Available online at www.sciencedirect.com

SCIENCE DIRECT®

Mechanical Systems and Signal Processing 18 (2004) 645–657

Mechanical Systems and Signal Processing

www.elsevier.com/locate/jnlabr/ymssp

ART–KOHONEN neural network for fault diagnosis of rotating machinery

B.S. Yang*, T. Han, J.L. An

School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea

Received 28 November 2002; received in revised form 14 April 2003; accepted 9 May 2003

Abstract

In this paper, a new neural network (NN) for fault diagnosis of rotating machinery which synthesises the theory of adaptive resonance theory (ART) and the learning strategy of Kohonen neural network (KNN), is proposed. For NNs, as the new case occurs, the corresponding data should be added to their dataset for learning. However, the 'off-line' NNs are unable to adapt autonomously and must be retrained by applying the complete dataset including the new data. The ART networks can solve the plasticity–stability dilemma. In other words, they are able to carry out 'on-line' training without forgetting previously trained patterns (stable training); it can recode previously trained categories adaptive to changes in the environment and is self-organising. ART–KNN also holds these characteristics, and more suitable than original ART for fault diagnosis of machinery. In order to test the proposed network, the vibration signal is selected as raw inputs due to its simplicity, accuracy and efficiency. The results of the experiments confirm the performance of the proposed network through comparing with other NNs, such as the self-organising feature maps (SOFMs), learning vector quantisation (LVQ) and radial basis function (RBF) NNs under the same conditions.The diagnosis success rate for the ART–Kohonen network was 100%, while the rates of SOFM, LVQ and RBF networks were 93%, 93% and 89%, respectively.

Keywords: Artificial neural network; Fault diagnosis; Rotating machinery; Vibration signal; Feature extraction

1. Introduction

Presently, the fault diagnosis is increasingly intelligent with wide applications of artificial neural networks (NNs). However, 'off-line' NNs are unable to adapt well to unexpected changes in the

*Corresponding author. Tel.: +82-51-620-1604; fax: +82-51-620-1405.
E-mail address: bsyang@pknu.ac.kr (B.S. Yang).

0888-3270/03/$ - see front matter
doi:10.1016/S0888-3270(03)00073-6

environment. Furthermore, the data of the dataset used to train networks need be added, as new fault occurs. In this case, the 'off-line' network requires to be retrained using the complete dataset. This can result in a time-consuming and costly process [1]. In the real world, although part of fault signals can be obtained, it is very difficult to compose the training dataset representing the features of all faults. Nobody knows what will happen next time. These characteristics limit the applications of 'off-line' NNs in fault diagnosis field. The NNs for fault diagnosis of machinery are required to learn gradually the knowledge in operating process, and to have the adaptive function expanding the knowledge continuously without the loss of the previous knowledge during learning new knowledge. A human brain is able to learn many new events without necessarily forgetting events that occurred in the past. So we want an intelligent system capable of adapting 'on-line' to changes in the environment, the system should be able to deal with the so-called stability–plasticity dilemma [2–5]. That is, the system should be designed to have some degree of plasticity to learn new events in a continuous manner, and should be stable enough to preserve its previous knowledge, and to prevent new events destroying the memories of prior training. As a solution to this problem, the adaptive resonance theory (ART) networks were developed, and have been applied with some success to real-time training and classification [6]. The ART network is an NN that self-organises stable recognition codes in real time in response to arbitrary sequences of input patterns, and is a vector classifier as the mathematical model for the description of fundamental behavioural functions of the biological brain such as the learning, parallel and distributed information storage, short- and long-term memory and pattern recognition.

The Kohonen neural network (KNN) also is called self-organising feature map network (SOFM); it defines a forward two-layer NN that implements a characteristic non-linear projection from the high-dimensional space of sensory or other input signals onto a low-dimensional array of neurons [7–9]. The KNN consists of three major steps: competition, co-operation and adaptation. In the first step, the network compares the output values with the input vector according to a chosen discriminating function. Among the output neurons, only one particular neuron with the closest relationship to the input vector is picked up and labelled as the winning (best-matching) neuron. Once the winning neuron is picked up, the next step is to select those neurons within a predefined neighbourhood. Only the weights of those neurons defined within the topological neighbourhood of the winning neuron will be updated. The synaptic weights of neurons outside the neighbourhood will remain unchanged. As the winning neuron best matches the input vector in the sense of the Euclidean distance, the above learning strategy is able to move the synaptic weight vectors towards the distribution of the input vectors.

