Available online at www.sciencedirect.com
ScienceDirect
Talanta 71 (2007) 848–853
Talanta
www.elsevier.com/locate/talanta

Artificial neural network-based transformation for nonlinear partial least-square regression with application to QSAR studies

Yan-Ping Zhou, Jian-Hui Jiang *, Wei-Qi Lin, Lu Xu, Hai-Long Wu, Guo-Li Shen, Ru-Qin Yu *

State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, P R.China

Received 18 February 2006; received in revised form 22 May 2006; accepted 22 May 2006
Available online 27 June 2006

Abstract

In the present study a new version of nonlinear partial least-square method based on artificial neural network transformation (ANN-NLPLS) has been proposed. This algorithm firstly transforms the training descriptors into the hidden layer outputs using the universal nonlinear mapping carried by an artificial neural network, and then utilizes PLS to relate the outputs of the hidden layer to the bioactivities. The weights between the input and hidden layers are optimized by a particle swarm optimization (PSO) method using the criterion of minimized model error via PLS modeling. An *F*-statistic is introduced to determine automatically the number of PLS components during the weight optimization. The performance is assessed using a simulated data set and two quantitative structure–activity relation (QSAR) data sets. Results of these three data sets demonstrate that ANN-NLPLS offers enhanced capacity in modeling nonlinearity while circumventing the overfitting frequently involved in nonlinear modeling.

Keywords: QSAR; Artificial neural network transformation-based nonlinear partial least-square; Particle swarm optimization; 2-Aryl(heteroaryl)-2,5-dihydropyrazolo[4,3-*c*]quinolin-3-(3*H*)-ones; Antitumor agents

1. Introduction

Partial least square (PLS) [1,2], a popular linear modeling technique, has been extensively used in quantitative structure–activity relation (QSAR) studies, since it can provide good remedial measures to the problems of correlated inputs and limited observations. However, only linear relation can be extracted from data. In the practice of QSAR studies, PLS used to show deteriorated performance as compared to nonlinear algorithms, as in such situations the relationships involved are essentially nonlinear. Consequently, it is desirable to have an efficient technique that can model nonlinear relations.

To tackle the issue of nonlinearity, one way is to develop nonlinear versions of PLS by incorporating nonlinear features into the linear PLS framework. Nowadays, several nonlinear versions of the PLS method have been developed. Examples of such nonlinear variants of PLS include quadratic PLS (QPLS) [3], spline PLS (SPLS) [4], neural network PLS (NNPLS) [5] and kernel PLS (KPLS) [6]. The basic idea of QPLS, SPLS and NNPLS is to use a nonlinear function based on the quadratic function, spline and artificial neural network (ANN), respectively, to replace the inner linear relation in PLS. However, QPLS is often incapable of providing adequate flexibility for modeling complex nonlinear relationship due to the predefined form of the quadratic function. In contrast, SPLS and NNPLS are expected to be sufficiently flexible for fitting varying nonlinearity, but these two algorithms are prone to overfitting in that extra flexibility might be adopted to fit the model errors. Unlike the aforementioned variants of nonlinear PLS, KPLS firstly transforms the original input data into a feature space using a nonlinear kernel function and then establish a linear PLS model in this space. The user-defined kernel mapping of the original variables might also tether the variability of the nonlinear transformation, leading to underfitting of the model in cases when the nonlinear relation among the data is too complex.

Another way to model complex nonlinear relationships is to resort to a nonlinear regression method such as ANN [7,8]. ANN

* Corresponding authors. Tel.: +86 731 8822577; fax: +86 731 8822782.
E-mail addresses: jianhuijiang@hnu.cn (J.-H. Jiang), rqyu@hnu.cn (R.-Q. Yu).

0039-9140/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2006.05.058

can be viewed as a universal model-free approximator that can represent any nonlinear function within sufficient accuracy by seeking proper combination of several sigmoid functions [9]. For this reason, ANNs are often taken to solve problems with nonlinear relationships. However, it is frequently criticized for the proneness of overfitting the training data. To combat the overfitting encountered in ANN, in this laboratory, a chaotic dynamic system has been introduced into the genetic algorithm to train ANN to formulate the CGANN algorithm [10] and a hybridized particle swarm optimization (PSO) approach to perform the network training (HPSONN) [11].

