

APPLICATION OF GREY RELATION ANALYSIS AND RBF NETWORK ON GRINDING-CONCENTRATION'S SOFT SENSING

Yong Li
Cheng Shao

*Institute of Advanced Control Technology,
Dalian University of Technology, China*

Abstract: As an internal parameter of ball milling process, grinding concentration (GC) is not able to be measured directly. According to ball mill's principle, based on radial base function (RBF) network, a soft sensor method is presented to estimate the ball mill's GC. And a novel grey relation analysis method as a way to determine the secondary variables is proposed. Simulation results demonstrate that the design possesses high accuracy and can meet the practical demands. *Copyright © 2005 IFAC*

Keywords: radial base function networks, soft sensing, data processing, optimization, disturbance

1. INTRODUCTION

Mineral processing consists of four steps -- crushing, grinding, classifying, floating. Among them, grinding, being an imperative working procedure to fine ore, guarantees fluently running of the following steps. In the past, the control of grinding process was mainly on the basis of measurements of some external parameters, such as feed rate, water flow, input power and so on, which resulted in poor control performance, unqualified products and high cost of steel and power. In recent years, through plentiful experiments, some internal parameters of ball mill -- the main equipment of grinding process -- have proved their vital effect on quantity of production and granularity of products. Grinding concentration (GC) is one of the most important parameters. Presently, among numerous grinding control methods based on ball mill's internal parameters' optimization, most consider only charge ratio of me-

dia, some also take account of ratio of material to media. However, almost none of them conceive GC, which results in their poor performance when these designs are put to use in practice.

The efficiency of ball mill and the qualified products' quantity can be greatly improved, as well the cost of resources being reduced, by optimal control of GC. Meanwhile, combining GC's and other internal parameters' optimal control can essentially enhance the control quality of grinding process, realizing genuine automation of grinding process. However, because the grinding process is too much complex on mechanism and interfered by many factors, there are no effective direct methods yet to measure GC online. Some indirect methods, such as empirical formula theme (Cui, 1994) based on external parameters -- mill's size, throughput per unit per hour, retuning sand flow -- of ball mill and three-factors means deriving from mathematics mechanism, can't achieve satisfying results for either lacking enough external parameters or unobtainable mechanism knowledge and heavy burden of calculation.

Artificial neural network (ANN) is a data based non-mechanism modeling method. It possesses strong ability to approximate nonlinear function. Nowadays, with the arising and prevalence of soft-sensor technology, using the ANNs to set up soft-sensor models has gotten extensive attention. The author has done some research work on ANN, actually on RBF network, and presents RBF

¹The work was supported by State Science & Technology Pursuing Project of China (2001BA204B01); Supported by Science & Technology Pursuing Project of Liaoning Province

Y. Li is with Institute of Advanced Control Technology, Dalian University of Technology, China. Corresponding phone: 086-411-84702131; fax: 086-411-84707579; e-mail: leeyon2004@yahoo.com.cn

network-based theme to predict the GC of ball mill in practical processes according to ball mill's external measurable parameters, as Fig. 1.

