International Journal of Engineering and Technology Innovation, vol. 14, no. 1, 2024, pp. 103-114

Prediction of Distribution Network Line Loss Rate Based on Ensemble Learning

Jian-Yu Ren[1], Jian-Wei Zhao[2,*], Nan Pan[2], Nuo-Bin Zhang[2], Jun-Wei Yang[3]

[1]Measurement Center, Yunnan Power Grid Co., Ltd., Kunming, China

[2]Faculty of Civil Aviation and Aeronautical, Kunming University of Science & Technology, Kunming, China

[3]LongShine Technology Group Co., Ltd., Wuxi, China

Received 18 September 2023; received in revised form 27 November 2023; accepted 28 November 2023

DOI: https://doi.org/10.46604/ijeti.2023.12869

Abstract

The distribution network line loss rate is a crucial factor in improving the economic efficiency of power grids. However, the traditional prediction model has low accuracy. This study proposes a predictive method based on data preprocessing and model integration to improve accuracy. Data preprocessing employs dynamic cleaning technology with machine learning to enhance data quality. Model integration combines long short-term memory (LSTM), linear regression, and extreme gradient boosting (XGBoost) models to achieve multi-angle modeling. This study employs regression evaluation metrics to assess the difference between predicted and actual results for model evaluation. Experimental results show that this method leads to improvements over other models. For example, compared to LSTM, root mean square error (RMSE) was reduced by 44.0% and mean absolute error (MAE) by 23.8%. The method provides technical solutions for building accurate line loss monitoring systems and enhances power grid operations.

Keywords: line loss rate prediction, integrated learning, K-means clustering, LSTM, XGBoost

1. Introduction

In recent years, environmental and energy issues have become increasingly important. Energy is considered a major factor in achieving sustainable development, and energy saving and reduction in consumption have become long-term strategic goals for power grid operating companies. With the social and economic developments, electricity consumption has been rising year by year, and the problem of line loss in distribution networks has become increasingly prominent [1]. Line loss rate refers to the power network loss of electric energy accounted for the percentage of electric energy supplied to the power network. It reflects the planning level of distribution lines, as well as the level of the power grid company operation and management. Simultaneously, to a certain extent, the high and low line loss rates represent the high and low levels of power transmission technology [2]. Therefore, the effective prediction of the line loss rate of distribution networks has become a key issue in this field of research.

Currently, most studies online loss focus on theoretical calculations of line loss and their management methods. Moreover, the volatility and complexity inherent in the line loss data from power supply station areas render traditional analytical methods for grid line loss prediction inadequate for in-depth exploration of the data and the underlying patterns of line loss. Consequently, the prediction of grid line loss using traditional methods tends to be inaccurate. Traditional approaches often lack the necessary depth in their analytical methods, hindering accurate forecasts of line losses within the power grid. Subsequently, the prediction

* Corresponding author. E-mail address: 202214506235@stu.kust.edu.cn

English language proofreader: Chih-Wen Teng

of grid line loss is inaccurate. With the deep study of artificial intelligence algorithms and machine learning, the integration of machine learning methods and their application to the prediction of power grid line loss is becoming a trend, and the ability to realize the accurate and fast prediction of reasonable line loss has become a major problem that needs to be solved urgently [3-4]. This study proposes a line loss rate prediction method based on machine learning and integration ideas.

The study first identifies and cleanses the original "dirty data," and the common data cleaning methods are statistical 3σ criterion, box plots, and clustering methods based on machine learning, local anomaly factors, isolated forests, and deep learning methods [5-8]. Due to the diversity of line loss problems, false alarms, omissions, and other problems in the detection process, the above methods present a significant human impact on anomaly data detection. In the process of anomalous data detection, false alarms, and omissions are inevitable due to the problems' diversity, and in the process of anomalous data detection, places such as line loss size, clustering categories, thresholds, and so on need to be analyzed by human beings, and the detection results are affected by human beings to a greater extent [9-10]. On the other hand, data, in the collection process, inevitably produce multiple errors or missing attribute values, resulting in data errors, missing data, data inconsistency, and various data problems.

