Data Fusion Based Phase Space Reconstruction from Multi-Time Series

Rongxi Wang[1], Jianmin Gao[1], Zhiyong Gao[1*], Xu Gao[1], Hongquan Jiang[1] and Le Cui[2]

[1]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China*
[2]*Shaanxi Weihe Coal Chemical Corporation Group Ltd, Weinan, 714000, China*
zhygao@mail.xjtu.edu.cn

Abstract

Focused on the problem of imperfect information in the process of reconstruction from single time series, a new technology for phase space reconstruction from multi-time series based on the data fusion is proposed. Firstly, the methods Cao and mutual information are used to select the reconstruction parameters, time delay and embedded dimension; secondly, the social cognitive optimization algorithm is brought to calculate the weights for each variable; thirdly, an adaptive weighted fusion estimating method is applied for data fusion; lastly, the effectiveness of the methods mentioned in this paper is demonstrated by the analysis results of one case study of real chemical plant data sets, and the proposed methods in this paper can improve the completeness of the information of the reconstructed phase space, which is also a good foundation for further analysis of complex system.

Keywords: data fusion, multi-time series, phase space reconstruction, adaptive weighted fusion estimating

1. Introduction

The complex electromechanical system always has hundreds or even more monitoring variables for state monitoring, and a large number of time series are generated at the same time. Time series analysis is one of the most important tools for capturing the features of the dynamic system. Because of the complex coupling relationship among the variables, it is difficult to obtain the characters of the complex electromechanical system from time series analysis directly, which is an obstacle for further researching of complex system too. Reconstruct the original dynamic characters from time series has become to a research focus of industry and academic.

In order to get more information from the time series, the embedding theorem was proposed by Tokens in 1981, which can be described as follows: for an infinitely long and noise-free scalar time series, $\{x(n)\}$, which has a chaotic attractor with dimensions d, it always can find an embedding phase space with the dimension property of m from topological invariant perspective, so long as the relationship $m \geq 2d+1$ is established[1].The embedding theorem is the theoretical foundation of the non-linear time series analysis, and the phase space reconstruction from single time series has been researched sufficiently [1]-[6]. In theory, a well phase space can be reconstructed from single time series if the reconstruction parameters, delay time and embedding dimension, are

---

*Corresponding author: zhygao@mail.xjtu.edu.cn

selected properly. Actually, the information of phase space is imperfect because of the limited and noisy samples. In other words, the phase space reconstruction from single time series always cannot reflect the original characters of the dynamic system accurately.

Many researchers have realized the drawbacks of the phase space reconstruction from single time series, and have begun to consider and develop the reconstruction technology based on multi-time series. Garcia [7] proposed a novel phase space reconstruction method from multi-time series based on different delay times, which made a breakthrough for phase space reconstruction from multi-time series; Cong [8] proposed a new method based on Bayes estimation theory. The work mentioned above either based on the hypothesis of the data samples obeyed a known distribution or extended the multi-variables in the phase space, which increased the dimensions of the reconstructed phase space and the computational effort. The selection of the reconstruction parameters of the current work is just the generalizing from single time series to multi-time series, which did not resolve the problem of imperfect information perfectly.

Combined the data fusion and phase space reconstruction, a novel reconstruction method for multi-time series is proposed from a new perspective. Firstly, select and preprocess the variables who contain more information than others to reconstruct the phase space, such as normalization and noise reduction; secondly, an adaptive weighted fusion estimating method which has the advantages of fusion the redundant and complementary information is applied to fuse each dimension of each phase space reconstructed by first step, and finally, a new phase space which contains more information of the original dynamic system is reconstructed. A case of real plant is studied. Through the comparison of the graphs reconstructed from single and multi-time series, we can find that the phase space reconstructed by the methods proposed in this article contains the main characters of the variables presented in reconstruction, and can get more information of the original dynamic system than traditional methods.