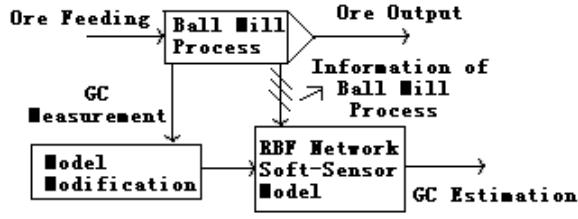


Fig. 1. Structure of GC soft sensor

2. GREY RELATION ANALYSIS (GRA)

Grey relation analysis (Deng, 2002) is the primary content of grey theory. Compared with traditional similarity analysis methods, such as correlation coefficient method, absolute subtrahend algorithm and so on (Yang, 2001), GRA is more concise and needs little calculation. Furthermore, GRA has no strict demand on the number of samples as well as on typical distributing rule. Also its results of quantitative analysis and qualitative analysis are generally coincident. So in the cases of analyzing grey system with uncertain knowledge, GRA shows incomparable superiority. Grey relation analysis' basic principle is to tell the incidence degrees (IDs) of different factors on the basis of curves' or serial data's resemblant degrees. Grey incidence degrees describe the correlations between factors' varying tendencies, such as magnitudes, directions and velocities of variations, in the procedure of system's operation. If there are correlations between the trends of variations of diverse factors, a certain IDs can be used to expound their relation. According to the magnitudes of factors' IDs, the correlations between the factors are able to be numerically determined, and multifactor advantage analysis is within reach by using relation matrix. Presently, there are a lot of themes can be used to calculate ID, such as Deng-style ID, slope ID, similarity ID (Ma, 2000) and T-shaped ID (Tang, 1995). These themes have been pervasively employed in many fields such as industry control, agriculture, management, etc.

However, in some industries, using existing GRA methods isn't proper, sometimes even false. For example, In mineral process, almost all the experiment data are vibratory, but Deng-style ID (Deng, 2002), being a usual applied way, described as equation (1) (2) (where x_i and x_j represent two initialized variables, x_i measurable and x_j not; i and j are the serial numbers of two kinds of variables; k is the serial number of samples; m delegates the number of samples; μ_{ij} is the correlation coefficient of two variables; r_{ij} is ID), only takes account of the distance between two factors-series, doesn't consider the swing and the varying trends of them.

$$\mu_{ij}(k) = \frac{\min_j \min_k |x_i(k) - x_j(k)| + \alpha \max_j \max_k |x_i(k) - x_j(k)|}{|x_i(k) - x_j(k)| + \alpha \max_j \max_k |x_i(k) - x_j(k)|} \quad (1)$$

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m \mu_{ij}(k) \quad (2)$$

And other ID solutions, such as slope ID and similarity ID, fail to solve the calculating imbalance between plus ID and minus ID, as well the ambiguous affiliation of zero ID.

So for conquering above shortcomings, an improved solution which, based on the varying trend and resemblant degree of data curves, can be used to analyze the correlation degree between variables is proposed by the author. It's named as partial correlation incidence degree (PCID). Its computation is shown as follows:

$$\mu_{ij}(k) = \frac{1}{1 + \left| |\Delta x_i(k)| - |\Delta x_j(k)| \right|} \quad (3)$$

$$\Delta x_i(k) = \frac{x_i(k+1) - x_i(k)}{x_i(1)} \quad (4)$$

$$\Delta x_j(k) = \frac{x_j(k+1) - x_j(k)}{x_j(1)} \quad (5)$$

Defining ξ_k as sign factor, then:

If $\Delta x_i(k)\Delta x_j(k) > 0$ and $\Delta x_i(k) = \Delta x_j(k) = 0$, then x_i and x_j have the same varying trend at point k , $\xi_k = 1$, and ID of point k is plus;

If $\Delta x_i(k)\Delta x_j(k) = 0$ and $\Delta x_i(k) \neq \Delta x_j(k)$, then x_i and x_j have no relation each other at point k , $\xi_k = 0$, and there is no contribution to total ID;

If $\Delta x_i(k)\Delta x_j(k) < 0$, then x_i and x_j have the contrary varying trend at point k , $\xi_k = -1$, and ID of point k is minus.

$$r_{ij} = \left| \frac{1}{m-1} \sum_{k=1}^{m-1} \xi_k \mu_{ij}(k) \right| \quad (6)$$