Thus, using certain data cleaning methods to correct these problems or complete the deficiencies improves the overall availability of data and data quality [11-12]. In this study, a dynamic cleaning method based on machine learning for anomalous missing data in distribution networks is drawn upon and adopted to achieve the purpose of data cleaning [9]. The adopted method significantly improves the limitations of missing data replenishment, which requires manual settings and low replenishment efficiency, and enables accurate and automatic selection of anomaly identification thresholds, effectively detecting anomalous data in distribution networks, taking into account both accuracy and speed of replenishment of missing data.

The traditional methods for calculating the theoretical line loss rate of a station area include the current calculation method, the load curve method, the node voltage method, the square root current method, the equivalent resistance method, and so on. However, it is difficult for the traditional line loss calculation methods to clarify the deeper relationship between the massive data and the line loss value of the network since they are usually of low computational accuracy and poor real-time performance, which are at greater variance with the line loss model of the distribution network [13-15]. As for the line loss prediction of the power system, the above methods, to some extent, address the network line loss prediction problems, compared with the traditional methods, to improve the prediction accuracy.

However, some issues are still present; the prediction of real-time is better but the number of the considered indicators is minor and the prediction results of the existence of chance are limited or can only be predicted when the data set is complete; they can't cover all the indicators [16-19]. Moreover, there are multiple problems in the face of large and complex data streams of low-voltage station power supply: The prediction accuracy is not high [20-21]. In this study, firstly, data features were extracted using K-means clustering, which simplifies the process of calculating the line loss in the station area and increases the calculation efficiency.

Additionally, the time series prediction model was established using the long short-term memory (LSTM) network, and recent research shows that the use of LSTM effectively solves nonlinear and time series problems, and the accuracy of the LSTM algorithm is higher than that of backpropagation (BP) neural network neural networks and recurrent neural networks (RNN), particularly in cases where large amounts of data are involved [22-24]. Subsequently, the multiple linear regression model was established according to the clustering results, and the linear regression equation prediction calculation method has the advantages of being a simple model, accurate calculation, and strong model interpretation ability [25]. Finally, the extreme gradient boosting (XGBoost) algorithm was used to integrate the prediction results of LSTM and linear regression to achieve the prediction of the line loss rate. The XGBoost, integrated with the boosting method, increases the regularization term, which effectively prevents the occurrence of overfitting in addition to being more efficient [26-27].

This research introduces a novel approach to forecasting electric grid line loss rates through a method rooted in data preprocessing and model integration. By incorporating concepts of integrated learning and machine learning, the study proposes an integrated methodology that involves dynamic cleaning technology to enhance data quality. Additionally, the approach integrates LSTM, linear regression, and XGBoost models from diverse perspectives. Through weighted fusion and optimization of multiple models, the line loss prediction results surpass those of single-model predictions, achieving heightened accuracy, improved generalization ability, and enhanced fitting clarity. This innovative methodology contributes a fresh perspective to line loss prediction in the field. Implemented on the PyCharm platform using the Python programming language, the simulation utilizes the hardware environment of an 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz processor.

2. Method Brief Introduction

This section outlines machine learning-based approaches for dynamic data cleansing in analytical processing. It includes data loading, the local outlier factor (LOF) algorithm for outlier detection, and nuanced missing value imputation strategies. Subsequent sections (2.2 to 2.5) cover integrated prediction methods, including K-means clustering for data classification, LSTM networks for time-series modeling, linear regression based on K-means, and XGBoost-driven model fusion for enhanced predictive accuracy. These methodologies collectively form a robust framework for comprehensive data analysis and forecasting.

2.1. Dynamic data cleansing methods based on machine learning

Data preprocessing consists of four steps: data loading, outlier detection, missing value imputation, and feature normalization. Outlier detection uses the LOF algorithm. The idea of this algorithm is to calculate the LOF of each sample and use the local deviation of the relative density of the surrounding points to measure the degree of abnormality. Hence, if the set of all points in the distance field is denoted and the local attainable density in it is denoted as well, then the value of LOF is as follows:

$$LOF_k(x_i) = \frac{\sum_{x_i \in N_k(x_i)} lrd_k(x_i)}{\left|N_k(x_i)\right| lrd_k(x_i)} \tag{1}$$

After calculating the LOF values for each sample, a Gaussian mixture model (GMM) was used to cluster the LOF values and automatically determine the abnormality detection threshold. When filling in missing values, according to the criteria of consecutive missing data in each fifteen-minute scale, those with more than ten consecutive missing data were judged as long-term missing, those with less than ten consecutive missing data were judged as local missing, the least squares regression (LSR) was used to fill in the missing values for local missing, and the random forest (RF) regression algorithm was employed to fill in the missing values for long-term missing.

