CHEMICAL ENGINEERING TRANSACTIONS
VOL. 66, 2018
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Songying Zhao, Yougang Sun, Ye Zhou

ISBN 978-88-95608-63-1; **ISSN** 2283-9216
DOI: 10.3303/CET1866139

Analytical Chemometrics Research Based on Artificial Intelligence Algorithm

Guohua Zou

School of Software, East China University of Technology, Nanchang 330013, China
ghzou@ecit.cn

This paper conducted analytical chemometrics research based on artificial intelligence algorithms. It combined the Ant Colony Optimization (ACO) with the fuzzy clustering to perform clustering twice to determine the hidden layer nodes of the radial basis function (RBF) network. By introducing the concept of fuzzy mathematics, it performed flexible partitioning of ant colony clustering algorithms. The ant colony optimization-fuzzy c-means-radial basis function network (ACO-FCM-RBF) was used to obtain satisfactory results for the simultaneous determination of bismuth and zirconium. In this paper, a new color system of DBS-chlorophosphonazo-zirconium was established, which lays the foundation for the simultaneous determination of bismuth and zirconium. This paper also established a new method of analytical chemometrics, compared with the test results of RBF network method, the ACO-FCM-RBF had better calculation result accuracy.

1. Introduction

Nowadays, the boundaries between disciplines have long been rather vague, multi disciplines infiltrate with each other to establish interdisciplinary subject has become a feature of the progress of today's subjects. Chemometries is a new branch of chemistry that appears at the intersection of chemistry, computer science, mathematics, and statistics (Kowalkowski et al., 2006). It effectively uses computer science, mathematics, statistics, and other theories and measures of relevant subjects to optimize the chemical measurement process and obtains practical chemical information from chemical measurement information as much as possible.

Artificial intelligence algorithms are a kind of iterative search algorithms that simulate or explain the development, evolution and variation of some natural phenomena (Cavus, 2010; Chen et al., 2015). The advantages of artificial intelligence algorithms are: they can perform global search, have fast convergence speed, high numerical accuracy, good generality, etc. (Hetmani et al., 2015; Kargarian et al., 2014; Kim et al., 2017). Currently widely used intelligent algorithms include Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), and neural network algorithm (Zhang et al., 2018). Radial Basis Function (RBF) neural network is a kind of forward-type network with excellent performance, it has been receiving much attention and has been widely used, and there have been many improvements and developments in recent years. The determination of hidden layer nodes in RBF networks is directly related to the accuracy of network prediction. Therefore, many scholars have proposed different determination methods. This paper connected the ACO with the fuzzy clustering, and used the ACO-FCM-RBF network system for the simultaneous detection of zirconium and bismuth to construct a new analytical chemometrics measures.

2. Basic theory and algorithm

2.1 Basic theory of ACO

ACO is a "natural" algorithm that appeared and inspired by the specific actions of various creatures in the natural world. It comes from the analysis of various actions of ant colonies (Ding et al., 2008; Ding et al., 2003; Saidi-Mehrabad et al., 2015). Ant colony optimization (ACO) is a key content of this system. The theory of ant colony system is shown in Figure 1 for details:

Please cite this article as: Zou G., 2018, Analytical chemometrics research based on artificial intelligence algorithm, Chemical Engineering Transactions, 66, 829-834 DOI:10.3303/CET1866139

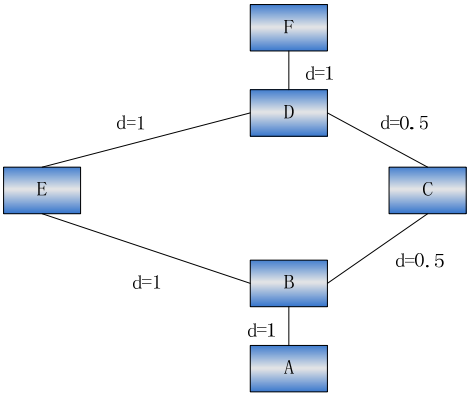

Figure 1: Principle of ACO

As shown in the figure, if A is an ant nest, F is a food source. 16 ants crawled from A to F, and then the 16 ants returned to A from F. At t=1, 16 ants are at B and D; at t=2, 8 ants are at B and D; 16 ants are at E. At this time, the distribution of the pheromones concentration on each route is: τ_{BE}=8, τ_{ED}=8, τ_{AB}=16, τ_{CD}=16, τ_{BC}=16, τ_{FD}=16, for the amount of pheromones on the routes, BCD is two times of BED. Therefore, with the progress of time, most ants pick the route BCD, by the end, all ants make the same selection to achieve the optimization of the process.

