

Research Article

Fault Diagnosis Method of Polymerization Kettle Equipment Based on Rough Sets and BP Neural Network

Shu-zhi Gao,¹ Jie-sheng Wang,² and Na Zhao¹

¹ College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China

² School of Electronic and Information Engineering, University of Science & Technology Liaoning, Anshan 114044, China

Correspondence should be addressed to Jie-sheng Wang; wang_jiesheng@126.com

Received 25 March 2013; Revised 27 June 2013; Accepted 2 July 2013

Academic Editor: Jyh-Horng Chou

Copyright © 2013 Shu-zhi Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Polyvinyl chloride (PVC) polymerizing production process is a typical complex controlled object, with complexity features, such as nonlinear, multivariable, strong coupling, and large time-delay. Aiming at the real-time fault diagnosis and optimized monitoring requirements of the large-scale key polymerization equipment of PVC production process, a real-time fault diagnosis strategy is proposed based on rough sets theory with the improved discernibility matrix and BP neural networks. The improved discernibility matrix is adopted to reduce the attributes of rough sets in order to decrease the input dimensionality of fault characteristics effectively. Levenberg-Marquardt BP neural network is trained to diagnose the polymerize faults according to the reduced decision table, which realizes the nonlinear mapping from fault symptom set to polymerize fault set. Simulation experiments are carried out combining with the industry history datum to show the effectiveness of the proposed rough set neural networks fault diagnosis method. The proposed strategy greatly increased the accuracy rate and efficiency of the polymerization fault diagnosis system.

1. Introduction

Polyvinyl chloride (PVC) is one of the five largest thermoplastic synthetic resins, and its production is second only to the polyethylene (PE) and polypropylene (PP). PVC is a kind of general colophony, which is good in quality and is widely used. It has good mechanical properties, antichemical properties and it is corrosion-resistant and difficult to burn. With vinyl chloride monomer (VCM) as a raw material, the suspension method to produce polyvinyl chloride (PVC) resin is a kind of typical batch chemical production process. PVC polymerization process is a complex control system with multivariable, uncertain, nonlinear, and strong coupling.

Polymerization kettle is the key equipment of the PVC production process, where vinyl chlorides go on the polymerization reaction to generate polyvinyl chloride [1]. Whether the polymerization kettle can run steadily is directly related to the working conditions of the PVC production device. On the other hand, the motor, reducer, and machine seal are key equipments to ensure that the polymerization kettle device runs normally. Once they failed to work, the serious losses

will be brought to the PVC polymerizing process. Therefore, the earlier diagnosis of the fault type and location of polymerization kettle can avoid the huge economic losses which are caused by the parking of polymerization kettle, which has the important practical significance to improve the product quality and reduce the production costs.

Rough set (RS) theory is a kind of knowledge mining theory and a mathematical tool to describe imperfection and uncertainty [2]. The neural network can effectively distinguish fault patterns brought out by sensor failures, the mismatch between process and model, noises and disturbances [3–5]. The main goal of the data-driven fault diagnosis is to realize the fault diagnosis and isolation by finding the hidden fault pattern and the relationship between data and fault pattern. The hybrid fault diagnosis method by combining RS theory and neural network takes advantage of the two technologies, which takes the rough set method as the input system of neural network to simplify the complexity of the neural network and improve the precision and efficiency of fault diagnosis [6, 7]. And the utilizing of neural networks as a rear information recognition system is to make up the deficiency

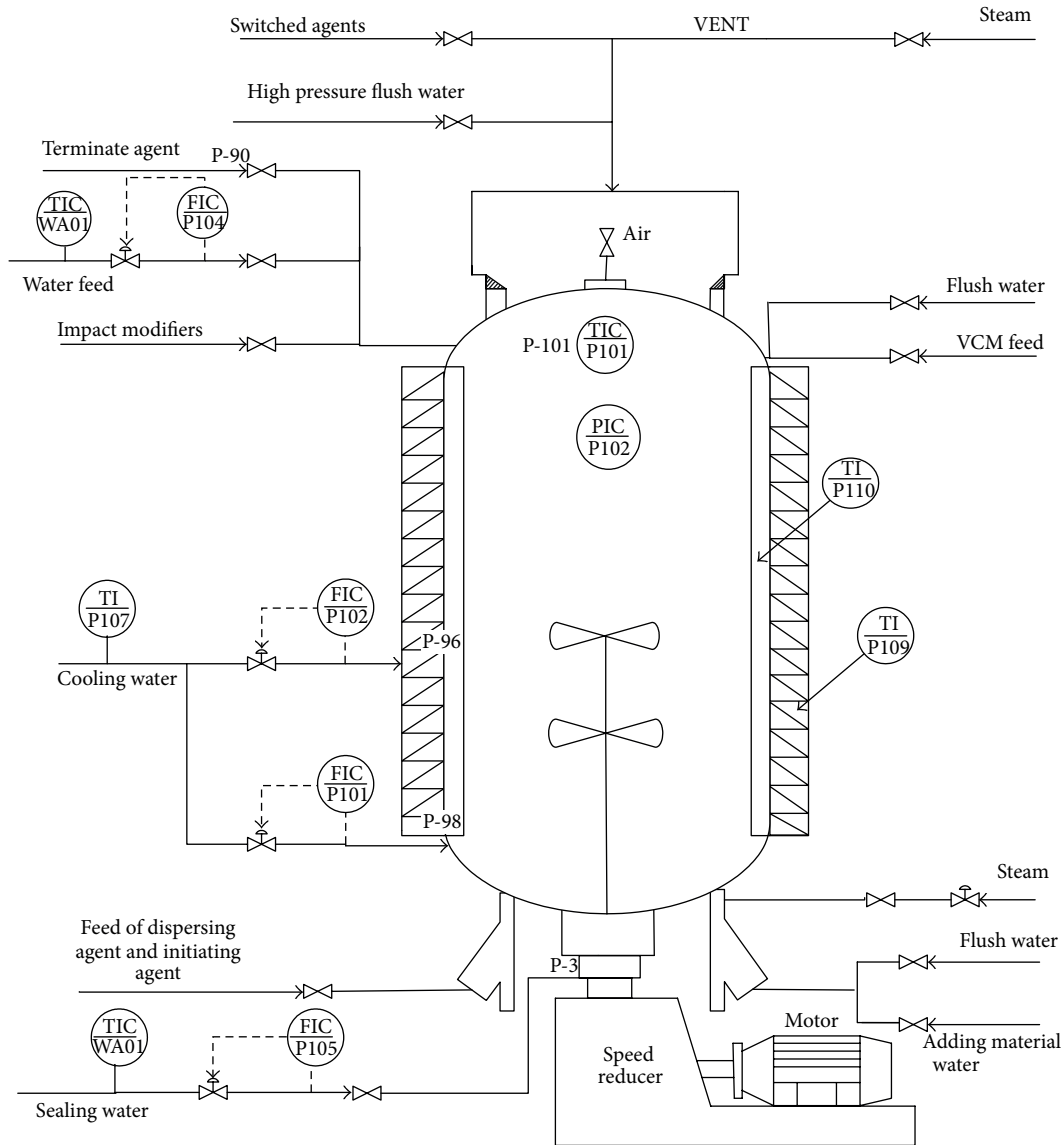


FIGURE 1: Technique flowchart of polymerization apparatus.

of the rough set, which is widely used in many industry fields, such as electric power, metallurgy, and chemical industry [8–11].