In this paper, motivated by the property that proper combination of sigmoid functions can approximate successfully any complex nonlinear relationships [9] and the power of PLS to combat overfitting of linear models in the presence of excessive variability in the inputs, we develop a novel nonlinear PLS approach based on an optimized ANN transformation of the input variables, ANN-NLPLS, in order to effectively model the nonlinear data. In ANN-NLPLS, firstly, the original variables are transformed from the input layer into the hidden layer using an ANN. Then, the PLS is utilized to relate the outputs of the hidden layer to the bioactivities. To adaptively adjust the nonlinear transformation of the original descriptors, the particle swarm optimization algorithm [11–13] is used to optimize the weights involved in ANN with *F*-statistic to extract latent variable automatically [13]. The ANN-NLPLS, PLS and ANN training by back-propagation (BP-ANN) algorithms have been firstly implemented for simulation data and then for two QSAR sets, i.e., 5-*N*-substituted-2-(substituted benzenesulphonyl) glutamines and 2-aryl(heteroaryl)- 2,5-dihydropyrazolo[4,3-*c*]quinolin-3-(3*H*)-ones. Results of these three data sets reveal that ANN-NLPLS provides superior performance to BP-ANN and PLS, and it can not only model well the nonlinear relations in the data, but also be resistant to the overfitting problem.

2. Theory

2.1. Artificial neural network-based transformation for nonlinear partial least-square regression

In the present study, we develop a new variant of nonlinear PLS, artificial neural network-based transformation for nonlinear partial least-square (ANN-NLPLS) regression, combining the advantages of ANN to model nonlinearity and PLS to combat overfitting. This method aims to simultaneously model nonlinearity and circumvent overfitting problem. In ANN-NLPLS, the role of ANN is to map the original descriptors into the hidden layer outputs. In general, such a nonlinear transformation of the original input variables produce a set of new feature ones that not only comprise the information for modeling the bioactivities, but also include additional variability to overfit the model errors. To combat the excessive variability in the feature variables, the PLS regression is then employed to relate the feature variables to the bioactivity. The procedure of ANN-NLPLS is described as follows.

Suppose $\boldsymbol{X}$ is the original descriptor input matrix (the elements in each row represents the descriptors describing one chemical compound), the outputs of the hidden layer can be obtained by the following expression:

$$\boldsymbol{O} = f(\boldsymbol{X}\boldsymbol{W} + \mathbf{1}\boldsymbol{\theta}^{\mathrm{T}}) = \frac{1}{1 + \mathrm{e}^{-(\boldsymbol{X}\boldsymbol{W} + \mathbf{1}\boldsymbol{\theta}^{\mathrm{T}})}} \quad (1)$$

where $\mathbf{1}$ represents a column vector with all elements equal to 1, $\boldsymbol{\theta}$ is the bias vector of the hidden layer, f represents the sigmoid function, and $\boldsymbol{W}$ is the weight matrix connecting the input and hidden layers. The hidden layer outputs $\boldsymbol{O}$ can be regarded as a set of extracted feature variables obtained via the ANN-based nonlinear transformation. Then, the PLS procedure is employed to model the relation between the extracted feature variables $\boldsymbol{O}$ and the bioactivities of the compounds.

In this paper, in order to make the algorithm flexible to approximate any nonlinearity, the weights involved in ANN-NLPLS are adaptively optimized by PSO with error as objective function.

2.2. Particle swarm optimization

PSO [14–17] is a stochastic global optimization method, simulating the social behavior of bird flocking. It is a population-based optimization tool, exploring the problem space by a population of individuals or particles. In PSO, each single solution represents a particle in the search space. The system is initialized with a population of stochastic solutions and searches for optima via updating generation. Each individual in PSO moves in the search space with a velocity guiding the flying of the particle, keeping track of the current optimum particles. The detailed description on the PSO algorithm can be found elsewhere [18]. In this paper, the PSO algorithm was utilized to optimize the weights using error as fitness function by PLS modeling. Each particle involved in PSO is encoded to a real string, the bits of which stand for the weights between the input and hidden layers. During the weight optimization, the PLS algorithm acts as the modeling method and the PLS components are automatically determined by the *F*-statistic defined as follows:

