

Postprint: Railway Signaling Equipment Fault Prediction Based on KNE-BPNN

Authors: Li Chenguang, Qiao Shuai, Yang Xiaojie, Xie Weifan, Li Chuanzi, Li Junhong

Date: 2018-06-19T00:00:00+00:00

Abstract

To address practical issues such as frequent faults in railway signal equipment, low operational efficiency, and the lack of effective fault prediction methods, a fault prediction model for signal equipment based on K-means–Neighborhood Approximate Conditional Entropy and BP Neural Network (KNE-BPNN) is proposed. First, a K-means clustering-based example reduction algorithm is employed to reduce redundant examples in the equipment fault decision table. Second, a neighborhood approximate conditional entropy attribute reduction method is utilized to reduce non-essential attributes in the fault decision table after example reduction. Finally, the BP neural network is trained using the sample set after both example and attribute reduction, and model prediction is performed until the output results satisfy the preset conditions. Experimental results demonstrate that the prediction accuracy and generalization performance of the KNE-BPNN fault prediction model both meet the practical requirements of signal equipment management.

Full Text

Preamble

Fault Prediction for Communication and Signal Equipment Based on KNE-BPNN

Li Chenguang¹, Qiao Shuai¹, Yang Xiaojie¹, Xie Weifan², Li Chuanzi^{1†}, Li Junhong^{1†}

(1. College of Mathematics & Information Science, Hebei Normal University, Shijiazhuang 050024, China;

2. School of Information Science & Engineering, Southeast University, Nanjing 211189, China)

Abstract: To address practical challenges such as frequent failures, low operational efficiency, and the lack of effective fault prediction methods for railway communication and signal (C&S) equipment, this paper proposes a fault prediction model for C&S equipment based on K-means-neighborhood approximate conditional entropy and BP neural network (KNE-BPNN). First, a sample reduction algorithm based on K-means clustering is employed to reduce redundant samples in the equipment failure decision table. Second, the neighborhood approximate conditional entropy attribute reduction method is used to eliminate non-essential attributes in the fault decision table after sample reduction. Finally, the BP neural network is trained using the reduced sample set, and the model is iteratively trained until its output meets the expected requirements. Experimental results demonstrate that the prediction precision and generalization performance of the KNE-BPNN fault prediction model satisfy the actual requirements of C&S equipment management.

Key words: neural network; neighborhood rough set; approximate conditional entropy; reduction attributes; fault prediction; K-means method

0 Introduction

Communication and signal (C&S) equipment, as a primary component of railway transportation systems, plays a crucial role in ensuring railway safety 错误! 未找到引用源. In recent years, China's railway infrastructure has been gradually improved, with annual increases in railway fixed-asset investment and continuous technological innovation. Today, China's railway construction has entered a stage of rapid development, which imposes higher demands on the stability and reliability of C&S equipment. Due to the complex internal structure of C&S equipment, failures occur frequently. Reducing these failures and ensuring safe and stable railway system operation requires research into effective equipment fault prediction methods during operation, enabling timely maintenance and reducing personnel injuries and economic losses caused by equipment failures. This represents an important and practical problem.

Given the high research value of equipment fault prediction, many experts and scholars have conducted extensive research on this issue, yielding many significant results. For instance, studies have utilized support vector machines (SVM), neural networks, grey systems, fuzzy systems, time series, and real-time expert systems for fault prediction 错误! 未找到引用源. Among these, artificial neural networks possess self-learning, self-organizing, and nonlinear capabilities that overcome early artificial intelligence limitations in pattern recognition and unstructured processing, demonstrating excellent intelligent characteristics 错误! 未找到引用源.