Therefore, a kind of fault diagnosis method of polymerization kettle based on the improved discernibility matrix and BP neural network is put forward in this paper. The improved discernibility matrix is used to reduce the fault diagnosis attributes of polymerization kettle to simplify the input dimension of the fault characters. Then LM-BP neural network is used to diagnose the fault pattern of polymerization kettle in order to improve the accuracy and efficiency of fault diagnosis.

The paper is organized as follows. In Section 2, the technique of PVC polymerizing process is introduced. In Section 3, the RS-NN fault diagnosis method of polymerization kettle equipment is summarized. In Section 4, the improved LM-BP neural network algorithm is introduced in

details. The simulation experiments are discussed in details in Section 5. Finally, the conclusion illustrates the last part.

2. Technique of PVC Polymerization Process

A flowchart of the typical PVC polymerization kettle production process is shown in Figure 1 [12].

In polyvinyl chloride (PVC) polymerizing process, all kinds of raw materials and auxiliary agents are placed into the reaction kettle. They are fully and evenly dispersed under the function of stirring. Then, we begin to ventilate the cooling water to the clipset of the reaction kettle and baffle plate constantly in order to remove homopolymer. When the conversion rate of VCM reaches a certain value, there is a proper pressure drop. Then, the reactions are terminated and the finished product is created. The discussed polymerizing reaction selects SG-5 as the example. 26 ton VCM is fed into

the polymerization kettle. The conversion rate is about 85%, the reaction temperature is 55.4°C, and the heat released from PVC polymerizing reaction is 1600 KJ/Kg.

3. RS-NN Fault Diagnosis Method of Polymerization Kettle

3.1. Structure of Polymerization Kettle Fault Diagnosis System.

The structure of the proposed RS-NN fault diagnosis system of polymerization kettle is shown in Figure 2. Firstly, a group of fault samples is utilized to train BP neural network to determine the structure and parameters of the network. After training, the classification of fault patterns is carried out in accordance with the given symptoms to realize the nonlinear mapping from the fault symptom set to the fault set.

3.2. Formation of Fault Information Table. A 70 M³ polymerization kettle is utilized to set up the RS-NN fault diagnosis system. The condition attributes of the decision table are the reducer vibration value of polymerization kettle (S_1 , mm), the stirring current (S_2 , A), the pressure of mechanical seal (S_3 , MPa), the operating pressure (S_4 , MPa), the stirring speed (S_5 , r/min), the reducer temperature (S_6 , °C), the operating temperature of polymerization kettle (S_7 , °C), and the mechanical seal temperature (S_8 , °C), whose corresponding variables are denoted as a, b, c, d, e, f, g , and h , respectively.

The fault of polymerization reactor includes the motor fault, the deceleration machine fault, and the machine seal fault. And the main mechanical seal failure forms of machine seal faults are the gland-shaft damage and components damage of machine seal. Assume D is the decision attribute in accordance with the direct reasons of the faults; that is to say, $D = 0$ stands for the normal working conditions of the polymerization kettle, $D = 1$ stands for the motor fault, $D = 2$ stands for the reducer fault, $D = 3$ stands for gland-shaft fault of the polymerization machine seal, and $D = 4$ stands for the fault of polymerization component. The corresponding BPNN outputs of the five working conditions in the polymerization kettle production process are the normal working condition (0000), the motor fault (0001), the deceleration machine fault (0010), the machine seal gland-shaft fault (0100), and the fault of machine seal component (1000), respectively. The historical data of the polymerization kettle fault diagnosis decision system are shown in Table 1.

3.3. Discrete Decision Table of Fault Diagnosis. Because the rough set theory can only deal with the discrete attribute values, the data of the fault diagnosis decision-making system have to be dealt to the discrete values. The paper discrete the continuous attributes based on the expert experiences. Thus each critical value of the condition attribute intervals is listed as follows:

S_1 : “1” represents the interval (0, 0.5], “2” represents the interval (0.5, ∞);

S_2 : “1” represents the interval (0, 175], “2” represents the interval (175, ∞);

S_3 : “1” represents the interval (0, 1.5], “2” represents the interval (1.5, ∞);

S_4 : “1” represents the interval (0, 1.1), “2” represents the interval [1.1, 1.3], “3” represents the interval [1.3, 1.4], “4” represents the interval (1.4, 1.6], “5” represents the interval (1.6, ∞);

S_5 : “1” represents the interval (0, 91), “2” represents the interval [91, 98], “3” represents the interval (98, 105], “4” represents the interval (105, ∞);

S_6 : “1” represents the interval (0, 40], “2” represents the interval (40, 80], “3” represents the interval (80, 90], “4” represents the interval (90, ∞);

S_7 : “1” represents the interval (0, 60), “2” represents the interval [60, 67], “3” represents the interval [67, 69], “4” represents the interval (69, 75], “5” represents the interval (75, ∞);

S_8 : “1” represents the interval (0, 80], “2” represents the interval (80, ∞).

The discrete results are shown in Table 2.

3.4. Attribute Reduction Based on Improved Discernibility Matrix. The attribute reduction method in rough set theory is one of the key research topics. Attribute reduction method based on the discernibility matrix [13, 14] is an important variant of rough set theory, whose main thought is to firstly use the discernibility matrix to derive the discernibility function and then solve the disjunctive paradigm, whose each paradigm is a reduction of the rough set. The concept of attribute 0-1 resolution matrix was put forward to transform the attribute reduction problem to the 0-1 matrix cover problem [15]. The concept and construction method of traditional discernibility matrix are based on the symbol-type information system with the complete attributes [16, 17]. A concept and configuration method of a flexible discernibility matrix were proposed to reduct attributes of the incomplete information system and Vague set attribute information system [18]. Aiming at the calculation complex problem of the traditional discernibility matrix, an attributes reduction algorithm based on the improved discernibility matrix is proposed, which have the superiority in tackling with the discernibility matrix with higher dimensions.