2.2. K-means clustering algorithm

The K-means algorithm was used to cluster the sample data to find a similar distribution pattern among different samples. The objective of the K-means algorithm is to classify the nth samples into cth families. Take $k \le n$, suppose the sample set $X = x_1, x_2, \cdots x_n$ is classified into families $C = c_1, c_2, \cdots c_k$, whose centers are $\mu = \mu_1, \mu_2, \cdots \mu_k$, then the K-means objective function is:

$$\arg\min c \sum_{i=1}^{k} \sum_{x \in C_i} \left|x - \mu_i\right|^2 \tag{2}$$

The clustering effect is achieved by iteratively solving the family and center in the objective function of the above equation. Use the contour coefficient to evaluate the clustering effect of different K values and select the optimal number of clusters.

2.3. LSTM timing modeling

LSTM networks are capable of learning long-term dependent features of time series data. Raw data are converted into a fixed-length sliding window input, where a window contains several consecutive time steps. Assuming the window size is w, the ith window is:

$$X^{(i)} = \left[x_i, x_{i+1}, \ldots, x_{i+w+1} \right] \tag{3}$$

where x_i is the input feature at the ith step.

These sliding window inputs are used to train the LSTM network, the structure of which consists of an input layer, multiple LSTM layers, and an output layer. In the input layer, the input of each time step is $x_t^{(i)}$. The LSTM layer learns the temporal characteristics of the time series by controlling the storage, forgetting, and output of information through the mechanisms of the forgetting gate, the input gate, and the output gate. The output layer is linearly transformed to obtain the final prediction $\hat{y}$.

The network loss function is the mean square error:

$$Loss = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)})^2 \tag{4}$$

where N is the number of samples and $y^{(i)}$ the target output of the ith sample. The LSTM model is trained by minimizing the loss function; the network parameters reflecting the time series characteristics are obtained.

2.4. Linear regression prediction based on K-means clustering establishment

Based on the K-means clustering results, a linear regression model is trained separately for each cluster to learn the features of the data within the cluster. The linear regression model assumes that the target value y and the features x satisfy the following criteria:

$$y = x^T \beta + \varepsilon \tag{5}$$

Solve for the model parameters β by minimizing the mean square error $J(\beta)$:

$$J(\beta) = \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - x^{(i)T} \beta)^2 \tag{6}$$

The model predictions are:

$$\hat{y} = x^T \hat{\beta} \tag{7}$$

2.5. XGBoost-based model fusion weighted prediction

Model fusion combines the temporal features of the LSTM network and the prediction results of the linear regression model to improve the prediction's accuracy. Given that the output features of the LSTM network are $\hat{y}_{lstm}$ and the predicted output of linear regression is $\hat{y}_{reg}$, they are concatenated in the sample dimension to form a new feature:

$$z = \left[\hat{y}_{lstm}, \hat{y}_{reg} \right] \tag{8}$$

Subsequently, weighted fusion is performed using the XGBoost algorithm, which fits a decision tree to the new features via the additive model and the concept of leaf weights:

$$Obj = \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i) + \Omega(f) \right] + \gamma T \tag{9}$$

where l is the loss function, f the leaf weights, T the number of leaf nodes, Ω the regularity term, and γ the regularization parameter.

3. Parameter Analysis

When constructing a data-driven prediction model, optimizing the hyperparameters is crucial to the model's performance. In this study, control variable experiments were designed to obtain the optimal parameter configurations for the three key parameters of the proposed model, i.e., the number of clusters, the sliding window size, and the number of LSTM training rounds.