In practical applications, sample data for prediction often exhibits high dimensionality, large volume, and complex relationships that are difficult to process. Inputting such data directly into neural network prediction models can lead to slow convergence, low fitting accuracy, and weak generalization ability. Rough

set theory-based attribute reduction can extract potential, high-precision, and well-classified data from complex datasets 错误! 未找到引用源. Although rough set attribute reduction can eliminate redundant data and improve model training speed, invalid and redundant samples remain in the sample set, causing over-learning and over-fitting in prediction models. Sample reduction algorithms based on K-means clustering [5,6] can effectively reduce redundant samples and accelerate model learning speed.

Currently, railway C&S departments primarily rely on manual methods for C&S equipment management, which suffers from cumbersome procedures, low efficiency, and poor multi-user sharing capabilities. Simultaneously, railway C&S equipment has complex structures with aging components, making equipment failures highly probable. How to professionally manage C&S equipment and improve its safety and reliability has become an urgent issue for railway C&S departments.

To address these problems, this paper proposes a method for equipment fault data reduction based on both samples and attributes, and constructs a model for C&S equipment fault prediction. First, K-means clustering is applied for sample reduction in the decision table. Second, the neighborhood approximate conditional entropy method 错误! 未找到引用源 is used for attribute reduction in the reduced decision table. Finally, a KNE-BPNN fault prediction model for C&S equipment is constructed (KNE-BPNN denotes BP neural network with K-means and neighborhood approximate decision entropy). Experimental results show that this model exhibits good performance in equipment fault prediction and can provide effective ideas and methods for subsequent intelligent C&S equipment management.

1 K-Means Sample Reduction

Real-world sample datasets often contain many irrelevant samples that cause over-learning and over-fitting, reducing model generalization ability. Therefore, a sample reduction algorithm based on K-means clustering is employed to reduce the sample set, eliminating irrelevant and redundant samples to improve the generalization ability of trained models.

1.1 K-Means Clustering Algorithm

K-means clustering is a traditional clustering algorithm whose primary objective is to partition n observation samples into K clusters, minimizing the distance between each observation and its cluster mean. The algorithm first randomly selects K observations as initial cluster centers, then assigns each sample to the nearest cluster based on distance to cluster centers. Next, it calculates each cluster's mean and updates it as the new cluster center, iteratively adjusting sample assignments until cluster centers stabilize. K-means typically uses a squared error criterion function that encourages K clusters to converge and become compact [5,6]. The K-means clustering algorithm flowchart is shown in

[Figure 1: see original paper].

The K-means clustering algorithm proceeds as follows:

Input: Observation sample set $X = \{x_1, x_2, \dots, x_n\}$ and number of clusters K .

Output: K clusters with minimized squared error for each cluster.

- a) Set iteration count $i = 1$, randomly select K observation samples as initial cluster centers $C_j, j = 1, 2, \dots, K$.
- b) Calculate the distance between each observation sample and all cluster centers. For sample x_i , compute $DIS(x_i, C_j), i = 1, 2, \dots, n$.
- c) If $DIS(x_i, C_h) = \min\{DIS(x_i, C_j)\}$ for $j = 1, 2, \dots, K$, then assign x_i to cluster h .
- d) Update cluster center values: $C_j = (1/|C_j|) \sum_{x \in C_j} x, j = 1, 2, \dots, K$.
- e) Calculate the error squared criterion function $J = \sum_{i=1}^n \sum_{j=1}^K \{C_j\} \|x_i - C_j\|^2$.
- f) If $|J[i] - J[i+1]| < \epsilon$ (where ϵ is a 极小值, generally 10^{-4}), the algorithm terminates; otherwise, return to step b).

1.2 Sample Reduction Algorithm

Sample reduction is performed based on the K-means clustering algorithm. First, a two-dimensional equipment fault decision table is constructed from the sample set (as shown in Table 1, where x represents samples, c are condition attributes, and D represents decision attributes, with 0 indicating no failure and 1 indicating failure). Second, the number of decision categories is specified as the cluster number K , and data subsets are partitioned according to decision categories (excluding decision attributes). The mean of each data subset is calculated and designated as the cluster center for each cluster. Then, the K-means clustering algorithm adjusts the clusters and outputs K clusters. Finally, the proportion of samples from different decision categories in each cluster is calculated, and samples with smaller proportions are reduced. Using the K-means clustering-based sample reduction algorithm can effectively reduce irrelevant and redundant samples, minimizing their impact on model prediction and improving model generalization ability.