3.4.1. Discernibility Matrix

Definition 1. Assume $S = (U, A, VA, f)$ is a decision table, $C \cup D = A$, $C \cap D = \Phi$, $D \neq \Phi$, $U = \{x_1, x_2, \dots, x_n\}$, C is the condition attributes set, D is the decision attributes set, and $a(x)$ is a value of the sample x in attribute a . So the discernibility matrix C_D is defined as follows:

$$C_{ij} = \begin{cases} \{a | a \in A, a(x_i) \neq a(x_j)\}, & D(x_i) \neq D(x_j), \\ 0, & D(x_i) = D(x_j), \\ 1, & D(x_i) \neq D(x_j), \\ & a(x_i) = a(x_j). \end{cases} \quad (1)$$

It can be seen from the definition of a discernibility matrix that C_{ij} is a set composed by all attributes which can

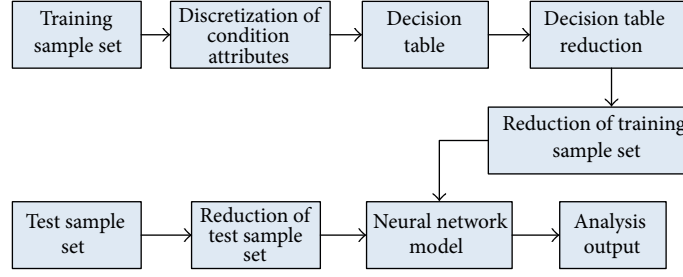


FIGURE 2: Fault diagnosis system configuration based on rough sets neural network.

TABLE 1: Historical data of polymerizer.

Sample (U)	Historical datum of polymerizer								Diagnosis type (D)
	a, S_1	b, S_2	c, S_3	d, S_4	e, S_5	f, S_6	g, S_7	h, S_8	
1	0.38	119.1	1.55	1.45	93.10	42.6	76.13	29.52	4
2	0.44	188.8	1.00	1.51	93.70	78.9	65.44	82.64	3
3	0.39	138.9	1.01	1.78	93.10	92.3	66.83	29.39	1
4	0.38	133.4	0.88	1.34	92.80	59.9	60.32	83.50	0
5	0.40	140.1	1.61	0.86	56.49	41.5	54.21	29.43	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
100	0.38	138.1	1.15	1.35	91.85	78.8	52.39	29.99	0

distinguish the sample X_i from X_j . For all $X_i, X_j \in U$, $C_{ij} \subseteq A$, which is in accordance with the following properties: (1) $C_{ii} = 0$; (2) $C_{ij} = C_{ji}$; (3) $C_{il} \subseteq C_{ij} \cup C_{jl}$.

The discernibility matrix is symmetry along the main diagonal. So in the analyzing process, its upper triangular or lower triangular part is only considered. When the decision attributes of the two samples are the same, the element of the corresponding discernibility matrix is 0. When the decision attributes of the two samples are different and are distinguished by certain conditions, the element values of the corresponding discernibility matrix are the different condition attributes in two samples.

3.4.2. Attribute Reduction Algorithm Based on Improved Discernibility Matrix. The attributes reduction algorithm based on the discernibility matrix and logic operations can be used to obtain all the possible attribute reduction results of the decision table, which is actually to reduce the attributes combination situations into logic formula. If there is an element in the discernibility matrix, whose value is the set containing elements with single attribute, the attribute is the necessary attribute of the two samples which distinguish the matrix elements and it is the only one that can distinguish the attributes of the two samples. Attribute set composed of the attributes in the discernibility matrix is actually the relative attribute kernel of the decision table system. The attribute reduction algorithm based on the improved discernibility matrix is described as follows.

Step 1. Calculate the discernibility matrix C_D of the decision table.

Step 2. Find out the single element in discernibility matrix, retain it to the kernel of the attribute reduction, and change all the elements C_{ij} containing it into zero.

Step 3. The elements in the discernibility matrix C_{ij} ($C_{ij} \neq 0$, $C_{ij} \neq 1$), whose values are not 0 or 1, are established as the corresponding disjunction logic expressions:

$$L_{ij} = \bigvee_{a_i \in C_{ij}} a_i. \quad (2)$$

Step 4. Put all the disjunction logic expressions L_{ij} into conjunction operation to obtain

$$L = \bigwedge_{C_{ji} \neq 0, C_{ji} \neq 1} L_{ij}. \quad (3)$$

Step 5. Convert the conjunctive normal L into the disjunctive normal form:

$$L' = \bigvee_i L_i. \quad (4)$$

Step 6. Output attribute reduction results. Each conjunction of disjunctive normal form is in accordance with one attribute reduction. So the condition attributes set after reduction is composed of all attributes in every conjunction.

3.4.3. Validation of the Improved Discernibility Matrix Algorithm. The validation experiments on the improved discernibility matrix algorithm are carried out by aiming at the rough attribute reduction of the fault diagnosis system for polymerization kettle. The condition attributes S_1, S_2, S_3 , and S_4 are,

TABLE 2: Fault diagnosis decision table of polymerizer.

Sample (U)	Condition attribute								Decision attribute (D)
	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	
1	1	1	2	4	2	2	5	1	4
2	1	2	1	3	2	1	1	1	0
3	1	2	1	3	1	2	1	1	0
4	1	1	2	5	2	4	2	1	1
5	1	2	1	4	2	2	2	2	3
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
40	1	1	2	4	2	2	5	1	4

respectively, the reducer vibration value of polymerization kettle (S_1 , mm), the stirring current (S_2 , A), the pressure of mechanical seal (S_3 , MPa), and the operating pressure (S_4 , MPa), whose corresponding variables are denoted as a , b , c , and d . The variable of fault types (D) is e . The discrete decision

table is shown in Table 3, in which the condition attribute set is $C = \{a, b, c, d\}$ and the decision attribute set is $D = \{e\}$.