3.1. Selection of the number of clusters K

The K-means clustering algorithm needs to specify the number of clusters K. Setting different values of K leads to the allocation of samples to different clusters. Moreover, the number of linear regression models changes, ultimately affecting the predictive effect of the model. To determine the optimal value of K, the Silhouette Coefficient is used as the evaluation index, and its value ranges from [-1, 1]. Note that the larger the value means the better the clustering effect. Fix the other parameters, change K from 2 to 10, and record the values of the contour coefficient of the model under different K.

The results are shown in Fig. 1. It can be observed that the contour coefficients show a trend of increasing and subsequently decreasing as the value of K increases, and reaches the maximum value when K = 3. This indicates that the optimal clustering effect is obtained when the samples are clustered into 3 classes. Considering the computational time cost, this study chose to set K = 3 for model training and prediction.

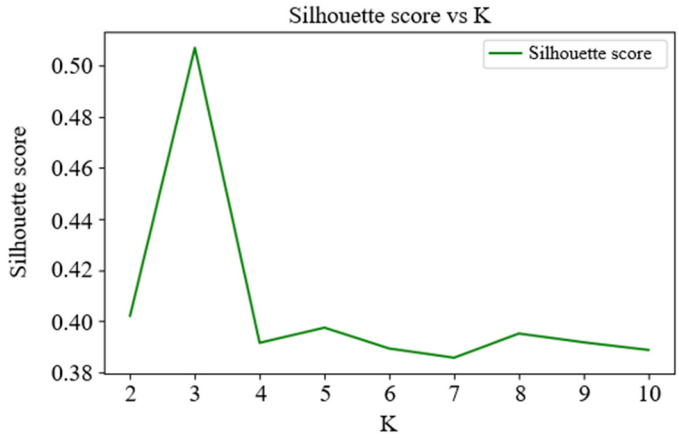

Fig. 1 Plot of contour coefficients corresponding to different values of K

3.2. Selection of sliding window size

The sliding window size is one of the substantial hyperparameters when constructing LSTM models, which controls the history length of the input sequence of a single training sample. If the window size is too small, it is unable to contain enough historical information to learn the data pattern of the time series; while if the window is too large, it introduces more redundant or even irrelevant information, increasing the difficulty of model training. To obtain the optimal window size, several LSTM models with window sizes ranging from 2 to 15 were trained and evaluated, and their root mean square error (RMSE) metrics on the test set were recorded. The specific results are shown in Fig. 2. It is clear that the smallest RMSE is achieved when the window size is 10. Considering the dependence characteristics of the time series and the computational cost, this study chose to set the sliding window size of 10 as the final hyperparameter of the LSTM model.

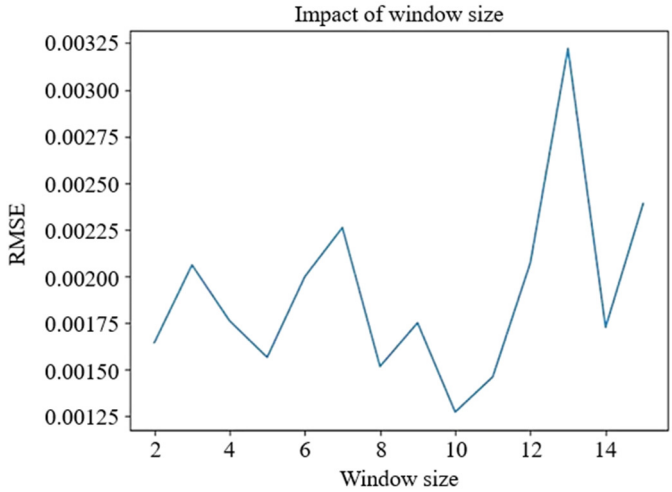

Fig. 2 Plot of RMSE values corresponding to different window sizes

3.3. Determination of the number of training rounds for LSTM neural networks

For neural network models, generally, increasing the number of training rounds improves the fitting ability of the model. However, it may lead to overfitting problems. To acquire the optimal number of training rounds, the loss function value Loss and the validation set error validation loss were recorded when the LSTM model was trained for 100 rounds, and the results are shown in Fig. 3 and Fig. 4. It is obvious that the loss function value tends to stabilize after 60 rounds, and the model converges; while the validation loss has slight fluctuations, it is generally maintained in a small stable interval. Considering the cost of training time, this study chose the number of training rounds for the model to be 60, and this hyperparameter setting makes the model achieve the best balance between performance and efficiency.

