J. Meas. Sci. Instrum., 2022, 13(3): 371-378 http://xuebao.nuc.edu.cn
DOI: 10.3969/j.issn.1674-8042.2022.03.012 jmsi@nuc.edu.cn

Turnout fault diagnosis based on DBSCAN/PSO-SOM

YANG Juhua[1], LI Xutong[1,2], XING Dongfeng[2,3], CHEN Guangwu[2,3]

(1. *School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China*;
2. *Gansu Provincial Key Laboratory of Traffic Information Engineering and Control, Lanzhou 730070, China*;
3. *Automatic Control Research Institute, Lanzhou Jiao tong University, Lanzhou 730070, China*)

Abstract: In order to diagnose the common faults of railway switch control circuit, a fault diagnosis method based on density-based spatial clustering of applications with noise (DBSCAN) and self-organizing feature map (SOM) is proposed. Firstly, the three-phase current curve of the switch machine recorded by the micro-computer monitoring system is dealt with segmentally and then the feature parameters of the three-phase current are calculated according to the action principle of the switch machine. Due to the high dimension of initial features, the DBSCAN algorithm is used to separate the sensitive features of fault diagnosis and construct the diagnostic sensitive feature set. Then, the particle swarm optimization (PSO) algorithm is used to adjust the weight of SOM network to modify the rules to avoid "dead neurons". Finally, the PSO-SOM network fault classifier is designed to complete the classification and diagnosis of the samples to be tested. The experimental results show that this method can judge the fault mode of switch control circuit with less training samples, and the accuracy of fault diagnosis is higher than that of traditional SOM network.

Key words: turnout; fault diagnosis; density-based spatial clustering of applications with noise (DBSCAN); particle swarm optimization (PSO); self-organizing feature map (SOM)

0 Introduction

Turnout is one of the important basic equipment in the railway, and the operation safety of the train is directly affected by the fault of turnout. At present, the point of failure of turnout is analyzed by current curve and power curve recorded in signal centralized monitoring system. With the rapid construction of the high-speed railway, solely relying on the experience of the field maintenance personnel to judge the fault state of the switch equipment has not met the current development needs.

With the development of intelligent diagnostic technology, many new techniques are applied to fault diagnosis. Wu et al.[1] established the bearing model by fault tree to identify the mechanical failure. Taheri et al.[2] identified six fault modes of radiator by means of image processing and neural network. In the terms of the fault diagnosis of railway turnouts, many scholars have carried out more researches. Wang et al.[3] combined the gray correlation degree with the neural network, the common failure modes of switch machine have high diagnosis accuracy. In this method, a large amount of sample data are needed to train the neural network, and the diagnostic ability is poor in the case of small samples. Zhao et al.[4] used the fisher criterion to filter the useful features and the common fault modes of switch machine can be correctly diagnosed by the gray correlation degree. Shi et al.[5] extracted the residual of the signal and we established the fault detection and diagnosis (FDD) model, and various algorithms are introduced to process prediction and diagnose the switch failure modes in parallel. A great diagnostic result is gained by this method with less expert knowledge. Lee et al.[6] used the channel to rotate the audio information to judge the early equipment failure. In this method, the external modules is needed to collect the audio information, which is difficult to use on a large scale. Eker et al.[7] established a failure mode diagnosis classifier for the common switch control circuit based on support vector machine (SVM), and SVM can only implement two classifications. In the case of multi-

Received date: 2020-11-19

Foundation items: High Education Research Project Funding (No. 2018C-11); Natural Science Fund of Gansu Province (Nos. 18JR3RA107, 1610RJYA034); Key Research and Development Program of Gansu Province (No. 17YF1WA158)

Corresponding author: CHEN Guangwu (cgwyjh1976@126.com)

mode classifications, several SVMs are needed.

In our study, based on the track action current curve recorded by the signal centralized monitoring system, an intelligent diagnosis technology based on density-based spatial clustering of applications with noise (DBSCAN) and self-organizing feature map (SOM) network is proposed, and seven kinds common failure modes of switch control circuit are diagnosed and classified. This method does not require a large number of samples to train the classifier network, and the diagnosis accuracy is high, which reduces the risk of human factors causing misjudgment of fault modes and provides reference for fault location and maintenance of switch control circuits. The current curve is segmented according to the principle in Section 1. The diagnosis method based on DBSCAN and SOM is detailed in Section 2. Test and analysis are presented in Section 3. The conclusion is presented in Section 4.