In this paper, we proposed a fault diagnosis network, the adaptive resonance theory–Kohonen neural network (ART–KNN), which does not destroy the initial learning and can adapt the additional training data that are suitable for fault diagnosis of rotating machinery. The validity of ART–KNN is examined through the experimental results.

2. ART–Kohonen neural network (ART–KNN) algorithm

The characteristics of ART networks are suitable for condition monitoring and fault diagnosis. There are two general classes of ART networks: ART1 and ART2 and ART3. The ART1 is for classifying the binary input patterns, while the ART2 and ART3 are for the binary and decimal input patterns.

But the ART networks have some disadvantages for the fault diagnosis. The input patterns to input layer are normalised before passing through the adaptive filter that is defined between input layer and discernment layer. Because the absolute values of input signals represent only the image brightness and sound level for the image and sound classifications, the relative value normalised is important to analyse the image and sound discrimination. But, the absolute value of vibration signal is important information for the fault diagnosis. When it is normalised, some important information to detect the faults may be lost. At the same time, the ART2 and ART3 adequately control the noise of input signal; the initial signal becomes fussy after filtering. So the features of fault signals are destroyed to some degree.

In this paper, the proposed ART–KNN combines the theory of ART with Kohonen's learning strategy to realise machinery fault diagnosis. The architecture of ART–KNN is shown in Fig. 1. It is similar to ART1's, excluding the adaptive filter. ART–KNN is also formed by two major subsystems: the attentional subsystem and the orienting subsystem. Two interconnected layers, discernment layer and comparison layer, which are fully connected both bottom-up and top-down, comprise the attentional subsystem. The application of a single input vector leads to patterns of neural activity in both layers. The activity in discernment nodes reinforces the activity in comparison nodes due to top-down connections. The interchange of bottom-up and top-down information leads to a resonance in neural activity. As a result, critical features in comparison are reinforced, and have the greatest activity. The orienting subsystem is responsible for generating a reset signal to discernment when the bottom-up input pattern and top-down template pattern mismatch at comparison, according to a similarity. In others words, once it has detected that the input pattern is novel, the orienting subsystem must prevent the previously organised category neurons in discernment from learning this pattern (via a reset signal). Otherwise, the category will

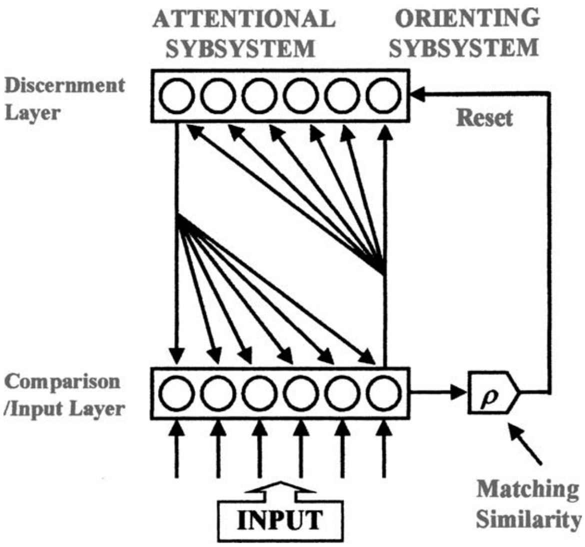


Fig. 1. Architecture of the ART–KNN network.

become increasingly non-specific. When a mismatch is detected, the network adapts its structure by immediately storing the novelty in additional weights. The similarity criterion is set by the value of the similarity parameter. A high value of the similarity parameter means than only a slight mismatch will be tolerated before a reset signal is emitted. On the other hand, a small value means that large mismatches will be tolerated. After the resonance check, if a pattern match is detected according to the similarity parameter, the network changes the weights of the winning node.

The learning strategy is introduced by the KNN. The Euclidean distances of all weights between input vector X and each neuron of the discernment layer are evaluated as the similarity given by the following equation, the smallest one becomes the winning neuron:

$$\|B_J - X\| < \|B_J - X\| \quad (j, J = 1, 2, \ldots, n;\ j \neq J) \tag{1}$$

where B_J is the weight of the jth neuron in the discernment layer, B_J is the weight of the winning neuron.

After producing the winning neuron, input vector X returns to the comparison layer. The absolute similarity S is calculated by

$$S = \frac{\|B_J\| - \|B_J - X\|}{\|B_J\|}. \tag{2}$$

If B_J and X in Eq. (2) are same, $\|B_J - X\|$ is equal to 0, and S is 1. The larger the Euclidean distance between B_J and X is, the smaller S is. A parameter ρ is introduced as the evaluation criterion of similarity. If $S > \rho$, it indicates that the Jth cluster is sufficiently similar to X. So X belongs to the Jth cluster. In order to make the weight more accurate to represent the corresponding cluster, the weight of the Jth cluster is improved by the following equation:

$$B_J = (n * B_{J0} + X)/(n + 1) \tag{3}$$

where B_J is the enhanced weight, B_{J0} is the origin weight, and n is changed time.