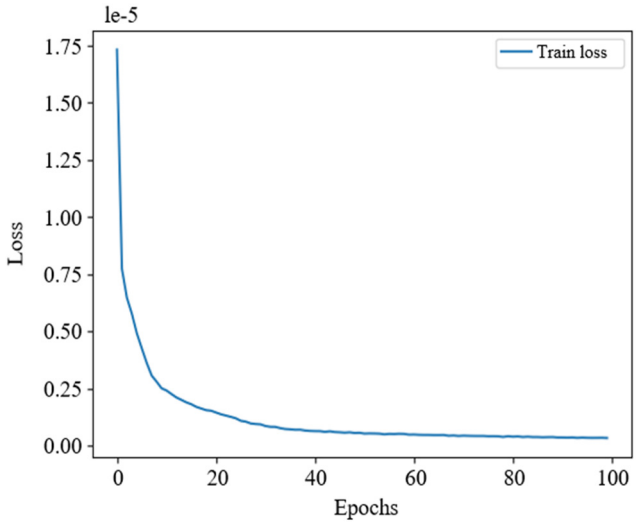

Fig. 3 Plot of training rounds versus train loss

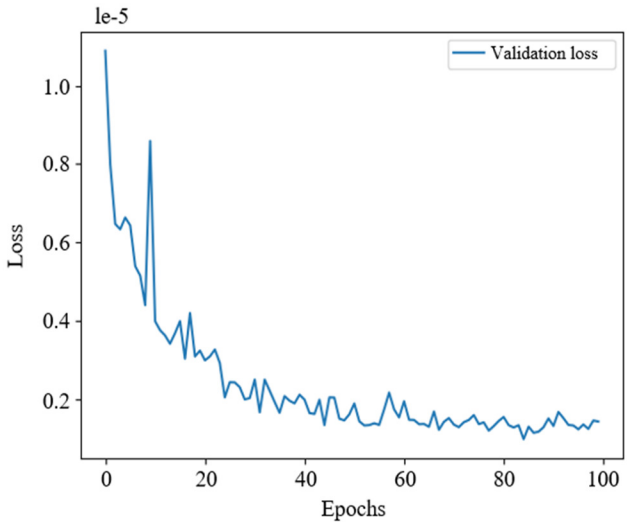

Fig. 4 Plot of training rounds against validation loss

Through the targeted hyperparameter optimization experiments described above, this study identified the optimal parameter configurations for the prediction model. Compared with manual empirical selection of parameters, this data-based approach with systematic evaluation of multiple sets of hyperparameters optimizes the performance of complex models.

4. Prediction Algorithm Simulation

An integrated prediction model based on linear regression and LSTM time series modeling: To enhance the accuracy of line loss rate prediction, this study established an integrated prediction process for line loss rate. The process encompassed data preprocessing, feature extraction, prediction modeling, and model integration, and the process framework is illustrated in Fig. 5.