1 Analysis of action process of switch machine

At present, most of the switch machines used in the speed-increasing section of the railway are AC switch machines. We mainly analyze the action current curve data of ZYJ7 AC switch machine with single machine traction. The three-phase current curve of the ZYJ7 AC switch machine in normal mode is shown in Fig. 1.

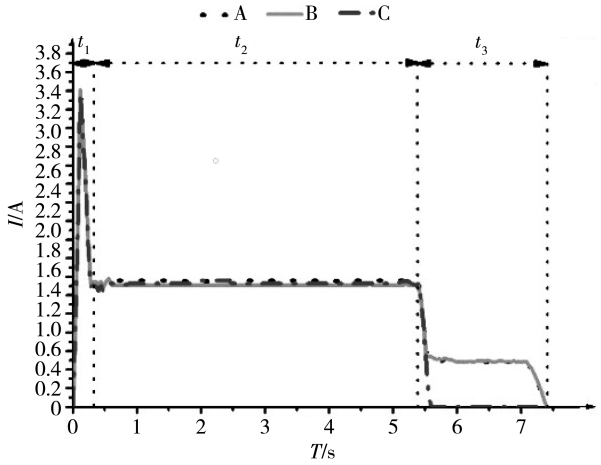


Fig. 1 Three-phase current curve of ZYJ7 AC swith machine

As shown in Fig. 1, the current curve during the rotation of switch can be split into three stages.

1) In t_1 stage, 1DQJ is sucked up, the ballast current curve starts to be recorded, and then 2DQJ is sucked up, the action current curve shows a large peak, and the ballast starts to rotate.

2) In t_2 stage, switch machine acts and the rotation current curve is relatively straight, the three-phase current value is balanced, the normal working current of the ZYJ7 switch machine is not more than 1.8 A, and the action time is slightly different depending on the model of equipment, usually between in the range of 6 s—12 s.

3) In t_3 stage, the rotation of the switch is completed, and the start circuit is disconnected. The 1DQJ self-closing circuit enters the slow release state. The two-phase current is not zero, and the current curve forms a "small tail" current.

In the signal centralized monitoring system, the sampling period of the switching process current is 400 ms, and the ZYJ7 switch machine has about 190 samples under normal conditions points. The 7 kinds of switch control circuit common faults are shown in Table 1.

Table 1 Switch control circuit failure modes

Mode	Phenomenon	Causes
A1	The recorded three-phase current is only 0.5 A	The first start rely sucks up, but the second start rely does not return
A2	"Small tail" current lasts too long after the switch is completed	Phase failure protector failure causes the self-closing circuit of the first start rely to be disconnected
A3	The current of one phase is reduced to 0, and the other two phases are increased	Insufficient oil pressure on the rail of the switch leads to prolonged operation of the circuit
A4	Turnout action time reaches about 13 s	Poor circuit cable causes three-phase current loop abnormality
A5	Turnout action time lasts only about 1 s	Protection relay failure
A6	Sudden rise in three-phase current	Mechanical jamming of switch machine
A7	The value of one phase current is dithered	Synchronous circuit failure of three-phase current

2 Design of active disturbance rejection depth controller

The fault diagnosis process of the switch control circuit mainly includes two parts: feature extraction and fault classification. Feature extraction mainly extracts feature data based on action current curve. The sensitive features are separated by DBSCAN, and the sensitive fault feature set is constructed. Fault classification is mainly based on SOM neural network, and its weight rules are adjusted by particle swarm optimization (PSO). After the training is completed, the SOM network divides the test samples according to learning rules and then completes fault

diagnosis. The structure of the diagnostic system is shown in Fig. 2.

