of the relationship between electroosmotic flow and separation parameters in capillary zone electrophoresis using artificial neural networks and experimental design

Ya Xiong Zhang[a,1], Hua Li[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, 61137 Brno, Czech Republic*

Received 21 May 2004; received in revised form 28 July 2004; accepted 11 August 2004
Available online 1 October 2004

Abstract

The prediction of migration time of electroosmotic flow (EOF) marker was achieved by applying artificial neural networks (ANN) model based on principal component analysis (PCA) and standard normal distribution simulation to the input variables. The voltage of performance, the temperature in the capillary, the pH and the ionic strength of background electrolytes (BGE) were applied as the input variables to ANN. The range of the performance voltage studied was from 15 to 27 kV, and that of the temperature in the capillary was from 20 to 30 °C. For the pH values studied, the range was from 5.15 to 8.04. The range of the ionic strength investigated in this paper was from 0.040 to 0.097. The prediction abilities of ANN with different pre-processing procedure to the input variables were compared. Under the same performance conditions, the average prediction error of the migration time of the EOF marker was 5.46% with RSD = 1.76% according to 10 parallel runs of the optimized ANN structure by the proposed approach, and that of the 10 parallel predictions of the optimal ANN structure for the different performance conditions was 12.95% with RSD = 2.29% according to the proposed approach. The study showed that the proposed method could give better predicted results than other approaches discussed.

Keywords: Artificial neural networks; Principal component analysis; Standard normal distribution simulation; Electroosmotic flow; EOF prediction

1. Introduction

Capillary electrophoresis (CE) has been used extensively in analysis of amino acids, peptides, proteins, drugs, nucleic acids, inorganic ions, chiral compounds and even living cells [1]. For its shorter time of analysis, higher separation efficiency and lower consumption of reagents and samples, CE is used as an alternative and complementary technique to high performance liquid chromatography. Capillary zone electrophoresis (CZE) shows its unique merits for the separation of charged compounds. Electroosmotic flow (EOF) is defined as a liquid flow induced by an external electric field along a charged surface [2]. In CE, capillaries made of fused-silica are generally used. SiOH groups on the capillary wall will be ionized and form SiO^- groups depending on the pH of the background electrolyte (BGE) solutions [3–5]. As a result, the capillary wall is negatively charged. In the powerful electric field, the positively charged BGE flow migrates from anode to cathode. Generally speaking, this is the EOF in CE. Many separation parameters show their influence on EOF. EOF is controlled by the electrokinetic's zeta potential [6–7]. Zeta potential in capillary was determined to be depended on pH value of BGE [8–10]. A theory relating the change of concentration of amphiphile in the BGE to the variation of EOF is presented [11]. Stead et al. [12] studied the dependence of EOF on electric field strength, pH of the BGE and the content of organic modifier. Locascio et al. [13]

* Corresponding author. Tel.: +86 29 8830 2942; fax: +86 29 8830 3448.
E-mail addresses: huali@nwu.edu.cn (H. Li), zhangyaxiongabc.student@sina.com (Y.X. Zhang).
[1] Tel.: +86 29 8830 2942; fax: +86 29 8830 3448.

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.08.016

investigated the effect of material of microfluid channels, external electric field strength and non-specific protein adsorption on the EOF. Ionic strength of BGE is another parameter that affects the EOF in CE. Tseng et al. [14] studied the rules of the effect of ionic strength on EOF in different BGE solutions. Ye et al. [15] studied the EOF depending on ionic strength and pH of eluent. The properties of EOF have been studied under different separation parameters including nature and concentration of electrolyte, ionic strength, pH of the BGE and solvent composition [16]. Higher temperature will decrease the viscosity of the eluent, decrease the zeta potential and also increase pH of eluent. Increasing the temperature in capillary will increase the velocity of EOF. Walhagen et al. [17] discovered the linear relationship between EOF value and square root of T (temperature by centigrade unit) in their research work. Cahours and Morin [18] carefully investigated the influence of the three physiochemical parameters (temperature, ionic strength, and content of organic modifier in buffer) on EOF in capillary electrochromatography. From the research results, an increase in ionic strength results in a decrease in EOF. This paper also reported that higher temperature generates an increase in EOF. From the papers cited above, it is clear that EOF in CE is mainly influenced by pH of BGE, ionic strength in the buffer, applied external electric field, and temperature of CE performance.