$$F = \frac{\mathrm{RSS}_{n-1}/(m-n)}{\mathrm{RSS}_n/(m-n-1)} \quad (2)$$

where m is the number of compounds in the training set and n is the number of the latent variable extracted from the $\boldsymbol{O}$ matrix, RSS refers to the sum of squared residuals. The F value can be identified by referring to the F-distribution table. When the F value is less than or equal to the predefined value, that is, the newly extracted latent variable does not reduce the model error significantly, the latent variable will be abnegated and the procedure of extracting the latent variable will cease. On the contrary, the procedure for extracting latent variable will continue.

3. Data sets

3.1. Simulated data

A 10-dimensional data set ($\boldsymbol{X}$, $\boldsymbol{y}$) including 150 samples was generated using Monte Carlo simulation. $\boldsymbol{X}$ was a matrix with all the elements in the interval (−1, 2), and $\boldsymbol{y}$ was obtained by

$\boldsymbol{y} = (\exp(\boldsymbol{X}) \circ \boldsymbol{X}) \times \boldsymbol{a} + \boldsymbol{X} \times \boldsymbol{b} + \boldsymbol{c} + \boldsymbol{e}$, where $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ were the coefficients obtained by random numbers uniformly distributed in [0, 1], $\boldsymbol{e}$ was the errors generated using normal random numbers with a standard deviation of 0.001, and the operator "○" represented the Hadamard product. Ninety-two samples were randomly selected to make up the training set, the remaining 58 objects comprising the test set.

3.2. Data set 2

In order to evaluate the performance of the proposed ANN-NLPLS method in QSAR studies, 53 5-*N*-substituted-2-(substituted benzenesulphonyl) glutamines as antitumor agents [19,20] were used as a data set. The biological activity (BA) represents the percentage inhibition of tumor cell count, expressed as $BA = (1 - T/C) \times 100\%$, where T is the cell count of the test and C is that of the control. We stochastically divided the data set into a training set of 38 compounds and a test set of 15 compounds. Using Cerius^2 3.5 software system, over 70 descriptors representing the chemical structure were calculated as the original variables. On the other hand, three descriptors used by Jha and co-workers [20] were also introduced in the family of variables. As to each compound, seven variables selected from the original over 73 variables by the modified PSO algorithm [18] were used for PLS, BP-ANN and ANN-NLPLS modeling, including two dipole moments (Dipole-mag, Dipole-Y) [21], number of hydrogen bond donors (Hbond donor), superdelocalizability (Sr), indicator parameter I (I takes value 1 for the presence of *n*-butyl at R_5 position, otherwise zero), sum of molar reflectivity values at R_1, R_2, R_3, R_4 positions of the benzene ring (ΣMR) and *E*-state indices (S-ssNH) [22–26]. In the symbol S-ssNH, 'S' represents the electronic topological state of atom, 'ss' stands for the two single bonds of the –NH–group, and 'NH' refers to the formula of the hydride group.

3.3. Data set 3

A set of 2-aryl(heteroaryl)-2,5-dihydropyrazolo[4,3-*c*] quinolin-3-(3*H*)-ones (PQs) with their corresponding affinities to benzodiazepine receptor (BzR) [27,28] were used as another data set to further check the validity of the newly developed algorithm. The data set was randomly divided into two subsets, the training set of 39 compounds and the test set of 13 compounds. In addition to the molecular descriptors listed in [27], more than 70 descriptors calculated by Cerius^2 3.5 software system were also utilized to describe the structures of these compounds. From these descriptors, nine parameters were selected by the modified PSO for multivariate modeling for each compound [18], including three dipole moments (Dipole-mag, Dipole-X, Dipole-Y) [21], density, principal moment of inertia (PMI-Y) [29], Jurs charged partial surface area descriptor (Jurs-TPSA) [29], $\sigma_{R'}$, MR_{R8} and I. The descriptor $\sigma_{R'}$ denotes a negative role for electron-attracting groups in the 3- and 4-position. MR_{R8} refers to the molar reflectivity of 8-substituent. I is an indicator parameter (1/0) for the 7-substituted compounds, that is, I equals to 1, when R is present at 7-position.