2.2 Basic theory of fuzzy clustering

Chapter 2 Clustering refers to the automatic classification of data with different degrees of similarity, so as to minimize the similarities between different classes, and maximize the similarity within the same class. The most widely used means clustering algorithm is the C-means clustering algorithm. The fuzzy C-means clustering (FCM) algorithm is an improvement of the common C-means clustering algorithm to make it fit better with the RBF networks. The algorithm is shown in Figure 2.

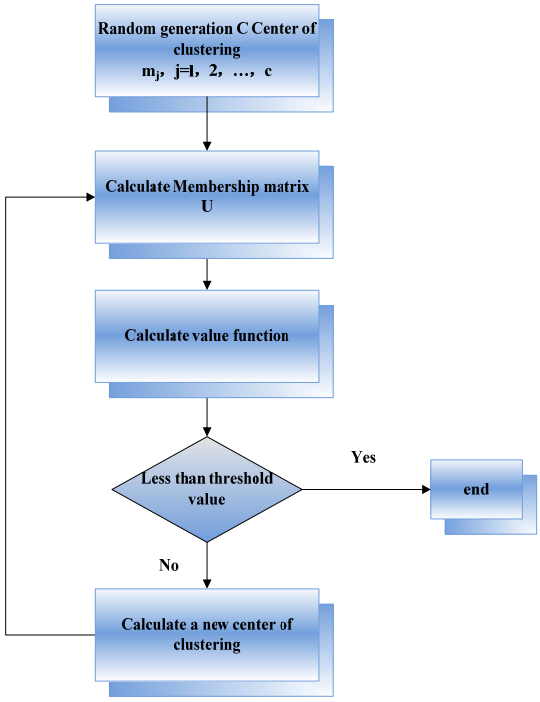

Figure 2: Improved fuzzy clustering methods

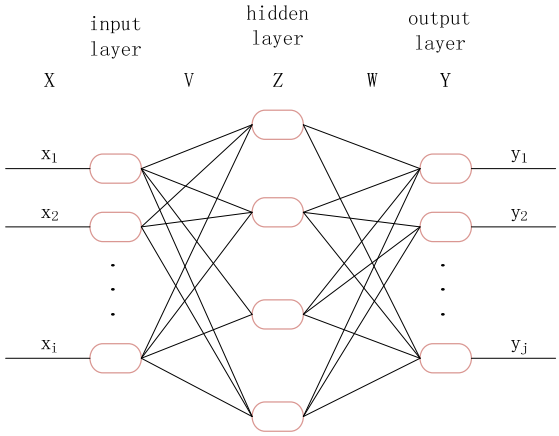

Figure 3: Structure of RBF neural network

2.3 Basic theory of RBF neural network

For Radial Basis Function (RBF) neural network, it is a forward artificial neural network and is composed of three layers of neurons: input layer, hidden layer, and output layer, its basic idea is: to use RBF as a basis of the hidden module to form the space range of the hidden layer, and to map the input vector straight into the