The specific implementation of the K-means clustering-based sample reduction algorithm is as follows:

- a) Construct a two-dimensional decision table and set the number of decision categories in the decision table as the cluster number K .
- b) Partition the data into subsets according to decision categories and calculate the mean of each subset (excluding decision attributes), designating these means as cluster centers C_j .

- c) Adjust clusters using the K-means clustering algorithm to obtain K clusters.
- d) Calculate the proportion of samples from different decision categories in each cluster and reduce samples with smaller proportions. The sample reduction algorithm flowchart is shown in [Figure 2: see original paper].

2 Neighborhood Rough Set Theory

2.1 Rough Sets

Rough set theory, proposed by Polish mathematician Z. Pawlak in 1982, is a mathematical tool for handling uncertain and incomplete knowledge. Rule extraction-based attribute reduction is one of its important research areas. Existing attribute reduction methods mainly include positive region preservation-based decision-theoretic rough set attribute reduction [错误! 未找到引用源](#), discernibility matrix-based attribute reduction [错误! 未找到引用源](#), and information view-based attribute reduction [错误! 未找到引用源](#). However, classical rough set theory only applies to discrete data. For numerical data, discretization is required beforehand, which may lead to loss of important information and affect subsequent reduction effectiveness [错误! 未找到引用源](#).

2.2 Neighborhood Rough Sets

The neighborhood rough set reduction model is built on neighborhood equivalence relations and can effectively handle numerical data without discretization, maximizing the preservation of sample set classification performance [错误! 未找到引用源](#).

A dataset for classification learning can be defined as a five-element decision system $NDS = (U, C, D, V, f)$, where U is a universe (non-empty finite set), C represents condition attributes (non-empty finite set), D represents decision attributes (non-empty finite set), $V = \{a \in C \cup D\}$ is the attribute value domain, and $f: U \times (C \cup D) \rightarrow V$ is an information function that assigns an information value to each object for each attribute, i.e., $u \in U, a \in C \cup D, f(u, a) \in V$.

Definition 1 [13]: For any object $u \in U$ and attribute subset $B \subseteq C$, the neighborhood of u on B is $\delta_B(u) = \{u' \in U, \Delta(u, u') \leq \delta\}$, where $\Delta(u, u')$ represents the distance between two objects, typically used to measure similarity. For $u = (f(u, a_1), f(u, a_2), \dots, f(u, a_N))$ and $u' = (f(u', a_1), f(u', a_2), \dots, f(u', a_N))$, the Minkowsky distance in N -dimensional space is $\Delta_M(u, u') = (\sum_{i=1}^N |f(u, a_i) - f(u', a_i)|)^{1/M}$. Common Minkowsky distance functions include three types: when $M = 1$, it's Manhattan distance; when $M = 2$, Euclidean distance; when $M = \infty$, Chebychev distance. This paper primarily uses Euclidean distance.

Definition 2 [14,15]: In neighborhood decision system $NDS = (U, C, D, V, f)$, for any $B \subseteq C$, the neighborhood relation on U is: $\delta_B = \{(x, y) \in U \times U \mid$

$\Delta_B(x, y) \leq \delta$. The indiscernibility relation determined by decision attribute D is $IND(D) = \{(x, y) \in U \times U \mid a \in D, f(x, a) = f(y, a)\}$. The partition of U by decision attribute D is $U/IND(D) = \{Y_1, Y_2, \dots, Y_m\}$. For $B \subseteq C$, the neighborhood of u on B is $\delta_B(u)$. The neighborhood lower approximation of D about B is $BND(D) = \{i=1\}^m BND(Y_i)$, where $BND(Y_i) = \{u \mid \delta_B(u) \cap Y_i \neq \emptyset, u \in U\}$. The neighborhood upper approximation of D about B is $\overline{BND}(D) = \{i=1\}^m \overline{BND}(Y_i)$, where $\overline{BND}(Y_i) = \{u \mid \delta_B(u) \cap Y_i \neq \emptyset, u \in U\}$.