Based on the definition of the discernibility matrix, C_D is described as follows:

$$\begin{bmatrix} 0 \\ 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ d & d & d & bd & cd & ad & 0 \\ bd & bd & bd & bcd & bcd & abd & bd & 0 \\ cd & bd & cd & bcd & d & acd & cd & 0 & 0 \\ abd & abd & ab & ad & abcd & bd & abd & ad & abcd & 0 \\ cd & cd & cd & bcd & d & acd & 0 & bcd & d & abcd & 0 \\ ad & ad & ad & abd & acd & d & 0 & abd & abcd & bd & 0 & 0 \end{bmatrix}. \quad (5)$$

Therefore, the final attributes reduction results according to the discernibility matrix are $\{a, d\}$ or $\{b, d\}$, namely, $\{S_1, S_4\}$ or $\{S_2, S_4\}$. Because the single element of the above matrix is d , according to the improved discernibility matrix algorithm, the attribute kernel is $d(S_4)$. So all combinations containing d are set as zero to get the simpler discernibility matrix:

$$\begin{bmatrix} 0 \\ 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ d & d & d & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d & 0 & 0 & 0 & 0 \\ 0 & 0 & ab & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d & 0 & 0 & 0 & d & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & d & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (6)$$

The matrix obtained quickly is described as follows:

$$L = d \wedge d \wedge d \wedge d \wedge (a \vee b) \wedge d \wedge d \wedge d = d \wedge (a \vee b),$$

$$L' = (a \wedge d) \vee (b \wedge d).$$

(7)

So the core attribute is $d(S_4)$ and the attribute reduction result is $\{a, d\}$ or $\{b, d\}$, namely, $\{S_1, S_4\}$ or $\{S_2, S_4\}$. The above reduction results based on two methods show the correctness of the proposed algorithm and the proposed method has characteristics of simple calculation and less errors. It especially can have the superiority when processing the discernibility matrix with higher dimension.

3.4.4. Attribute Reduction by Utilizing the Improved Discernibility Matrix. In order to facilitate the data reduction, Table 2 is divided into two separate decision tables, which respectively include attributes $\{S_1, S_2, S_3, S_4\}$ and $\{S_5, S_6, S_7, S_8\}$. Because one shortcoming of the rough set theory applied in the fault diagnosis of polymerization kettle is that it is sensitive to noise to produce the inconsistent rules, for the rules extraction of compatible system, the reliability and the support degree of each rule are calculated. Then by setting two thresholds, when reliability and the support degree are larger than the corresponding threshold, the rule is acquired, or the rule is deleted. By getting rid of the repetition and incompatible lines of the decision table, the partition decision (Table 4) is obtained.

According to the improved discernibility matrix algorithm, it is concluded that the core attribute is $\{S_4\}$ and

TABLE 3: Initial decision-making table.

Sample (U)	Condition attribute				Decision attribute (D)
	a, S_1	b, S_2	c, S_3	d, S_4	
1	1	1	1	1	0
2	1	1	1	3	0
3	1	1	1	2	0
4	1	2	1	3	0
5	1	1	2	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
12	2	1	1	5	1

TABLE 4: Partition decision table.

Sample (U)	Condition attribute				Decision attribute (D)
	a, S_1	b, S_2	c, S_3	d, S_4	
1	1	1	1	1	0
2	1	1	1	3	0
3	1	1	1	2	0
4	1	2	1	3	0
5	1	1	2	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
12	2	1	1	5	1

the reduction is $\{S_1, S_4\}$ or $\{S_2, S_4\}$. Similarly, another block decision table is shown in Table 5. So the core attributes and the reduction are $\{S_5, S_6, S_7\}$, namely, $\{e, f, g\}$.

Finally, by using the attribute reduction method based on the improved discernibility matrix, the attributes reduction of the decision table is carried out to obtain the minimum attribute reduction $\{S_1, S_4, S_5, S_6, S_7\}$ or $\{S_2, S_4, S_5, S_6, S_7\}$.

4. BP Neural Network

4.1. Improved BP Neural Network. Back propagation (BP) neural network is a kind of multilayer forward neural network. BP algorithm has merits of parallel processing, distributed storage, and adaptive learning, but there are also slow convergence speed and local minimum [19, 20]. Many proposed improved BP learning algorithm may be divided into two categories: (1) the methods based on the gradient descent including momentum BP algorithm (MOBP) and variable leaning-rate BP algorithm (VLBP); (2) the methods based on the numerical optimization including conjugate gradient BP algorithm (CGBP) and Levenberg-Marquardt BP algorithm (LMBP).

4.1.1. Heuristic Improved BP Algorithm

(1) *Momentum Back Propagation Algorithm (MOBP).* The basic BP learning algorithm adjusts weights only along the gradient descent direction of error in t moment and does not consider the direction before t moment. So the learning process often has oscillation and converges slowly. In order

TABLE 5: Partition decision table.

Sample (U)	Condition attribute				Decision attribute (D)
	e, S_5	f, S_6	g, S_7	h, S_8	
1	2	2	1	1	0
2	2	2	2	1	0
3	1	2	1	1	0
4	2	1	1	1	0
5	2	2	2	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
13	2	2	5	1	4

to improve the training speed, a momentum component is added into the weights adjustment formula:

$$\Delta w(t) = \eta [a\Delta w(t-1) - (1-a)\delta O], \quad (8)$$

where w is the weight matrix, O is the output vector, η is the learning ratio, and $a \in [0, 1]$ is the momentum coefficient. Define the momentum item $a\Delta w(t-1)$ to reflect the prior adjustment experiences. $a = 0$ means that the weights adjustment is only concerned with the current negative gradient. $a = 1$ means that the weights adjustment depends on the negative gradient in the last cycle. The momentum component can be regarded as a low-pass filter, which smoothes the oscillation in the learning process and improves the convergence velocity.

In addition, when the error gradient occurred the local minimum, though $\Delta w_{ijk} = -\eta(\partial E/\partial w_{ijk}) = -\eta\delta_{ik}o_{ij} \rightarrow 0$, $\Delta w(t-1) \neq 0$ can make it jump out of the local minimum and speed up the iterative convergence speed.

(2) *Variable Leaning-Rate Back Propagation Algorithm (VLBP).* In the gradient descent method, the learning speed has great influence on the whole training process. That is to say the training success or not mainly depends on the selection of learning speed. If it can improve learning speed in relatively flat surface and reduce it when slope increases, the convergence speed can be improved. This kind of VLBP rule is described as follows:

$$\Delta w(t) = \eta [a\Delta w(t-1) - (1-a)\delta O]. \quad (9)$$

- If mean square error (in the whole training set) increases to be larger than the predefined percentage ξ (typical value is 1–5%), the weight update is cancelled. Then the learning speed is multiplied by a factor ρ ($0 < \rho < 1$) and the momentum coefficient a is set to zero.
- If mean square error (in the whole training set) decreases after weights update, the weight update is accepted. Then the learning speed is multiplied by a factor $\beta > 1$. If the momentum coefficient a was to be set zero, it will be back to the previous value.
- If the growth of the mean square error is less than ξ , the weights update is accepted, but learning speed

TABLE 6: Performance indexes of four kinds of improved BP training algorithm.

