J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 7963--7978

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Bayesian network-based multimodal large model optimization of speech text and its fault prediction capability in power industry

Haitao Yu[1], Xuqiang Wang[2,✉], Jian Zheng[2], Tianyi Liu[1], Yongdi Bao[1]

[1] *State Grid TJ Information & Telecommunication Co., Ltd., Tianjin, 300140, China*
[2] *State Grid Tianjin Electric Power Company, Tianjin, 300010, China*

ABSTRACT

Speech-text multimodal large model as a key tool in the operation of the power industry, its fault prediction performance directly affects the operational safety of mechanical equipment, this paper designs a detailed scheme for the optimization of its performance. Firstly, the structural design of the unimodal model is discussed, and the audio classifier based on Wav2Vec2 and the text classifier based on BERT are used to pre-train the model. Based on the above foundation, a multimodal model is introduced, with the cross-attention mechanism as the fusion strategy, so that the different modal information in the deep neural network is fused with each other, thus improving the accuracy and robustness of the recognition task. After completing the fault feature extraction task, on the premise of introducing the relevant theory of BNN, the structure of BBN is optimized, and after fusing the HC algorithm, BIC and annealing idea, the fault diagnosis method based on the improved BBN network is constructed by combining the fault feature extraction method in the electric power industry and the optimized BBN method. The effectiveness of the method is verified through simulation experiments. The prediction accuracy of this paper's method for nine categories of fault data is above 90% at a high level, and the prediction accuracy of faults in some categories can reach 100%. The multimodal model fusion strategy proposed in this paper significantly improves the performance of fault feature recognition, in addition, the fault diagnosis method based on the improved BBN reduces the computational volume of the model and improves the fault prediction ability of the model.

Keywords: Multimodal fusion, Cross-attention mechanism, BBN structure optimization, Fault prediction

✉ Corresponding author.
E-mail address: mypaper2021@126.com (X. Wang).
Received 10 March 2024; Revised 15 May 2024; Accepted 05 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-435

1. Introduction

Under the continuous improvement of social economy and people's living standards, the demand for household electricity and industrial electricity is increasing, and the direction of the development of the power system is also gradually approaching large-scale and intelligent. In the context of the increasing scale and complexity of the structure of the power system, the risk of failure in it is also correspondingly elevated, attracting people's attention and attention [1-2]. Nowadays, digital technology has been widely used in the power industry. At present, the research on power system fault diagnosis is mainly based on some information to identify, after the occurrence of faults, the use of power system in the structure, important devices and other information to identify the type of fault, the location of the fault in the power system to judge, and the possible causes of the failure to be analyzed, which illustrates the importance of the practical significance of the fault diagnosis research [3-6]. With the expansion of the scale of the power system, a large amount of information will enter the system after a fault occurs, and this information, in addition to the normal fault information, may contain a large amount of redundant information and error information, so that the operation and control personnel can accurately analyze and judge the faults that occur [7-8]. So that the timeliness and efficiency of fault handling can be realized to reduce the losses caused by faults, which requires a complete set of fault diagnosis system to judge and analyze the faults [9-10]. In order to guarantee the stability, reliability, economy and quality of the power system and other aspects of the operational requirements, on the one hand, to achieve the power system failure rate control, on the other hand, to shorten the repair time after the occurrence of faults, can be in-depth discussion of the power system fault characterization and fault diagnosis methods and related theories, fault prediction methods of the power system for related research [11-12]. Optimization of the multimodal large model of speech text in the power industry based on Bayesian network can improve the accuracy of the diagnosis results of large-scale power grids under the situation of large amount of missing data, which can help to improve the fault prediction ability of the power industry, and then ensure the safe and stable operation of the power system [13-15].

In this paper, we first use a pre-trained model based on Transformers encoder and a BERT model for speech and text feature coding, respectively, and use an extended cross-attention mechanism for multimodal fusion. Based on the optimization needs of Bayesian network structure, a random restart hill-climbing algorithm is incorporated to optimize the Bayesian network structure based on the Bayesian information criterion. And the annealing idea is integrated in this optimization algorithm, and finally the fault feature extraction method for power industry proposed in this paper is combined with the optimized Bayesian network structure method to propose a fault prediction method based on improved BBN. The efficient performance of fault feature extraction and fault prediction of this paper's model is verified by using case analysis respectively, which is of certain reference value for the upgrading research of fault prediction technology in the power industry.

2. Overview

In modern society, electricity is essential in people's lives, and its status in the development of the national economy is also higher and higher, and people's lives are more and more dependent on electricity, and at the same time, the stability and safety of the power system has begun to become a common concern. Literature [16] based on data-driven establishment of multiple power equipment failure prediction maintenance scheduling model, with multiple power transformer maintenance as an example of experimental analysis, the results verify the superiority of the proposed maintenance scheduling model in reducing costs and improving system stability. Literature [17] focuses on various fault prediction and localization methods in traditional distribution networks, smart grids and microgrids, aiming to find out more accurate and fast fault prediction and localization methods for distribution networks to improve reliability, fast restoration, optimization of power consumption and customer satisfaction. Literature [18] proposed a wind turbine fault prediction and diagnosis method based on SCADA data, and empirical analysis confirmed that the proposed method can predict the remaining service life of a generator up to 18 days in advance and has high accuracy in diagnosing specific types of generator faults. Literature [19] analyzes the early fault detection and classification of industrial equipment with respect to three fault types: spikes, bias and jammed sensor data, in which a powerful Extreme Gradient Enhancement (XGBoost) machine learning method is utilized, and the experimental results show the effectiveness of XGBoost in reducing the number of false alarms and improving the reliability of early fault detection, and the research provides a reference value for the field of predictive maintenance in the IoT industry. The research provides a reference value for the field of predictive maintenance in the IoT industry. Literature [20] proposed a power transformer fault prediction method based on transformer oil spectroscopy and data mining analysis, and experimentally verified that the proposed method can effectively and accurately predict low-energy discharges, high-energy discharges, and thermal faults in power transformers at temperatures above 700°C. The optimal absorbance-wavelength combination that can improve the prediction accuracy was also determined. Literature [21] proposed a hybrid data mining method for power system fault classification and prediction based on clustering, association rules and stochastic gradient descent in order to accurately categorize and predict power system fault types, and verified the feasibility of the method through an example analysis, which has global optimality and high prediction accuracy.