EOF is an important phenomenon in CE. Mobile phase is driven through the capillary column in capillary electrochromatography by EOF [19]. The solutes are transported by EOF in capillary electrokinetic chromatography [20]. Kopp et al. [21] summarized the three ways by which EOF benefits miniaturized analysis system. However, EOF has to be controlled under some conditions in order to optimize the separation in CE. Capillary wall coating and addition of organic modifier in BGE are the main ways to adjust EOF [22–24]. Moreover, external electric field [25,26] and light [27] were also applied to control EOF.

Although the influences of some physicochemical parameters during separation performance by CE on EOF have been investigated in many papers, a mathematical model describing the relation between the value of EOF and the performance parameters has not been accomplished. A detailed mathematical model and explanation of EOF is given by Tavares and McGuffin [28]. However, explicit physicochemical description of EOF is too complicated for practical use and even Tavares' description did not take into account all parameters. On the other hand, Smoluchowski equation [29] only gives an indirect correlation between the value of EOF and the corresponding performance parameters.

The main purpose of this paper is to model the relationship between EOF value and the corresponding performance physicochemical parameters using ANN in CZE. In order to select the input variables for ANN, principal component analysis (PCA) algorithm is applied. ANN is also called "soft modeling" tool, without knowing or establishing an explicitly defined mathematical model [30]. ANN has been applied in CZE [31–33], electrokinetic micellar chromatography [34], chiral separation by CE [35], and increasing precision of analysis in CE [36,37] successfully.

2. Theory

2.1. Principal component analysis

Principal component analysis (PCA) is a statistic technique to extract information from multivariate data sets. To do this, the linear combinations of original variables are developed, which are called principal components. The greatest amount of variability of the original multivariate data set is represented by the first component, and the second component explains the maximum variances of the residual data set. Then, the third one will describe the most important variability of the next residual data set, and so on. According to the theory of least squares, the eigenvectors of all principal components are orthogonal each other in multi-dimensional data space. Generally speaking, p principal components are enough to account for the most variance in an m-dimensional data set, where p is the number of important principal components of the data set, and m means the number of all the principal components in the data set. It is obvious that p is less than m. So PCA is generally regarded as a data reduction technique. That is to say, a multi-dimensional data set can be projected to a lower dimension data space without loss most of the information of the original data set by PCA. Statheropoulos et al. [38] and Dong and McAvoY [39] described the algorithm of PCA in detail in their papers.

The selection of input variables to ANN is necessary to avoid "over fitting" [40] in the case of many input parameters offered. As a linear technique for dimensionality reduction, PCA can transform the input case from its original form (points in m-dimensional space) to its new form (points in p-dimensional space), where p is less than m. During the process, most amount of the variability of the original input cases is retained. Using the corrected input cases in a lower dimension, smaller ANN is applied in the performance of prediction. Since PCA is a linear projection technique, we achieved the projection of input set by linear ANN, which has the same number of input and output nodes.

2.2. Artificial neural networks

ANN is a kind of information processing computer program. 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 [41–43]. Despite different algorithms of ANN have been developed, error back-propagation (BP) algorithm is one of the most widely used. In this paper, ANN based on BP algorithm is applied to model the relationship between EOF value and corresponding separation parameters. The theory of BP ANN is given briefly here.

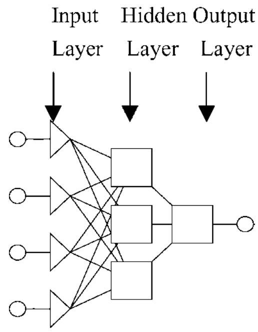

Fig. 1. The optimal architecture of ANN (Section 4.1.1).

BP ANN is composed of some logic units and the connection weights between the units. ANN is divided into three levels in order to understand the process of information processing. The three levels are called input layer, hidden layer(s) and output layer respectively. There are some logic units in each layer. The logic units are the basic information-processing unit in ANN. The simple architecture of BP ANN is shown in Fig. 1.

The relationship of the input value of unit i in input layer and that of unit h in hidden layer is

$$\varphi_h = \sum_{i=1}^{m} s_i \omega_{ih} + \theta_h \tag{1}$$

where s_i is the value input to logic unit i in the input layer, φ_h is the value input to logic unit h of the hidden layer. In Eq. (1), ω_{ih} represents the connection weights between unit h and i, θ_h is called the input bias of the unit h, while m is the number of logic units in the input layer, and it is adjustable according to the problem to be studied.