Improved BP algorithm	MOBP algorithm	VLBP algorithm	CGBP algorithm	LMBP algorithm
Iterations	Without reaching a designated error limit in predetermined steps	416	175	15
Training error		0.00946039	0.00938337	0.009994105

TABLE 7: Testing sample data set.

Sample (U)	Historical datum of polymerizer								Diagnosis type (D)
	a, S_1	b, S_2	c, S_3	d, S_4	e, S_5	f, S_6	g, S_7	h, S_8	
1	0.37	119.2	1.57	1.43	93.15	42.4	76.19	29.57	4
2	0.42	200.1	1.09	1.39	92.83	38.5	58.84	29.42	0
3	0.42	211.8	0.94	1.34	56.69	78.8	59.62	29.34	0
4	0.42	130.2	1.53	2.02	98.82	99.2	63.51	29.55	1
5	0.44	188.6	1.01	1.57	93.71	78.7	65.45	82.64	3
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
10	0.55	186.6	1.15	1.11	94.73	38.3	61.35	36.81	2

remains unchanged. If the momentum coefficient a was to be set zero, it will be back to the previous value.

The heuristic BP learning algorithm will improve the convergence speed for some problems. But these methods have two main shortcomings. First of all, these improvements need to set more parameters (such as ρ, β, ξ , etc.), but the standard BP algorithm only needs a parameter (learning speed), which results in the performance of the algorithm sensitive to the changes of these parameters. The second shortcoming of these algorithms is that they can find the solution of the problem but not convergence for some applications.

4.1.2. Numerical Optimization Technique. The steepest descent method, the conjugate gradient method, and the Newton method are three widely used techniques in numerical optimization. The steepest descent method is the simplest algorithm, but the speed of convergence is slower. The convergence speed of Newton method is more quickly, but it needs to calculate Hessian matrix and its inverse matrix. The conjugate gradient method is compromise in that it does not need to calculate second derivative and still has the characteristics of quadratic convergence.

(1) Conjugate Gradient Back Propagation Algorithm (CGBP). The conjugate gradient method is an improved gradient method, which can improve the shortcomings that the oscillation of gradient method is too big and the convergence is poor. Its basic idea is to calculate the conjugate direction of the negative gradient direction and the previous searching direction in order to speed up the training speed and improve the training accuracy. All of the conjugate gradient methods adopt the negative gradient direction as the initial convergence direction:

$$p_0 = -g_0, \quad (10)$$

where g_k is error gradient direction; that is to say,

$$g_k = \nabla F(x) \Big|_{x=x_k} \quad (11)$$

$F(x)$ stands for the weight error surface in the form of mean square sum. Then the learning speed a_k is selected to minimize the function along the searching direction:

$$x_{k+1} = x_k + a_k p_k. \quad (12)$$

Then the conjugate direction is used as a new search direction. Usually the previous search direction is added to the current gradient direction:

$$p_k = -g_k + \beta_k p_{k-1}. \quad (13)$$

If the algorithm does not obtain convergence, return to (13). The selection of β_k derived various conjugate gradient methods, such as the following.

The conjugate gradient method of Fletcher-Reeves is defined as

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}. \quad (14)$$

The conjugate gradient method of Polak-Ribiere is defined as

$$\beta_k = \frac{\Delta g_k^T g_k}{g_{k-1}^T g_{k-1}}. \quad (15)$$

(2) Levenberg-Marquardt Back Propagation Algorithm (LMBP). The principle of Newton algorithm is to seek the approximated quadratic stagnation of $F(x)$, which represents the weight error surface in the form of mean sum square. Assume

$$\nabla F(x) = A_k \Delta x_k + g_k. \quad (16)$$

Then, the gradient of the function $F(x)$ to Δx_k is defined as follows:

$$\nabla F^2(x) = A_k. \quad (17)$$

TABLE 8: Fault diagnosis results of polymerizer.

Test sample	1	2	3	4	5	6	7	8	9	10
BP output	0.9988	0.0000	0.0000	0.0016	0.0012	0.0000	0.0004	0.0018	0.0000	0.0012
	0.0039	0.0000	0.0002	0.0122	0.9833	0.0001	0.9848	0.0132	0.0002	0.0000
	0.0004	0.0011	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9837
	0.0008	0.0010	0.0026	0.9986	0.0019	0.0002	0.0033	0.9978	0.0002	0.0000
Real code	1000	0000	0000	0001	0100	0000	0100	0001	0000	0010
Diagnosis type	Gland-shaft fault	Normal	Normal	Motor fault	Mechanical seal fault	Normal	Gland-shaft fault	Motor fault	Normal	Reducer fault

The bold values represent the fault diagnosis results of BP neural network by using test samples.

TABLE 9: Fault diagnosis results.

Test sample	1	2	3	4	5	6	7	8	9	10
BP output	0.9468	0.0212	0.0080	0.0271	0.0001	0.0359	0.0001	0.0282	0.0117	0.0644
	0.0000	0.0032	0.0090	0.0000	0.9895	0.0008	0.9978	0.0000	0.0057	0.0541
	0.0007	0.0004	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0001	0.9295
	0.0165	0.0000	0.0000	0.9765	0.0000	0.0000	0.0001	0.9732	0.0000	0.0000
Diagnosis type	Gland-shaft fault	Normal	Normal	Motor fault	Mechanical seal fault	Normal	Gland-shaft fault	Motor fault	Normal	Reducer fault

The bold values represent the fault diagnosis results of BP neural network by using test samples.

If (17) is set to zero, then obtain

$$\Delta x_k = -A^{-1} g_k. \quad (18)$$

The Newton algorithm is defined as follows:

$$x_{k+1} = x_k - A^{-1} g_k, \quad (19)$$

where $A_k = \nabla^2 F(x)$ is the secondary derivative of the current weight error function.

