Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/070503

Prediction of Gas Emission Based on Partial Correlation Analysis and SVR

Li Yang and Chengcheng Liu*

School of Economics & Management, Anhui University of Science and Technology, Huainan 232001, China

Received: 7 Dec. 2012, Revised: 21 Apr. 2013, Accepted: 23 Apr. 2013
Published online: 1 Sep. 2013

Abstract: The prediction model of gas emission is established based on partial correlation analysis and support vector regression (SVR) in order to accurately predict gas emission of working face under the condition of small samples. Not only are the problems of small samples and nonlinear prediction effectively resolved by applying SVR, but also the main control factors of gas emission are selected by applying partial correlation analysis method, which can reduce variables space dimension of the model to improve prediction accuracy. Through empirical analysis, the superiority of the model is proved by prediction results that are quite close to the measured values.

Keywords: Gas Emission, Partial Correlation, SVR

1 Introduction

Mine gas emission refers to the total gas from coal seam and dropping coal (rocks) pouring into the wind in unit time. It is the foundation of mine ventilation design and gas disaster prevention [1, 2]. Meanwhile, it also affects the safety of staff in mine production. To strengthen prediction method study of gas emission is to accurately predict gas emission as gas emission of working face accounted for 40 – 80% of the total gas emission, which is very important in improving safety production situation of coal mine in China.

Many traditional statistical theories have been applied to predicting gas emission at home and abroad, such as regression analysis method, fuzzy comprehensive evaluation method, gray correlation analysis method and neural network. These theories need the limit condition that sample size tends to be infinite. In actual data collection, samples can not be collected enough in order that the errors of prediction results are unsatisfactory.

Support vector machine (SVM) is the best theory of solving small samples prediction at present, and there are many influencing factors of gas emission of working face. So, prediction model of gas emission is constructed based on partial correlation analysis and SVR. Not only does the prediction model effectively resolve small samples and nonlinear prediction problems, but also makes full use of partial correlation analysis method to select main control factors of influencing gas emission index of working face in order to reduce variable space dimension, which can improve prediction accuracy of the model.

2 Prediction Model of Emission Based on Partial Correlation Analysis and SVR

2.1 Partial Correlation Analysis

Relationship of some correlated variables is relatively complex. There exists simple correlation between any two variables in varying degrees, and this relationship is also influenced by other variables. Therefore, only to eliminate the influence of other variables is to study the correlation between two variables, which can reflect their related nature and close degree. And partial correlation analysis is a statistical analysis method which fixes other variables to study correlation of two variables. In variables, to study relationship between two variables under the condition of keeping other variables unchanged is called partial correlation. Statistics which can express nature and degree of partial correlation of two variables are called partial correlation coefficient. Given variables x, y and z,

* Corresponding author e-mail: y321212@163.com

if z is fixed, partial correlation coefficient of x and y can be expressed as:

$$r_{xy,z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}} \tag{1}$$

where, r_{xy} , r_{xz} and r_{yz} are simple correlation coefficients. $r_{xy,z}$ is the statistical index of correlation degree of x and y , and $-1 \leq r_{xy,z} \leq 1$. $\left|r_{xy,z}\right|$ is close to 1, which means that the correlation is more close, and vice versa. $|r| = 1$ means perfect correlation, and $r = 0$ means perfect un-correlation. Due to controlling the other of variables, partial correlation coefficient of two variables is not 0, which can't explain that the coefficient is not 0 overall. So, the correlation coefficients should be inspected by t inspection method. The formula is expressed as:

$$t = \frac{r\sqrt{n-k-2}}{\sqrt{1-r^2}} \tag{2}$$

where, n is sample size, and k is the number of fixed variables. t statistics obeys the t distribution of $n-k-2$ degree of freedom.

2.2 SVR Principle

SVM [3–10] is based on the principle of VC dimension and structural risk minimization, which is a tool of solving machine learning problem by means of optimization method. The following describes ε-support vector regression (ε-SVR) model.

Given training set $T = \{(x_1,y_1), \cdots, (x_l,y_l)\} \in (R^n \times y)^l$ with input vector $x_i \in R^n$ and corresponding output $y_i \in y = R\,(i = 1, \cdots, l)$. So, support vector regression function can be expressed as:

$$g(x) = w \cdot x + b \tag{3}$$

where, w is weight vector, b is called offset. Accordingly, using $y = g(x)$ deduces the output value of any input x. The y is a real-valued.