Suppose the two variables-series x_i and x_j have m_1 points in the same tendency, numbered as $\{t_1 \cdots t_{m_1}\}$; m_2 points in irrelative tendency,

numbered as $\{s_1 \cdots s_{m_2}\}$; m_3 points in contrary tendency, numbered as $\{d_1 \cdots d_{m_3}\}$. And

$$P_{ij} = \frac{1}{m-1} \sum_{k_1=t_1}^{t_{m_1}} \xi_{k_1} \mu_{ij}(k_1) \quad (7)$$

$$Z_{ij} = \frac{1}{m-1} \sum_{k_2=s_1}^{s_{m_2}} \xi_{k_2} \mu_{ij}(k_2) \quad (8)$$

$$N_{ij} = \frac{1}{m-1} \sum_{k_3=d_1}^{d_{m_3}} \xi_{k_3} \mu_{ij}(k_3) \quad (9)$$

$$m_1 + m_2 + m_3 = m - 1 \quad (10)$$

$$r_{ij} = \left| \frac{1}{m-1} \sum_{k=1}^{m-1} \xi_k \mu_{ij}(k) \right| = |P_{ij} + N_{ij}| \quad (11)$$

Then P_{ij} , Z_{ij} and N_{ij} respectively represent plus ID, zero ID and minus ID.

When $P_{ij} > |N_{ij}|$, the two variables are in the relation of plus correlation, which means they vary in the similar trend. The correlation degree can be determined by the magnitudes of P_{ij} and r_{ij} ;

when $r_{ij} = Z_{ij} = 0$, there is no relation between the two variables; When $P_{ij} < |N_{ij}|$, the two variables are minus correlative, which indicates they are in the contrary changing trend. And the correlation degree are confirmed by $|N_{ij}|$ and r_{ij} . What is worth noting is, in the processes of estimation and prediction, minus relationship also is a sort of correlation. Through erecting appropriate model, minus relational variables can embody each other.

3. RBF NETWORK

Artificial neural network (ANN) is a nonlinear, large scale and dynamic system. It's characterized by distributed store of information and collateral data processing. ANN system possesses collective-computation capability, adaptive learning ability and broad error tolerance. Researchers, by applying ANNs, can realize function approximation, data clustering, pattern identification, optimization calculation, etc. So ANNs are widely being applied in many information-processing fields, such as artificial intelligence, automation, robot, statistics and so on. And on the part of function approximation, RBF network is one of main ways.

RBF network is a partial approximation network. It has the ability to accept overlapping domains knowledge and partial modification capability of

biology. Its essential principle is to transform the original low-dimension space into high-dimension linear dividable space and conduct clustering by the RBFs of hidden layer. For RBF's denseness in space, RBF network can approximate any function in the form of combination.

In the course of modeling, efficient training is the key. Generally, RBF's training is divided into two steps. In the first step, nonsupervision learning is applied to train the RBF network for pinpointing the number and locations of RBF s' centers; in the second step, supervision learning is applied to train the weights of linear output layer.

There are some methods being usually used to locate the centers of RBFs, such as orthogonal least squares (OLS) (Chen, et al, 1991), K-mean clustering, competitive learning, etc. Among them, OLS is more efficient. It derives from Linear Regression Model and has the ability to choose the required hidden nodes continually in the process of weights learning. Using OLS as training algorithm, RBF network can converge more quickly and get the unique solution of the optimization.

$\{x_n, d(n)\}$ ($n=1, 2, \dots, N$) represents a sample group. Where N is the number of samples; $x_n \in R^n$ is seen as the input vector and the output vector is assumed as $d(n) \in R^1$. Then the desired output is depicted as:

$$d(n) = \sum_{j=1}^M v_j(n) \omega_j + e(n) \quad (12)$$

Where $j = 1, 2, \dots, M$; M is the number of hidden units; ω_j is the weight from the j^{th} hide unit to output; $e(n)$ is the residual error; $v_j(n)$ figures the regression operator which can be expanded as follow:

$$v_j(n) = G(\|x_n - c_j\|) \quad (13)$$

$(n = 1, 2, \dots, N; j = 1, 2, \dots, M)$

In which c_j represents the center vector of the j^{th} hide unit; and Gaussian function is delegated by G . So (12) also equals to the matrix form:

$$d = VW + e \quad (14)$$

In the equation, $V = [V_1, V_2, \dots, V_M]$, $V_j = [V_j(1), V_j(2), \dots, V_j(N)]^T$, V delegates a regression matrix.