2. Phase Space Reconstruction from Multi-Time Series

2.1. The Mathematic Model of Phase Space Reconstruction

For an arbitrary time series $\{x_1, x_2, x_3, ..., x_{N-1}, x_N\}$, the phase space reconstructed through embedding theory is

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & ...? & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & ...? & x_{2+(m-1)\tau} \\ ...? & ...? & ...? & \\ x_{N_m-1} & x_{N_m-1+\tau} & ... & x_{N-1} \\ x_{N_m} & x_{N_m+\tau} & ... & x_N \end{bmatrix} \tag{1}$$

Where N_m is the number of phase point $N_m = N - (m-1)\tau$, in which, τ is the delay time, and it is the integral multiple of sampling period usually, and m is the number of embedding dimension. Currently, the accepted methods for reconstruction parameters selection are mutual information and Cao method [9]. So, the methods are applied to select the reconstruction parameters in our work too.

We extend the concepts of reconstruction from single time series to multi-time series. We mark $x(t) = [x_1(t), x_2(t), ..., x_s(t)]^T$ as the multivariable discrete time series of complex

system, in which, $t = 1, 2, ..., N$ and $x_1(t), x_2(t), ..., x_s(t)$ are the values of s variables at time t. The essence of phase space reconstruction from multi-time series is to select the perfect parameters of delay time τ and embedding dimension m for the s variables, and fuse the phase space reconstructed from single variable, and get final phase space Z.

2.2. Adaptive Weighted Fusion Estimating Based Data Fusion

As a new research field of information science, data fusion was proposed and launched in 1970s. Data fusion has been more widely adopted in wireless sensor network and multi-target tracking fields.

The adaptive weighted fusion estimating can be described as follows: one object is measured by N sensors at the same time, and the separate weighting factors are assigned to different sensors. The goal of adaptive weighted fusion estimating is to get the optimal fusion result by the optimal weighting factors and the values measured by sensors based on the constraint of minimum total mean square error.

The following processes assume a set of measured data $X_1, X_2, ..., X_N$, and they are independent of each other. Computing the variances, $\sigma_1^2, \sigma_2^2, ..., \sigma_N^2$ and weighting factors, $W_1, W_2, ..., W_N$, of each datum, the estimation value is $\bar{X} = \sum_{p=1}^{N} W_p X_p$, and the key point of data fusion is searching the minimal value of fitness function $\sigma^2 = \sum_{p=1}^{N} W_p^2 \sigma_p^2$ under the constraints.

$$st. \begin{cases} 0 \leq W_p \leq 1, \quad p = 1, 2, ..., N \\ \sum_{p=1}^{N} W_p = 1 \end{cases} \tag{2}$$

The process is a typical problem of multi-object optimization, and the social cognitive algorithm is applied to getting the optimal solutions.

2.3. Weight Optimization Based On Social Cognitive Algorithm

In the past few decades, some intelligent algorithms such as ant colony algorithm and genetic algorithm had been proposed, and these algorithms solved some complex problems. Most of these algorithms are based on the insect social. Considered the real life, the intelligence of human is the highest among the animals. One person can learn from the activities of others and the results of these activities, and this is the most important point that human intelligence is higher than other animals. At 2002, XIE [10] proposed the social cognitive optimization algorithm based on the social cognitive theory. Reference 12 [11] and reference 13 [12] applied this algorithm to the solving of complex non-linear problem and multi-properties optimization. The social cognitive algorithm is used to optimize the weightings of variables selected to reconstruct the phase space.

The social cognitive algorithm contains four concepts, knowledge points, library, learning agents and neighborhood searching. The detail process of this algorithm is as follows:

Step 1: Initialization

a) Create all the N_{pop} knowledge points in library randomly (including the positions and its level of every knowledge point)

b) Allocate a knowledge point in library for every learning agent randomly, but it is not allowed to allocate one knowledge point to many learning agents repeatedly.

Step 2: Vicarious learning

a) Model selection: select two or more knowledge points from library randomly (ordinary two is ok), but these selected knowledge points cannot repeat with learning agent itself, then we need to select a better knowledge point from these knowledge points based on the principle of competitive alternatives (compared with the adaptive value function *F(X)* of every knowledge points)

b) Observational learning: compare the levels of selected knowledge points with that of processes by itself, then choose the better knowledge point as central point, and the worse knowledge points as the reference point, and the learning agent moves between the two points to search a new knowledge point based on the principle of neighborhood searching, and add the new points into the library.

Step 3: Library refreshment: remove N_c knowledge points with the worst levels from the library.

Step 4: Repeat the procedures from Step 2 to Step 4 till satisfy all stop conditions (for examples, the result reached number of iterations which fixed ahead, or reached precision made ahead).