The φ_h is transformed to the output value of unit h with “sigmoid function”, which is the most widely applied.

$$O_h = \frac{1}{1 + \mathrm{e}^{-\varphi_h}} \tag{2}$$

where O_h is the output value of the unit h.

About logic unit o in the output layer, similar formulas are still valid.

$$\varphi_o = \sum_{h=1}^{n} s_h \omega_{ho} + \theta_o \tag{3}$$

$$O_o = \frac{1}{1 + \mathrm{e}^{-\varphi_o}} \tag{4}$$

The goal of the training of ANN is to minimize the error between target and calculated values by adjusting the connection weights and biases. The error is given by Eq. (5):

$$E = \frac{1}{2} \sum_{p=1}^{P} \sum_{m=1}^{M} (a_{pm} - o_{pm})^2 \tag{5}$$

The symbol M is the number of logic units in output layer, and P is the number of cases. a_{pm} and o_{pm} are the predicted and target values, respectively.

3. Experimental

3.1. Chemicals and buffers

Benzene was used as the marker of EOF. Ethanol was the solvent for benzene. The content of benzene in ethanol was 4% (v/v). Twice distilled water was used for preparing solutions. The stock solution of phosphate buffer at pH 5.29, pH 6.24, and pH 8.04 were prepared by mixing appropriate volumes of 1/15 mol dm^{-3} KH_2PO_4 and 1/15 mol dm^{-3} Na_2HPO_4 solutions. The buffer being used is a kind of Sørensen buffer. The solid reagent of potassium chloride was used to control and adjust the ionic strength in the corresponding solutions. The stock phosphate buffer with different pH and ionic strength was used as BGE respectively. In Section 4.3, the BGE was prepared by mixing appropriate volumes of 0.2 mol dm^{-3} NaOH and 0.2 mol dm^{-3} KH_2PO_4 solutions. All the chemicals used in this study were analytical reagent grade.

3.2. Instrumentation

A P/ACE 5500 CE instrument (Beckman-Coulter, USA) equipped with a photo diode array detector (DAD) was employed to perform all the experiments and collect the experimental data. The response time of the detector is 1 s. The processing and evaluation of the experimental results were accomplished using P/ACE workstation software (version 2.0). The data acquisition rate was four data points per second. The range of the scanning wavelength of the detector is from 190 to 600 nm. The maximum absorbance of EOF marker (benzene) is at 190 nm in the experimental conditions. Uncoated fused-silica capillary was purchased from Yongnian Optic Fiber Plant (Hebei, China). The total length (from the inlet side to the outlet side) of the capillary is 57 cm, and its length from the inlet side to the detector is 50 cm. The inner diameter of the capillary is 75 μm. A precision digital pH meter (Beijing, China) with a glass electrode as an indicator electrode and a saturated calomel electrode as a reference electrode was used to measure the pH values of different buffers at 25 °C.

The capillary was rinsed by twice distilled water and corresponding BGE for 5 min respectively prior to each run. When all the runs had been performed daily, the capillary was washed for 2 min with twice-distilled water, and for 2 min with high-pressure purified nitrogen. The pressure of the purified rinsing nitrogen was 137.84 kPa in all performance. All the values of EOF were determined with one capillary. The migration times of the EOF marker were based on the peak maximum.

The injection time for each run was 1 s with pressure injection approach. Other operating parameters were arranged

Table 1
The OA_9 (3^4) orthogonal array for experimental design

Voltage (kV)	Temperature (°C)	pH[a] (18 °C)	pH[b] (25 °C)	Ionic strength
15	20	8.04	8.02	0.08
20	20	5.29	5.42	0.04
25	20	6.24	6.26	0.06
15	25	6.24	6.38	0.04
20	25	8.04	7.79	0.06
25	25	5.29	5.15	0.08
15	30	5.29	5.34	0.06
20	30	6.24	6.28	0.08
25	30	8.04	8.04	0.04

[a] Where are the theoretical pH values at 18 °C.
[b] Where are the measured pH values at 25 °C.

according to OA_9 (3^4) orthogonal array. The array is shown in Table 1. The ramp time of all the performance voltage was 0.17 min.