3. Multimodal Large Model of Speech-Text in Power Industry

3.1. *Structural design of the unimodal model*

The Mel Frequency Cepstrum Coefficient (MFCC) features are extracted from the raw audio to form a spectrum [22], and then the features are further feature extracted and classified using 2D convolution. In this paper, the use of recurrent neural networks or Transformers class of networks that have stronger ability to extract serialized information will have better results.

Therefore, this paper chooses Wav2Vec2. In the pre-training stage of the model, self clustering and mask self-supervision are used for pre-training, for the input audio features, randomly masking off some frames to the encoder, and using the Transformers-based encoder structure to predict the actual representations at the masks in an on-line clustered code list with corresponding quantization matches. For the Chinese downstream task, this paper uses the starting point of the English pre-training model and the starting point of the Chinese pre-training model for fine-tuning, respectively. The Bi-directional Encoder Representation Technique (BERT) model is used to encode, obtain sentence vectors and classify them to get the predicted results.

3.1.1. Wav2Vec2 based audio sequence classifier. The Contextual Transformers encoder is a module that follows Wav2vec2 [23] for deeper processing of the output of the feature encoder. This module uses the Transformer architecture. A convolutional layer is also used as relative position embedding. Finally, the output is processed by layer normalization.

The Quantization module is used to discretize the output of the feature encoder into a finite set of speech representations. The module also uses the Gumbel Softmax method.

Gumbel Softmax allows discrete codebook entries to be selected in a fully microscopic manner. Use the pass-through estimator and set up a G hard Gumbel softmax operation. Feature Encoder Output z maps to $l \in \mathbb{R}^{G\times V}$ Logits and selects the v th codebook entry as the probability of group g:

$$p_{g,v} = \frac{\exp\left(l_{g,v} + n_v\right)/\tau}{\sum_{k=1}^{V} \exp\left(l_{g,k} + n_k\right)/\tau} \tag{1}$$

The objective function for pre-training consists of two parts: the contrast loss function and the codebook diversity loss function.The Objective function can be expressed as the following equation:

$$L = L_m + \alpha L_d \tag{2}$$

where L_m is the contrast loss, L_d is the codebook diversity loss, and α is a hyperparameter.

In the contrast learning task, given the context network output c_t on the masked time step t, the model needs to recognize the true quantized potential speech representation q_0 among a set of quantized candidate representations $\tilde{q} \in Q$, including q, and K interference terms that are uniformly sampled from other masked time steps of the same discourse. The loss function L_m is defined as:

$$L_m = -\log \frac{\exp\left(sim\left(c_t, q_t\right)/\kappa\right)}{\sum_{i-\lambda_i} \exp\left(sim\left(c_t, \tilde{q}\right)/\kappa\right)} \tag{3}$$

where the cosine similarity between the contextual representation and the quantized potential speech representation is computed.

$$sim(a,b) = \frac{a^T b}{\|a\|\|b\|} \tag{4}$$

By maximizing each codebook p_i in which the codebook distribution does not contain gumbel noise or temperature, theloss is shown below:

$$L_d = \frac{1}{GV}\sum_{k=1}^{G} -H\left(\bar{p}_k\right) = \frac{1}{GV}\sum_{k=1}^{G}\sum_{\gamma=1}^{V} \bar{p}_{k,\gamma} \log \bar{p}_{k,\gamma} \tag{5}$$

3.1.2. BERT-based text classifier. BERT [24] uses a bidirectional Transformer encoder as its infrastructure. An unsupervised pre-training method is used to learn a generic language representation. In the pre-training phase, the input to the BERT model is a text sequence containing both (MLM) tasks and (NSP) tasks. After pre-training is completed, the BERT model can be trained on specific tasks by means of fine-tuning. In the fine-tuning phase, the BERT model is trained on a task-specific dataset and the model parameters are adjusted according to the labels of the pending tasks.

In order to convert the word-level vector representation into a sentence-level vector representation, an average pooling approach can be used here. That is, by adding up all contextually relevant word vectors and dividing by the total number of words in the sentence, the final hidden layer state sequence $H = \{h_0, h_1, \ldots, h_n\}$, sentence vector V, for a model of length n, can be represented as:

$$V_i = \frac{1}{n}\sum_{i=0}^{n} h_i \tag{6}$$

After obtaining the vector representation of the sentence, it can be fed into a fully connected layer for classification. In this way, BERT can effectively capture the semantic information of the sentence.

3.2. Modal fusion strategies using cross-attention mechanisms

3.2.1. Attention mechanisms. Through the attention mechanism, the model can enhance or diminish attention to different locations in the input sequence based on the importance of those locations, thus improving the performance of the model.

3.2.2.Self-attention mechanisms. In the traditional self-attention mechanism, the model treats each position in the input sequence as the model treats each position in the input sequence as a query, key, and value, and uses this information to compute an attention weight for each position.

$$Output = Soft\max\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{7}$$

3.2.3. Cross-attention mechanisms. Cross-attention [25] is a commonly used attention mechanism in multimodal tasks, which extends the self-attention mechanism.

For this task, cross-attention can be introduced in the encoder to handle two different input types, speech and text. The operation flow within the modal fusion layer of the cross-attention mechanism is shown in Fig. 1.