Definition 3 [7]: In neighborhood decision system $NDS = (U, C, D, V, f)$, for $X \subseteq U$ and $B \subseteq C$, the neighborhood approximation accuracy of X under relation B is $\gamma_B(X) = |BNB(X)| / |X|$, where $|X|$ represents the cardinality of set X. When $X \neq \emptyset$, $0 \leq \gamma_B(X) \leq 1$.

Definition 4 [7]: For neighborhood decision system $NDS = (U, C, D, V, f)$, from Definition 1 we know $\delta_B(u)$ is the neighborhood of u on attribute set B, and $U/IND(D) = \{Y_1, Y_2, \dots, Y_m\}$ is the partition of U by decision attribute D. From Definition 3, $\gamma_B(Y_i)$ is the neighborhood approximation accuracy of Y_i under neighborhood relation δ_B . Then the neighborhood approximate conditional entropy of decision attribute D relative to attribute set B is:

$$NACH(D|B) = -\sum_{i=1}^m (|Y_i| / |U|) \times \sum_{u \in Y_i} \{ |U| (|\delta_B(u) \cap Y_i| / |\delta_B(u)|) \times \log(|\delta_B(u) \cap Y_i| / |\delta_B(u)|) \}$$

Definition 5 [7]: For neighborhood decision system $NDS = (U, C, D, V, f)$, from Definition 1 we know $\delta_B(u)$ is the neighborhood of u on attribute set B, and $U/IND(D) = \{Y_1, Y_2, \dots, Y_m\}$ is the partition of U by decision attribute D. From Definition 3, $\gamma_B(Y_i)$ is the neighborhood approximation accuracy of Y_i under neighborhood relation δ_B . Then the internal attribute importance of attribute a $\in B$ relative to decision attribute D is:

$$IMP(a, C, D) = NACH(D|B) - NACH(D|B - \{a\})$$

Definition 6: In neighborhood decision table system $NDS = (U, C, D, V, f)$, if $NACH(D|B) = NACH(D|C)$, then B is called a reduct of C relative to D.

Definition 7 错误! 未找到引用源: Given neighborhood decision system $NDS = (U, C, D, V, f)$, for a $\in C$, if $NACH(D|C) \neq NACH(D|C - \{a\})$, then a is called a core attribute of C relative to D.

Definition 8 错误! 未找到引用源: For a neighborhood decision system $NDS = (U, C, D, V, f)$, the external attribute importance of attribute set $M \subseteq C$ is:

$$IMP(M, C, D) = NACH(D|M) - NACH(D|C)$$

Property 1 错误! 未找到引用源: Given a neighborhood decision system $NDS = (U, C, D, V, f)$, using Euclidean distance to calculate $\delta_C(u)$ on condition attribute set C, we know $U/IND(D) = \{Y_1, Y_2, \dots, Y_m\}$. Then the neighborhood approximate conditional entropy satisfies: $0 \leq NACH(D|C) \leq \log_2(|U|)$.

Property 2 错误! 未找到引用源: In neighborhood decision table system $NDS = (U, C, D, V, f)$, for $M \subseteq C$, we have $NACH(D|M) \geq NACH(D|C)$.

Property 3 错误! 未找到引用源: In neighborhood decision table system $NDS = (U, C, D, V, f)$, when $NACH(D|B) = NACH(D|C)$, attribute $a \in C - B$ is considered unnecessary relative to D .