3.3. Data processing

The calculations of ANN and those of PCA were performed by Trajan ANN software version 3.0 (Durham, UK) on a Pentium III personal computer. The standard normal distribution data were produced using a library function of MATLAB 6.5.

4. Results and discussion

Benzene was chosen to be the marker of EOF because it is a neutral molecule without any charge during the experimental process. So only EOF in the CE system gives its contribution to the migration time of the marker, of which migration time can be used as a measurement of EOF.

4.1. EOF model under the same performance conditions

In order to establish a “soft model” of EOF depending on the performance parameters in CZE by ANN, the marker sample was run in CE instrument twenty times under the same condition. The temperature of performance was 25 °C, the applied voltage was 20 kV, the ionic strength of BGE was 0.06, and the pH of BGE was 7.79 (measured value, theoretical value at 18 °C is 8.04). The observed migration time of the marker in each run is shown in Table 2.

4.1.1. Prediction of migration time using corresponding performance parameters by ANN

The parameters such as temperature in the capillary, ionic strength and pH value of BGE, and separation voltage show their strong influence on EOF. So it is reasonable that ANN can apply these parameters as input variables to predict the migration time of the EOF marker. Three-level ANN architecture was used in this paper. The input data of the ANN

Table 2
Prediction results of migration time by ANN using different pre-processing procedure to input variables under the same performance conditions

Target value (min)	Predicted value (ANN) (min)	Difference (pred − exp) (min)	Predicted value (ANN + PCA) (min)	Difference (pred − exp) (min)	Predicted value (ANN + RDS[a]) (min)	Difference (pred − exp) (min)	Predicted value (ANN + RDS[a] + PCA) (min)	Difference (pred − exp) (min)
3.629	3.299	−0.3305	3.269	−0.360	3.276	−0.3531	3.242	−0.3869
3.475	3.299	−0.1765	3.269	−0.206	3.295	−0.1801	3.241	−0.2338
3.446	3.299	−0.1475	3.269	−0.177	3.279	−0.1667	3.243	−0.2032
3.425	3.299	−0.1265	3.269	−0.156	3.288	−0.1372	3.240	−0.1851
3.412	3.299	−0.1135	3.269	−0.143	3.289	−0.1234	3.240	−0.1720
3.388	3.299	−0.08946	3.269	−0.119	3.288	−0.1003	3.243	−0.1451
3.321	3.299	−0.02246	3.304	−0.01652	3.291	−0.0301	3.242	−0.07948
3.271	3.299	0.02754	3.269	−0.001995	3.301	0.03043	3.240	−0.03061
3.229	3.299	0.06954	3.304	0.07548	3.291	0.0621	3.241	0.0121
3.179	3.299	0.1195	3.269	0.09001	3.289	0.11	3.241	0.06222
3.142	3.299	0.1565	3.269	0.127	3.288	0.1456	3.240	0.09803
3.104	3.299	0.1945	3.269	0.165	3.283	0.1793	3.241	0.1372
3.067	3.299	0.2315	3.269	0.202	3.284	0.2173	3.240	0.1735
3.033	3.299	0.2655	3.269	0.236	3.289	0.2564	3.242	0.2089
3.004	3.299	0.2945	3.269	0.265	3.286	0.282	3.242	0.2380
2.996	3.299	0.3025	3.269	0.273	3.282	0.2857	3.241	0.2455
3.017	3.299	0.2815	3.269	0.252	3.280	0.2633	3.242	0.2245
3.150	3.299	0.1485	3.269	0.119	3.289	0.1385	3.239	0.08852
3.237	3.299	0.06154	3.269	0.03201	3.291	0.05352	3.241	0.003657
3.313	3.299	−0.01446	3.269	−0.04399	3.289	0.02366	3.241	−0.07172
E_{pred}[b] (%)		5.72		5.52		5.62		5.44
E_{apred}[c] (%)		5.63		5.52		5.58		5.46
RSD[d] (%)		1.88		0.54		1.58		1.76

[a] RDS means random distribution simulation.
[b] E_{pred} means prediction error of one training case.
[c] E_{apred} means average prediction error of the training cases.
[d] RSD means relative standard deviation.