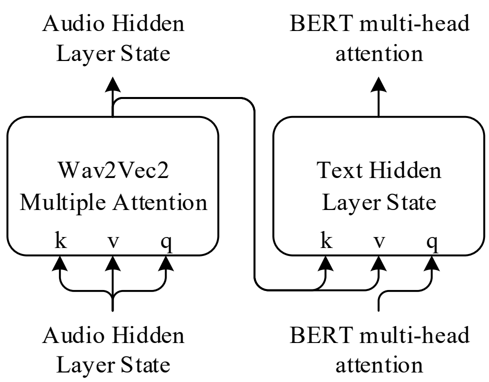

Fig. 1. Interfocus mechanism mode fusion layer operation

Suppose there are L layers, where the l nd layer contains a BERT layer and a Wav2Vec2 layer. Let $BERT_l$ denote the BERT model of layer l and $Wav2Vec22_l$ denote the Wav2Vec2 model of layer l. For a text sequence $T=\{t_1^0,t_2^0,\cdots,t_n^0\}$ and an audio signal $A=\{a_1^0,a_2^0,\cdots,a_m^0\}$, where n and m denote the lengths of the text sequence and the audio signal, respectively. First, the output of the audio layer $H_A^l=\{h_{A,1}^l,h_{A,2}^l,\cdots,h_{A,m}^l\}$ is obtained by self-attention coding the hidden layer state of the audio signal from layer $l-1$ at layer 11:

$$H_A^l = Wav2Vec2_l\left(\left(a_1^{l-1},a_2^{l-1},\cdots,a_m^{l-1}\right)\right. \tag{8}$$

Next, the output of the audio layer is used as the keys (Key) and values (Value) of the text layer, and the text sequence is used as a query (Query) for cross-attention coding:

$$\begin{aligned} Q^l &= BERT_l\left(t_1^{l-1},t_2^{l-1},\cdots,t_n^{l-1}\right) \\ H_T^l &= CrossAttention\left(Q^l,H_A^l,H_A^l\right) \\ &= Soft\max\left(\frac{Q^l H_A^l}{\sqrt{d}}\right)H_A^l \end{aligned} \tag{9}$$

The text sequence and the audio signal are jointly encoded. The obtained $H_A^l=\{h_{T,1}^l,h_{T,2}^l,\cdots,h_{T,n}^l\}$ represents the output of the text content in layer l.

Finally, the hidden layer state $H_{T,n}^L$ of the last text layer is input into a linear layer after average pooling in the sequence dimension and a Sigmoid function is used to obtain the probabilistic result of the rejection p:

$$p = Sigmoid\left(Linear\left(\frac{1}{n}\sum_{i=0}^{n} h_{T,i}^L\right)\right) \tag{10}$$

where Linear denotes a linear layer that maps the hidden layer states to a binary classification result.

4. Fault diagnosis method based on improved BBN

4.1. BBN network

Bayes' theorem describes the relationship between 2 conditional probabilities, usually the probability of event A conditional on the occurrence of event B is different from the probability of event B conditional on the occurrence of event A, but there is a definite relationship between the two. Conditional probability is specific, remember the probability of event A occurs as $P(A)$, event B occurs as $P(B)$, then in the B event occurs under the premise of A event occurs the probability of the conditional probability that is the conditional probability, recorded as $P(A \mid B)$.

$$P(A \mid B) = \frac{P(AB)}{P(B)} \tag{11}$$

The Bayesian formula is based on conditional probability, where $P(B \mid A)$ is used to find $P(A \mid B)$:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)} \tag{12}$$

BBN is a typical directed graph model. In this paper, the BIC method in the score-based structure search algorithm combined with the improved HC algorithm is selected for BBN structure learning.

The BIC is based on Bayes' theorem and takes into account the trade-off before the complexity of the model and the fit of the data.The formula for the BIC is shown in equation (13), where L represents the likelihood function of the data, k represents the number of estimated parameters in the model, and n represents the amount of data.

$$BIC = -2\ln(L) + k * \ln(n) \tag{13}$$

The scoring formula based on Bayesian scoring criterion is shown in equation (14), where m denotes the sample size.

$$Score_{BIC} = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} m_{ijk} \log_2 \frac{m_{ijk}}{m_{ij^*}} - \sum_{i=1}^{n} \frac{q_i\left(r_i - 1\right)}{2} \log_2 m \tag{14}$$

The set BIC score for variable X_i is:

$$BIC\left(\left(X_i, \pi\left(X_i\right)\right)\right) \middle| D = \sum_{j=1}^{q_i}\sum_{k=1}^{r_i} m_{ijk} \log_2 \frac{m_{ijk}}{m_{ij\cdot}} - \frac{q_i\left(r_i - 1\right)}{2} \log_2 m \tag{15}$$

Then the final BIC score formula for variable X_i is obtained:

$$Score_{BIC} = \sum_{i=1}^{n} BIC\left(\left(X_i, \pi\left(X_i\right)\right) ID\right) \tag{16}$$

4.2. Improved HC algorithm combined with BIC for learning BBN structures

The HC algorithm is a classical heuristic among many BBN structure learning methods, and is essentially a greedy algorithm. The algorithm combines the scoring criterion to make local adjustments to the current structure of the BBN, and contains a total of three methods: increasing edges, decreasing edges, and reversing edges. When carrying out these three operations, firstly, we need to ensure that the generated new network structure does not contain loops, and then according to the scoring criterion to calculate the scoring of each step to generate a new network, and select the optimal structure scoring, if the new scoring is smaller than the scoring of the previous network structure, the operation is valid, and based on the operation, the network structure is updated, and vice versa, the search is stopped and the current model is returned.