Regression problem can be transformed into the problem of minimizing structure risk function by introducing insensitive ε loss function. Insensitive loss function is expressed as:

$$\begin{aligned} c(x,y,g(x)) &= |y - g(x)|_{\varepsilon} \\ &= \max\{0, |y - g(x)| - \varepsilon\} \end{aligned} \tag{4}$$

where, ε is a positive that is taken in advance, which reflects the accuracy of function approximation.

In order to solve w and b , relaxation variables $\xi_i \geq 0$ and $\xi_i^* \geq 0$ are introduced. So, the original problem is expressed as:

$$\begin{aligned} &\min_{w,b,\xi^{(*)}} \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &s.t. \begin{cases} ((w \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i\,,\ i = 1,2,\cdots,l \\ y_i - ((w \cdot x_i) + b) \leq \varepsilon + \xi_i^*\,,\ i = 1,2,\cdots,l \\ \xi_i^{(*)} \geq 0\,,\ i = 1,2,\cdots,l \end{cases} \end{aligned} \tag{5}$$

where, $\xi_i^{(*)} \geq 0$ means $\xi_i \geq 0$ and $\xi_i^* \geq 0$; constant C is the penalty factor, which controls the degree of punishment beyond error ε.

For nonlinear regression, the sample is mapped to a high dimension space. And then regression estimation function is constructed in this space. In the solution actual problem, according to the question chooses the appropriate kernel function which must satisfy Mercer theorem. Question (5) can be transformed into its dual problem by applying Lagrange optimization method:

$$\begin{aligned} &\min_{\alpha^{(*)} \in R^{2l}} \tfrac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)\left(\alpha_j^* - \alpha_j\right)K(x_i,x_j) \\ &+\varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) \\ &s.t. \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i^{(*)} \leq C,\ i = 1,2,\cdots,l \end{cases} \end{aligned} \tag{6}$$

$\overline{\alpha}^* = (\overline{\alpha}_1, \overline{\alpha}_1^*, \overline{\alpha}_2, \overline{\alpha}_2^*, \cdots, \overline{\alpha}_l, \overline{\alpha}_l^*)^T$ can be obtained by solving expression 6 and $\overline{b}$ is calculated to construct decision function:

$$y = g(x) = \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)K(x_i \cdot x) + \overline{b} \tag{7}$$

3 Empirical Analysis

3.1 Factor Analysis Influencing Gas Emission of Working Face

Gas emission of working face is the variable which is disturbed by multiple factors [11–13], such as occurrence condition of coal seam, gas geological conditions, mining conditions etc. Its influencing factors have gas content in coal seam, embedding depth of coal seam, coal seam thickness, angle of coal seam, face length, advance speed, coal seam interval and mining intensity. Samples are shown in table 1.

3.2 Predicting Gas Emission

(1) Selecting Main Controlling Factors by Applying Partial Correlation Analysis Method For there are many factors influencing gas emission of working face, main

Tab.1 Sample Set of Gas Emission

Number	*Gascontent incoalseam* $/m^3\cdot t^{-1}$	*Embedding depthofcoal seam/m*	*Coalseam thickness/m*	*Angleof coalseam/*$\circ$	*Face length/m*	*Advance speed* $/m\cdot d^{-1}$	*Coalseam interval/m*	*Mining intensity* $/t\cdot d^{-1}$	*Gasemission* $/m^3\cdot t^{-1}$
1	1.92	408	2.0	10	155	4.42	20	1825	3.34
2	2.15	411	2.0	8	140	4.16	22	1527	2.97
3	2.14	420	1.8	11	175	4.13	19	1751	3.56
4	4.80	634	6.5	12	175	2.92	15	3620	8.51
5	2.40	456	2.2	15	160	4.51	20	2104	4.17
6	3.22	516	2.8	13	180	3.45	12	2242	4.60
7	4.34	607	6.1	9	165	2.77	17	3087	7.68
8	3.35	531	2.9	9	165	3.68	13	2288	4.78
9	3.61	550	2.9	12	155	4.02	14	2325	5.23
10	3.68	563	3.0	11	175	3.53	12	2410	5.56
11	4.62	629	6.4	13	170	2.80	19	3456	8.04
12	4.03	604	6.2	9	180	2.64	16	3354	7.80
13	2.80	527	2.5	17	180	3.28	11	1979	4.92
14	2.58	432	2.3	10	145	4.67	17	2078	3.62
15	4.67	640	6.3	11	175	2.75	15	3412	7.95
16	2.43	450	2.2	12	160	4.32	16	1996	4.06
17	3.16	544	2.7	11	165	3.81	13	2207	4.92
18	4.21	590	5.9	8	170	2.85	18	3139	7.24