2.3 Neighborhood Approximate Conditional Entropy-Based Rough Set Theory

Obtaining the minimal reduct of all condition attribute sets through neighborhood rough set theory is an NP problem. An effective approach to solving such problems is using heuristic methods. The heuristic addition strategy first obtains core attributes of condition attributes, then iteratively compares the attribute importance of each core attribute, adding the element with maximum attribute importance to the reduct set until $NACH(D|B) = NACH(D|C)$.

2.4 Neighborhood Approximate Conditional Entropy-Based Attribute Reduction Algorithm

The basic steps of the neighborhood approximate conditional entropy-based rough set attribute reduction algorithm are:

Input: Neighborhood decision system $NDS = (U, C, D, V, f)$

Output: Minimal reduct set B of neighborhood decision system.

- a) Initialize reduct set $B = \emptyset$, let $B = \emptyset$.
- b) For $u \in U$, calculate $\delta_C(u)$.
- c) Calculate the neighborhood approximate conditional entropy $NACH(D|C)$ of decision attribute D about C .
- d) For $a \in C$, calculate $IMP(a, C, D)$. If $IMP(a, C, D) > 0$, add a to reduct set B , i.e., $B = B \cup \{a\}$.
- e) If $B \neq \emptyset$, calculate $NACH(D|B)$. If $NACH(D|B) = NACH(D|C)$, proceed to step g); otherwise, go to step f). If $B = \emptyset$, directly execute step f).
- f) For $a \in C - B$, calculate the external attribute importance $IMP(a, B, D)$. If $IMP(a, B, D) = \max\{IMP(a, B, D) \mid a \in C - B\}$ (if multiple attributes satisfy this condition, randomly select one), then let $B = B \cup \{a\}$ and jump to step e).
- g) Output attribute reduct set B and terminate the algorithm.

The flowchart of the neighborhood approximate conditional entropy-based rough set attribute reduction algorithm is shown in [Figure 3: see original paper].

3 KNE-BPNN-Based Fault Prediction Model

The specific steps of the KNE-BPNN fault prediction model are as follows:

- a) Construct an initial fault decision table using historical equipment failure monitoring data.
- b) Use the K-means clustering method to cluster samples in the decision table into two categories (since sample data decision results only have two types: failure and no failure), calculate the proportion of samples from different decision categories in each cluster, and reduce samples with smaller proportions.
- c) Apply the neighborhood approximate conditional entropy reduction algorithm to perform attribute reduction on the decision table after sample reduction, removing redundant attributes.
- d) After steps b) and c), a new reduced decision table is generated. Input its data into the BP neural network, and through repeated training and parameter tuning, determine the settings for each layer of the neural network and the number of neuron nodes in each layer. Finally, use the tuned neural network for equipment fault prediction.
- e) Output result analysis. Input the sample test set into the trained neural network model to obtain prediction results for sample fault conditions. Analyze the output results. If the prediction target is not achieved, continue adjusting hyperparameters in the model or increasing training iterations until the output meets preset conditions.

The KNE-BPNN-based fault prediction model uses K-means sample reduction and neighborhood approximate conditional entropy attribute reduction algorithms to reduce the fault decision table, effectively eliminating redundant data while ensuring the sample data retains better classification performance. Inputting the reduced sample set into the neural network can improve training efficiency and prevent over-learning and over-fitting. [Figure 4: see original paper] shows the workflow of the KNE-BPNN fault prediction model, with specific steps as follows:

- a) Fault data decision table construction. Collected C&S equipment failure monitoring data is processed by adding missing values and replacing abnormal values to initialize a two-dimensional fault decision table.
- b) Sample reduction. The K-means clustering method clusters samples in the initial fault decision table into N clusters according to the number of decision categories N in the decision table. Then, the proportion of sample sets from different decision categories in each cluster is calculated, and samples with smaller proportions are removed to complete sample reduction of the fault decision table.
- c) Attribute reduction. First, calculate the neighborhood approximate conditional entropy of all condition attribute sets in the fault decision table, initialize attribute reduct set B as empty, and add attributes with internal importance greater than 0 to the reduct set. Then iteratively calculate the neighborhood approximate conditional entropy of attribute reduct set

B to determine if it equals the approximate conditional entropy of all attributes. If equal, output; otherwise, calculate the external importance of all condition attributes except the attribute reduct set, add the attribute with maximum external importance to the attribute reduct set, and iterate again until the approximate conditional entropy of the attribute reduct set equals that of the condition attribute set.