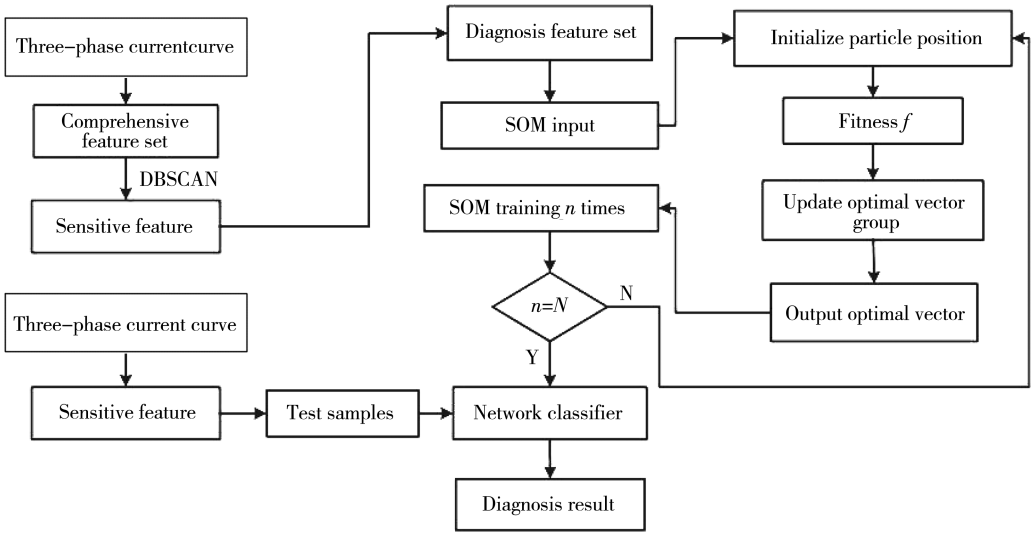


Fig. 2 Structure of diagnostic system

2.1 Sensitive feature extraction based on DBSCAN

DBSCAN is a kind of widely used methods on data clustering. Its disadvantage is that parameters ε and *Minpts* are difficult to determine. Zhao et al.[8] identified the distribution of ships by DBSCAN. Sun et al.[9] analyzed the regular pattern of population movement to solve the traffic problem caused by population growth. Kim et al.[10] proposed a data density based on quadtree structure to find the similar clusters and achieve adaptive parameters selection. In this study, ε and *Minpts* are determined by DBSCAN, and main steps are as follows.

1) Initial sample set is D, and neighborhood parameters(ε, *Minpts*) and the number of clusters $K=0$ are also initialized.

2) Find the ε-neighbor subsample set N_ε of the sample d_j by the distance, and the sample d_j is added to the core sample set Ω if the number of subsample set samples is greater than *Minpts*.

3) If the current cluster core sample set is empty and the unvisited sample set is empty, the algorithm ends and the clustering result $C=\{C_1, C_2, \cdots, C_k\}$ is obtained. Otherwise, new cluster is generated, $k=k+1$ and turns to step 2.

Based on the segmentation curve of the switch machine, the feature parameters in Table 2 are calculated for each phase current in each stage.

The feature set of single-phase is recorded as $\{T_i \mid [T_{i,1}^1, T_{i,1}^2, \cdots, T_{i,3}^3, T_{i,3}^4]\}$. The comprehensive feature set of three-phase current is recorded as $\{P_{j,n}^l \mid [P_{1,1}^1, P_{1,1}^2, P_{1,2}^1, \cdots, P_{3,4}^2]\}$ and calculated by

$$P_{j,n}^1 = \frac{1}{3}\sum_{i=1}^{3} T_{i,j}^n, \tag{1}$$

$$P_{j,n}^2 = \sum_{i=1}^{3} (T_{i,j}^n - P_{j,1}^n), \tag{2}$$

where $I_{i,j}(k)$ represents the kth sample data of stage j of phase i current, and $N_{i,j}$ represents the number of sampling points of the curve in this phase.

Table 2 Formulae of feature parameters

Feature	Formula
Difference of max and min	$T_{i,j}^1 = \max(I_{i,j}(k)) - \min(I_{i,j}(k))$
Duration	$T_{i,j}^2 = t_j$
Average value	$T_{i,j}^3 = \frac{1}{N_{i,j}}\sum_{k=1}^{N_{i,j}} I_{i,j}(k)$
Stand deviation	$T_{i,j}^4 = \frac{1}{N_{i,j}}\sum_{k=1}^{N_{i,j}} \sqrt{(I_{i,j}(k) - T_{i,j}^3)^2}$

In summary, the three-phase current integrated feature set is recorded as $\{P_{j,m} \mid [P_{1,1}, P_{1,2}, P_{1,3}, \cdots, P_{3,8}]\}$, where $P_{j,m}$ represents the mth comprehensive feature of the jth order in the feature set.