The algorithms used in this study were written in Matlab 5.3 and run on a personal computer (Intel Pentium processor 4/2.66 GHz 256 MB RAM). The descriptors were calculated by using Cerius^2 3.5 software system that is a comprehensive molecular modeling and simulation package used for materials science and life science.

4. Results and discussion

4.1. Simulated data

To demonstrate the performance of the proposed algorithm, the ANN-NLPLS algorithm was applied to the analysis of the

Table 1
Results of simulated data using ANN-NLPLS compared with those obtained by PLS and BP-ANN

Data set	*R* (correlation coefficient)			RMSE (root mean squared error)		
	Method 1[a]	Method 2[b]	Method 3[c]	Method 1[a]	Method 2[b]	Method 3[c]
Training set	0.8951	0.9357	0.9740	3.9825	3.0968	2.0241
Test set	0.8934	0.8512	0.9437	4.2854	4.8525	3.1939

[a] Modeling by PLS.
[b] Modeling by BP-ANN.
[c] Modeling using ANN-NLPLS.

Table 2
Results of QSAR analysis of antitumor agents using ANN-NLPLS compared with those obtained by PLS and BP-ANN

Data set	*R* (correlation coefficient)			RMSE (root mean squared error)		
	Method 1[a]	Method 2[b]	Method 3[c]	Method 1[a]	Method 2[b]	Method 3[c]
Training set	0.7804	0.8495	0.8702	0.1229	0.1209	0.1024
Test set	0.7998	0.7771	0.8449	0.1232	0.1471	0.1075

[a] QSAR study by PLS.
[b] QSAR study by BP-ANN.
[c] QSAR study using ANN-NLPLS.

simulated data including large number of samples with comparison to conventional procedures, PLS and BP-ANN. A monitoring set was selected from the training samples for the training of BP-ANN so as to reduce the possibility of overfitting the training data. The statistical results of these three algorithms are shown in Table 1. From Table 1 one observes that ANN-NLPLS compares favorably with PLS and offers better generalization than BP-ANN.

4.2. Data set 2

As a comparison, PLS was firstly utilized for modeling the bioactivities of the antitumor agent data using the selected seven variables. The statistical results of PLS are listed in Table 2. The correlation coefficient (R) for the training set and the test set are 0.7804 and 0.7998, respectively. The correlation between the calculated and observed values of antitumor activity is shown in Fig. 1a. From Fig. 1a and Table 2, it can be seen that the correlation was rather poor and the modeling error is quite high, which indicates the deficiency of the PLS in QSAR studies and the presence of the complex or unknown nonlinearity in the data set. In addition, the presence of nonlinearity in this data set was proven by a runs test [30] that yielded a statistical value 2.25, larger than the critical value 1.96.

Due to the presence of nonlinearity among the antitumor agent data, we built the model using ANN learned by back-propagation (BP-ANN) to further examine the correlation between the molecular structure and the antitumor activity based on the same descriptors as PLS. To reduce the possibility of BP-ANN to overfit the training data, a monitoring set including 10 samples was selected from the training samples for training the BP-ANN. The statistical results of the BP-ANN algorithm are also shown in Table 2. From Table 2, one can obtain that the correlation coefficient of the training set by BP-ANN is significantly larger than that of the test set, indicating the BP-ANN model exhibits overfitting. The calculated versus the observed bioactivities obtained by BP-ANN are shown in Fig. 1b. As shown in Fig. 1a, b and Table 2, the correlation was rather poor and both the PLS and BP-ANN are not sufficiently accurate to model this data set.