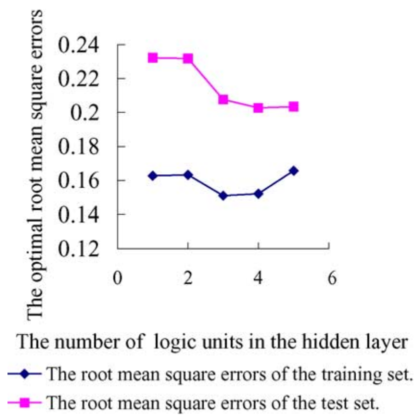

Fig. 2. The optimal root mean square errors vs. the number of logic units in the hidden layer of ANN (Section 4.1.1).

architecture was the values of the four parameters mentioned, while the migration time of the EOF marker was the only output value. The number of logic units in the hidden layer was changed from 1 to 5. To each different ANN architecture, in order to avoid "over training", the data set containing 20 cases was divided into the training set and test set at random. The value of root mean square error of each ANN architecture at their best training epochs in 5000 iterative epochs is shown in Fig. 2. It can be concluded that three logic units in hidden layer were sufficient to obtain low root mean square error values. So the optimized ANN structure was 4:3:1.

During the training process of the optimal structure found, the learning rate and the learning momentum were 0.6 and 0.3 respectively. The training of ANN was stopped when there was no improvement on the best root mean square error value searched by Trajan software automatically during the training process of the maximum value of epochs. However, for the weights of the network were reset randomly for each run, different runs of the ANN software by the trained network may result in different predicted results. So the average prediction errors were calculated by 10 parallel runs on the optimized network structure. The predicted results are shown in Table 2.

The prediction errors were calculated by Eq. (6):

$$\text{Prediction error} = \frac{\sqrt{\sum_{p=1}^{P}(i_{p\text{true}} - i_{p\text{pred}})^2}}{\sqrt{\sum_{p=1}^{P}(i_{p\text{true}})^2}} \qquad (6)$$

In the formula, $i_{p\text{true}}$ is the pth target migration time, and $i_{p\text{pred}}$ is the pth predicted one. P is the number of the cases.

4.1.2. Prediction of migration time using corresponding performance parameters with PCA pre-processing procedure by ANN

The goal of using PCA to select the input variables of ANN in this paper is to avoid "over fitting" and improve the precision of EOF marker migration time prediction. After processing the original data set by PCA, only one important principal component was identified. The principal component represented nearly 100% variability of the original data set. So the principal component can represent the four original input variables. The principal component was calculated by a 4:1 linear network and was regarded as a corrected form of input variables. Using the new form of input variable, an optimized BP ANN with the structure of 1:1:1 was obtained. The prediction results of the simple BP network are given in Table 2. The results show that the pre-processing procedure of PCA improved the prediction ability of ANN.

4.1.3. Prediction of migration time using corresponding standard normal distribution modified performance parameters with PCA pre-processing procedure by ANN

In CE system, some of its operating parameters have normal distribution noise in their values or signals, i.e., the data of the controlled parameters may have some drifts to their theoretical values. In order to obtain better predicted results, standard normal distribution drift data were added to the values of applied temperature and voltage. The drift data were some pseudo-random numbers generated by a library function of MATLAB 6.5. Two important principal components were extracted from the corrected data set. The two principal components calculated using a 4:2 linear network were input to an optimized 2:3:1 BP ANN. The predicted results are better than those acquired in Sections 4.1.1 and 4.1.2.

We also investigated the predicted results from the data set added standard normal distribution drift values without any pre-processing approach. It was observed that ANN also gave better predicted results this time than ANN without any pre-processing approach to the input variables. In this section, the calculated results are also shown in Table 2.

In summary, comparing the predicted results from different pre-processing method for the original data sets, adding standard normal distribution drift to some input variables with PCA procedure can give better predicted results by ANN.

4.2. EOF model under the different performance conditions

The behavior of EOF under different electrophoretic conditions was also studied. Since multiple parameters influence the properties of EOF, an OA_9 (3^4) orthogonal array was employed to design the experiments.