The selection of operator V_j is the key point when solving (14). OLS orthogonalizes V_j by using orthogonal optimum seeking method ($1 \leq j \leq M$), and analyzes the contribution of V_j to the reduction

of residual error. After that, those operators possessing apparent contribution are retained and those poor ones are kicked out. Finally, M , the number of hidden nodes, as well as that of V_j , is determined. The concrete inference is depicted as follows:

Orthogonal triangle decomposition of V is given as:

$$V = UA \quad (15)$$

In which A is in the form of $M \times M$ up-triangle matrix, its diagonal cells are 1.

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1M} \\ 0 & 1 & a_{23} & \cdots & a_{2M} \\ 0 & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & 1 & a_{M-1,M} \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (16)$$

U is $M \times N$ matrix with all its column orthogonal each other.

$$U^T U = H \quad (17)$$

H is a diagonal matrix, whose diagonal cell is h_j .

$$h_j = u_j^T u_j = \sum_{n=1}^N u_j^2(n) \quad (18)$$

Substitute equation (15) into (14), then

$$d = UAW + e = Ug + e \quad (19)$$

$$g = AW \quad (20)$$

Its least square solution is:

$$g = H^{-1}U^T d \quad (21)$$

Or

$$g_j = \frac{u_j^T d}{u_j^T u_j} \quad (1 \leq j \leq M) \quad (22)$$

In the equation, g_j is a component of vector g . g and W ought to satisfy the following equation:

$$AW = g \quad (23)$$

Traditional orthogonalizing method Gram-Schmidt is put into use here. Each time one column of A will be calculated and orthogonalized as follow:

$$u_j = v_j$$

$$\alpha_{jk} = \frac{u_j^T v_k}{u_j^T u_j} \quad (24)$$

$$u_k = v_k - \sum_{i=1}^{k-1} \alpha_{ik} u_i$$

$$(1 \leq j \leq k; k = 2, \dots, M)$$

According to U 's orthogonality, while vectors g and e are irrelevant, corresponding output energy can be depicted as:

$$d^T d = \sum_{j=1}^M g_j^2 u_j^T u_j + e^T e \quad (25)$$

Both sides of above equation divide $d^T d$, then

$$1 - \sum_{j=1}^M \varepsilon_j = q \quad (26)$$

$$\varepsilon_j = \frac{g_j^2 u_j^T u_j}{d^T d} \quad (1 \leq j \leq M) \quad (27)$$

In which ε_j represents error compression ratio, while $q = e^T e / d^T d$ is relative quadric error. By equation (26), the bigger ε_j is, the smaller q is.

By equation (22), g_j only has relation with d and u_j , so does ε_j . For d is known, regressive factor u_j can be selected to make ε_j as big as possible based on equation (27). And seeking

$$j_k = \max_j \{ \varepsilon_j, 1 \leq j \leq M, j \neq j_1, \dots, j \neq j_{k-1} \} \quad (28)$$

The corresponding regressive factor u_{j_k} has the biggest contribution to ε_j . Analogically, every factor u_j can be worked out one after another till ε_j reaches following demand:

$$1 - \sum_{j=1}^M \varepsilon_j > \rho \quad (29)$$

Where $0 < \rho < 1$ is supposed tolerated error. And finally, the number of regressive factors M is confirmed. Meanwhile, on the basis of equation (20), the weight W can be figured out by LS.

Through above inference, after OLS orthogonalizes the regressive matrix, regressive factors are irrelevant each other, so pattern adjustment only brings the least disturbance to the existing section. Hence as far as function approximation is concerned, OLS is better than k-means clustering and random selection

on the part of computation time and approximating performance.