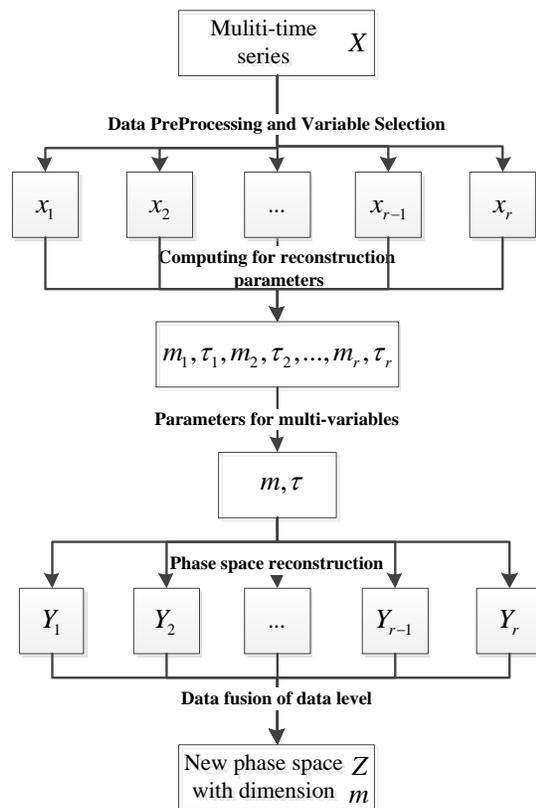## 3. Process of Phase Space Reconstruction from Multi-Time Series

Figure 1.The Processes of Phase Space Reconstruction from Multi-Time Series

Figure 1 shows the processes of phase space reconstruction from multi-time series, and there are five steps.

Step 1: Data preprocessing and variable selection. The main work of this step is to

normalize the original data and select the r variables to reconstruct the phase space.

Step 2: Computing the reconstruction parameters for each time series. Compute the delay times $\tau_1, \tau_2, ..., \tau_r$ and embedding dimensions $m_1, m_2, ..., m_r$ of variables selected in step 1, and mutual information and Cao method are applied.

Step 3: Computing the reconstruction parameters of multi-time series. In order to embed each component to the same phase space and guarantee the unfolding of each variable with no-distortion, max embedding dimension and min delay time of all variables are taken as the reconstruction parameters of multi-time series. This step can be described using the following two formulas briefly.

$$m = \max(m_1, m_2, ..., m_r) \tag{3}$$

$$\tau = \min(\tau_1, \tau_2, ..., \tau_r) \tag{4}$$

Where m is the embedding dimension and τ is the delay time for multi-time series.

Step 4: Phase space reconstruction from each variable. Do the phase space reconstruction from every component, and the reconstruction parameters decided by step 3 are used. The ith phase space reconstructed from ith time series, $x_i(t) = (x_{i,1}, x_{i,2}, ..., x_{i,N})$, is as follows:

$$Y_i = \begin{bmatrix} x_{i,1}, & x_{i,1+\tau}, & ..., & x_{i,1+(m-1)\tau} \\ & & ... & \\ x_{i,k}, & x_{i,k+\tau}, & ..., & x_{i,k+(m-1)\tau} \\ & & ... & \\ x_{i,M}, & x_{i,M+\tau}, & ..., & x_{i,M+(m-1)\tau} \end{bmatrix} \tag{5}$$

Where $i = 1, 2, ..., r$, and k is the index of random point in time series, and M, who equals $N - (m-1)\tau$, is the total number of phase points. $\{x_{i,k}, \quad x_{i,k+\tau}, \quad ..., \quad x_{i,k+(m-1)\tau}\}$ is the kth phase point of the ith variable.

Step 5: Data-based fusion. An adaptive weighted fusion estimating method is used to fuse the each dimension of every phase space, and the final phase space is

$$Z = \begin{bmatrix} z_1, & z_{1+\tau}, & ..., & z_{1+(m-1)\tau} \\ & & ... & \\ z_k, & z_{k+\tau}, & ..., & z_{k+(m-1)\tau} \\ & & ... & \\ z_M, & z_{M+\tau}, & ..., & z_{M+(m-1)\tau} \end{bmatrix} \tag{6}$$

Where $z_k = \sum_{i=1}^{r} x_{i,k} * w_{i,k}$, and z_k is the kth coordinate point of final phase space, and $x_{i,k}$ is kth value of the ith variable, and $w_{i,k}$ is the weight of $x_{i,k}$. $\{z_k, \quad z_{k+\tau}, \quad ..., \quad z_{k+(m-1)\tau}\}$ is the kth phase point of Z.