The Newton method is fast in convergence speed, but in each iteration, it requires to calculate the second derivative Hessian matrix of $F(x)$, which makes the calculation amount become very large. If $F(x)$ is represented in the form of quadratic sum, namely,

$$F(x) = \sum_{i=1}^N e_i^2(x) = e^T(x) e(x) \quad (20)$$

then the j th gradient component is calculated by

$$[\nabla F(x)]_j = \frac{\partial F(x)}{\partial x_j} = 2 \sum_{i=1}^N e_i(x) \frac{\partial e_i(x)}{\partial x_j}. \quad (21)$$

Therefore, the gradient can be written in the following matrix form:

$$2J^T(x) e(x), \quad (22)$$

where

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \frac{\partial e_N(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \frac{\partial e_N(x)}{\partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial e_N(x)}{\partial x_1} & \frac{\partial e_N(x)}{\partial x_2} & \dots & \frac{\partial e_N(x)}{\partial x_n} \end{bmatrix}, \quad (23)$$

where $J(x)$ is a Jacobian matrix. So the Hessian matrix can approximately be replaced by the following matrix:

$$H = J^T J. \quad (24)$$

The gradient is defined as follows:

$$g = J^T e, \quad (25)$$

where J contains the first order derivative of the network training error and it is a function of the weights and threshold values. e is the error vector of the network training. So the training of the LMBP algorithm is described as

$$x_{k+1} = x_k - [J^T(x_k) J(x_k) + \mu_k I]^{-1} J^T(x_k) e(x_k). \quad (26)$$

The characteristic of this algorithm is that when μ_k increases, it is close to the steepest descent method with a small learning speed

$$x_{k+1} \approx x_k - \frac{1}{\mu_k} J^T(x_k) e(x_k) = x_k - \frac{1}{2\mu_k} \nabla F(x). \quad (27)$$

When, in (26), μ_k decreases to zero, it becomes to the Newton method with an approximate Hessian matrix. In the

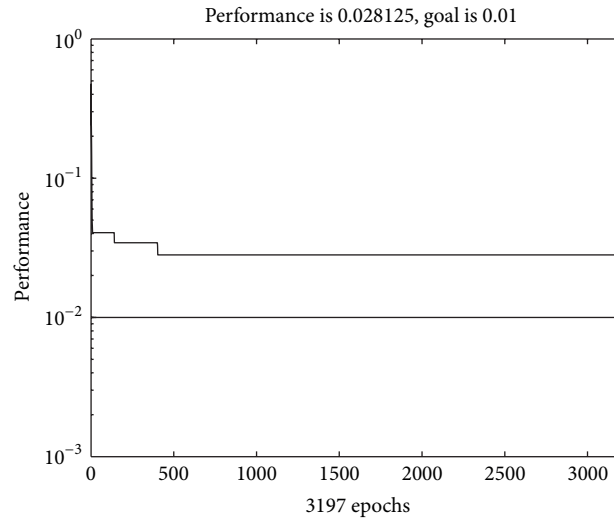
TABLE 10: Fault diagnosis results.

Test sample	1	2	3	4	5	6	7	8	9	10
BP output	0.9957	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007
	0.0000	0.0033	0.0174	0.0162	0.9767	0.0008	0.9604	0.0173	0.0012	0.0000
	0.0001	0.0210	0.0006	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.9993
	0.0000	0.0000	0.0036	0.9998	0.0000	0.0000	0.0001	0.9965	0.0000	0.0000
Diagnosis type	Gland-shaft fault	Normal	Normal	Motor fault	Mechanical seal fault	Normal	Gland-shaft fault	Motor fault	Normal	Reducer fault

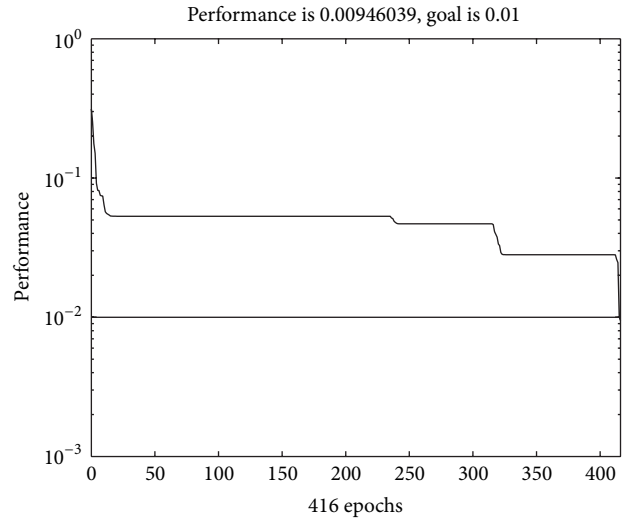
The bold values represent the fault diagnosis results of BP neural network by using test samples.

TABLE 11: Performance contrast of fault diagnosis results under 10-fold cross validation.

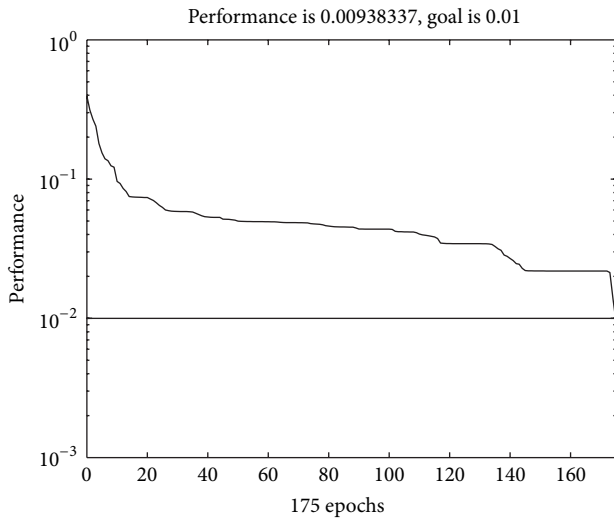
Attribute set	Average training error	Average training step	Average accuracy (%)
$\{S_1, S_4, S_5, S_6, S_7\}$	7.12×10^{-5}	21	98.8
$\{S_2, S_4, S_5, S_6, S_7\}$	2.37×10^{-4}	37	97.2
$\{S_1 \sim S_8\}$	5.15×10^{-5}	134	99.1



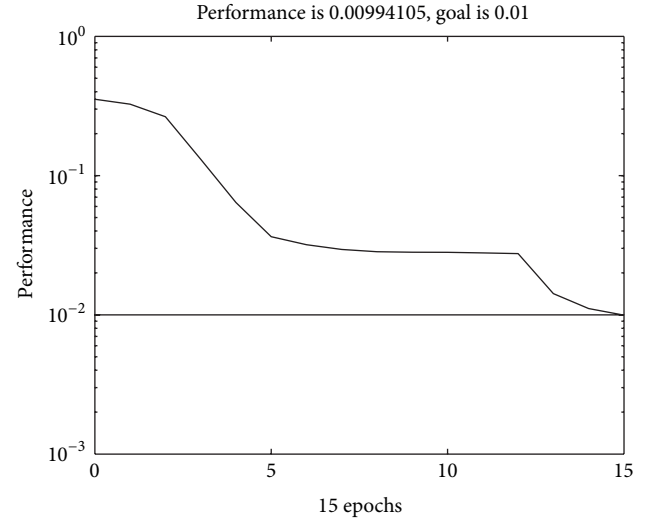
(a) MOBP algorithm



(b) VLBP algorithm



(c) CGBP algorithm



(d) LMBP algorithm

FIGURE 3: Training process based on four training LMBP algorithms.