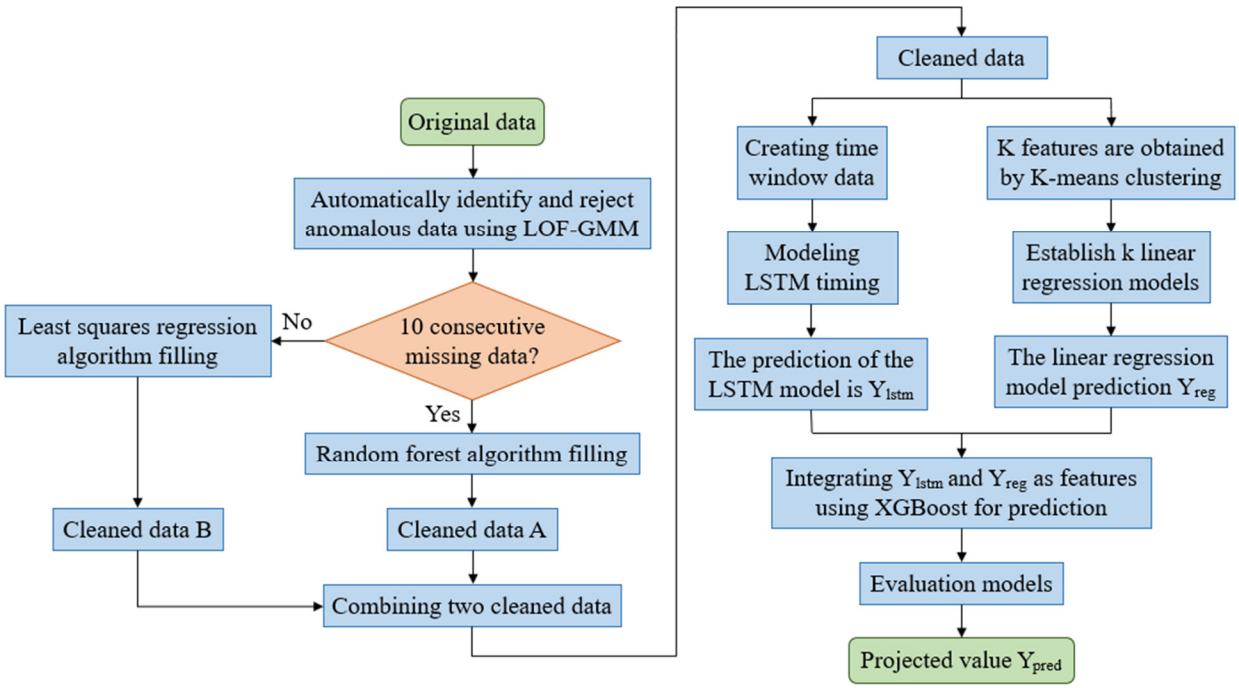

Fig. 5 Flowchart of algorithm prediction

4.1. Data preprocessing

Aiming at the problem of outliers and missing values in the original line loss data, this study adopted the LOF-GMM and LSR-RF algorithms to preprocess data. First, the LOF value of each sample was calculated. Subsequently, the sorted LOF values were clustered based on GMM to automatically determine the LOF anomaly judgment threshold. According to the threshold value, the abnormal samples were judged and rejected, and the abnormal blank samples were generated.

Secondly, the length of missing intervals in the original data was counted, and the data with more than ten consecutive missing items in a fifteen-minute scale were judged to be long-term missing, and those with less than ten items were judged to be short-term missing. Moreover, the LSR algorithm was used to predict and fill in the local missing items, and the RF algorithm was employed to predict and fill in the long-term missing items. After data preprocessing, the cleaned and usable line loss sample data were obtained.

4.2. Feature engineering

To fully extract the sample features, the cleaned samples were first clustered using the K-means algorithm, and the number of clustering categories K was determined by evaluating the profile coefficients. The classification labels were added as new features of the samples. Afterward, a time window of w was set, and the samples were generated in chronological order to generate window samples of w length.

Based on the LSTM network, the input window samples were used for temporal modeling to predict the line loss rate at the current moment, and the prediction results were added to the sample as new features. After the above processing, rich feature sample data were acquired.

4.3. Establishment of sub-model for prediction

According to the K-means clustering results, the samples are divided into K classes. Corresponding to the K classes of samples, a linear regression model was built to fit each class of samples to predict the line loss rate. For the test samples, according to their categories, the corresponding trained single model was selected for prediction. Subsequently, the sample data, LSTM prediction results, and all single-model prediction results were integrated as new samples.

4.4. Model fusion weighted prediction

The prediction results of the two sub-models were input as the features in the XGBoost model for training, to realize the weighted fusion of the prediction results of the single models and produce the integrated prediction results. The line loss rate was set as the target in the XGBoost model, and the other features were the inputs. After several rounds of training, an improved integrated regression model was obtained. Finally, using the obtained XGBoost integrated model, the test samples were predicted, and the final line loss rate prediction value y_{pred} was output.