On the contrary, as $S < \rho$, it means that X is much different with the Jth cluster. Thus, there is no cluster that matches X in the original network. The network needs one more neuron to remember this new case by resetting in the discernment layer. The weight of the new neuron is given by

$$B_{n+1} = X. \tag{4}$$

3. Diagnosis system using ART–KNN

3.1. System structure

The fault diagnosis system is shown in Fig. 2. The system mainly consists of three sections: data acquisition, feature extraction and fault diagnosis. The raw time signal is obtained by the accelerometer from the machinery fault simulator, shown in Fig. 3, Then the features of the data are extracted through the discrete wavelet transform and feature extraction algorithms [13]. Wavelet transform is more effective than FFT in terms of data compression and is highly tolerant to the presence of additive noise and drift in the sensor responses. Feature extraction algorithms

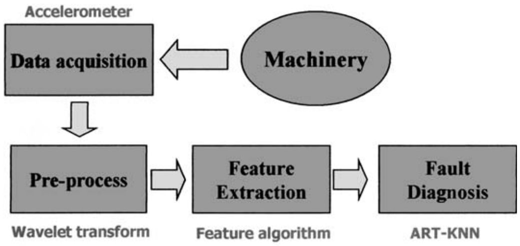


Fig. 2. Architecture of fault diagnosis system.

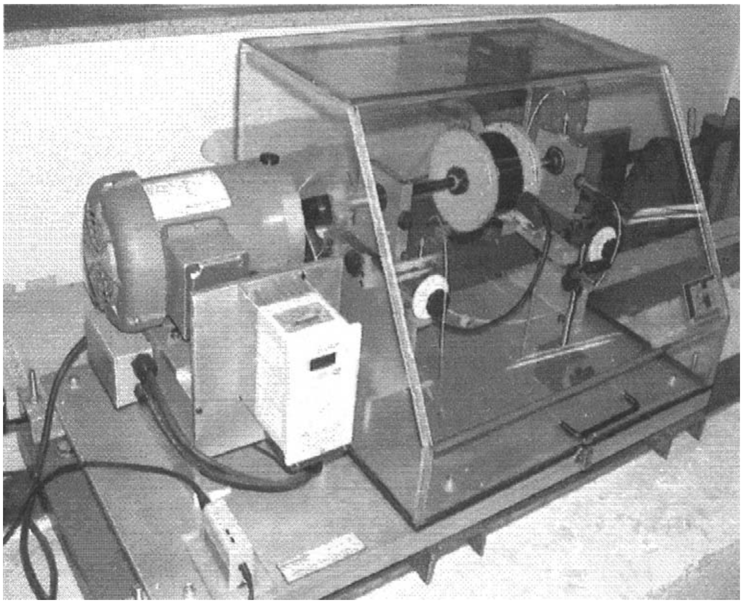


Fig. 3. Machinery fault simulator.

make data quantity from the view of statistics. Finally, the ART–KNN is trained and used to classify the faults of machinery.

3.2. Data acquisition

Experiments were performed on a machinery fault simulator, which can simulate the most common faults, such as misalignment, unbalance, resonance, ball bearing faults and so on. Its running span is from 0 to 6000 rpm. The schematic of the test apparatus is shown in Fig. 3. It mainly consists of a motor, a coupling, bearings, discs and a shaft.

The analysed faults are the bearing faults and structural faults, such as unbalance and misalignment. The faulty bearings were rolling element bearings that damaged on an inner race, an outer race, a ball and the combination of these faults. The misalignment faults, parallel misalignment and phase misalignment, were simulated by adjusting the simulator plane highness and degree. Adding an unbalance weight on the disc at the normal condition creates the unbalance.

A radial acceleration was picked up from an accelerometer located at the top of the right bearing housing. The shaft speed was obtained by one laser speedometer.

A total of eight conditions were tested: four types of bearing faults (inner race, outer race, ball and multiple), two misalignments (parallel and angular), one unbalance and one normal condition. Each condition was measured 20 times continuously. The frequency of used signal is 5000 Hz and the number of sampled data are 16 384. A mobile DSP analyser performed the data acquisition, and the data were collected into a notebook computer. Parts of the condition signals are shown in Fig. 4.