The current curve comprehensive feature data are extracted, and 7 sample data in 8 modes (7 fault modes and a normal mode) are extracted, and the data are normalized by

$$q_{j,m}^{(a,b)} = \frac{p_{j,m}^{(a,b)} - \min\limits_j(p_{j,m}^{(a,b)})}{\max\limits_j(p_{j,m}^{(a,b)}) - \min\limits_j(p_{j,m}^{(a,b)})}. \tag{3}$$

The DBSCAN is used to analyze each comprehensive feature in set Q, the feature data need to be transformed into two-dimensional data as

$$(d_x^{(a,b)}, d_y^{(a,b)}) = (q_{j,m}^{(a,b)}, q_{j,m}^{(a,b)}). \tag{4}$$

The parameter *Minpts* is set as 7 (the number of samples). As shown in Eq. (5), the parameter ε is set as the maximum Euclidean distance between the data in same mode, where k and l are the two different samples in mode a.

$$\varepsilon = \max_{a=1}^{8}\left(\sqrt{(d_x^{(a,k)} - d_x^{(a,l)})^2 + (d_y^{(a,k)} - d_y^{(a,l)})^2}\right). \tag{5}$$

According to the clustering analysis of each feature in the Q, the sensitive features combination is $[P_{1,3}, P_{2,4}, P_{2,6}, P_{3,3}]$, and clustering results are shown in Figs. 3—6 and Table 3.

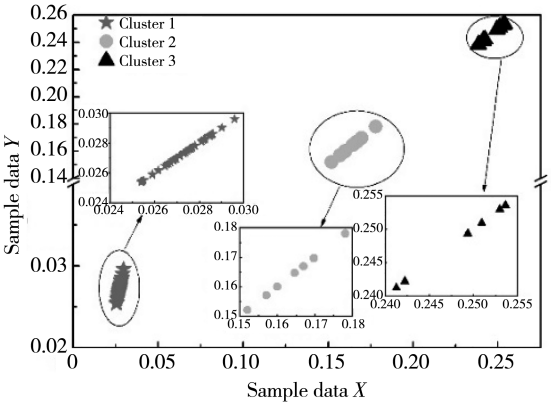


Fig. 3 Clustering result of $P_{1,3}$

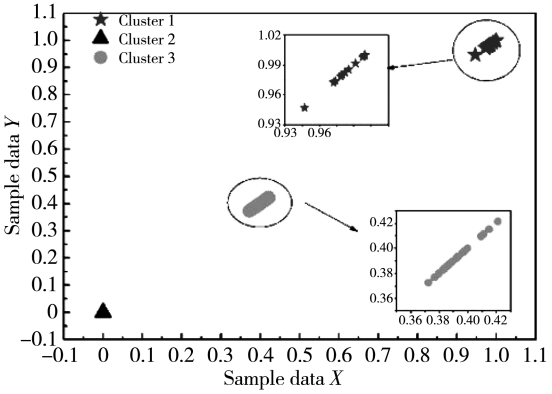


Fig. 4 Clustering result of $P_{2,4}$

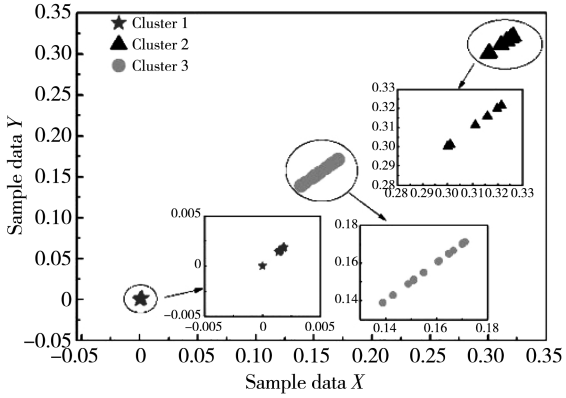


Fig. 5 Clustering result of $P_{2,6}$

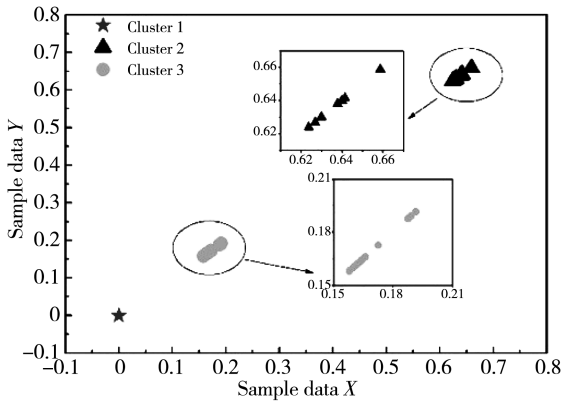


Fig. 6 Clustering result of $P_{3,3}$

Table 3 Standard feature set

Mode	$P_{1,3}$	$P_{2,4}$	$P_{2,6}$	$P_{3,3}$
A0	0.027 7	0.392 3	0.001 4	0.161 5
A1	0.240 1	0	0	0
A2	0.025 4	0.385 4	0.152 0	0.630 8
A3	0.028 5	0.396 2	0.166 0	0
A4	0.026 4	0.393 8	0.150 0	0.169 2
A5	0.146 9	0	0	0
A6	0.026 9	1	0.001 5	0
A7	0.026 5	1	0.310 0	0