For improving the QSAR modeling of the antitumor agents, ANN-NLPLS was also investigated using the same variables as BP-ANN and PLS. The statistical results of ANN-NLPLS, together with those obtained by PLS and BP-ANN, are shown in Table 2. The correlation between the calculated and observed bioactivities is depicted in Fig. 1c. As one can see from Table 2, ANN-NLPLS gives a correlation coefficient of 0.8702 and a root mean squared error (RMSE) of 0.1024 for the training set and a correlation coefficient of 0.8449 and a RMSE of 0.1075 for the test set. Compared with PLS, ANN-NLPLS provides enhanced performance for both the training set and the test set, indicating that the proposed algorithm has the ability to model nonlinearity more effectively than the PLS. When refers to the comparison with BP-ANN algorithm, as shown in Table 2, the ANN-NLPLS offers a slightly better correlation coefficient for the training set while much higher correlation coefficient for the test set,

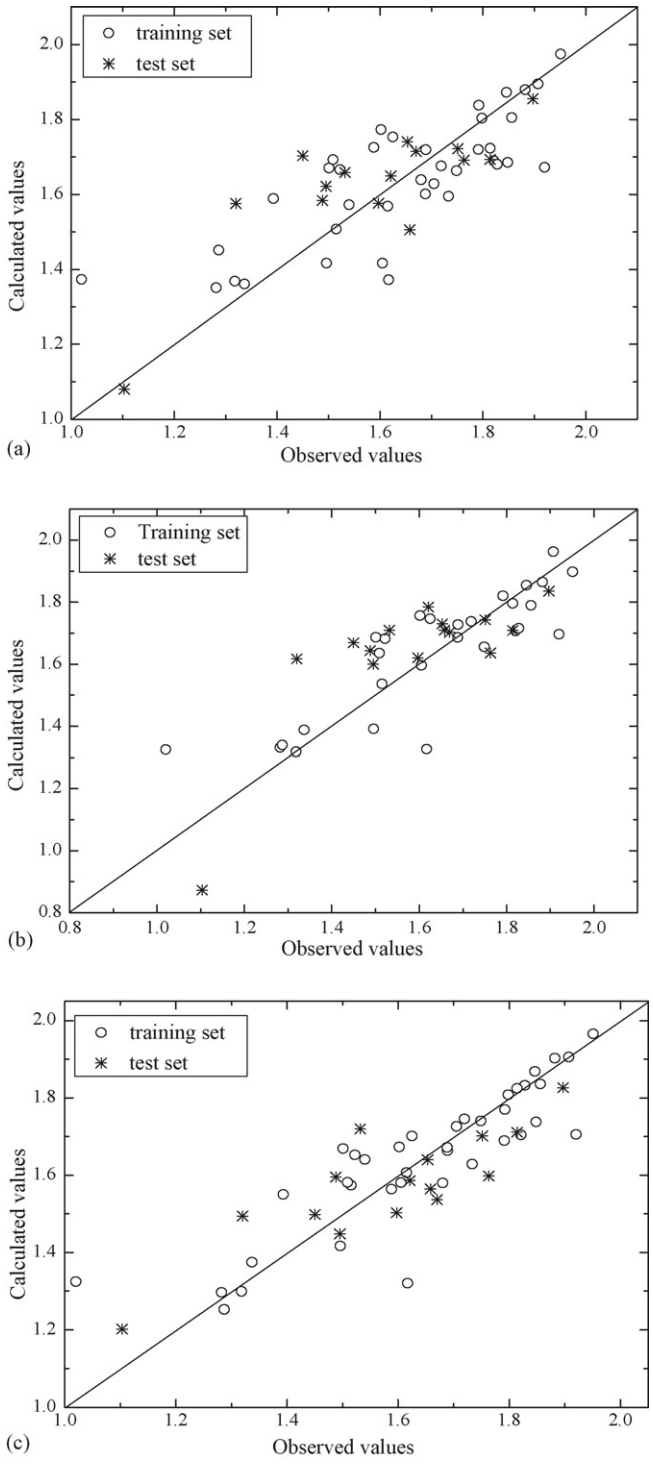

Fig. 1. (a) Calculated and observed values of bioactivity of antitumor agents by PLS modeling. (b) Calculated and observed values of bioactivity of antitumor agents by BP-ANN modeling. (c) Calculated and observed values of bioactivity of antitumor agents by ANN-NLPLS modeling.