It can be concluded from the description and analysis of the HC algorithm that, due to its greedy search, it is easy to make the BBN structure fall into a local optimum, and this problem is related to the directed acyclicity of BBN itself. Based on this, this paper proposes a method based on the annealing idea of RRHC algorithm combined with BIC to learn the BBN structure.

The method first utilizes the HC algorithm for structure learning of BBN to generate the initial BBN.Subsequently, the HC algorithm is applied recurrently and the BIC is used to compute a score for each newly generated BBN structure.A smaller BIC value indicates a better model, and thus a progressively decreasing value of the BC score can be obtained. When the HC algorithm obtains a locally optimal solution for the BBN, the scoring value of its next network structure may will rise, which also means that the number of edges in the BBN may increase. At this time, the RRHC algorithm is used to re-select the network structure learning starting point and continue to learn the network structure. When the learning process reaches another local optimal solution, if the BIC score value of the optimal solution at this point is lower than the BIC score value of the previous optimal solution, this indicates that the new local optimal solution is closer to the global optimal solution, and therefore the iteration continues. The process draws on the annealing idea to determine whether to accept a new solution based on the Metropolis criterion. If the new solution is accepted, the algorithm does not stop by reaching the local optimum, but continues to iterate. The iteration will continue until the BIC score value stabilizes or the BIC score value of the last local optimal solution is higher than the BIC score value of the previous local optimal solution. Eventually, the network structure whose BIC scoring value stabilizes, or the network structure at the penultimate local optimal solution, is selected as the approximate global optimal solution. The flowchart of the method is shown in Fig. 2.

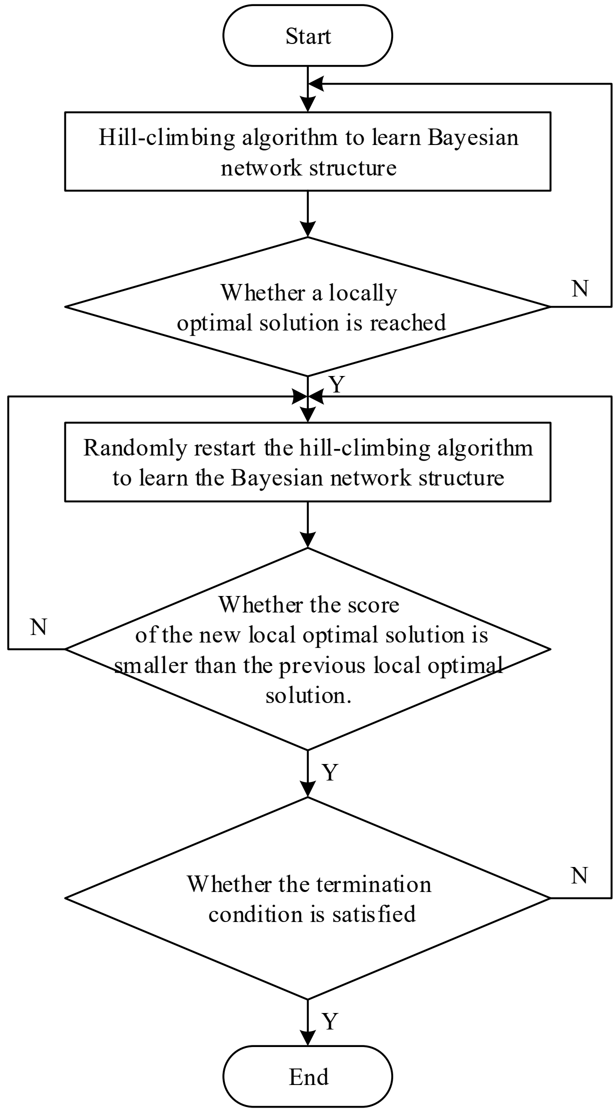

Fig. 2. Improved RRHC algorithm

4.3. Construction of fault diagnosis method based on improved BBN

In this section, the fault feature extraction method based on LMD, time-frequency domain features and PCA and the BBN structure learning method based on the improved HC algorithm are used to construct the rotating machinery fault diagnosis method based on the improved BBN. The software used is Pycharm Community Edition.According to the previous analysis, the flowchart of the rotating machinery fault diagnosis method based on improved BBN is shown in Fig. 3.

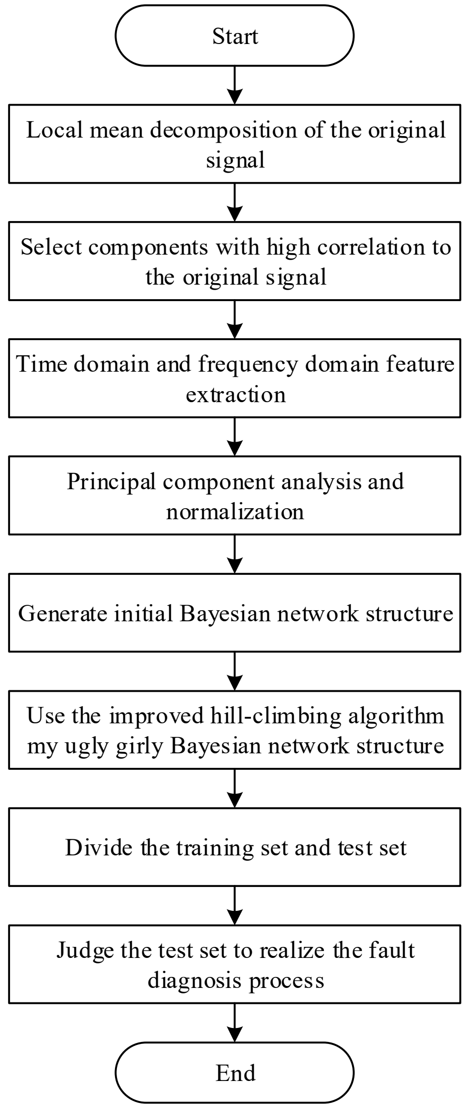

Fig. 3. Shows the flow chart of the rotating mechanical fault of BBN

The fault diagnosis method is divided into three main parts: data processing, structure optimization and fault diagnosis.