4. SOFT-SENSOR TECHNOLOGY (SST)

Owing to the reasons of technologies and economy, there are a lot of important parameters can't be measured easily or even at all in industry processes. These parameters, however, need to be controlled strictly because of their strong relation with the quality of products and the efficiency of production. Being a new means to solve this problem, SST (or soft-analyzer technology) has gotten increasing attention, and it has been applied to real-time evaluation, failure redundance, intelligent modification and compound multi-employment. Currently, making use of computer systems to construct soft-sensor has been put into practical projects (Mejdell and Skogestad, 1993; Yu, et al, 2000).

5. PREDICTION OF GC

In allusion to the need of on-line measurement of ball mill's GC in grinding process, and considering that there are no satisfying methods which can be used in practice, the author, comprehensively applying the techniques mentioned in above chapters, conducts the research on soft-sensor of GC based on RBF network.

5.1 Variables Selection

There are many measurable parameters and variables in the procedure of grinding. According to the principle of soft-sensor, taking a ball mill in gauge of $\phi 350 \times 450$ mm as experiment equipment, the paper analyzes the correlations between the measurable variables of grinding process and GC by PCID introduced above in view of grinding system being a typical grey system.

The result of correlation analysis shows, when the satisfaction threshold value is assigned to 0.6, the ball mill's external parameters— sound, pressure, input power (IP) and rotational speed ratio of ball mill (RSRBM)—have qualified ID with the primary variable GC. Therefore, combining practical experience, the author selects these four variables as the secondary variables of soft sensor model. Table 1 shows the original data (Wang, 2001).

Table 1 Part of experiment data

Data	Pressure (N)	Sound (DB)	IP (W)	RSRBM (%)	GC (%)
1	417	73.5	139.83	67	78
2	428	77.4	143.54	75	82
3	400	70.9	146.31	75	74
4	514	84.2	183.53	83	80
5	465	85.2	783.58	85	70
6	511	84.7	188.8	85	76

7	427	80.1	155.95	85	82
8	474	84.3	170.74	79	72
9	484	89.2	162.02	71	76

5.2 RBF Network Model

Taking the sound, pressure, IP and RSRBM as the inputs of input layer, Gaussian functions as the hidden layer's functions, GC as the output, the author builds a RBF network.

In the experiment, there exist some noises in gauging signals, so, before training, the data in Table 1 are denoised by wavelet transform to enhance training quality. Then, for eliminating the dimensional difference between different variables, original data are normalized. Comparing different normalizing methods, according to the features of data, the author chooses the extremum normalization which makes training speed much quicker than using ordinary standard error normalization.

After erected, the network is trained by using OLS algorithm introduced above. And the training goal is set to 1×10^{-6} . Results of the experiment, shown in Table 2, indicate that RBF network can well predict the GC with the maximum relative error smaller than 1.2% and mean relative error smaller than 0.8%. (In Table 2, real values of GC in Table 1 have been denoised by wavelet analysis)

Table 2 Part of estimation value and error

Data	Real value of GC (%)	Estimation of GC (%)	Relative error absolute value (%)
1	78.563	79.113	0.70
2	78.795	78.054	0.94
3	75.535	74.979	0.74
4	77.826	78.136	0.40
5	73.697	74.327	0.85
6	75.497	75.017	0.64
7	77.944	77.321	0.80
8	74.446	75.264	1.10
9	74.764	74.131	0.85

6. DISCUSSION

For comparison, the author also constructs model using a three layers back-propagation (BP) network, selecting the fastest Levenberg-Marquardt as the learning algorithm, and setting the same goal as RBF network.

The process of the network's training and Fig. 2 indicate that, comparing with BP network, RBF network has better precision. Moreover, because RBF network's output layer is linear transfer, using OLS algorithm can escape converging to partial minimum, which is a big bug of BP network.

Meanwhile, RBF network is a partial approximation network, while BP network is universal approximation network, so the former can approximate the partial characters of functions more easily than the latter, and when getting new samples, the former is capable of converging more quickly.