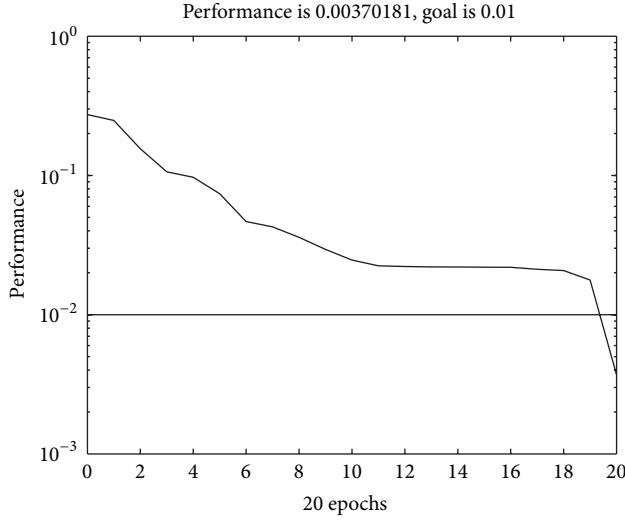


FIGURE 4: Network training process.

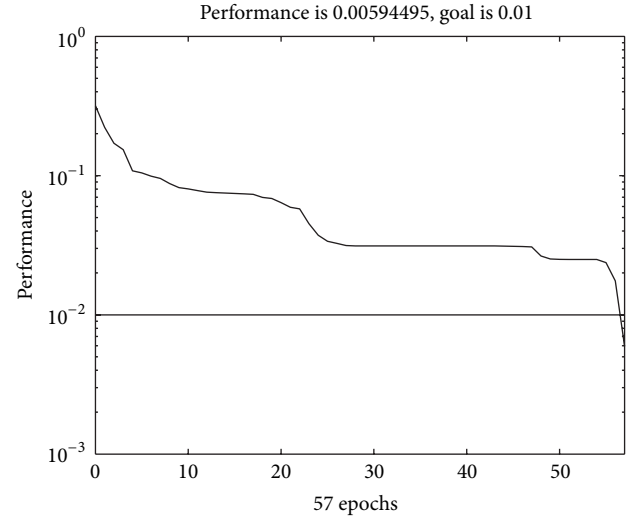


FIGURE 6: Network training process.

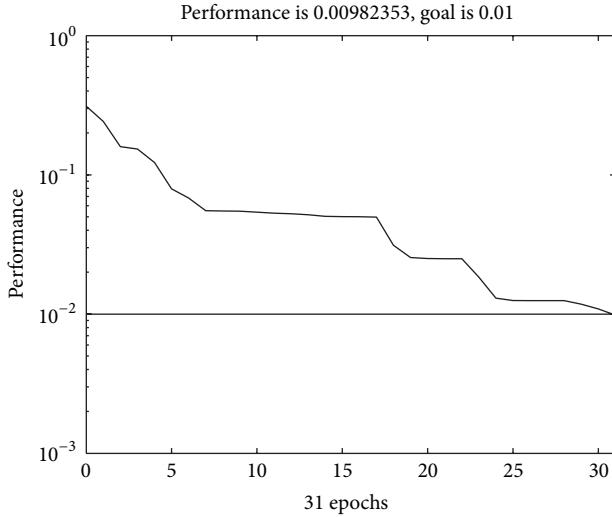


FIGURE 5: Network training process.

iterative process utilizing LMBP algorithm, the value of μ will be reduced if the training is successful; otherwise, it will increase. When the squared error sum is reduced to a predefined value, the algorithm is considered to converge.

4.2. Simulation of Improved BP Neural Network. Four kinds of the improved BP training algorithm (MOBP, VLBP, CGBP, and LMBP) are adopted to carry out the simulation experiments for polymerization data, and a performance comparison is analyzed in details in order to guideline the selection of the BP learning algorithm for the fault diagnosis system of the polymerizer kettle. The maximum number of training cycles is 5000 and the error limit is 0.001. The training samples are selected as follows. $P = [1\ 4\ 2\ 2\ 5; 1\ 3\ 2\ 1\ 1; 1\ 3\ 1\ 2\ 1; 1\ 5\ 2\ 4\ 2; 1\ 4\ 2\ 2\ 2]^T$; $T = [1\ 0\ 0\ 0; 0\ 0\ 0\ 0; 0\ 0\ 0\ 0; 0\ 0\ 0\ 1; 0\ 1\ 0\ 0]^T$. The training process of the four algorithms for the above training samples is shown in Figures 3(a)–3(d). The

performance indexes of four kinds of improved BP training algorithm are listed in Table 6.

It can be seen from the above simulation results, for the provided training samples, the network training methods based on the improved BP learning algorithm can converge to the specified error after designating 5000 iterations in addition to MOBP algorithm. On the other hand, the LMBP learning algorithm has a great advantage in training speed. So it may be used in the discussed polymerization fault diagnosis system.

5. Simulation Experiments

5.1. Network Training Simulation. In this paper, a typical three-layer BP neural network is used for the polymerizer fault diagnosis system, in which the number of nodes in the input layer is 5, the number of nodes in the output layer is 4, and through experiments comparison the number of nodes in the hidden layer nodes is 9. So the structure of the LMBP neural network of polymerizer fault diagnosis system is $N(5, 9, 4)$. The learning rate lr is 0.01 and the training expectations error is 0.0001. When testing the polymerizer fault diagnosis system based on the rough sets neural network, the data in Table 2 with the attributes $\{S_1, S_4, S_5, S_6, S_7\}$ are selected as the training samples and the data in Table 7 with same condition attributes are selected as the test samples for training and testing the LMBP neural network. The network training steps for the test data are shown in Figure 4, and the fault diagnostic results are shown in Table 8.