4.2.1. Prediction of migration time using corresponding performance parameters by ANN

According to the OA_9 (3^4) orthogonal array, nine experiments with different performance parameters were carried out. The data set including nine cases was divided into training set and test set randomly. The optimal structure of BP

Table 3
Prediction results of migration time by ANN using different pre-processing procedure to input variables under the different performance conditions

Target value (min)	Predicted value (ANN) (min)	Difference (pred − exp) (min)	Predicted value (ANN + PCA) (min)	Difference (pred − exp) (min)	Predicted value (ANN + RDS[a]) (min)	Difference (pred − exp) (min)	Predicted value (ANN + RDS[a] + PCA) (min)	Difference (pred − exp) (min)
7.596	4.297	−3.299	4.687	−2.909	4.294	−3.302	7.149	−0.4471
8.317	8.341	0.0237	8.392	0.07533	8.347	0.03015	8.135	−0.1820
4.554	4.473	−0.08139	4.700	0.1459	4.444	−0.1096	4.780	0.2257
7.621	7.818	0.1975	7.616	−0.00544	8.395	0.7744	8.213	0.5919
4.217	4.336	0.1191	4.648	0.4307	4.362	0.1450	4.567	0.3499
5.654	5.652	−0.0016	5.648	−0.005868	5.659	0.005482	5.638	−0.01557
8.729	8.619	−0.1096	8.536	−0.1928	8.605	−0.1238	8.152	−0.5774
4.696	4.710	0.01433	4.760	0.06399	4.681	−0.01523	4.865	0.1690
2.192	4.381	2.189	4.644	2.452	4.406	2.214	4.352	2.160
E_{pred}[b] (%)		20.96		20.27		21.42		12.70
E_{apred}[c] (%)		20.78		18.37		21.12		12.95
RSD[d] (%)		0.85		1.31		0.94		2.29

[a] RDS means random distribution simulation.
[b] E_{pred} means prediction error of one training case.
[c] E_{apred} means average prediction error of the training cases.
[d] RSD means relative standard deviation.

ANN was 4:2:1. The predicted results at the best iterative times are shown in Table 3.

4.2.2. Prediction of migration time using corresponding performance parameters with PCA pre-processing procedure by ANN

Under these conditions, for only nine cases in the experiments, it is necessary to reduce the number of input variables of ANN to avoid "over fitting" [40]. The four input variables were projected to three-dimensional space using the pre-processing procedure of PCA. Three important principal components were sufficient to represent nearly all the information (or variability) in the original data set. We calculated the three main principal components on a 4:3 linear network. Then the three corrected input variables were input to an optimal 3:1:1 BP network. The calculated results of the BP ANN are given in Table 3.

4.2.3. Prediction of migration time using corresponding standard normal distribution modified performance parameters with PCA pre-processing procedure by ANN

Standard normal distribution data drifts were also added to the performance temperature and applied voltage for the same reason as above. The pseudo-random numbers were produced using the same method. According to the PCA procedure, three main principal components were found in the original data set. The corrected input parameters to BP ANN were calculated applying a 4:3 linear network. A 3:1:1 optimal BP network was constructed. All the predictions were performed using this BP network. The calculated results are also given in Table 3.

The experimental results show that the pre-processing procedure of PCA can improve the prediction results. The simulation of random drift to input parameters led to worse prediction results without PCA procedure. However, the random simulation with PCA pre-processing procedure gave the best predictive ability in this paper.

4.3. Prediction of migration time of EOF marker under other experimental conditions

The different "soft" models based on ANN with (or without) PCA pre-processing procedure and random simulation approach were applied to predict the migration time of EOF marker under other performance conditions. The BGE was prepared by mixing appropriate volume of 0.2 mol dm^{-3} NaOH and 0.2 mol dm^{-3} KH_2PO_4 solutions. The average prediction errors of different ANN models in 10 parallel runs, the pH values, the ionic strength of BGE, the other corresponding performance parameters, and the estimated migration time of the EOF marker were shown in Tables 4 and 5.

In this paper, the effect of the material of the capillary and that of the state of the capillary surface on EOF had not been studied. Some uncontrollable alternations on the capillary surface induce the little changes in EOF, which leads to some not always understandable migration time shifts [44]. In order to minimize the effect of the variations of the physicochemical properties of the capillary surface, the performances of each separation condition (under the same or different experimental conditions, respectively) were carried out in the same batch run respectively. The performance parameters of the cases to test the predictive ability of the proposed ANN models were not included in the training sets being used. The performance conditions of the testing cases were selected randomly. Under the same separation conditions, fifteen cases were applied to train the networks. Other three cases were used to test the predictive ability of the ANN models proposed. The predicted results are shown in Table 4. For the different separation conditions, 10 cases were employed to the "learning" of the ANN, and the trained ANN model