- d) Neural network prediction model construction ([Figure 5: see original paper] shows the neural network structure). First, determine the number of input layer nodes and output layer nodes based on sample dimension and number of categories. Second, determine the number of hidden layers based on sample data volume and dimension, then set a node number range for hidden layer neurons (e.g., set hidden layer neuron count $N \in (5, 50]$ where N is a positive integer), uniformly select numbers from this range as hidden layer neuron nodes. Determine the final hidden layer neuron count through repeated experiments.
- e) Output result analysis. Input the sample test set into the trained neural network model to obtain final fault condition predictions. Analyze the output results. If prediction targets are not met, continue adjusting model hyperparameters or increasing training iterations until results satisfy preset conditions.

4 Experiments

4.1 Initial Decision Table

This paper uses monitoring data from railway switch equipment circuits as the sample set for the KNE-BPNN fault prediction system, as shown in . Data from is initialized into a decision table $NDS = (U, C, D, V, f)$, where condition attributes $C = \{\text{temperature, power supply voltage, transmission voltage, transmission frequency, reception 1.1 voltage, reception 1.2 voltage, interference voltage, fault condition}\}$ and decision attribute $D = \{\text{fault condition}\}$ (equipment failure = 1, normal operation = 0). Not all factors in the initial decision table affect switch equipment failure, but K-means sample reduction and neighborhood approximate conditional entropy attribute reduction can maximally retain samples with significant impact on switch failures. Using reduced samples for equipment fault prediction can improve model prediction accuracy and generalization ability.

4.2 K-Means Sample Reduction

Before applying the K-means sample reduction algorithm, the number of clusters K and each cluster's center C must be determined. From the K-means sample reduction algorithm, the number of clusters K should be set to the number of decision attribute categories in the decision table. To ensure most samples are clustered into their actual categories, the sample set is first standardized (scaled

by removing mean and dividing by standard deviation). Second, the sample set is partitioned into subsets by decision category, then the mean of each subset is calculated and designated as each cluster's center C , i.e.:

$$C_k = (1/|M_k|) \sum_{x \in M_k} x, \text{ where } M_k = \{x \mid x \in U, D(x) = D_k\}, k = 0, 1, \dots, K-1, K \geq 2$$

From analyzing decision attribute D in the fault decision table, we observe two cases: equipment failure and normal operation, represented by 1 and 0 respectively, so $K = 2$. The sample set is then standardized (standardization results shown in , with data 保留两位小数). Data standardization maintains differences between attributes while balancing their magnitudes. After partitioning by decision category and calculating subset means, the cluster centers are:

$$\begin{aligned} C_0 &= (-0.28, -2.38, -0.11, -0.51, -0.55, -0.53, -0.63) \\ C_1 &= (0.27, 2.36, 0.18, 0.52, 0.56, 0.55, 0.65) \end{aligned}$$

After determining cluster centers, samples are clustered until centers stabilize. From [Figure 6: see original paper] and , among 15 samples, samples numbered 5, 7, and 13 were misclassified. To avoid over-learning and over-fitting, these three samples should be removed.

The clustering results are shown in .