2.2 PSO-SOM network fault classifier

SOM is a self-organizing and self-learning network. Liu et al.[11] utilized SOM network to classify the common faults of vacuum circuit breaker. Song et al.[12] sovled the travel salesman problem (TSP) by an advanced SOM network. However, the problem of "dead neurons" in SOM is neglected in above study. In order to avoid this phenomenon, the PSO algorithm is introduced to optimize the SOM network training process and prevent the occurrence of "dead neurons". The topology of traditional SOM network is shown in Fig. 7.

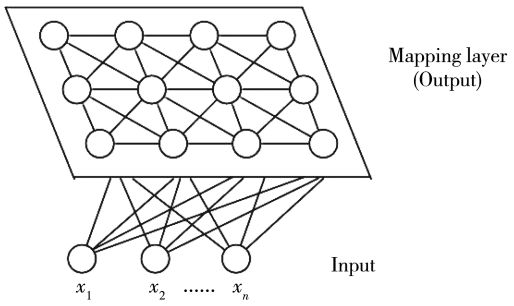


Fig. 7 Topology of SOM

In the traditional SOM network learning algorithm, the connection weight between the input layer and the mapping layer is adjusted by

$$\Delta w_{ij} = \eta(t)(x_i(t) - w_{ij}(t)), \tag{6}$$

where $\eta(t)$ is a constant between 0 and 1, which

shortens time until $\eta(t)=0$.

It can be seen from the above formula that the correction of the connection weight changes with time. If the initial connection time of each neuron connection weight is too long, the neuron will not be able to win over time, and this kind of neurons will be called "dead neurons".

In this study, the PSO is introduced to prevent the occurrence of the "dead neurons". The optimization process is as follows.

1) Initialize PSO particles with weight vectors $\boldsymbol{W}_j$ of SOM;

2) The distance between input vectors and weight vectors is defined as the fitness and expressed as

$$f=\sum_{i=1}^{n}\|\boldsymbol{W}_j-X_i\|. \tag{7}$$

3) The optimized output of PSO is assigned to each connection weight of the network, and the SOM network is iterated n times.

4) Repeat steps 1 to 3 until the target number of iterations is reached.

In this study, the SOM network iteration number n is set to 20, which means the network training iteration is 20 times, and the weight vector $\boldsymbol{W}_j$ is optimized by the PSO once. By cross training of two kinds of algorithms, the occurrence of "dead neurons" can be effectively avoided.

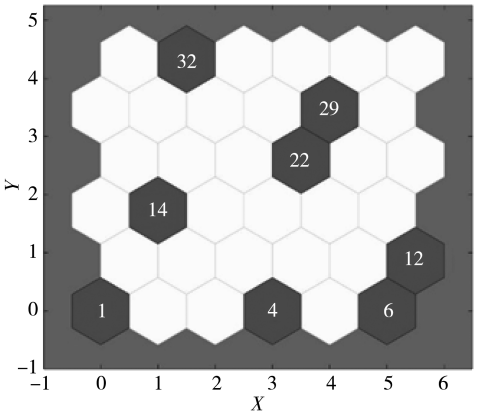
3 Experiment

Based on the above 56 of sample data, the standard fault feature data in Table 4 are used to be the SOM network input vectors. x_1, x_2, x_3, x_4, the structure of network mapping layer is set as 6×6. When the number of training steps N is 50, 100, 200 and 500, respectively, the output and the training time of SOM network are shown in Table 4

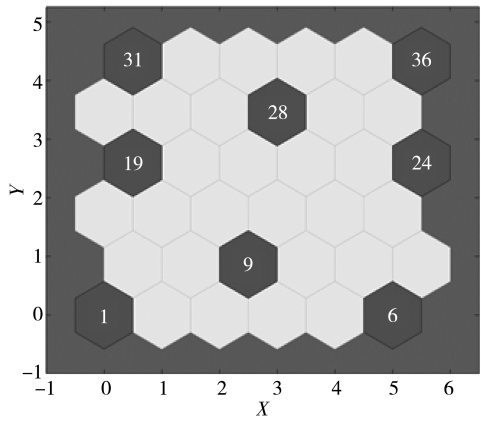
Table 4 Classification results of network in each N