3.3. Feature extraction [13]

Firstly, one-dimensional (1-D) discrete wavelet transform was used to decompose the time signal into three levels. Then the transformed signal and the original signal are estimated by eight feature parameters (mean, standard deviation, rms, shape factor, skewness, kurtosis, crest factor

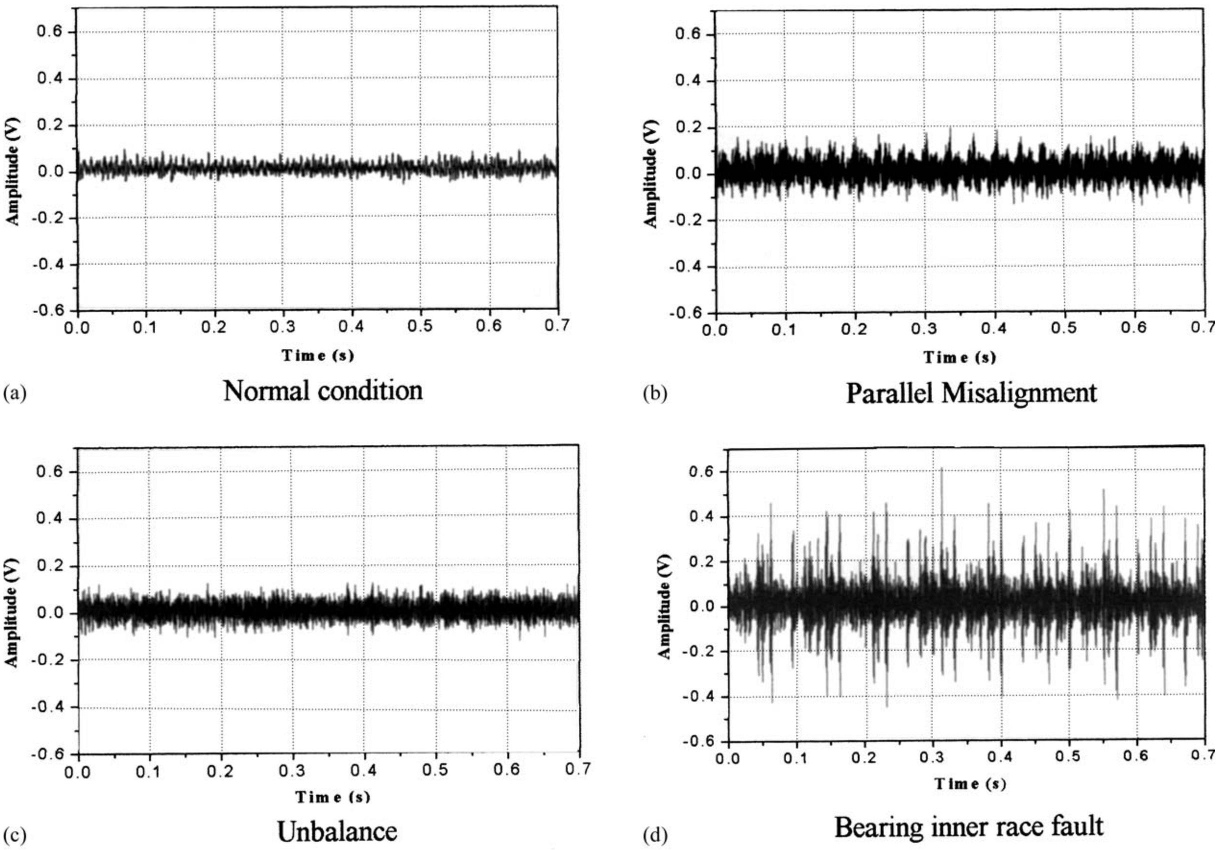


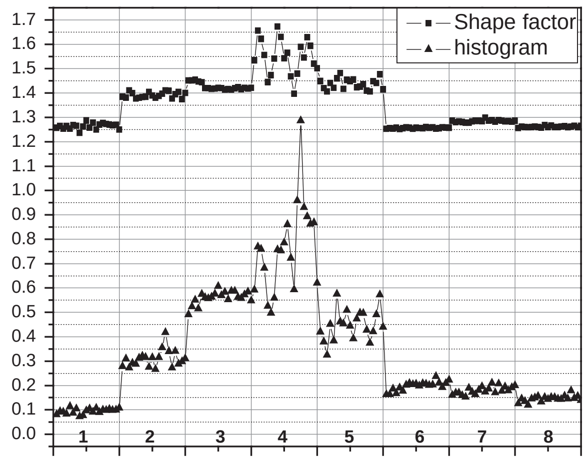
Fig. 4. The vibration signals from the machinery fault simulator.

and entropy estimation). Finally, a total of 32 feature parameters were obtained, shown in Fig 5. However, having too many feature parameters is a burden of networks, much time is needed to calculate the results. Usually, five to ten parameters are good from the view of calculation time and accuracy [10,11]. In order to solve this problem, a new parameter evaluation technique is proposed to select eight parameters that can well represent the fault features from 32 parameters.