Table 4
The predicted results for the migration time of the EOF marker using the different ANN models under the same performance conditions

Voltage (kV)	Temperature (°C)	Ionic strength	pH	Target value	Predicted value (ANN + PCA + RDS[a])	Predicted value (ANN + PCA)	Predicted value (ANN + RDS[a])	Predicted value (ANN)
27	20	0.05564	5.80	6.617	6.995	7.136	7.043	7.251
27	20	0.05564	5.80	6.767	6.680	7.136	6.988	7.251
27	20	0.05564	5.80	7.592	7.151	7.136	7.196	7.251
E_{pred}[b] (%)					4.84	6.46	5.13	7.15
E_{apred}[c] (%)					6.03	10.49	7.41	10.52
RSD[d] (%)					1.28	0.84	1.81	0.070

[a] RDS means random distribution simulation.
[b] E_{pred} means prediction error of the test cases (not included in the training cases).
[c] E_{apred} means average prediction error of the training cases.
[d] RSD means relative standard deviation.

Table 5
The predicted results for the migration time of the EOF marker using the different ANN models under the different performance conditions

Voltage (kV)	Temperature (°C)	Ionic strength	pH	Target value	Predicted value (ANN + PCA + RDS[a])	Predicted value (ANN + PCA)	Predicted value (ANN + RDS[a])	Predicted value (ANN)
20	30	0.09685	7.30	5.863	5.367	8.168	5.520	5.093
27	30	0.09274	7.14	4.500	5.102	5.187	5.220	5.071
20	25	0.06774	6.57	9.279	8.056	8.282	7.969	11.11
E_{pred}[b] (%)					12.23	21.95	12.93	17.42
E_{apred}[c] (%)					5.84	16.73	7.92	17.35
RSD[d] (%)					1.23	6.35	2.64	3.30

[a] RDS means random distribution simulation.
[b] E_{pred} means prediction error of the test cases (not included in the training cases).
[c] E_{apred} means average prediction error of the training cases.
[d] RSD means relative standard deviation.

predicted the results of the other three cases. Table 5 gives the predicted results of each ANN models. All of the architectures of ANN were optimized. Obviously, ANN with PCA pre-processing procedure and standard normal distribution simulation to the input variables shows its better predictive ability.

5. Conclusions

Migration time of suitable EOF marker in CZE can be predicted using pH and ionic strength of BGE, applied voltage, and performance temperature as input parameters to BP ANN. A “soft model” based on ANN was applied to describe the relationship between EOF and the corresponding separation parameters in CZE. In addition, different pre-processing approaches to the input variables of ANN were investigated. It was shown that ANN with PCA pre-processing procedure and standard normal distribution noise to some input parameters could give better predicted results.

The use of ANN approach for prediction of EOF values by suitable performance parameters in capillary electrophoresis seems to be a promising tool for control and optimization of EOF in CE analysis.

Acknowledgement

The financial support of the National Natural Science Foundation of China (No. 20075021) is gratefully acknowledged.

References

[1] N.M. Maier, P. Franco, W. Lindner, J. Chromatogr. A 906 (2001) 3.
[2] C. Yang, C.B. Ng, V. Chan, J. Coll. Interf. Sci. 248 (2002) 524.
[3] A.K. Kolstad, P.Y.T. Chow, F.F. Cantwell, Anal. Chem. 60 (1988) 1565.
[4] A. Davidowicz, W. Janusz, J. Szeypa, A. Waksmundski, J. Coll. Interf. Sci. 115 (1987) 555.
[5] I. Altug, M.L. Hair, J. Phys. Chem. 71 (1967) 4260.
[6] P.Y.T. Chow, F.F. Cantwell, Anal. Chem. 60 (1988) 1569.
[7] D.M. Vermeulen, F.F. Cantwell, Anal. Chem. 65 (1993) 1360.
[8] Ch. Schwer, E. Kenndler, Anal. Chem. 63 (1991) 1801.
[9] W. Schützner, E. Kenndler, Anal. Chem. 64 (1992) 1991.
[10] M.M. Dittmann, G.P. Rozing, J. Chromatogr A 744 (1996) 63.
[11] O. Stålberg, J. Ståhlberg, J. Chromatogr. A 776 (1997) 311.
[12] D.A. Stead, R.G. Reid, R.B. Taylor, J. Chromatogr. A 798 (1998) 259.
[13] L.E. Locascio, C.E. Perso, C.S. Lee, J. Chromatogr. A 857 (1999) 275.
[14] W.L. Tseng, M.M. Hsieh, S.J. Wang, H.T. Chang, J. Chromatogr. A 894 (2000) 219.