Mode	SOM				PSO-SOM			
	50	100	200	500	50	100	200	500
A0	29	13	14	9	16	9	20	14
A1	2	32	32	4	36	19	32	32
A2	1	16	22	25	31	6	25	8
A3	31	25	29	1	16	28	8	17
A4	29	13	12	14	16	1	14	20
A5	2	31	4	12	30	31	30	29
A6	30	4	1	24	2	24	1	3
A7	30	1	6	32	1	36	3	1
time/s	0.81	1.21	3.55	6.89	1.02	1.64	4.1	8.72

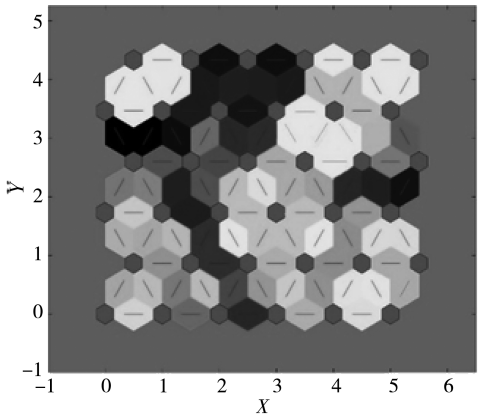
It can be seen from Table 5 that when $N=200$, the wining neurons can be completely separated in SOM, and in PSO-SOM net network, the wining neurons can be completely separated when $N=100$. In terms of network training time, the training time of PSO-SOM is slightly longer than that of SOM due to the optimization of weights by adding PSO algorithm. Overall, the classification ability of PSO-SOM is better than that of SOM when N_{SOM} and $N_{\text{PSO-SOM}}$ equal. When the classification effects are similar ($N_{\text{SOM}}=200$, $N_{\text{PSO-SOM}}=100$), the training time of PSO-SOM is shorter. Otherwise, the occurrence of "dead neurons" can be avoided by the optimization of PSO. The wining neurons of the two kinds of networks are shown in Fig. 8.



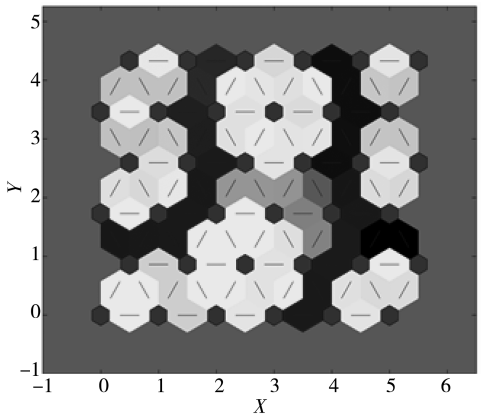
(a) SOM wining neuron in 200



(b) PSO-SOM wining neurons in 100



(c) SOM neuron distance



(d) PSO-SOM neuron distance

Fig. 8 Training results of SOM and PSO-SOM

It can be seen from the Figs. 8(a) and 8(b) that there is no direct connection between two wining neurons in the PSO-SOM network but there is direct connection in SOM network. There is a possibility of misjudgment when the unknown samples are classified. Compared with the PSO-SOM network, the classification effect is better. Figs. 8(c) and 8(d) show the neurons distance of the SOM network and PSO-SOM network, respectively, and the darker color area indicates that the Euclidean distance between the two neurons is large, and the shallower distance indicates that the distance is smaller, which can be used as an auxiliary judgment in the sample classification. The above figures are used to determine the relationship between each mode classification and mapping neurons as shown in Table 5.

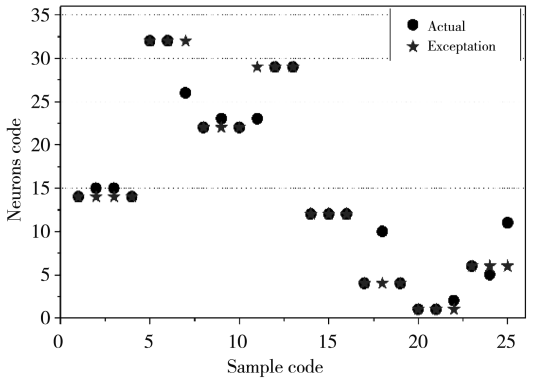
Table 5 Mapping neuron classification

Mode	Mapping layer neuron code	
	SOM	PSO-SOM
A0	14,15	9,3,4,8,10,15,16
A1	32,31,25,26	19,13,14,20
A2	22	6,5,11,12
A3	29	28,21,22,27,29,,33,34
A4	12,18	1,2,7
A5	4,10	31,32
A6	1,2	24,23,18
A7	6,5	36,35
AX	3,7,8,9,11,13,16,17,19, 20,21,23,24,27,28,30, 33,34,35,36	17,25,26,30