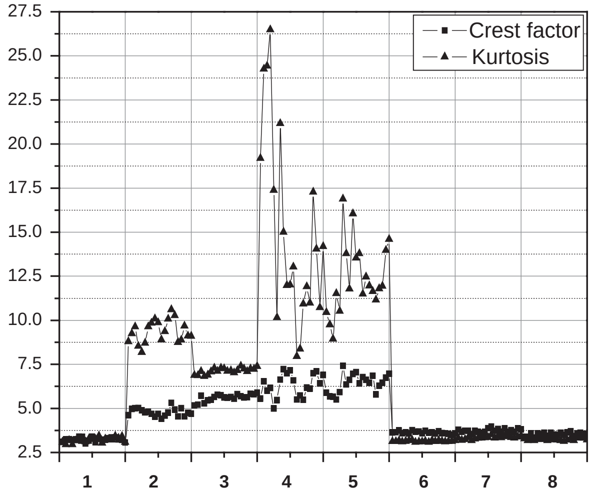
Step 1: Calculating the average distance of the same condition data ($d_{i,j}$) and then getting the average distance of eight conditions (d_{ai}).

The equations can be defined as follows:

$$d_{i,j} = \frac{1}{N \times (N-1)} \sum_{m,n=1}^{N} |p_{i,j}(m) - p_{i,j}(n)|; \quad (m,n = 1,2,\ldots,N,\ m \neq n) \tag{5}$$



(a) Shape factor and histogram



(b) Crest factor and kurtosis

Fig. 5. Feature extraction of the time signal.

where N is the number of the same condition ($N = 20$), $p_{i,j}$ is the eigenvalue, $d_{i,j}$ is the average distance of the same condition, i and j represent the number of parameters and conditions, respectively. Here

$$d_{ai} = \frac{1}{M} \sum_{j=1}^{M} d_{i,j} \tag{6}$$

where M is the number of different conditions ($M = 8$).

Step 2: Calculating the average distance between different condition data (d'_{ai})

$$d'_{ai} = \frac{1}{M \times (M-1)} \sum_{m,n=1}^{M} |p_{ai,m} - p_{ai,n}| \quad (m, n = 1, 2, \ldots, M;\ m \neq n) \tag{7}$$

where d'_{ai} is the average distance of different conditions data, and $p_{ai,j}$ is the average value of the same condition 20 data:

$$p_{ai,j} = \frac{1}{N} \sum_{n=1}^{N} p_{i,j}(n) \quad (n = 1, 2, \ldots, N). \tag{8}$$

Step 3: Calculating the ratio d_{ai}/d'_{ai}.

Step 4: Selecting the eight feature parameters α_i from large value to small value (smaller the d_{ai}, the better; on the contrary, bigger the d'_{ai}, the better. So bigger α_i represents the feature well.):

$$\alpha_i = d'_{ai}/d_{ai} \tag{9}$$

where α_i is the effectiveness factor of features shown in Fig. 6.

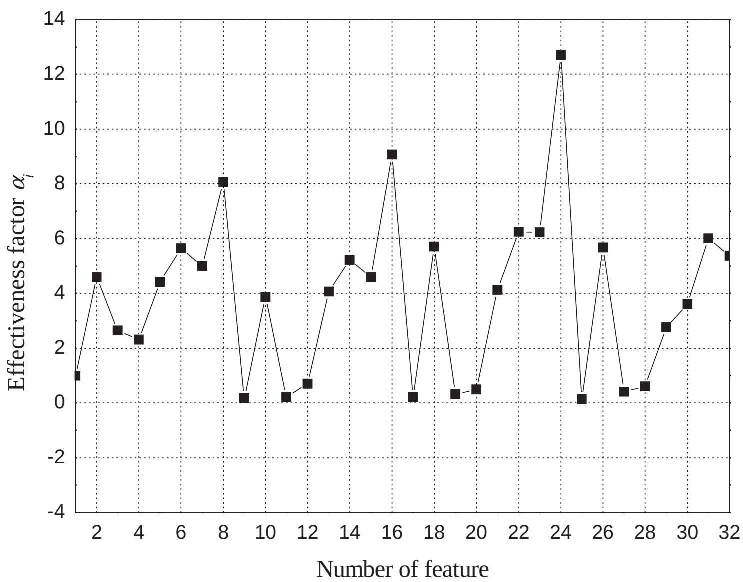


Fig. 6. Effectiveness factor of features.