demonstrating that this new algorithm is capable of resisting the overfitting problem. From Fig. 1a–c, one can find that the ANN-NLPLS shows much smaller deviations in the bioactivity estimation as compared to those gained from PLS and BP-ANN techniques. These results verify that ANN-NLPLS is compared

favorably with PLS and BP-ANN. This should be due to the ANN-NLPLS method taking advantages of the merits of both PLS and ANN, that is, proper combination of many sigmoid functions can model effectively complex and unknown nonlinearity and PLS can maximize the covariance between data sets. And unlike BP-ANN, it does not using simple least-square (LS) regression to relate the outputs of the hidden layer to the bioactivities but using PLS which has better generalization ability than simple LS, so it is less vulnerable to overfitting problem. Moreover, the time required to operate this algorithm is just several minutes.

4.3. Data set 3

For further checking the performance of the proposed ANN-NLPLS algorithm, we applied this newly developed technique to calculate the BzR affinities of 52 PQs, and the PLS and BP-ANN were also tested as comparisons. The nine selected descriptors were used in these three models. The results of PLS and BP-ANN are shown in Table 3 and Fig. 2. The correlation coefficient of 0.9095 and 0.8847 are obtained from PLS for the training set and test set, respectively. The correlation between the observed and the calculated affinities by PLS is shown in Fig. 2a. As to BP-ANN, 10 samples picked from the training samples are used as the monitoring set for training BP-ANN to mitigate the probability of overfitting the training data. The model constructed by BP-ANN gave a correlation coefficient of 0.9427 and a RMSE of 0.4186 for the training set. Generalization of this model to the test compounds yielded a correlation coefficient of 0.8193 and RMSE of 0.5783. One notices that the correlation coefficient of the training set is much higher than that of the test set and the RMSE for the training set is much smaller than the test set, indicating that the ANN seems to generate an overfitted model. Fig. 2b shows the correlation between the observed and calculated values of affinity by BP-ANN. As shown in Fig. 2a, b and Table 3, the modeling errors are quite high and deviations in affinity prediction are rather large. Moreover, BP-ANN displays a sign of overfitting. It seems that PLS, together with BP-ANN, are not sufficiently accurate to predict the BzR affinities of PQs.

For this data set, the results of the ANN-NLPLS algorithm are also listed in Table 3. It is observed that ANN-NLPLS gives a correlation coefficient of 0.9755 for the training set and 0.9536 for the test set, indicating that the ANN-NLPLS method has good generalization ability. When compared with the BP-ANN model, the ANN-NLPLS provided comparable correlation coefficient for the training set but much higher correlation coefficient for the test set. By using the ANN-NLPLS, the RMSE for the traing set is reduced from 0.4976 by PLS and 0.4186 by BP-ANN to 0.2633. As compared to the RMSE values of 0.4482 by PLS and 0.5783 by BP-ANN for the compounds in the test set, a RMSE of 0.3238 is obtained by ANN-NLPLS, demonstrating the superior performance of ANN-NLPLS to PLS and BP-ANN methods. The affinity correlation plot obtained by ANN-NLPLS is revealed in Fig. 2c. A comparison of Fig. 2a–c shows that the best results are obtained by ANN-NLPLS.