[15] M.L. Ye, H.F. Zou, Z. Liu, J.Y. Ni, J. Chromatogr. A 887 (2000) 223.
[16] M.C. de Beuaistre, O. Mothieu, J. Randon, J.L. Rocca, J. Chromatogr. A 971 (2002) 199.
[17] K. Walhagen, K.K. Unger, M.T.W. Hearn, J. Chromatogr. A 893 (2000) 401.
[18] X. Cahours, P. Morin, J. Chromatogr. A 845 (1999) 203.
[19] J.H. Knox, I.H. Grant, Chromatographia 24 (1987) 135.
[20] B. Potoček, B. Maichel, B. Gaš, M. Chiari, E. Kenndler, J. Chromatogr. A 798 (1998) 269.
[21] M.U. Kopp, H.J. Crabtree, A. Manz, Curr. Opin. Chem. Boil. 1 (1997) 410.
[22] C. Finkler, H. Charrel, H. Engelhardt, J. Chromatogr. A 822 (1998) 101.
[23] H. Burt, D.M. Lewis, K.N. Tapley, J. Chromatogr. A 739 (1996) 367.
[24] D. Volgger, A.J. Zemann, G.K. Bonn Jr., M.J. Antal, J. Chromatogr. A 758 (1997) 263.
[25] V. Kašička, Z. Prusik, P. Sázelová, M. Chiari, I. Mikšĺk, Z. Deyl, J. Chromatogr. B 741 (2000) 43.
[26] V. Kašička, Z. Prusík, P. Sázelová, T. Borth, E. Brynda, L. Machová, J. Chromatogr. A 772 (1997) 221.
[27] J. Moorthy, C. Khoury, J.S. Moore, D.J. Beebe, Sens. Actuat. B 75 (2001) 223.
[28] M.F.M. Tavares, V.L. McGuffin, Anal. Chem. 67 (1995) 3687.
[29] P.R. Monika, M.D. Gerard, J. Chromatogr. A 744 (1996) 63.
[30] D.L. Massart, B.G.M. Vandeginste, S.M. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988.
[31] J. Havel, E.M. Peňa-Méndez, A. Rojas-Hernández, J.-P. Doucet, A. Panaye, J. Chromatogr. A 793 (1998) 317.
[32] M. Farková, E.M. Peňa-Méndez, J. Havel, J. Chromatogr. A 848 (1999) 365.
[33] V. Dohnal, M. Farková, J. Havel, Chirality 11 (1999) 616.
[34] J. Havel, J.E. Madden, P.R. Haddad, Chromatographia 49 (1999) 481.
[35] V. Dohnal, H. Li, M. Farková, J. Havel, Chirality 14 (2002) 509.
[36] P. Polášková, G. Bocaz, H. Li, J. Havel, J. Chromatogr. A 979 (2002) 59.
[37] H. Li, Y.X. Zhang, P. Pavla, J. Havel, Acta Chimica Sinica 60 (2002) 1264.
[38] M. Statheropoulos, A. Pappa, P.K. Tzanis, H.L.C. Meuzelaar, Anal. Chim. Acta 401 (1999) 35.
[39] D. Dong, T.J. McAvoY, Comput. Chem. Eng. 20 (1996) 65.
[40] I.V. Teko, G.I. Poda, J. Med. Chem. 36 (1993) 811.
[41] J. Zupan, J. Gasteiger, Anal. Acta Chim. 248 (1991) 1.
[42] B.G. Sumpter, C. Gettino, D.W. Noid, Annu. Rev. Phys. Chem. 45 (1994) 439.
[43] B.G. Sumpter, D.W. Noid, Annu. Rev. Mater. Sci. 26 (1996) 223.
[44] P.S. Kopplin, A.V. Garmash, A.V. Kudryavtsev, F. Menzinger, I.V. Perminova, N. Hertkorn, D. Freitag, V.S. Petrosyan, A. Kettrup, Electrophoresis 22 (2001) 77.