Fig. 6 shows the computation results of the effectiveness factor α_i of 32 feature parameters. From the results, eight feature parameters are selected such as shape factor, crest factor and standard error of the time signal, mean and entropy of wavelet transform level 1, shape factor and standard error of wavelet transform level 2, entropy of wavelet transform level 3, and are used as the input vectors of the network for fault diagnosis.

3.4. Fault diagnosis

The characteristics of ART–KNN are training and diagnosis together. In this paper, 160 signals acquired from the experiments and eight feature parameters are used for the training and diagnosis. The procedure of the training and diagnosis are shown in Table 1. In the table, A is the number of input data, B is the attribute label of each condition defined in Table 2, C is the input and output attributes in the network, and D is the neuron number of the network.

In the beginning of the ART–KNN, it is empty. So when the first input vector enters the network, the symbol "?" appears to ask the attribute of input signal, then produces a new neuron 1 to point at the new cluster 1. When the next vector enters the network, it compares with the only cluster, 1; cluster 1 wins then returns to compare with this input vector. The similarity S matches the criteria and the vector is considered as one condition. Depending on Eq. (3), the weight is improved. The following may be deduced by analogy. Usually, one condition needs many neurons to study a few times due to the complex conditions. The detailed processing is demonstrated in Table 3.

From Table 3, we notice that condition 4, bearing ball defect, uses many neurons to learn. The reason is that its vibration signal is diverse waveform due to complex fault mechanism. In order to understand the relationships of criterion parameter ρ, number of neurons and classification success rate, Figs. 7 and 8 are used to explain it.

The equation of classification success rate CSR is defined by

$$CSR = C/(T - N) \times 100\% \tag{10}$$

where C is the number of accurate classification, T is the number of total data, and N is the number of generated neurons. $(T - N)$ means the number of used data for test, which is equal to the input data number minus the training data number.

From Fig. 7, the number of neurons follows the criterion parameter ρ in the discernment layer. As $\rho = 1$, the number of neurons reaches 160, it indicates all data are used to train the network, the test data do not exist. It is meaningless in the real world. Synthesising Figs. 7 and 8, one conclusion can be got that $\rho < 0.95$, the number of neurons is around 20; but CSR is too low; $\rho > 0.96$, CSR rises, used neurons increase; $\rho > 0.962$, the classification success rate reaches 100%. The network needs not many neurons for this case, comparing with other networks. In view of the relationship of ρ, N and CSR, the proposed range is 0.95–0.98. In this paper, ρ is 0.962; the corresponding number of neurons is 27, which accords with the results of experiments.

The general trend of CSR as shown in Fig. 8 is increasing with ρ. However, it is not continuous. Each cluster is composed of many neurons with the same property, and the cluster region becomes the summation of total neuron region representing its region. The number of neurons is directly proportional to ρ. Because each neuron region becomes small and the number of neurons increases with increasing ρ, the region of the cluster changes becomes bigger or smaller depending