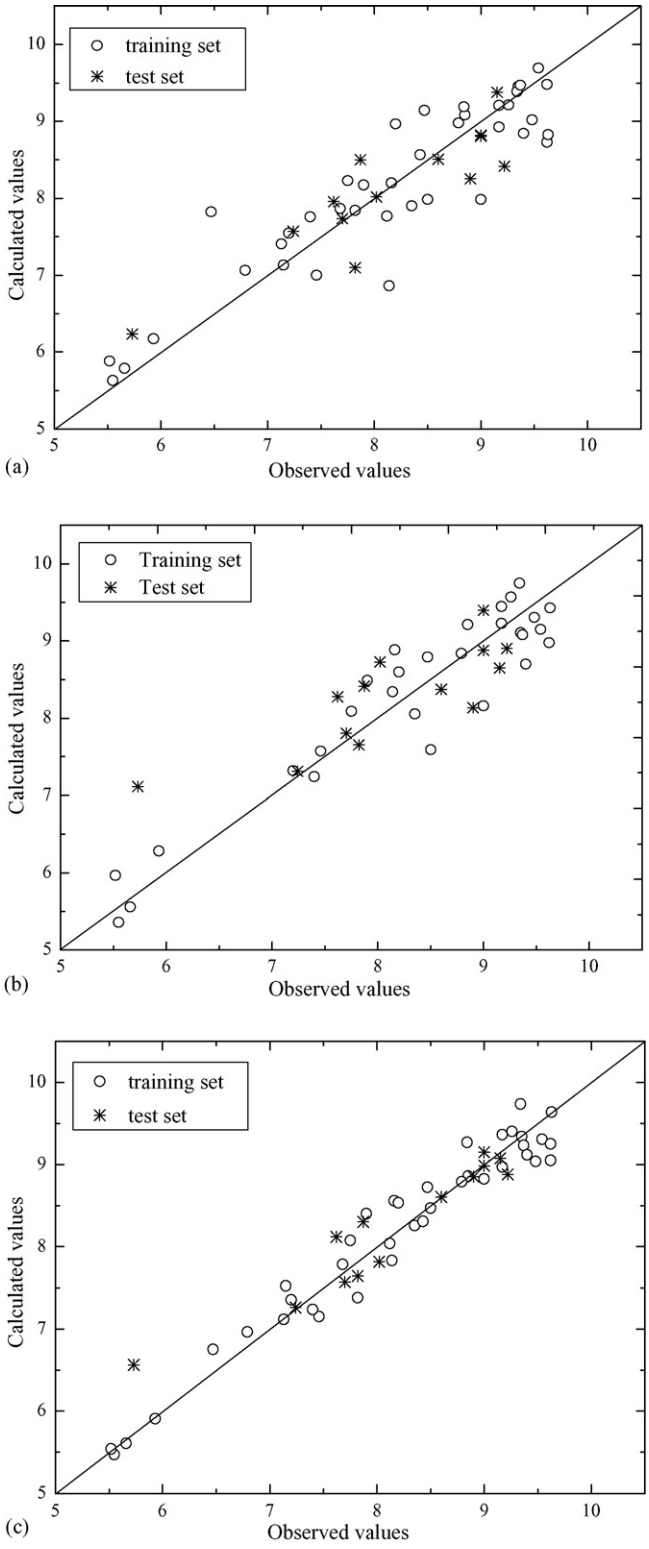

Fig. 2. (a) Calculated and observed values of affinity of 2-aryl(heteroaryl)-2,5-dihydropyrazolo[4,3-*c*]quinolin-3-(3*H*)-ones (PQs) by PLS modeling. (b) Calculated and observed values of affinity of PQs by BP-ANN modeling.(c) Calculated and observed values of affinity of PQs by ANN-NLPLS modeling.

Table 3
Results of QSAR analysis of PQs using ANN-NLPLS compared with those obtained by PLS and BP-ANN

Data set	*R* (correlation coefficient)			RMSE (root mean squared error)		
	Method 1[a]	Method 2[b]	Method 3[c]	Method 1[a]	Method 2[b]	Method 3[c]
Training set	0.9095	0.9427	0.9755	0.4976	0.4186	0.2633
Test set	0.8847	0.8193	0.9536	0.4482	0.5783	0.3238

[a] QSAR study by PLS.
[b] QSAR study by BP-ANN.
[c] QSAR study using ANN-NLPLS.

5. Conclusions

In this paper, a new version of nonlinear PLS algorithm (ANN-NLPLS) has been proposed by introducing the idea of nonlinear transformation in ANN. The idea behind ANN-NLPLS is that the descriptors are first transformed from the input layer into the hidden outputs following the ANN architecture, and then a model is constructed on the outputs of the hidden layer and the bioactivities of the chemical compounds by PLS. PSO has been applied to adaptively adjust the weights involved in ANN-NLPLS on the basis of PLS modeling with the *F*-staticstic to decide the number of latent variables automatically. Its performance has been demonstrated with a simulation data and two QSAR data sets and compared with those of PLS and BP-ANN. Results reveal that ANN-NLPLS offers a substantial improvement in the ability to model the nonlinearity while circumvent the overfitting frequently involved in nonlinear modeling.