Tab.2 Partial Correlation Analysis of Gas Emission and Its Influencing Factors

		Gascontent incoalseam	*Embeddingdepth ofcoalseam*	*Coalseam thickness*	*Angleof coalseam*	*Facelength*	*Advancespeed*	*Coalseam interval*	*Miningintensity*
Gas emission	Correlation	0.557	0.746	0.696	0.218	0.578	0.170	-0.048	0.541
	Sig.(2-tailed)	0.026	0.008	0.017	0.341	0.062	0.618	0.889	0.085
	df	9	9	9	9	9	9	9	9

controlling factors influencing gas emission are selected by applying partial correlation analysis method. So, the results that gas emission and its influencing factors are analyzed with SPSS soft are shown in table 2.

According to partial correlation theory, the bigger the absolute value of partial correlation coefficient is, the greater effect of the factor is. In table 2, gas emission is influenced by its influencing factors. Where, partial correlation coefficients of gas content in coal seam, embedding depth of coal seam, coal seam thickness, face length and mining intensity are 0.557p0.746p0.696p0.578p6.2%p0.541 respectively. They with gas emission showed very significant correlation. So, prediction model is established with five main controlling factors as input variables of SVR.

(2) Predicting Gas Emission by Applying -SVR Data of gas content in coal seam, embedding depth of coal seam, coal seam thickness, face length and mining intensity are standardized, which is shown in table 3. And then selecting Gauss RBF kernel function

$$K(x,x') = \exp\left(-\left\|x-x'\right\|^2 \Big/ \sigma^2\right), (\sigma > 0)$$

as ε -SVR kernel function, its algorithm is realized with Matlab 10.0.

Choosing 1 to 12 data as training samples and inspecting 11 to 13 samples are to predict values of 14 to 16 samples in table 3. In order to have intuitive comparison, original samples in table 1 are made the same regression prediction. The results are shown in table 4.

In table 4, predicted results based on partial correlation analysis and ε-SVR are quite close measured values. The relative error of original factors prediction results and measured values is about 3.21%, while the relative error of regression prediction with partial correlation analysis to select factors is about 0.96%, so that the superiority of the model in predicting gas emission is demonstrated.

Tab.3 Data of Standardization

Number	$Gascontentincoalseam$ $/m^3 \cdot t^{-1}$	$Embeddingdepth$ $ofcoalseam/m$	$Coalseamthickness/m$	$Facelength/m$	$Miningintensity/t \cdot d^{-1}$	$Gasemission$ $/m^3 \cdot t^{-1}$
1	-1.4884	-1.4826	-0.9108	-0.9395	-0.9969	3.34
2	-1.2473	-1.4457	-0.9108	-2.2078	-1.4443	2.97
3	-1.2577	-1.3349	-1.0176	0.7516	-1.1080	3.56
4	-0.7964	-1.1872	-0.7506	-1.7850	-0.6170	8.51
5	-0.9851	-0.8918	-0.8040	-0.5167	-0.5779	4.17
6	-0.1252	-0.1532	-0.4836	1.1744	-0.3707	4.60
7	-0.5657	-0.0178	-0.6438	1.1744	-0.7656	7.68
8	0.0111	0.0315	-0.4302	-0.0939	-0.3016	4.78
9	0.2837	0.2653	-0.4302	-0.9395	-0.2461	5.23
10	0.3571	0.4254	-0.3768	0.7516	-0.1185	5.56
11	0.9129	0.7577	1.1719	0.3288	0.9762	8.04
12	0.7241	0.9300	1.3321	1.1744	1.2990	7.80
13	1.0492	0.9670	1.2787	-0.0939	0.8981	4.92
14	1.5315	1.2993	1.4923	0.7516	1.6984	3.62
15	1.3952	1.3732	1.3855	0.7516	1.3861	7.95
16	-0.9536	-0.9656	-0.8040	-0.5167	-0.7401	4.06
17	-0.1882	0.1915	-0.5370	-0.0939	-0.4233	4.92
18	1.3428	1.2378	1.4389	0.3288	1.4522	7.24

Tab.4 Prediction Results

Number	$Measuredvalue$	Original factors		Factors after selection	
		Prediction value	Relative error /%	Prediction value	Relative error /%
16	4.06	4.1560753	2.37	4.0442183	0.39
17	4.92	5.1086426	3.83	5.0040323	1.71
18	7.24	6.9923685	3.42	7.4754783	0.78