Table 1
Classification procedure and results of ART–Kohonen network

A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1	1	?:1	1	41	3	?:3	5	81	5	?:5	18	121	7	?:7	24
2	1	1	1	42	3	3	5	82	5	5	18	122	7	7	24
3	1	1	1	43	3	3	5	83	5	5	18	123	7	7	24
4	1	1	1	44	3	3	5	84	5	5	18	124	7	7	24
5	1	1	1	45	3	3	5	85	5	?:5	18	125	7	7	24
6	1	1	1	46	3	3	5	86	5	?:5	19	126	7	7	24
7	1	1	1	47	3	3	5	87	5	5	19	127	7	7	24
8	1	1	1	48	3	3	5	88	5	5	18	128	7	7	24
9	1	1	1	49	3	3	5	89	5	5	18	129	7	7	24
10	1	1	1	50	3	3	5	90	5	5	18	130	7	7	24
11	1	?:1	2	51	3	3	5	91	5	5	18	131	7	?:7	25
12	1	1	1	52	3	3	5	92	5	5	19	132	7	7	24
13	1	1	2	53	3	3	5	93	5	5	18	133	7	7	25
14	1	1	2	54	3	3	5	94	5	5	18	134	7	7	25
15	1	1	2	55	3	3	5	95	5	5	18	135	7	7	25
16	1	1	2	56	3	3	5	96	5	5	18	136	7	7	25
17	1	1	1	57	3	3	5	97	5	5	19	137	7	7	24
18	1	1	1	58	3	3	5	98	5	5	20	138	7	7	24
19	1	1	1	59	3	3	5	99	5	5	19	139	7	7	24
20	1	1	1	60	3	3	5	100	5	5	19	140	7	7	24
21	2	?:2	3	61	4	?:4	6	101	6	?:6	21	141	8	?:8	26
22	2	2	3	62	4	4	6	102	6	?:6	22	142	8	8	26
23	2	2	3	63	4	?:4	7	103	6	6	22	143	8	8	26
24	2	2	3	64	4	?:4	8	104	6	6	21	144	8	8	26
25	2	2	3	65	4	4	7	105	6	6	21	145	8	8	26
26	2	2	3	66	4	?:4	9	106	6	6	21	146	8	8	26
27	2	2	3	67	4	?:4	10	107	6	6	21	147	8	8	26
28	2	2	3	68	4	?:4	11	108	6	6	21	148	8	8	26
29	2	2	3	69	4	4	6	109	6	6	21	149	8	8	26
30	2	2	3	70	4	?:4	12	110	6	6	21	150	8	8	26
31	2	?:2	4	71	4	?:4	13	111	6	?:6	23	151	8	?:8	27
32	2	2	3	72	4	?:4	14	112	6	6	21	152	8	8	26
33	2	2	4	73	4	?:4	15	113	6	6	22	153	8	8	27
34	2	2	4	74	4	4	15	114	6	6	22	154	8	8	27
35	2	2	4	75	4	4	13	115	6	6	22	155	8	8	26
36	2	2	4	76	4	?:4	16	116	6	6	22	156	8	8	26
37	2	2	3	77	4	?:4	17	117	6	6	21	157	8	8	26
38	2	2	3	78	4	4	13	118	6	6	21	158	8	8	26
39	2	2	3	79	4	4	11	119	6	6	21	159	8	8	26
40	2	2	3	80	4	4	11	120	6	6	21	160	8	8	26

on the space distribution of neurons with the same property. Then, if the distance of adjoining clusters is close to each other, the classification success rate will be increased or decreased locally.

The advantage of ART–KNN is validated, comparing with conventional networks, the SOFM, LVQ [10,11] and RBF [12] networks. Same data are used to compare these networks. Half the

Table 2
Attribute label of each condition

Condition	Normal	Bearing defect				Misalignment		Unbalance
		Outer race	Inner race	Ball	Complex	Angular	Parallel	
Label	1	2	3	4	5	6	7	8

Table 3
Number of neurons presenting each condition

Label	1	2	3	4	5	6	7	8	Total
Number of neurons	2	2	1	12	3	3	2	2	27

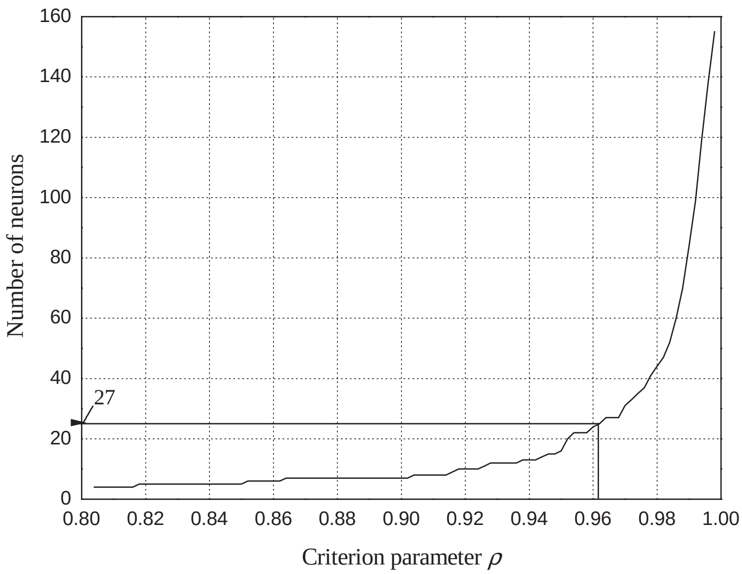


Fig. 7. Relationship of discernment layer neuron number and criterion parameter ρ.

data are used for training of the network, the rest for testing and getting the *CSR*. The results are as shown in Table 4. The maximum *CSR* for the ART–KNN was 100% and for the SOFM, LVQ and RBF networks, the *CSR*s were 93%, 93% and 89%, respectively.