hidden space. The mapping of the hidden layer space range to the output layer space range is linear, that is, the output is the linear weighting of the hidden module output.
For the RBF neural network, its topological structure is shown in Figure 3 in detail:
The connotations of layers of this network are as follows:
(I) For the first layer, it is the input layer, the input layer neurons only functions effectively as a linkage, and does not transform the signal.
(2) For the second layer, it is the hidden layer. Assume the correlation coefficient from the i-th neuron in the input layer to the j-th neuron in the hidden layer is v_{ij}($1\leq i\leq I, 1\leq j\leq H$), the correlation coefficient vector from input layer neurons to the j-th neuron in the hidden layer is $V_j=(v_{1j},v_{2j},\ldots,v_{ij})^t$ ($1\leq j\leq H$), the transition function of hidden layer neuron is a Gaussian function, and the current state of the corresponding input x of the j-th neuron in the hidden layer is: $Z_j = k\|X - V_j\| = \exp\left(-\sum x_i - v_{ij}\right)^2/(2\sigma_j^2)$, where σ_j($1\leq j\leq H$) is the specific width of the basis function of the j-th neuron in the hidden layer.
(3) For the third layer, it is the output layer. Assume the j-th neuron in the hidden layer is connected to the K-th neuron in the output layer, its specific actual coefficient is w_{jk}($1\leq j\leq H, 1\leq K\leq O$), and the correlation coefficient vector from the hidden layer neuron to the output layer K-th neuron is $W_k=(w_{1k},w_{2k},\ldots,w_{HK})^t$, ($1\leq K\leq O$).

3. Algorithm improvement

The research algorithm of RBF network focuses on: the nearest neighbor clustering research algorithm, self-starting research algorithm, ACO, randomized algorithm, and fuzzy clustering algorithm, this is used to select the main body of RBF. The paper uses fuzzy clustering and ACO to perform clustering twice to clearly determine the main body of RBF, it's called ACO-FCM-RBF network research algorithm.
The actual steps of the research algorithm of the RBF network are as follows:

3.1 Determine the main body of the basis function

The actual algorithm is described below:
I) First randomly selecting M typical points or selecting according to the experience. This paper uses the random selecting measures to determine the original typical points;
2) Initializing parameters: Assume N, ε_0, r, m, $\tau_s(0)$ = 0, β, α, M, P_0, η;
3) Calculating Euclidean distance d_{ij}: $d_{ij} = \|P(X_i - X_j)\|^2 = \sqrt{\sum_{k=1}^{m} P_k(x_{ik} - x_{jk})^2}$;
4) Calculating the amount of pheromones of each route: $\tau_{ij}(t) = \begin{cases}1, d_{ij}\leq r\\ 0, d_{ij} > r\end{cases}$
5) Calculating the probability combining X_i to X_j: $P_{ij} = \frac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{s\in S}\tau_{si}^{\alpha}(t)\eta_{ij}^{\beta}(t)}$
6) Judging whether $P_{ij}(t)\geq P_0$ is true; if it's true, proceed next step, else i=i+1, go to step 3;
7) Calculating the new clustering main body $\overline{vj} = \frac{1}{J}\sum_{k=1}^{J} X_k, X_k \in v_j$;
8) Calculating deviation errors and global errors: the deviation of the j-th clustering is:
$D_j = \sum_{k=1}^{J}\sqrt{\sum_{k=1}^{m} P_k(x_{kj} - v_{ji})^2}$
Where v_{ji} represents the i-th component in the j-th cluster. The global error of the calculation is: $\varepsilon = \sum_{j=1}^{k} D_j$.
9) Judging whether $\varepsilon\leq\varepsilon_0$ is true. If it's true, stop the calculation, and output the results; else go to step 3 and proceed effective iteration;
10) Through effective ant colony clustering, the actual number of clusters c (2≤c≤n) is obtained, and the conclusion of the ant colony clustering is taken as the original reflection point of fuzzy clustering, then further select the specific index weight m'(1<m'<∞);
11) Calculating the actual matrix U it belongs to;
12) Calculating the actual value function. If it is smaller than a certain threshold, or compare to its previous value function, its change value is smaller than a certain threshold, then the algorithm stops naturally;
13) Calculating the new clustering main body, and then return to step 11;
14) Effectively output the ultimate clustering main body.

3.2 Determine the basis function center width σ_j

The basis function center represents a certain measured quantity of the samples in the sub-sample set associated with each center, and many methods can be determined clearly. This paper uses the following formula to determine:

$$\sigma j = \sum_{j=1}^{N} d_{ij} u_{i\,j} \tag{1}$$

Where, d_{ij} is the actual Euclidean distance from sample x_i to the j-th clustering main body. U_{ij} is the specific degree of membership of sample x_i to the j-th cluster.