Similarly, the data in Table 2 with the attributes $\{S_2, S_4, S_5, S_6, S_7\}$ are selected as the training samples and the data in Table 7 with same condition attributes are selected as the test samples for training and testing the LMBP neural network. The network training steps for the test data are shown in Figure 5, and the fault diagnostic results are shown in Table 9. Finally, the data in Table 2 with the attributes $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$ are selected as the training samples and the data in Table 7 with same condition

attributes are selected as the test samples for training and testing the LMBP neural network. The network training steps for the test data are shown in Figure 6, and the fault diagnostic results are shown in Table 10.

5.2. Performance Comparison of Fault Diagnosis. Many performance evaluation methods have been proposed, of which various cross validations are most popularly used, such as 3-fold cross validation, 10-fold cross validation, leave-one-out cross-validation (LOOCV), and others. Here, we used the 10-fold cross validation to evaluate the performance of three condition attributes $\{S_1, S_4, S_5, S_6, S_7\}$, $\{S_2, S_4, S_5, S_6, S_7\}$, and $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$. The results by using 10-fold cross validation also are shown in Table 11. In Table 11, we list the average training error, training step, and average accuracy by using 10-fold cross validation based on the same samples described in Table 1.

It can be seen from Table 11, for the training error and training time of the discussed samples, the attributes reduction results based on the improved discernibility matrix algorithm are attributes set $\{S_1, S_4, S_5, S_6, S_7\}$ and $\{S_2, S_4, S_5, S_6, S_7\}$. So the network training steps are significantly lower than the training steps of attributes set $\{S_1-S_8\}$. This greatly reduced the training time and increased the fault diagnosis accuracy. In conclusion, the attribute set $\{S_1, S_4, S_5, S_6, S_7\}$ after the attribute reduction is more suitable for the polymerization fault diagnosis system, which reflects the efficiency of rough set neural network and the effectiveness and practicality of the proposed method are proved.

6. Conclusions

The improved attributes reduction algorithm based on discernibility matrix is proposed for rough set, which simplified the single property element of the matrix and the element containing the single attribute in mathematics to reduce the calculated amount and the error probability. On the other hand, the fault diagnosis method combining the attributes reduction algorithm based on the improved discernibility matrix and the LMBP neural network is applied to the fault diagnosis of the polymerizer process. Simulation results show that the proposed method has higher diagnosis accuracy and shorter training time.

Acknowledgments

This work was supported by China Postdoctoral Science Foundation (no. 20110491510), Program for Liaoning Excellent Talents in University (no. LJQ2011027), and Special Research Foundation of University of Science and Technology of Liaoning (no. 2011zx10).

References

- [1] S. Z. Gao, J. S. Wang, and X. W. Gao, "Modeling and advanced control method of PVC polymerization process," *Journal of Process Control*, vol. 23, no. 5, pp. 664–681, 2013.
- [2] X. Zhang, J. Zhou, J. Guo, Q. Zou, and Z. Huang, "Vibrant fault diagnosis for hydroelectric generator units with a new combination of rough sets and support vector machine," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2621–2628, 2012.
- [3] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, no. 1, pp. 696–711, 2012.
- [4] C. Ma, X. Gu, and Y. Wang, "Fault diagnosis of power electronic system based on fault gradation and neural network group," *Neurocomputing*, vol. 72, no. 13–15, pp. 2909–2914, 2009.
- [5] C. Wang, J. Zhou, H. Qin, C. Li, and Y. Zhang, "Fault diagnosis based on pulse coupled neural network and probability neural network," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14307–14313, 2011.
- [6] M. Demetgul, I. N. Tansel, and S. Taskin, "Fault diagnosis of pneumatic systems with artificial neural network algorithms," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10512–10519, 2009.
- [7] D. Bhalla, R. K. Bansal, and H. O. Gupta, "Integrating AI based DGA fault diagnosis using Dempster-Shafer Theory," *International Journal of Electrical Power & Energy Systems*, vol. 48, pp. 31–38, 2013.
- [8] Z. Chenglin, S. Xuebin, S. Songlin, and J. Ting, "Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9908–9912, 2011.
- [9] N. Li, R. Zhou, Q. Hu, and X. Liu, "Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine," *Mechanical Systems and Signal Processing*, vol. 28, pp. 608–621, 2012.
- [10] N. R. Sakthivel, V. Sugumaran, and B. B. Nair, "Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump," *Mechanical Systems and Signal Processing*, vol. 24, no. 6, pp. 1887–1906, 2010.
- [11] N. R. Sakthivel, V. Sugumaran, and B. B. Nair, "Automatic rule learning using roughset for fuzzy classifier in fault categorization of mono-block centrifugal pump," *Applied Soft Computing Journal*, vol. 12, no. 1, pp. 196–203, 2012.
- [12] J. S. Wang and Q. P. Guo, "D-FNN based soft-sensor modeling and migration reconfiguration of polymerizing process," *Applied Soft Computing*, vol. 13, no. 4, pp. 1892–1901, 2013.
- [13] Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Information Sciences*, vol. 179, no. 7, pp. 867–882, 2009.
- [14] Y. Zhao, Y. Yao, and F. Luo, "Data analysis based on discernibility and indiscernibility," *Information Sciences*, vol. 177, no. 22, pp. 4959–4976, 2007.
- [15] H.-X. Li and X.-Z. Zhou, "Heuristic attribute reduction based on 0-1 discernibility matrix," *Journal of Central South University (Science and Technology)*, vol. 40, no. 1, pp. 304–308, 2009.
- [16] X. Hu and N. Cercone, "Learning in relational databases: a rough set approach," *Computational Intelligence*, vol. 11, no. 2, pp. 323–338, 1995.
- [17] J. Qian, D. Q. Miao, Z. H. Zhang, and W. Li, "Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation," *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212–230, 2011.
- [18] D.-Q. Yan, K.-Q. Li, and Z.-X. Chi, "Soft discernibility matrix and attribute reduction for information systems," *Journal on Communications*, vol. 30, no. 8, pp. 45–50, 2009.
- [19] W. Xuanyin, L. Xiaoxiao, and L. Fushang, "Analysis on oscillation in electro-hydraulic regulating system of steam turbine

and fault diagnosis based on PSOBP,” *Expert Systems with Applications*, vol. 37, no. 5, pp. 3887–3892, 2010.

- [20] D. Ma, Y. Liang, X. Zhao, R. Guan, and X. Shi, “Multi-BP expert system for fault diagnosis of power system,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 3, pp. 937–944, 2013.