1) Data processing. Based on the structural principle of the unimodal model, the original signal is characterized by time and frequency domain feature extraction through this model and the designed multimodal fusion strategy based on the cross-attention mechanism, and the feature matrix is obtained. The feature matrix is used as the input of the fault diagnosis method in this chapter.

2) Structure optimization. Combine the idea of annealing with the improved HC algorithm (RRHC algorithm) for BBN structure learning, optimize the initial BBN structure, remove redundant edges, calculate lower BIC score values, and obtain a better network structure.

3) Fault diagnosis. The feature matrix after dimensionality reduction is discretized and the CPT of each node is calculated, combined with the generated globally optimal network structure, for the test set to achieve fault diagnosis.

5. Example analysis of fault prediction algorithms in the power industry

5.1. Multimodal feature extraction for power sector faults

In order to visually represent the feature extraction capability of the multimodal model designed in this paper, the two-dimensional feature vectors output from this model are plotted, and the results of the image presentation are shown in Fig. 4. As a comparison, for the output vectors of the multimodal model using the cross-attention mechanism, the principal component analysis (PCA) based on linear transformation was used for dimensionality reduction. PCA compresses the data space and projects the main components into two-dimensional rectangular coordinates, and the output is shown in Figure 5.

The multimodal model produces two-dimensional features with x-coordinates between (-80,80) and y-coordinates between (-80,80) and is more evenly distributed. It shows that different classes converge into different clusters and the overlap between different clusters is small. After being processed by the multimodal model, the mapped 2D features have good distinguishing characteristics between them. Therefore, the multimodal model based on the cross-attention mechanism has a powerful nonlinear feature extraction capability.

The 2D features generated by PCA have x-coordinates between (-9,9) and y-coordinates between (-6,7), but the distribution of the points is concentrated in a specific range, specifically, x-coordinates between (-3,3) and y-coordinates between (-3,2). It shows that there is a large overlap between the different clusters, in other words, the 2D features generated by PCA cannot distinguish between the different classes. This is due to the fact that PCA extracts semantic features based on linear transformations, thus limiting the ability to analyze the underlying text data. Therefore, in text mining, multimodal models based on cross-attention mechanism have better feature extraction capability than PCA.

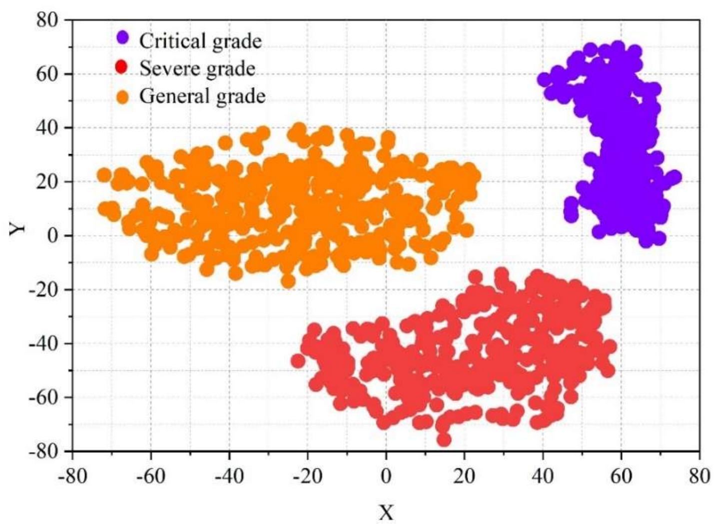

Fig. 4. The two-dimensional eigenvectors of the output of this article

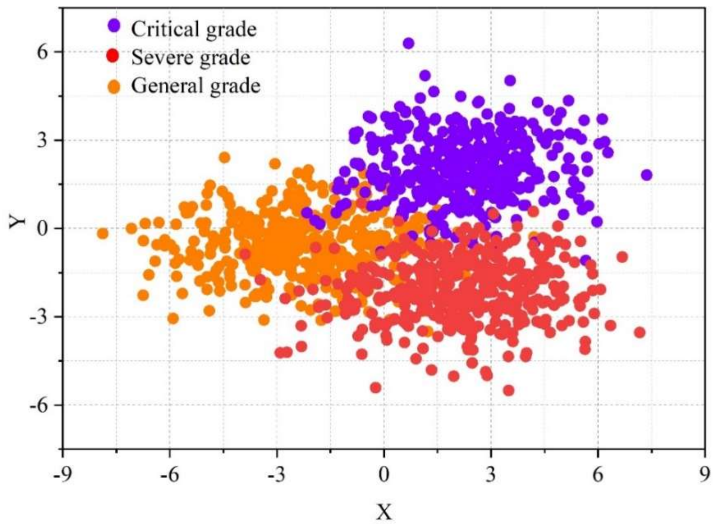

Fig. 5. The PCA performs the two-dimensional eigenvectors of the output

In addition, in order to intuitively and visually represent the semantic correlation between texts, the PCA method is used to reduce the dimensionality of the word vectors output by word2vec, and clustering is carried out in the two-dimensional space, and the clustering results are shown in Fig. 6. Among them, the words in the black round box are the input specified words, and the five words that are semantically closest to the specified words are shown. In the lower left corner of the figure, for the specified word "thermometer", it is semantically similar to "thermometer" and "oil thermometer", and the two-dimensional coordinates of these three are not far apart. The semantics of "thermometer" and "lid" are quite different, and their two-dimensional coordinates are far apart. And for the specified words such as "mild", "box lid" and "color change", the distance from their coordinates is closely related to the semantic correlation, therefore, the output of word2vec can be mapped to the Therefore, the output of word2vec can be mapped to the feature space, and the semantic relevance can be represented by the clustering distance of the words in the space.