Acknowledgement

This work was financially supported by National Natural Science Foundation of China (Grants No. 20375012, 20435010, 20205005).

References

[1] S. Wold, H. Martens, H. Wold, The multivariate calibration method in chemistry solved by the PLS method, in: A. Ruhe, B. KågstrÖm (Eds.), Proceedings on the Conference on Matrix Pencils, Lecture Notes in Mathematics, Springer-Verlag, Heidelberg, 1983, pp. 286–293.
[2] M.P. Freitas, J.A. Martins, Talanta 56 (2005) 182–186.
[3] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Chemom. Intell. Lab. Syst. 7 (1989) 53–65.
[4] S. Wold, Chemom. Intell. Lab. Syst. 14 (1992) 71–84.
[5] S.J. Qin, T.J. McAvoy, Comput. Chem. Eng. 16 (1992) 379–391.
[6] R. Rosipal, L.J. Trejo, J. Mach. Learn. Res. 2 (2001) 97–123.
[7] Y. Zhang, H. Li, A. Hou, J. Havel, Talanta 65 (2005) 118–128.
[8] B. Suchacz, M. Wesolowski, Talanta 69 (2006) 37–42.
[9] W.J. Melssen, L.M.C. Buydens, Anal. Proc. 32 (1995) 53–56.
[10] Q.Z. Lu, J.H. Jiang, R.Q. Yu, G.L. Shen, J. Chemput. Chem. 23 (2002) 1357–1365.
[11] Q. Shen, J.H. Jiang, C.X. Jiao, W.Q. Lin, G.L. Shen, R.Q. Yu, J. Comput. Chem. 25 (2004) 1726–1735.
[12] W.Q. Lin, J.H. Jiang, H.L. Wu, G.L. Shen, R.Q. Yu, J. Chem. Inf. Model. 45 (2005) 535–541.
[13] W.Q. Lin, J.H. Jiang, G.L. Shen, R.Q. Yu, J. Chem. Inf. Model. 45 (2005) 486–493.
[14] J. Kennedy, R. Eberhart, IEEE Int. Conf. Neural Netw. 4 (1995) 1942–1948.
[15] Y. Shi, R. Eberhart, IEEE World Congr. Comput. Intell. (1998) p69.
[16] M. Clerc, J. Kennedy, IEEE Trans. Evol. Comput. 6 (2002) 58.
[17] Y. Shi, R. Eberhart, Proc. Congr. Evol. Comput. (2001) 101.
[18] Q. Shen, J.H. Jiang, C.X. Jiao, G.L. Shen, R.Q. Yu, Eur. Pharm. Sci. 22 (2004) 145–152.
[19] K. Srikanth, C.A. Kumar, B. Ghosh, T. Jha, Bioorg. Med. Chem. 10 (2002) 2119–2131.
[20] S. Samanta, K. Srikanth, S. Banerjee, B. Debnath, S. Gayen, T. Jha, Bioorg. Med. Chem. 12 (2004) 1413–1423.
[21] J. Gasteiger, M. Marsali, Tetrahedron 36 (1980) 3219–3228.
[22] L.B. Kier, L.H. Hall, Pharm. Res. 7 (1990) 801–807.
[23] L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 31 (1991) 76–78.
[24] L.B. Kier, L.H. Hall, J.W. Frazer, J. Math. Chem. 7 (1991) 229–237.
[25] L.H. Hall, L.B. Kier, Quant. Struct.-Act. Relat. 10 (1991) 43–48.
[26] L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 35 (1995) 1039–1045.
[27] D. Hadjipavlou-Litina, R. Garg, C. Hansch, Chem. Rev. 104 (2004) 3751–3793.
[28] L. Savini, P. Massarelli, C. Nencini, C. Pellerano, G. Biggio, A. Maciocco, G. Tuligi, A. Carrieri, N. Cinone, A. Carotti, Bioorg. Med. Chem. 6 (1998) 389–399.
[29] D.T. Stanton, P.C. Jurs, Anal. Chem. 62 (1990) 2323–2329.
[30] V. Centner, O.E. de Noord, D.L. Massart, Anal. Chim. Acta 376 (1998) 153–168.