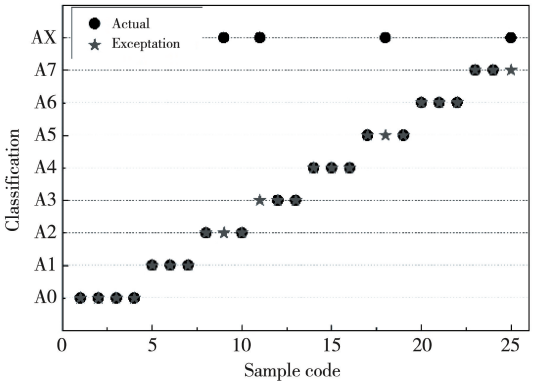
Note: AX means the sample is unrecognized.

As seen from Table 5 that PSO-SOM has two or more neurons in each mode, and SOM is much less, and the neurons belong to the unrecognized mode is far more than PSO-SOM. In the terms of mapping ability between input and output, PSO-SOM network is better than SOM network.

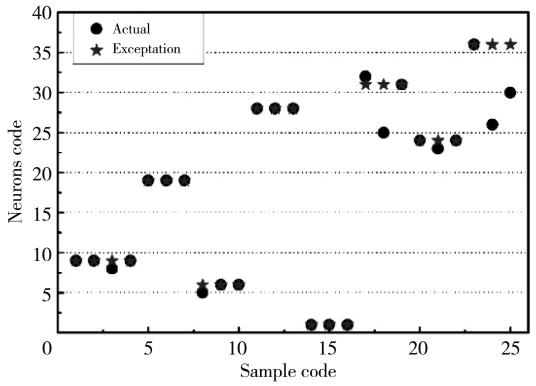
To verify the fault diagnosis ability of proposed method, 25 sets of test samples were brought into SOM network and PSO-SOM network, and diagnosis results are shown in Fig. 9.



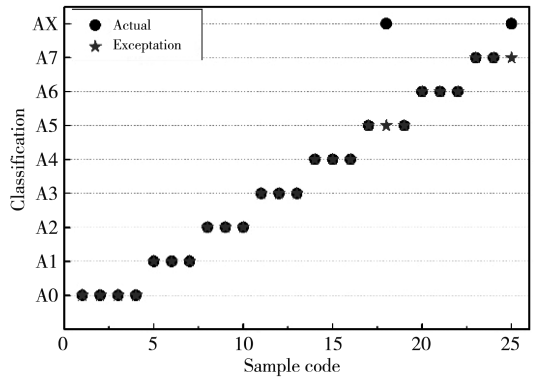
(a) Wining neurons of SOM



(b) Diagnosis result of SOM



(c) Wining neurons of PSO-SOM



(d) Diagnosis result of PSO-SOM

Fig. 9 diagnosis results of SOM and PSO-SOM

It can be seen from Figs. 9(c) and 9(d) that the sample wining neurons coded 3, 8, 17, 21 and 24 do are different from the expected positions, but they are similar to the desired neuron positional connections and can be classified into the correct type. The distance of the samples coded 18 and 25 between the correct neuron and output neuron does not reach the diagnostic threshold and is divided into unrecognized sample. In the terms of winning neurons and fault modes, the SOM network relies more on the distance between the neurons to help determine the samples classification, and the PSO-SOM network has higher mapping sensitivity.

Therefore, the PSO-SOM has higher mapping sensitivity than SOM. In the terms of diagnosis results, the diagnostic accuracy rate of SOM is 84%, the false rate is 0%, and the unrecognition rate is 14%, the diagnostic accuracy rate of PSO-SOM is 92%, the false rate is 0%, and the unrecognition rate is 8%. The PSO-SOM network has a higher fault diagnosis accuracy than SOM network.

4 Conclusions

1) Based on the current curve of the switch machine, a fault diagnosis method of the switch control circuit based on DBSCAN and PSO-SOM network is proposed. The experiment shows that the method has better recognition effect than SOM on the common turnout control circuit fault phenomenon, the diagnostic accuracy rate of 25 groups test sample is 92%, the false rate is 0%, and the unrecognition rate is 8%. The diagnostic accuracy is high.