4.3 Neighborhood Approximate Conditional Entropy Attribute Reduction

After sample reduction, to prevent slow model training and low learning efficiency, the neighborhood approximate conditional entropy attribute reduction algorithm is applied to condition attributes. From , the equipment fault decision table has 7 condition attributes. During attribute reduction, adjusting threshold δ can control the number of reduced attributes. Generally, the reduced attribute subset should contain more than half of the original attributes but be as small as possible, so the number of attributes $|B| \in [4, 7]$ and $|B|$ is integer. Calculations show that $\delta \in [0.38, 0.52]$ meets requirements. This interval is divided into 5 parts with 0.03 intervals: $[0.38, 0.41)$, $[0.41, 0.44)$, $[0.44, 0.47)$, $[0.47, 0.50)$, $[0.50, 0.52]$. One random number is selected from each interval, yielding 5 thresholds: $\delta_1 = 0.39$, $\delta_2 = 0.43$, $\delta_3 = 0.44$, $\delta_4 = 0.49$, $\delta_5 = 0.51$. The reduced sample sets from each threshold are input into neural networks with identical parameters, each undergoing 10 predictions, and the average is taken to select the attribute reduct set with better prediction results.

In , a_1, a_2, \dots, a_7 represent the 7 condition attributes in the switch fault decision table: temperature, power supply voltage, transmission voltage, transmission frequency, reception 1.1 voltage, reception 1.2 voltage, and interference voltage. The reduction results for δ_1 and δ_3 are consistent, as are those for δ_2 and δ_5 , so reduct sets from δ_1 , δ_2 , and δ_4 should be compared. Each reduced result is input into neural networks with identical parameters for training, with 10 predictions per reduct and averages taken. As shown in [Figure 7: see original paper],

after multiple neural network predictions, threshold $\delta_4 = 0.44$ yields the highest prediction accuracy, so the reduced sample set with condition attributes $\{a_1, a_2, a_3, a_4\}$ (temperature, power supply voltage, transmission voltage, transmission frequency) should be input into the neural network.

4.4 Comparison of Two Attribute Reduction Algorithms

To further verify the effectiveness of the neighborhood approximate conditional entropy attribute reduction algorithm, we compare it with the discernibility matrix attribute reduction algorithm 错误! 未找到引用源 from two perspectives: number of reduced attributes and prediction accuracy. The discernibility matrix attribute reduction is performed based on equal frequency binning discretization. Experimental results are shown in ~.

** Comparison of Reduction Counts**

** Condition Attributes Obtained by Two Algorithms**

** Prediction Accuracy Comparison of Two Reduction Algorithms Under BP Neural Network Model**

The experimental results demonstrate that the neighborhood approximate conditional entropy attribute reduction algorithm can effectively reduce redundant attributes in sample sets while achieving much higher prediction accuracy than the discernibility matrix attribute reduction algorithm. Therefore, the neighborhood approximate conditional entropy attribute reduction algorithm is a more effective reduction method.

4.5 BP Neural Network Prediction Model

With the deepening of deep learning research, scholars have designed specialized deep neural networks for specific problems, including CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory Network), and the recent research hotspot GANs (Generative Adversarial Networks). CNN is suitable for image data, RNN is commonly used for speech recognition and machine translation, LSTM is appropriate for processing and predicting time series events with relatively long intervals and delays, and GANs can perform image-to-text conversion and image denoising. Since current railway C&S equipment monitoring indicators (attributes) are limited, the data obtained through system monitoring is relatively small in scale and low in complexity. Applying these four networks for fault prediction would not improve prediction accuracy and might consume substantial time during the model learning phase. The BP neural network, due to its structural characteristics, is highly suitable for small-scale prediction problems, so this paper selects BP neural network to build the fault prediction model.

Considering that actual C&S equipment sample data has few attributes and the ideal classification function has low complexity, a 3-layer BP neural network model is prioritized for the C&S equipment fault prediction model. After K-means sample reduction and neighborhood approximate conditional entropy

attribute reduction, the sample' s condition attribute set contains 4 attributes, so the neural network input layer has 4 neurons. Since fault prediction is a binary classification problem, the output layer has 1 neuron. Through continuous training and comparison, the neural network achieves expected convergence speed and prediction accuracy when the hidden layer neuron count is set to 30. Additionally, the hidden layer activation function is set to ReLU to accelerate training speed, and the output layer uses Sigmoid function to ensure model outputs fall within [0,1].