4. The Application and Effects Analysis

A case of multi-time series generated by compressors of a real chemical plant is studied with the methods proposed in this article. The graphs reconstructed by single and multi-time

series are contrasted to illustrate the effectiveness of the proposed methods in this paper.

The compressors are the normal and important production equipment in chemical plants. The compressors have hundreds of variables, monitoring and controlling variables contained, such as pressure, temperature, flow, liquid level, vibration, rotating speed, switch, alarming signals and so on. A_ATI7611, inlet temperature of the booster, and PSE7655, rotating speed of air compressor, are selected as the objects to verify the effectiveness of our methods. The data sets in our example are sampled from the distributed control system of a chemical plant directly, and the continuous sampling was done once per minute in 48 hours. The total number of sampling is 2882.

Firstly, compute the reconstruction parameters with mutual information and Cao method for each variable respectively, and the results are shown as Table 1.
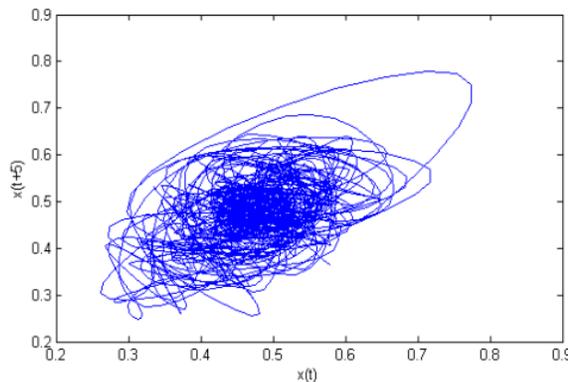
Table 1. Reconstruction Parameters of A_ATI7611 and PSE7655

Variable	Delay time	Embedding dimension
A_ATI7611	59	4
PSE7655	5	3

The graphs reconstructed by A_ATI7611 and PSE7655are shown as Figure 2 and Figure 3.
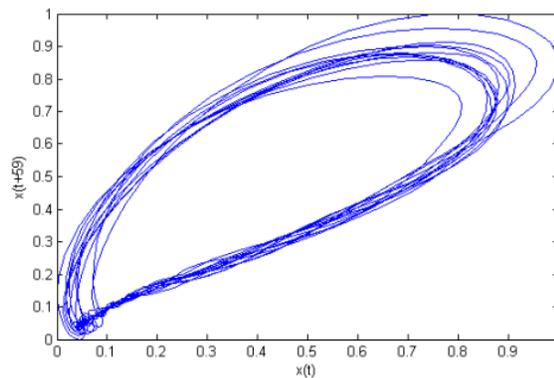
Figure 2. Reconstructed Graph of A_ATI7611
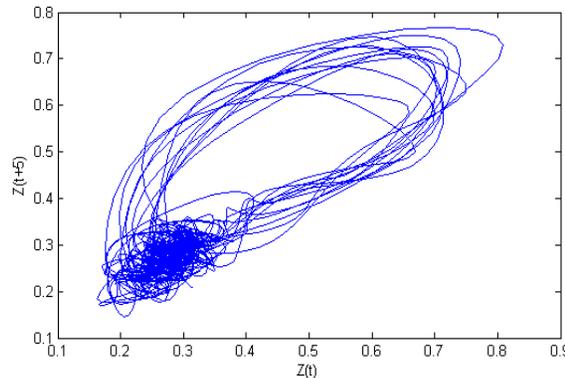

Figure 3. Reconstructed Graph of PSE7655

Obeyed the processes mentioned in section 3, the public embedding dimension m for the

two variables is4, and the delay time τ is 5. The phase space Y_1 and Y_2 are reconstructed with these parameters. The each dimension of Y_1 and Y_2 are divided into some segments with size of 480 points, and the variance is calculated for each segment too. The total number of each dimension is 2882-(4-1)*5=2867, so, There are 6 segments through the division above. The social cognitive algorithm is applied to optimize the weight of each segment of each dimension of Y_1 and Y_2 with the fitness function of minimum total mean square error. The weight of each segment is shown as Table 2, in which, V1 and V2 stand for the variables A_ATI7611 and PSE7655 respectively.