2) Based on the initial feature set, the diagnostic sensitive features are separated by DBSCAN, which reduces the complexity of the classification network and improves the training rate caused by high-dimensional features.

3) In order to avoid the phenomenon of "dead neurons" in SOM network, the PSO is introduced to optimize SOM network. Experiment shows that this method can improve the training rate and classification effect of SOM network.

4) According to the analysis process and experimental results, this method has reference significance for fault location and maintenance of important railway signal equipment.

References

[1] WU C X, CHEN T F, JIANG R, et al. A novel approach to wavelet selection and tree kernel construction for diagnosis of rolling element bearing fault. Journal of Intelligent Manufacturing, 2017, 38(6): 68-72.

[2] TAHERI G A, AHMADI H. An intelligent approach for cooling radiator fault diagnosis based on infrared thermal image processing technique. Applied Thermal Engineering, 2015, 87(1): 434-443.

[3] WANG R F, CHEN W B. Research on fault diagnosis method for S700K switch machine based on grey neural network. Journal of the China Railway Society, 2016, 38(6): 68-72.

[4] ZHAO L H, LU Q. Method of turnout fault diagnosis based on grey correlation analysis. Journal of the China Railway Society, 2014, 36(2): 69-74.

[5] SHI Z, LIU Z C, LEE J. An auto-associative residual based approach for railway point system fault detection and diagnosis. Journal of the International Measurement Confederation, 2018, 119(1): 246-258

[6] LEE J, CHOI H, PARK D, et al. Fault detection and diagnosis of railway point machines by sound analysis. Sensors, 2016, 16(4): 114-118.

[7] EKER O F, CAMCI F, KUMAR U. SVM based diagnostics on railway turnouts. International Journal of Performability Engineering, 2012, 8(3): 289-298.

[8] ZHAO L B, SHI G Y, YANG J X. An adaptive hierarchical clustering method for ship trajectory with DBSCAN algorithm. Navigation of China, 2018, 43(3): 53-58.

[9] SUN G D, LIU F, LIANG R H. Visual analysis of human movement: a function region perspective. Journal of Computer-aided Design & Computer Graphics, 2018, 30(6): 1073-1081.

[10] KIM J H, CHOI J H, YOO K H, et al. AA-DBSCAN: an approximate adaptive DBSCAN for finding clusters with varying densities. Journal of Supercomputing, 2019, 75(1): 142-169.

[11] LIU Y, CHEN L. Mechanical fault diagnosis of vacuum circuit breaker based on SOM. Transactions of China Electro-Technical Society, 2017, 32(5): 49-54.

[12] SONG J J, BAI Y P, HY H P. Study of TSP based on self-organizing map. Journal of Measurement Science and Instrumentation, 2013, 4(4): 353-360.

基于 DBSCAN/PSO-SOM 的道岔故障诊断

杨菊花[1]，李旭彤[1,2]，邢东峰[2,3]，陈光武[2,3]

（1. 兰州交通大学 交通运输学院，甘肃 兰州 730070；
2. 甘肃省高原交通信息工程及控制重点实验室，甘肃 兰州 730070；
3. 兰州交通大学 自动控制研究所，甘肃 兰州 730070）

摘　要： 为诊断铁路道岔控制电路中的常见故障，提出了一种基于数据密度聚类算法（Density-based spatial clustering of applications with noise，DBSCAN）与自组织特征映射网络（Self-organizing feature map，SOM）结合的诊断方法。利用微机监测系统记录转辙机三相电流曲线，以转辙机动作原理为标准对曲线分段处理并计算三相电流特征参数。针对初始特征维数较高的问题，以 DBSCAN 算法筛选故障诊断敏感特征，构建诊断敏感特征集。以粒子群优化算法（Particle swarm optimization，PSO）调整 SOM 网络权值修改规则从而避免网络出现“死神经元”，设计 PSO-SOM 网络故障分类器并完成待测样本分类诊断。实验表明，该方法在训练样本较少的情况下，能判断道岔控制电路故障模式。与传统 SOM 网络相比，其故障诊断准确率更高。

关键词： 道岔；故障诊断；基于数据密度聚类算法；粒子群算法；自组织特征映射

引用格式： YANG Juhua，LI Xutong，XING Dongfeng，et al. Turnout fault diagnosis based on DBSCAN/PSO-SOM. Journal of Measurement Science and Instrumentation，2022，13(3)：371-378. DOI：10.3969/j.issn.1674-8042.2022.03.012