4.5. Model evaluation

To comprehensively assess the prediction effect of the model, it is critical to establish evaluation indexes to quantitatively analyze the prediction results. In this study, root mean squared error (RMSE), mean absolute percentage error (MAPE), structural mean absolute percentage error (SMAPE), mean absolute error (MAE), and R^2 scores were used as the evaluation indexes, which were calculated as formulas [28-29].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(y_i - p_i)^2} \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{N}\left|\frac{y_i - p_i}{y_i}\right| \times 100\% \tag{11}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{N}\frac{\left|y_i - y_i^{'}\right|}{\left(\left|y_i\right| + \left|y_i^{'}\right|/2\right)} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^{'}\right| \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{'})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{14}$$

where n is the number of samples, y_i the actual value, y_i' the predicted value of the model, and $\bar{y}$ the mean value. RMSE reflects the absolute magnitude of the deviation of the predicted value from the real value. RMSE represents the absolute magnitude of the deviation between the predicted value and the real value; the smaller the value of RMSE, the smaller the deviation between the predicted result and the real value, and the better the model effect – MAPE – reflects the relative magnitude of the deviation between the predicted value and the real value.

On the other hand, the smaller the value of MAPE, the smaller the deviation between the predicted result and the real value, and the better the model effect SMAPE, a kind of improvement of MAPE – can deal with the situation that there is a zero value in both the real and predicted values. SMAPE is a kind of improvement of MAPE, which can deal with the situation that there is a zero value in the real value and the predicted value, and the smaller the value of SMAPE is, the better the model effect is. Note that MAE reflects the absolute magnitude of the prediction error, and the smaller the value of MAE, the better the model effect. The above indicators reflect the prediction effect of the model from different perspectives.

5. Discussion of Results

This section comprehensively evaluates the data preprocessing effectiveness and compares the performance of various predictive algorithms in the context of power grid line loss rate prediction. The preprocessing is assessed by contrasting the mean and standard deviation of data before and after cleansing, ensuring the cleaned data's mean falls within a specific range relative to the original. The prediction algorithms' efficacy, including ensemble learning, LSTM, K-means based linear

regression, and LightGBM, is compared using metrics like RMSE, MAPE, and R^2. The results highlight the proposed method's superiority in reducing prediction errors and improving accuracy, outperforming single-model approaches and demonstrating its robustness in handling complex data patterns in power distribution networks.

5.1. Results of data preprocessing

The cleaning effect was evaluated by comparing the mean and standard deviation of the data before and after cleaning. By comparing the mean value of the processed data and the mean value of the original data, it was judged whether the mean value of the cleaned data was between one-third and three times the mean value of the original data, to determine whether the cleaning effect was effective. Table 1 shows the mean and standard deviation before and after cleaning:

Table 1 Comparison of data indicators before and after cleaning

-	Original data mean	Original data std	Processed data mean	Processed data std
Total active power	0.962833	0.500733	0.964694	0.498141
Total reactive power	0.153359	0.083383	0.153490	0.082452
Power factor	0.974556	0.028935	0.974782	0.028416
Temperature	25.737705	1.433249	25.743130	1.424678
Ratio of line loss	0.005397	0.007037	0.005330	0.004767

5.2. Comparison of prediction algorithm results

As shown in Table 2, different prediction methods were used to predict the line loss rate under the same cleaning data method, and the prediction effect of different prediction models was evaluated by calculating indicators such as RMSE, MAPE, SMAPE, and so on.

Table 2 Comparison of prediction results of different prediction algorithms

-	RMSE	MAPE	SMAPE	MAE	R^2
Ensemble learning prediction method	0.00125	0.18264	1.25286	0.00070	0.94535
LSTM time series prediction method	0.00223	1.49750	1.13989	0.00092	0.82530
Linear regression prediction method based on K-means clustering	0.00298	0.36046	1.02593	0.00115	0.68698
LightGBM prediction method	0.00138	0.24211	1.29643	0.00074	0.90066

From the prediction results, it is clear that the method in this study has gained improvement over other single models in several key indicators. Specifically, the RMSE of this method is 0.00125, which is 44.0%, 58.2%, and 9.9% lower than those of the LSTM time-series prediction, linear regression prediction, and LightGBM prediction methods, respectively. Since the RMSE reflects the overall situation of the deviation of prediction value from the real value, this method minimizes the RMSE as a whole, which means that prediction errors are more concentrated and better predicted. This means the distribution of the prediction error is more concentrated and the prediction effect is better. As for the MAE, for this method, it is equal to 0.00070, which is 23.8%, 39.1%, and 4.8% lower than those of the above three methods, showing that the absolute deviation of the prediction error is also significantly reduced.