3.3 Specify the correlation weight w_{ij} from the hidden layer module to the output layer module

If we mark the corresponding expected output of x^u_i as y^u_i, then the average error measurement standard is a smaller square error, that is:

$$E = \frac{1}{2}\sum_{u=1}^{P}\sum_{i=1}^{M}(y_i^u - O_i^u)^2 \tag{2}$$

After determine the parameters of the basis function, E in the above formula belongs to the function of W_{ij}, and the solution can be obtained by using the least squares method.

This paper uses mathematical models and calculations, and uses the MATLAB language and C/C++ language to program the process.

4. Test results and discussion

4.1 Main equipment and reagents

Table 1: Main reagents and equipment

Reagents and equipments	SPEC	Concentration
ultraviolet and visible spectrophotometer	UV-265 (Shimadzu)	-
Bi	AR	5μ/L
$ZrOCl_2 \cdot 8H_2O$	AR	5μ/L
Cu	AR	5μ/L
DBS- chlorophosphonazo	AR	0.5g/L
H_2SO_4	AR	1.0mol/L
Ethanol	AR	99.5%

4.2 Test methods

Accurately absorb a certain amount of zirconium and bismuth standard solution in a 25-mL volumetric flask, and gradually add 0.5 g/LDBS-chlorophosphonazo 2.2 mL; l.0 mol/L sulfuric acid 7.0 mL; absolute ethanol 2.5 mL, add water and dilute to a certain scale, shake it to make it uniform, effective color reaction for 60 minutes, and then use the blank reagent as a control, use 1cm cuvette, in the 599nm ~ 656nm wavelength range, detect the absorbance every 3 nm, calculate by using ACO-FCM-RBF and RBF methods respectively.

The original parameters of ACO-FCM-RBF are: N=100, ε_0=0, r=0.l, m=2, $\tau_s(0)$=0, β=1, P_0=0.6, α=1, M=2 , η=0.9.

4.3 Results and discussion

4.3.1 Absorption spectrum

In the H_2SO_4 medium, the larger absorption wavelength of the blank reagent is substantially at 525 nm, and for the larger absorption wavelength of bismuth and the color developing agent complex, it is specifically at 638 nm; while for the larger absorption wavelength of zirconium and the color developing agent complex, it is specifically at 636 nm, their actual contrast reaches113 nm and 111 nm, respectively, the complex absorption wavelength is extremely overlapped. As shown in Figure 4.

4.3.2 Color conditions

In the luminosity study test, a new same color was developed, that is the same color with DBS-chlorophosphonazo-zirconium, and no relevant reports were found in the literature.

In a series of media such as perchloric acid, hydrochloric acid, phosphoric acid, sulfuric acid, and nitric acid, the sensitivity of color development in sulfuric acid medium is very high, Under the actual conditions of simultaneous detection, when the amount of 1.0 mol/L sulfuric acid is 0.5-12.0 mL, the absorbance value of the complex is large and very stable, the actual amount of sulfuric acid used in the test is 7.0mL.

Under the condition of simultaneous detection, in 1.0mol/L sulfuric acid medium, when the amount of 0.5g/L DBS-chlorophosphine solution is 1.0-3.0mL, the absorbance value reaches the maximum and is very stable. The test chooses to use color developing agent, and its actual amount reaches 2.2mL.

Increase the sensitivity through the use of alcohol to change the feature of the media, add a small amount of absolute ethanol into 1.0 mol/L sulfuric acid media, make the complex appear and move evenly, so as to achieve sensitivity increasing effect, the sensitivity increases by approximately 37%, when the amount of absolute ethanol reaches 0.510mL, the absorbance of the complex is large and very stable, the actual amount of ethanol used in the test is 2.5mL.

At room temperature, zirconium and bismuth reacted with DBS-chlorophosphonazo. After 60 minutes of color development, a blue-violet complex appeared. It was determined by the test that the complex was stable for at least 5.5 hours, and the absorbance remained unchanged.