However, what ought to be noted is RBF network will appear so called “curse of dimensionality” when the number of input is large enough. Hence sound training functions of hidden layers are supposed to be chosen for diverse applications (Meng, 2001).

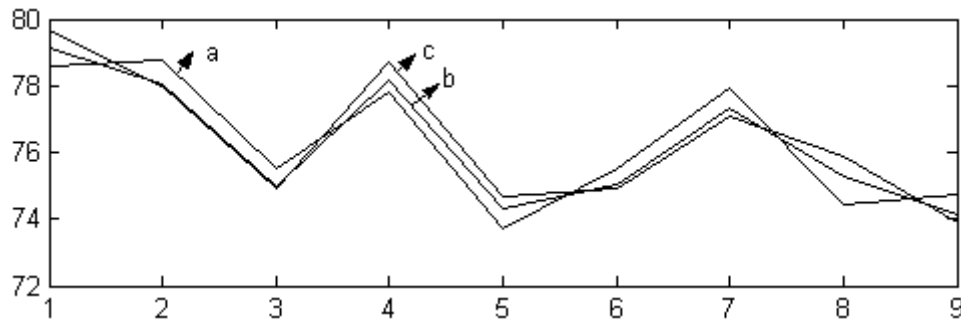


Fig. 2. Compare result of the two networks' estimation (a: Real value; b: Estimation of RBF network; c: Estimation of BP network)

7. CONCLUSION

According to the analyses of experiments and simulations, the proposed PCID method is proved to be efficient. And satisfying performance can be achieved when the soft-sensor method, combining RBF network and grey relation analysis, is applied to estimate the GC. And in this experiment, the mean relative error of evaluation is kept below 0.8%, which meets the need of practical control. So this design can solve the problem that two-factors and mechanism themes aren't able to meet the demand of accuracy. And also, it offers an efficient way for realizing the advanced control on grinding process, as well reaches the aims of prolonging the longevity of ball mill and saving resources. Moreover this design is more precious, as well as stronger and faster, than those based on BP network.

REFERENCES

- Chen, S., C.F.N. Cowan and P.M. Grant(1991). Orthogonal Least Squares Learning Algorithm for Radiate Basis Function Network. *IEEE Transactions on Neural Network*, March, **2(2)**, pp. 302-309.
- Cui, X.G. (1994). Discussion on $\phi 2.7 \times 3.6$ Meter Primary Ball Mill's Optimizing Grinding Concentration of Gongchangling concentrator. *Foreign Metal Ore Mineral Processing*, **1**, pp. 36-40.
- Deng, J.L. (2002). *Grey Theory Basis [M]*. pp. 122-154. HUST press, Wuhan, China.
- Ma, B.G. and G.Q. Cheng (2000). A Formula of Similarity Correlation Degree. *Systems Engineering-Theory&Practice*, (7), pp: 69-71.
- Mejdell, T. and S. Skogestad (1993). Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation. *AICHE J*, **39(10)**, pp. 1641-1653.
- Meng, H.F. (2001). Recursive Time- and Order-Update Algorithms for Radiate Basis Function

- Networks[DB/OL]. *UMI Microfilm 3023920*, <http://202.120.13.45/umi>
- Tang, W.X. (1995). The Concept and the Computation Method of T's Correlation Degree. *Mathematics Statistics and Management*, **14(1)**, pp. 34-38.
- Wang, Z.H. (2001). Study on Measurement Optimization of the Internal Parameters in Ball Mill. Ph.D. Dissertation, Dept. Mineral Process, North-East University, Shenyang, China, pp. 49-90.
- Yang, G.B. and Y.Y. Gao (2001). *Fuzzy Mathematics Theory and Application*. pp. 115-125. South China University of Technology Press, Guangzhou, China.
- Yu, J.S., A.L. Liu and K.J. Zhang (2000). *Soft Sensor Technology and Its Application in Petrochemical Industry*. pp. 56-104. Chemical Industry Press, Beijing, China.