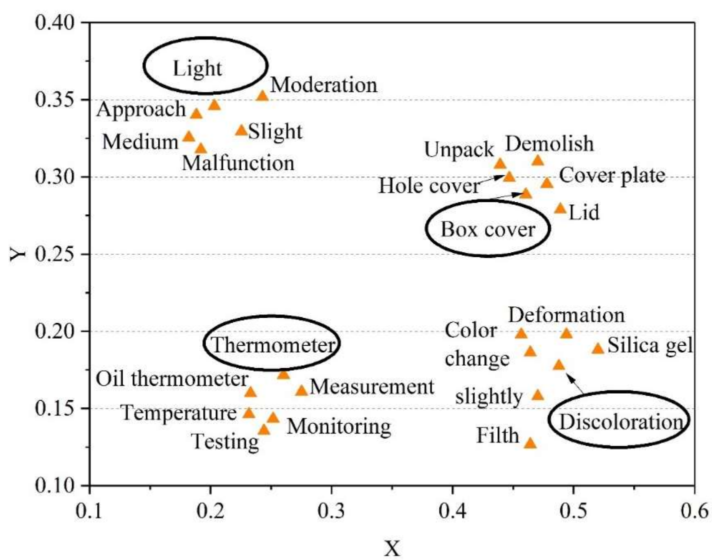

Fig. 6. Maps to the clustering results of two dimensional Spaces

5.2. Model Failure Prediction Capability Analysis

In order to illustrate the effectiveness of this method, a typical power system is analyzed as an example, and an extended Bayesian network is established for the failure of component B5. According to the principle of B5 associated relay protection action and the time limit of relay protection action, the time sequence of the associated relay protection action after the failure of component B5 can be obtained, and according to the method of identifying abnormal information, the determination of abnormal information is carried out.

The simulation software adopts Netica, which is a graphical interface and can realize the inference function of Bayesian network. According to the fuzzy rules of the alarm information to determine the abnormal information, the IFO node values of the component nodes that do not exist abnormal information are all taken as 1, and the original node probability value replacement is not carried out. The IFO node probability of component nodes with abnormal information is assigned a value and calculated after replacing the original node probability value, as a method to increase the number of trusted nodes in the network.

Thirty component failure scenarios (including bus faults, line faults, transformer faults, etc.) of this typical power system are verified and accurate results are obtained in all cases. By applying the proposed extended Bayesian network method for diagnosis, not only the faulty components can be obtained, but also the components with abnormal information can be determined. In order to visualize the probability change, the probability change of the simulation process of the 1st test is plotted and the change process is shown in Fig. 7.

The results show that the probability of B5 failure in the diagnostic results changes after each relay associated with B5 is actuated. However, when L4rs actuates, the probability of B5 failure decreases to 48.5%. This is because failures of other components can also cause L4rs to actuate, and the probability of a failure of B5 after a series of relay actuations is almost 100%. The above results show that the faulty component is B5 in test #1 of the simulation. By categorizing the anomaly information, the information about the mis-actuated and rejected components can also be obtained. The simulation process shows the characteristics of applying extended Bayesian network for fault diagnosis: the completeness of the information of the relays that are spatially close to the components has a great influence on the results, and the diagnosis results can be obtained from the action information of only a few relays in the individual tests, which greatly accelerates the diagnosis of faults.

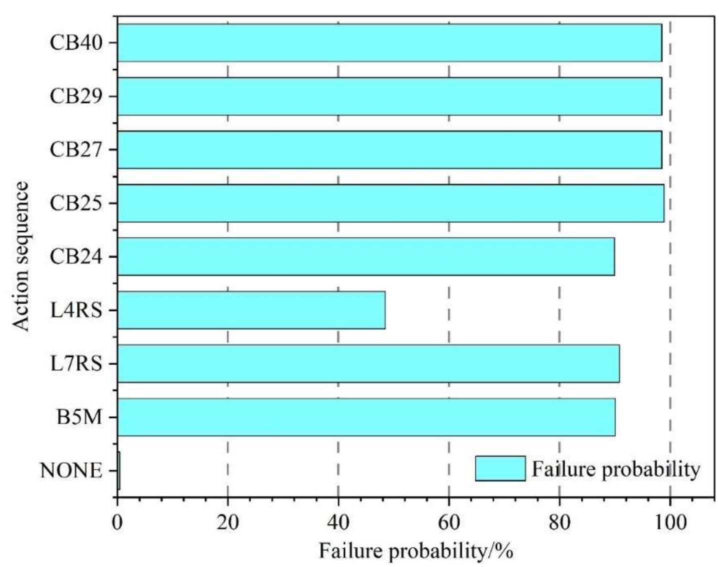

Fig. 7. B5 failure probability change process

The collected power system fault simulation data are processed, and each six-dimensional data sample is taken as the input, and the two-dimensional coordinate points are output, which are downgraded, and different colors represent different categories of data. Cluster analysis of the dimensionality reduction of the sample, the clustering results are obtained as shown in Figure 8, the 10 training sets obtained in the cluster analysis, respectively, the labeling of each training set, due to the occurrence of different types of faults, there will be some of the same signs appear, which results in the phenomenon of overlap of the category occurs, in addition to the system steady-state operation of the normal information, the remaining 9 categories for the system self-balancing set, as well as a variety of fault data sets, so it is necessary to take the sample data to test the type of fault data corresponding to the category labels. The clustering labels are shown in Table 1.