Table 2. Weight of Each Segment Based on Social Cognitive Algorithm

	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
	V1	V2	V1	V2	V1	V2	V1	V2
S_1	0.519	0.481	0.654	0.346	0.660	0.340	0.180	0.820
S_2	0.418	0.582	0.739	0.261	0.304	0.696	0.713	0.287
S_3	0.518	0.482	0.424	0.576	0.270	0.730	0.921	0.079
S_4	0.436	0.564	0.431	0.569	0.567	0.433	0.606	0.394
S_5	0.447	0.553	0.475	0.525	0.672	0.328	0.529	0.471
S_6	0.494	0.506	0.708	0.292	0.342	0.658	0.515	0.485

The last step is to reconstruct the each phase point of Z with the weights and the values of Y_1 and Y_2, and the reconstructed graph of Z is shown as Figure 4.

Figure 4. Reconstructed Graph of Fused A_ATI7611 and PSE7655

We can get some conclusions from these reconstructed graphs clearly. From the reconstructed graphs of single time series perspective, the reconstructed graph of A_ATI7611does not contain the main characters of the graph reconstructed from PSE7655, and vice versa, therefore, the information of the reconstructed phase space from single time series of A_ATI7611 and PSE7655 is imperfect. From the reconstructed graphs of multi-time series perspective, Figure 4 and the Figure 2 have the same outline as a whole, and Figure 4 has an eclipse profile which is the main character of Figure 2. In the lower-left corner of

Figure 4, there is a local character, and that is the main character of Figure 3.

The information can be measured by information entropy quantitatively, and they have an inverse relationship [13]. That is to say, the bigger of the information entropy, the less of the information, and vice versa. We calculated the each information entropy of each dimension for reconstructed phase space from original variables, A_ATI7611 and PSE7655, and the fused phase space, and the values of information entropy shown as Table 3.

Table 3. The Information Entropy Value of Each Dimension

	A_ATI7611	PSE7655	Fused Phase Space
D_1	0.399272	0.956936	0.375367
D_2	0.399272	0.955931	0.375367
D_3	0.399272	0.955931	0.378036
D_4	0.399272	0.955931	0.378725

Combined the comparison results of reconstructed graphics and information entropy values, we can get the following conclusions clearly: The phase space reconstructed by the methods proposed in this paper contains the characters of A_ATI7611 and PSE7655, and has more information than those reconstructed from single time series, which receives the goals of phase space reconstruction from multi-time series. These results illustrate the effectiveness of the methods of this article.

5. Conclusions

Actually, the phase space reconstruction of single time series cannot reconstruct all the characters of the complex systems because of the limit and noisy samples. Focused on this problem, a novel reconstruction method from multi-time series was proposed based on the data fusion, and the social cognitive algorithm and the adaptive weighted fusion estimating are used to optimize the weights of each component and fuse the different phase space, respectively. The analysis results of the real case show that: The reconstruction graph generated by multi-time series contains the main characters of all the present variables in the process of data fusion, and the information of phase space is more perfect than who reconstructed from single time series, which is a good foundation for further data analysis. The algorithms proposed in this paper combines the data fusion and phase space reconstruction, and can solve the data fusion problem of similar sensors, and it is not only but also a new thinking for heterogeneous sensors.

Acknowledgements

References

[1] Takens F., "Detecting strange attractors in turbulence", Dynamical Systems and Turbulence. Springer-Verlag, Berlin, **(1981)**.
[2] L. Yin and Y. He, "Adaptive chaotic prediction algorithm of RBF neural network filtering model based on

phase space reconstruction", Journal of computer, vol. 8, no. 6, **(2013)**.

[3]   R. Jing, Z. Q. Shan and W. R. Ran, "Neural network forecasting model for sunspots time series prediction based on phase space reconstruction", Computer simulation, vol. 31, no. 1, **(2014)**.

[4]   Z. Gao and N. Jin, "Complex network from time series based on phase space reconstruction", Chaos,   vol. 19, no. 3, **(2009)**.