After fault prediction model training, 50 test samples are selected to evaluate performance, yielding fault prediction results for the test set (shown in). The data shows 4 misclassified samples among 50 test samples (numbers 2, 10, 41, and 44).

5 Conclusion

This paper addresses practical problems of frequent C&S equipment failures, low operational efficiency, and lack of effective fault prediction methods by proposing a C&S equipment fault prediction model based on K-means clustering, neighborhood approximate conditional entropy, and BP neural network, providing effective ideas and methods for subsequent intelligent C&S equipment management.

Currently, equipment fault prediction primarily uses daily monitoring data. However, due to different railway transportation scheduling, failures of different equipment cause varying losses to railway operations. Future research should additionally establish equipment failure loss decision tables, integrating equipment failure losses with failure occurrence probabilities to form equipment maintenance urgency lists. Maintenance personnel can then inspect equipment according to urgency lists, minimizing personnel and property losses from equipment failures.

Meanwhile, with continuous innovation in railway intelligent technology, future railway intelligent systems can implement more comprehensive monitoring of C&S equipment, obtaining higher-dimensional and more complex data. As system applications proliferate, data volume will also increase. The BP neural network technology proposed in this paper may have unsatisfactory efficiency for processing multi-dimensional, complex, large-scale data, requiring deep neural networks to learn more features from complex data to achieve high-precision C&S equipment fault prediction.

References

- [1] Gong Yuanbin. Discussion about the problem and strategy according to emergency fault handling of railway electrical system [J]. Railway Signaling & Communication, 2012, 48 (01): 46-47.
- [2] Ai hong, Zhou Donghua. Fault prediction approach for dynamic system [J].

Journal of Huazhong University of Science and Technology: Natural Science Edition, 2009, 37 (S1): 222-225.

[3] Zhu Daqi. Principle and application of artificial neural network [M]. Beijing: Science Press, 2006.

[4] Guo Yu, Yang Yu. Equipment fault prediction based on grey rough set and BP neural network [J]. Application Research of Computers, 2017, 34 (09): 2642-2645.

[5] Wu Suhui, Cheng Ying, Zheng Yanning, et al. Survey on K-means Algorithm [J]. New Technology of Library and Information Service, 2011, 27 (05): 28-35.

[6] Hu Wei. Improved hierarchical K-means clustering algorithm [J]. Computer Engineering and Applications, 2013, 49 (02): 157-159.

[7] Zhang ning, Fan Nianbai. Heuristic attribute reduction based on neighborhood approximate conditional entropy [J]. Application Research of Computers, 2018, 5 (35): 1-2.

[8] Huang Guoshun. Positive region preservation reducts in decision-theoretic rough set models [J]. Computer Engineering and Applications, 2016, 52 (2): 165-169.

[9] Ge Hao, Li Longshu, Yang Chuanjian. Discernibility matrix-based reduct representation and quick algorithms [J]. Control and Decision, 2016, 31 (1): 12-20.

[10] Xu Xinying, Liu Haitao, Xie Jun, et al. Attribute reduction based on inconsistent neighborhood matrix under information view [J]. Control and Decision, 2016, 31 (1): 130-136.

[11] Hu Qinghua, Zhao Hui, Yu Daren. Efficient symbolic and numerical attribute reduction with neighborhood rough sets [J]. Pattern recognition and artificial intelligence, 2008, 21 (06): 732-738.

[12] Xie Lingling, Lei Jingshen, Xu Feifei. Vibrant fault diagnosis for hydro-turbine generating unit based on improved neighborhood rough sets and PNN [J]. Journal of Shanghai University of Electric Power, 2016, 32 (2): 181-187.

[13] Xu Zhangyan, Liu Zuopeng, Yang Bingru, et al. A quick attribute reduction algorithm with complexity of $\max(O(|C||U|), O(|C|^2|U|/|C|))$ [J]. Chinese Journal of Computers, 2006, 29 (03): 391-399.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.