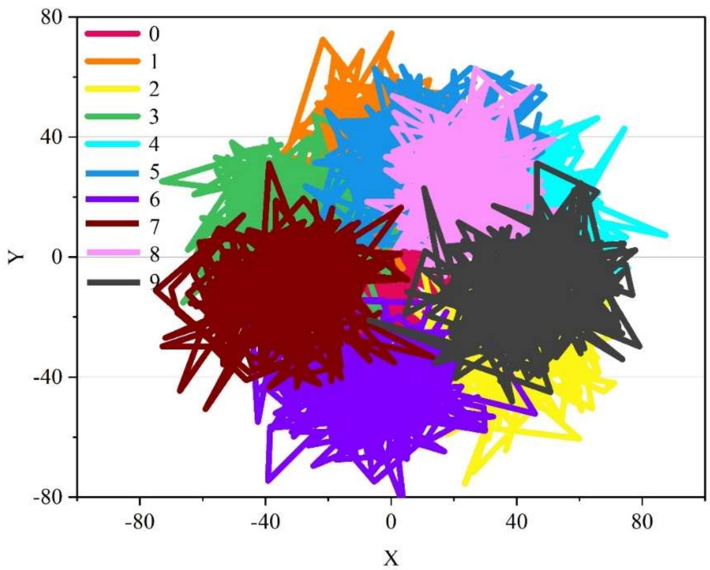

Fig. 8. Cluster result

Table 1. Different data tags

Data type	Single circuit fault	Two short circuit faults	Two short circuit ground fault	Three-phase fault	System self-balancing data	Steady state running data
Label category	1,6	4,7	5,8	2,3	9	0

After clustering the data labels are obtained, the training sets of different categories are obtained, the corresponding training sets are used to train the Bayesian network respectively after eliminating the steady state running data sets to obtain the Bayesian prediction model, the samples of the test set are selected and input into the Bayesian prediction model, and the confusion matrix of the predicted results is obtained by following the process of solving the Bayesian prediction model, the confusion matrix is shown in Fig. 9.

In the confusion matrix each column represents a different prediction category, and the sum of the column elements represents the total number of data in the category. The rows represent the category to which the data really belongs, the element sums represent the number of data instances, and the values in the diagonal lines represent the number of real data predicted for that category. The 10 training sets obtained in the cluster analysis, in which the steady state operation data were partially excluded from the training, the remaining 9 classes are the system self-balancing data set as well as the various fault data sets, the system self-balancing data has the highest number of samples, as can be seen in the prediction of Fig. 9, the 9th class has the highest number of correct predictions, 405, and the prediction accuracy is 100%. The prediction of the remaining 8 categories of fault data also yielded good results, and the category in which the test samples were categorized indicates that there is a likelihood that the fault corresponding to the category will occur. Since the samples are all obtained from Matlab simulation, with less perturbation and better sample divisibility, the overall prediction correctness for the test samples is relatively high, above 90%, and the overall prediction correctness meets the demand of the power industry.

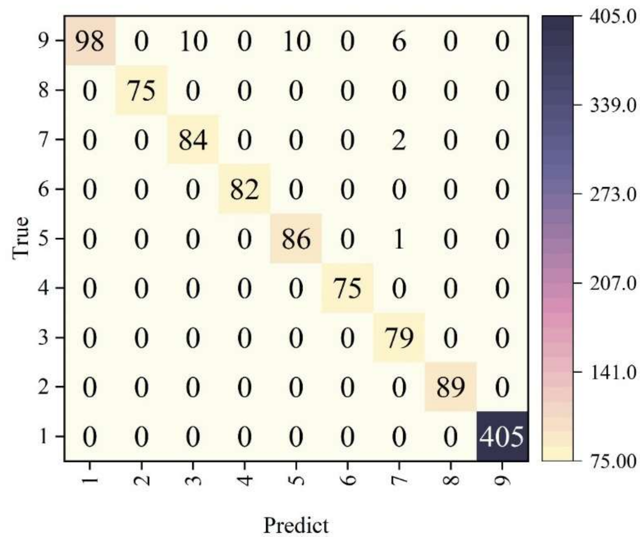

Fig. 9. The prediction results confuse the matrix

6. Conclusion

In this paper, a multimodal large model optimization method for speech text based on Bayesian network approach is proposed to be applied to system fault prediction in the power industry, and an arithmetic example analysis is carried out based on actual operation data. The main conclusions obtained are as follows.

1) The 2D features generated by this model are uniformly distributed, with different classes converging into different clusters, and there is almost no overlap between the clusters, while the 2D features generated by PCA are concentrated in a specific range, and there is a large overlap between the different clusters. As a result, the 2D features generated by PCA cannot distinguish feature classes. And the 2D features mapped by the model in this paper can distinguish feature categories better. It shows that the speech-text multimodal model in this paper has more powerful feature extraction ability for power industry fault features.

2) The prediction accuracy of this paper's model still reaches 100% on the basis of a large number of predicted faults in the 9th category. The prediction accuracies of the remaining 8 classes of fault data are also at a high level, indicating that the overall prediction accuracy of this paper's fault diagnosis method based on the improved BBN on the test samples reaches the design standard, and the optimization of the speech-text multimodal large model of the electric power industry based on the Bayesian network has been completed in a better way.

Funding

This work was supported by SGTJXT00SJJS2400175 (Information and Communication - Research and Development 2024-02) Research on Key Technologies of Power Artificial Intelligence Speech Semantic Large Model.

References

[1] Emmanuel, T., Mpoeleng, D., & Maupong, T. (2024). Power plant induced-draft fan fault prediction using machine learning stacking ensemble. Journal of Engineering Research, 12(2), 82-90.

[2] Patel, H. R., & Shah, V. A. (2018). Fault detection and diagnosis methods in power generation plants-the Indian power generation sector perspective: an introductory review. PDPU Journal of Energy and Management, 2(2), 31-49.