[5]   Y. Lv and J. Gao, "Study on several key state analysis technologies on distributed and complex electromechanical system supported by nonlinear data", Xi'an Jiaotong University, Xi'an, **(2013)**.

[6]   To A. C., Moore J. R. and Glaser S. D., "Wavelet denosing techniques with application to experimental geophysical data", Signal Processing, vol. 89, no. 2, **(2009)**.

[7]   L. Z. Bo, C. Z. Ming and J. K. Yu, "Determination of embedding parameters for phase space reconstruction based on improved C-C method", Journal of system simulation, vol. 19, no. 11, **(2007)**.

[8]   Garcia S. P. and Almeida J. S., "Multivariate phase space reconstruction by nearest neighbor embedding with different time delays", Physical Review E, vol. 72, no. 2, **(2005)**.

[9]   C. Rui, L. S. Lin and M. Rui, "An Approach to phase space reconstruction from multivariable data based on data fusion", Acta Physica Sinica, vol. 57, no. 12, **(2008)**.

[10] S. K. Palit, S. Mukherjee and D. K. Bhattacharya, "A high dimensional delay selection for the reconstruction of proper phase space with cross auto-correlation", Neurocomputing, vol. 113, no. 8, **(2013)**.

[11] X. X. Feng, Z. Wenjun and Y. Zhilian, "Social Cognitive Optimization for Nonlinear Programming Problem", Proceedings of International Conference on Machine Learning and Cybernetics, Beijing, China, Noverber4-5, **(2002)**.

[12] M. Li, W. R. Xi and C. Yanpin, "The Social cognitive optimization algorithm: Modifiability and Application", Proceedings of International Conference on E-Produce E-Service and E-Entertainment, Jiaozuo, China, Noverber7-9, **(2010)**.

[13] Z. Jianke, L. Sanyang, C. Yanpin and W. R. Xi, "Control of chaotic systems by chaotic social cognitive optimization", ICIC Express Letters, vol. 5, no, 3, **(2011)**.

[14] Z. Jiguo, "Singh. Information entropy-Theory and applications", China Water &Power Press, Beijing, **(2012)**.

Authors

Rongxi Wang, He received the B. Eng. and M. Eng. degrees in computer technology from Xi'an University of Posts &Telecommunications, Xi'an, China, in 2008 and 2011, respectively. Currently, he is working toward the PhD degree in mechanical engineering at Xi'an Jiaotong University, Xi'an, China. His current research interests include healthy condition evaluation of distributed and complex electromechanical system and the theory of complex network.
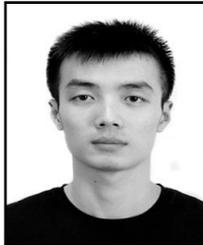
Jianmin Gao, He received his B.Eng., M. Eng. and PhD. degree from Xi'an Jiaotong University in 1982, 1989, 2000, respectively, majored in mechanical engineering. He worked at School of Mechanical Engineering of Xi'an Jiaotong University since 1982.He is a professor since 1999.His main research interests include digital manufacturing, quality control, ERP, contemporary integrated manufacturing, reliability and maintenance，quality control and measurement tech.

Zhiyong Gao, he is the vice professor of Xi'an Jiaotong University and he is working in the fields of manufacturing information engineering, information quality control and safety of process industry.

Xu Gao, He was born in 1984. Currently, he is working toward the Ph.D degree in mechanical engineering at Xi'an Jiaotong University, Xi'an, China. His main research interests include the real-time monitoring and condition analysis of power plant.

Hongquan Jiang, He received his B.Eng. degree of Mechanical Engineering from Jilin University in 2001, and earned the M.S degree and PhD degree from Xi'an Jiaotong University in 2005 and 2009, respectively, majored in mechanical engineering. He is a lecturer at the CIMS institute in the Department of Mechanical Engineering at Xi'an Jiaotong University. His main research interests include system reliability analysis, fault detection and diagnosis, quality control technology.

Le Cui, He received his B.Eng. degree from Shaanxi University of Science &Technology in 2002. He is working at Shaanxi Weihe coal chemical corporation group Ltd., and the information construction is his main work. His main research interests include the application of the information construction in coal chemical industry.