4 Conclusions

Gas emission prediction model is established based on partial correlation analysis and ε-SVR. 1) There exists correlation between gas emission of working face and its influencing factors, and each of the influencing factors has an impact on each other. Therefore, gas emission and its influence factors are analyzed by applying partial correlation analysis method to draw a conclusion that gas content in coal seam, embedding depth of coal seam, coal seam thickness, face length and mining intensity with gas emission had very significant correlation. And prediction model is established with five main controlling factors as input variables of ε-SVR. 2) Gas emission under the condition of small samples can be predicted accurately by applying the prediction model. Through empirical analysis, prediction results and measured values are quite close, which shows that this model is suitable for predicting gas emission in coal mine production.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (71071003), and the MOE Project of Youth Foundation of Humanities and Social Science (09YJC630004).

References

[1] Xue-hai Fu, Wen-ping Zhang, Ya-nan Zhou, Xu-fei Si, Hanga Wu. Technology and method of coal and gas outburst prediction during coal geological exploration. Procedia Earth and Planetary Science, **1**, 911-916 (2009).

[2] N. Taghizadeh, M. Mirzazadeh, A. Samiei Paghaleh, Exact solutions for the nonlinear Schrodinger equation with power law nonlinearity, Mathematical Sciences Letters, **1**, 7-16 (2012).

[3] Song Weihua, Zhang Hongwei. Regional prediction of coal and gas outburst hazard based on multi-factor pattern recognition. Procedia Earth and Planetary Science, **1**, 347-353 (2009).

[4] A. I. Aggour, F. E. Attounsi, Fuzzy Topological Properties on Fuzzy Function Spaces, Applied Mathematics & Information Sciences Letters, **1**, 1-5 (2013).

[5] Vapnik V. The Statistical Learning Theory. Jone Wileg, New York, (1998).

—

[6] A. Younes, Database Manipulation Operations on Quantum Systems, Quantum Information Review, **1**, 9-17 (2013).

[7] Vapnik V. N. An overview of statistical learning theory. IEEE Transactions on Neural Networks, **10**, 988-999 (1999).

[8] P. K. De, Amita Bhincher. Dynamic Programming and Multi Objective Linear Programming approaches. Applied Mathematics and Information Sciences, **5**, 253-263 (2011).
[9] Mangasarian, David Musicant. Lagrangian support vector machines. Journal of Machine Learning Research, **1**, 161-177 (2001).
[10] V. David Sánchez A. Advanced support vector machines and kernel methods. Neurocomputing, **55**, 5-20 (2003).
[11] Keerchi, Lin C J. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Computation, **15**, 1667-1689 (2003).
[12] Zachary B. Walters, Quantum Coherent Dynamics at Ambient Temperature in Photosynthetic Molecules, Quantum Physics Letters **1**, 21-33 (2012).
[13] Qi Wu. An intelligent forecasting model based on robust wavelet v-support vector machine. Expert Systems with Applications, **36**, 7604-7610 (2009).
[14] Reshma Khemchandani, Jayadeva, Suresh Chandra. Knowledge based proximal support vector machines. European Journal of Operational Research, **195**, 914-923 (2009).
[15] Enes Eryarsoy, Gary J. Koehler, Haldun Aytug. Using domain-specific knowledge in generalization error bounds for support vector machine learning. Decision Support Systems, **46**, 481-491 (2009).
[16] Yang Min, Yun-jia Wang, Yuan-ping Cheng. An incorporate genetic algorithm based back propagation neural network model for coal and gas outburst intensity prediction. Procedia Earth and Planetary Science, **1**, 1285-1292 (2009).
[17] Li Gao, Yanjun Hu, Guizhen Chen, Hongzhen Yu. Prediction of Gas Emission Volume Based on Wavelet Neural Network. IEEE Computer Society, **2**, 50-53 (2007).
[18] María B. Díaz Aguado and C. González Nieieza. Control and prevention of gas outbursts in coal mines, Riosa-Olloniego coalfield,Spain. International Journal of Coal Geology, **69**, 253-266 (2007).

Li Yang received the MS degree in Management science from University of Science and Technology of China in 2003, and the PhD degree in Management science from University of Science and Technology of China in 2011. He is currently a professor in Anhui University of Science and Technology. His research interests are in the areas of information systems and risk decision.

Chengcheng Liu is a M.Sc. student in Management science, Anhui University of Science and Technology. Her research interests are in the areas of information systems and decision analysis.