In the range of 0~20μg/25mL, bismuth conforms to the Lambert-Beer law, the actual linear equation is: A=0.01736+0.0l3l02C, r=0.9995. The apparent molar absorption coefficient is 7.9×10^4L·mol^{-1}cm^{-l}.

In the range of 0~10μg/25mL, Zirconium conforms to the Lambert-Beer law, the actual linear equation is: A=0.023816+0.00699C, r=0.9997. The apparent molar absorption coefficient is 2.6×10^4L·mol^{-1}cm^{-l}.

According to test measures, 10μg bismuth(zirconium)/25mL is determined, and the relative error of absorbance ≤5%, below are the co-existed ions (calculated with mg): Cu^{2+}(1.1), Ni^{2+}(0.9), Co^{2+}(0.5), Mg^{2+}(0.45), Mn^{2+}(0.4), KNO_3(1.5), Ca^{2+}(1.1), Si^{4+}(0.12), Zn^{2+}(5), ascorbic acid (6), which did not interfere the test.

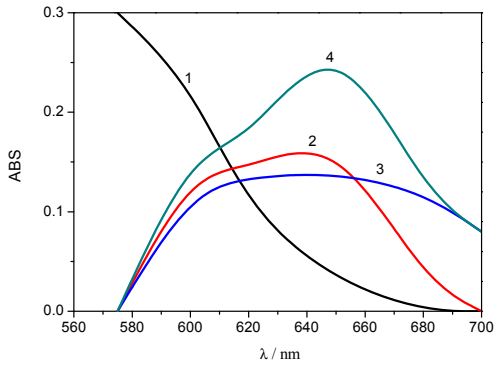

Figure 4: Absorption spectrum (1-blank, 2-Bi 10μg/25mL, 3-Zr 10μg/25mL, 4- Bi 10μg/25mL+ Zr 10μg/25mL)

4.4 Study of the synthetic sample

Table 2: Analytical results of mixed standard solution

No.	addition		measuredvaluebyRBF		measuredvaluebyACO-FCM-RBF	
	Bi	Zr	Bi	Zr	Bi	Zr
1	4.0	2.0	4.09	2.06	4.06	2.05
2	4.0	6.0	4.15	5.92	3.89	5.98
3	8.0	2.0	7.64	1.96	8.29	1.97
4	8.0	6.0	8.12	6.10	7.94	5.93
5	12.0	2.0	11.57	2.03	11.77	2.00
6	12.0	4.0	11.68	3.89	11.80	3.91
Relativemeanerror (%)			3.07	3.12	2.04	1.33

According to the test method, 9 sets of mixed standard solutions were matched with the orthogonal table L9 (3^4) as the training set, it was evenly distributed and neatly comparable; besides, 6 sets of samples were made as the test set. The absorbance value of the sample is taken as the input information, the amount of zirconium and bismuth in the sample is the network output, the detailed conclusion of the study is shown in Table 2.

4.5 The discussion of the algorithm

In this paper, the ACO-FCM-RBF algorithm and RBF algorithm were applied to the simultaneous detection of zirconium and bismuth, two new types of chemometrics measures were constructed and these two types of

algorithms were measured and compared with synthetic samples. Table 2 shows that the ACO-FCM-RBF algorithm error is significantly reduced compared to the RBF algorithm.

This paper used ACO to carry out each specific clustering, relied on the specific effects brought by the ant colony, and then used the fuzzy clustering to specifically expand secondary clustering. Although the ant colony clustering algorithm can present the optimal number of actual nodes, it is a kind of means clustering, in other words, the cluster brought by the ACO practically belongs to a kind of hard partitioning, so we specifically introduced the fuzzy clustering to make flexible fuzzy partitioning of the actual conclusion of the clustering in substantive sense.