[3] Gayam, S. R. (2022). Deep Learning for Predictive Maintenance: Advanced Techniques for Fault Detection, Prognostics, and Maintenance Scheduling in Industrial Systems. Journal of Deep Learning in Genomic Data Analysis, 2(1), 53-85.

[4] Atrigna, M., Buonanno, A., Carli, R., Cavone, G., Scarabaggio, P., Valenti, M., ... & Dotoli, M. (2023). A machine learning approach to fault prediction of power distribution grids under heatwaves. IEEE Transactions on Industry Applications, 59(4), 4835-4845.

[5] Chen, P. (2023, November). Fault Diagnosis and Prediction of Power System Based on Convolutional Neural Network. In 2023 International Conference on Computer Simulation and Modeling, Information Security (CSMIS) (pp. 701-704). IEEE.

[6] Ahmed, Q., Raza, S. A., & Al-Anazi, D. M. (2021). Reliability-based fault analysis models with industrial applications: a systematic literature review. Quality and Reliability Engineering International, 37(4), 1307-1333.

[7] Qian, L., Qingzhao, T., Qinghui, Z., & Tianqi, W. (2020, November). Research on power transformer fault prediction model based on lstm neural network. In 2020 International Conference on Robots & Intelligent System (ICRIS) (pp. 702-705). IEEE.

[8] Andresen, C. A., Torsæter, B. N., Haugdal, H., & Uhlen, K. (2018, September). Fault detection and prediction in smart grids. In 2018 IEEE 9th international workshop on applied measurements for power systems (AMPS) (pp. 1-6). IEEE.

[9] Wang, Y., Nie, L., Dang, Y., Deng, G., Li, J., Li, P., ... & Chen, S. (2021, August). Fault Prediction of Power Electronic Devices in Mobile UPS System. In 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA) (pp. 1552-1556). IEEE.

[10] Fan, J., & Yuan, W. (2021). Review of parametric fault prediction methods for power electronic circuits. Engineering Research Express, 3(4), 042002.

[11] Ling, J., Liu, G. J., Li, J. L., Shen, X. C., & You, D. D. (2020). Fault prediction method for nuclear power machinery based on Bayesian PPCA recurrent neural network model. Nuclear Science and Techniques, 31(8), 75.

[12] Klaar, A. C. R., Stefenon, S. F., Seman, L. O., Mariani, V. C., & Coelho, L. D. S. (2023). Optimized EWT-Seq2Seq-LSTM with attention mechanism to insulators fault prediction. Sensors, 23(6), 3202.

[13] Wang, B., Xiao, X., Guan, C., Xia, Z., Gao, Y., Xu, M., & Dong, L. (2024, July). An Intelligent Voice Interaction System Based on Intent Recognition and Large Model Technology in the Context of Electrical Inspection. In 2024 International Conference on Energy and Electrical Engineering (EEE) (pp. 1-6). IEEE.

[14] Kim, E., Huh, D. H., & Kim, S. (2022). Knowledge-based power monitoring and fault prediction system for smart factories. Personal and Ubiquitous Computing, 26(2), 307-318.

[15] Di, Y., Jin, C., Bagheri, B., Shi, Z., Ardakani, H. D., Tang, Z., & Lee, J. (2018). Fault prediction of power electronics modules and systems under complex working conditions. Computers in Industry, 97, 1-9.

[16] Geng, S., & Wang, X. (2022). Predictive maintenance scheduling for multiple power equipment based on data-driven fault prediction. Computers & Industrial Engineering, 164, 107898.

[17] Dashti, R., Daisy, M., Mirshekali, H., Shaker, H. R., & Aliabadi, M. H. (2021). A survey of fault prediction and location methods in electrical energy distribution networks. Measurement, 184, 109947.

[18] Zhao, Y., Li, D., Dong, A., Kang, D., Lv, Q., & Shang, L. (2017). Fault prediction and diagnosis of wind turbine generators using SCADA data. Energies, 10(8), 1210.

[19] Atassi, R., & Alhosban, F. (2023). Predictive Maintenance in IoT: Early Fault Detection and Failure Prediction in Industrial Equipment. Journal of Intelligent Systems & Internet of Things, 9(2).

[20] Fauzi, N. A., Ali, N. N., Ker, P. J., Thiviyanathan, V. A., Leong, Y. S., Sabry, A. H., ... & Mun, L. H. (2020). Fault prediction for power transformer using optical spectrum of transformer oil and data mining analysis. IEEE Access, 8, 136374-136381.

[21] Wang, Y., Wang, X., Wu, Y., & Guo, Y. (2020). Power system fault classification and prediction based on a three-layer data mining structure. Ieee Access, 8, 200897-200914.

[22] Sumair Aziz,Muhammad Umar Khan,Adil Usman,Muhammad Faraz,Yazeed Yasin Ghadi & Gabriel Axel Montes. (2025). Bearing faults classification using novel log energy-based empirical mode decomposition and machine Mel-frequency cepstral coefficients. Digital Signal Processing(PA),104776-104776.

[23] D. Grebenkin & I. Bondarenko. (2024). Wav2vec2 Without Attention: Do You Need Hopfield Networks for Self-Supervised Learning of Speech Representations?. Journal of Mathematical Sciences(1),1-8.

[24] Shiva Sattarpour,Ali Barati & Hamid Barati. (2024). EBIDS: efficient BERT-based intrusion detection system in the network and application layers of IoT. Cluster Computing(2),138-138.

[25] ZhifangDeng & YangdongWu. (2024). CATNet: A Cross Attention and Texture-Aware Network for Polyp Segmentation. International Journal of Imaging Systems and Technology(6),e23220-e23220.