4. Conclusions

In this paper, a new neural network for fault diagnosis of rotating machinery which synthesises the adaptive resonance theory (ART) and the learning strategy of Kohonen neural networks

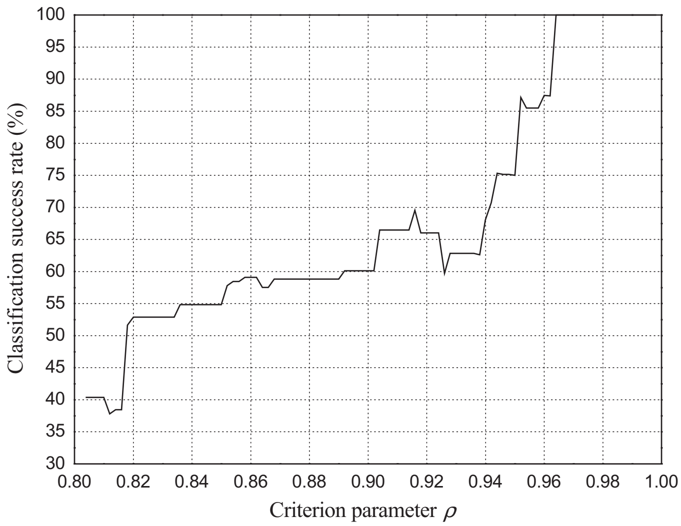


Fig. 8. Relationship of classification success rate and criterion parameter ρ.

Table 4
Comparison of classification success rate

Neural network	SOFM	LVQ	RBF	ART–Kohonen
Classification success rate (%)	93	93	89	100
Number of neurons	75	100	42	27

(KNN) is proposed. Under the same conditions, four different neural networks were evaluated for their success rate on fault diagnosis of rotating machinery. The diagnosis success rate of ART–KNN can reach 100%, while these success rates for SOFM, LVQ and RBF networks were 93%, 93% and 89%, respectively. It also can perform on-line learning without forgetting previous patterns. These make this approach very promising for the application in the real industry.

References

[1] E. Liobet, et al., Fuzzy ARTMAP based electronic nose data analysis, Sensors and Actuators B 61 (1999) 183–190.

[2] G.A. Carpenter, S. Grossberg, The ART of adaptive pattern recognition by a self-organizing neural network, IEEE Transactions on Computers 21 (3) (1988) 77–88.

[3] G.A. Carpenter, S. Grossberg, ART2: self-organization of stable category recognition codes for analog input patterns, Applied Optics 26 (23) (1987) 217–231.

[4] G.A. Carpenter, S. Grossberg, ART3: hierarchical search using chemical transmitters in self-organizing pattern recognition architectures, Neural Networks 4 (3) (1991) 129–152.

[5] G.A. Carpenter, S. Grossberg, A self-organizing neural network, IEEE Communications Magazine 30 (9) (1992) 38–49.

[6] J.W. Gardner, E.L. Hines, C. Pang, Detection of vapours and odours from a multisensor array using pattern recognition: self-organising adaptive resonance techniques, Measurement and Control 29 (1996) 172–177.

[7] T. Kohonen, E. Oja, O. Simula, A. Visa, J. Kangas, Engineering application of the self-organizing map, Proceedings of the IEEE 84 (10) (1996) 1358–1384.

[8] T. Kohonen, Self-Organizing Maps, Springer, New York, 1995.

[9] G.A. Carpenter, S. Grossberg, A massive parallel architecture for a self-organizing neural pattern-recognition machine, Computer Vision Graphics and Image Processing 37 (1987) 116–165.

[10] B.S. Yang, D.S. Lim, S.Y. Seo, M.H. Kim, Defect diagnostics of rotating machinery using SOFM and LVQ, Proceedings of Seventh International Congress on Sound and Vibration, Garmisch-Partenkirchen, Germany, July 4–7, 2000, pp. 567–574.

[11] B.S. Yang, D.S. Lim, J.L. An, Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transform, Proceedings of the 13th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management, Houston, USA, December 3–8, 2000, pp. 923–932.

[12] B.S. Yang, D.S. Lim, K.K. Kim, Vibration diagnosis system of rotating machinery using radial basis function neural network, Proceedings of the 4th International Conference on Quality Reliability Maintenance, University of Oxford, England, March 21–22, 2002.

[13] T. Han, Feature extraction of vibration signal for machinery condition monitoring, Master Thesis, Pukyong National University, 2002, pp. 1–61.