With respect to the determination methods of the hidden layer nodes in the RBF network, it usually uses the typical C-mean clustering algorithm. Although it's a kind of clustering algorithm with good performance, however, for this algorithm's selection sensitivity of the original clustering main body, how to pick the original clustering main body directly determines the ultimate clustering conclusion. If the selection of original clustering main body is not good, or it may be caught into the actual situation of local extremum. Generally, ACO will not trap in a local minimum situation, although for the fuzzy clustering algorithm, it is quite sensitive in the original clustering point, however, clustering under the premise of ant colony clustering can prevent this situation. For C-means clustering, the number is artificially determined. For the ACO clustering, it is determined specifically by the algorithm in the operation, not determined artificially, by the end of this papery, it determines that the number of nodes in the hidden layer module is 7. Because the algorithm can independently determine the actual node number in the hidden layer, which avoids the interference of human factors

5. Conclusion

(1) This paper constructed a new ant colony optimization-fuzzy c-means clustering-radial basis function neural network (ACO-FCM-RBF) model and established a new analytical chemometrics method.

(2) This paper made a comparison between the results of (ACO-FCM-RBF) model and RBF network method, the calculation results of ACO-FCM-RBF neural network were better in the accuracy.

(3) A new DBS-chlorophosphonazo-zirconium color system was established for the simultaneous detection of zirconium and bismuth using the spectrophotometric determination.

Reference

Cavus N., 2010, The evaluation of learning management systems using an artificial intelligence fuzzy logic algorithm, Advances in Engineering Software, 41(2), 248-254, DOI: 10.1016/j.advengsoft.2009.07.009.

Chen Y., Yang Y., Liu C., Li C., & Li L., 2015, A hybrid application algorithm based on the support vector machine and artificial intelligence: an example of electric load forecasting, Applied Mathematical Modelling, 39(9), 2617-2632, DOI: 10.1016/j.apm.2014.10.065.

Ding Y., Wu Q., Su Q., 2008, Chemical ant colony algorithm with supramolecular coefficient and multivariate calibration to calix[n]arenas (n = 4, 6, 8) supramolecular system, Journal of Chemometrics, 22(6), 366-371, DOI: 10.1002/cem.1119.

Ding Y.P., Su Q.D., Wu Q.S., 2003. Ant colony algorithm and optimization of test conditions in analytical chemistry, Chinese Journal of Chemistry, 21(6), 607-609, DOI: 10.1002/cjoc.20030210606.

Hetmaniok E., Słota D., Zielonka A., 2015, Restoration of the cooling conditions in a three-dimensional continuous casting process using artificial intelligence algorithms, Applied Mathematical Modelling, 39(16), 4797-4807, DOI: 10.1016/j.apm.2015.03.056.

Kargarian A., Raoofat M., Mohammadi M., Rezazadeh Z., 2014, Artificial intelligence-based loss allocation algorithm in open access environment, Journal of Energy Engineering, 140(2), 86-91, DOI: 10.1061/(asce)ey.1943-7897.0000172.

Kim S.H., Seo K., Kang S.H., Bae S., Kwak H.J., Hong J.W., Hwang Y., Kang S.M., Choi H.R., Kim G.Y., Lee Y., 2017, Study on feasibility for artificial intelligence (ai) noise reduction algorithm with various parameters in pediatric abdominal radio-magnetic computed tomography (ct), Journal of Magnetics, 22(4), 570-578, DOI: 10.4283/jmag.2017.22.4.570.

Kowalkowski T., Zbytniewski R., Szpejna J., Buszewski B., 2006, Application of chemometrics in river water classification, Water Research, 40(4), 744-752. DOI: 10.1016/j.watres.2005.11.042.

Saidi-Mehrabad M., Dehnavi-Arani S., Evazabadian F., Mahmoodian V., 2015, An ant colony algorithm (aca) for solving the new integrated model of job shop scheduling and conflict-free routing of agvs, Computers & Industrial Engineering, 86(C), 2-13, DOI: 10.1016/j.cie.2015.01.003.

Zhang X., Wang Y., Liu C., Chen Z., 2018, A novel approach of battery pack state of health estimation using artificial intelligence optimization algorithm, Journal of Power Sources, 376, 191-199, DOI: 10.1016/j.jpowsour.2017.11.068.