Furthermore, the MAPE of the present method is 0.18264, which is substantially reduced by 87.8%, 49.3%, and 24.6% compared with the other methods. Since the MAPE reflects the proportion of the prediction error relative to the target value, the present method minimizes this relative bias, and the prediction is more accurate. Moreover, although the SMAPE of the present method is slightly higher than that of the LSTM method, this is mainly because the SMAPE emphasizes more on the relative magnitude of the symmetric comparison error. Furthermore, the R^2 of the proposed method reaches 0.945, which is higher than that of other methods, indicating that the suggested method explains the variation of the target value in the optimal way. Analyzing all the indexes together, the present method is better than other single models in various key indexes and obtains the best prediction effect.

The method's quantitative improvements, as evidenced by reduced RMSE, MAE, and MAPE, along with an increased R^2, highlight its effectiveness in capturing underlying line loss patterns. The substantial decrease in both absolute and relative errors underscores the model's enhanced precision.

Furthermore, the method's outperformance against single models emphasizes its versatility for real-world applications. The comprehensive integration of LSTM, linear regression, and XGBoost contributes to its success in handling line loss intricacies. This multi-angle modeling approach enables superior adaptability to diverse data patterns, showcasing its resilience and predictive capabilities. In summary, the present method excels both quantitatively and qualitatively, demonstrating resilience, adaptability, and superior predictive performance in distribution network line loss rate prediction. Looking ahead, future work may explore further refinements to enhance the model's efficiency, consider the integration of emerging technologies, and extend its applicability to various power grid scenarios.

6. Conclusion

Distribution network line loss rate prediction is of high significance in improving the economic operation efficiency of power grids. However, the traditional prediction model has the limitation of low prediction accuracy. Thus, to improve the accuracy of line loss rate prediction in distribution networks, this study proposes a prediction of distribution network line loss rate based on ensemble learning. The contributions of this paper are as follows:

(1) In model construction, LSTM time series network, K-means clustering-based linear regression, and XGBoost algorithm are integrated for multi-angle modeling, merging predictive strengths.

(2) The effectiveness of the proposed model is evident in the achieved predictive performance. Through rigorous experimentation and analysis, the model successfully reduces the overall prediction error compared to traditional single models. This underscores the practical impact and effectiveness of the proposed methodology in enhancing the accuracy of distribution network line loss rate prediction.

(3) Key parameters including clustering categories, window size, and network training rounds are optimized through designed control variable experiments to tune predictive performance. Based on the experimental results, the optimal values for the key parameters were determined: a window size of 10, a training iteration of 60, and several clusters set at 3. These parameter values were obtained through careful experimentation and analysis, resulting in the best possible predictive performance.

The results demonstrate a significant reduction in overall prediction errors compared to single models. RMSE decreases by 44.0%, and MAE drops by 23.8% in comparison to LSTM, validating the effectiveness of the proposed approach. In summary, this study employs a multi-faceted optimization approach, encompassing data processing and model integration, resulting in enhanced predictive capabilities. This methodology offers an accurate technical solution for monitoring line loss rates in distribution networks.

Acknowledgments

This work was supported by the Science and technology project of Yunnan Power Grid Co., Ltd. under Grants YNKJXM20222483, YNKJXM2022534, and National Key Research and Development Program of China under Grants 2023YFB2407300.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] X. Chen, C. Song, and T. Wang, "Spatiotemporal Analysis of Line Loss Rate: A Case Study in China," Energy Reports, vol. 7, pp. 7048-7059, November 2021.

[2] J. Tulensalo, J. Seppänen, and A. Ilin, "An LSTM Model for Power Grid Loss Prediction," Electric Power Systems Research, vol. 189, article no. 106823, December 2020.

[3] Y. Shang, L. Kang, M. Zhang, X. Liu, and Y. Li, "Prediction Method of Electricity Stealing Behavior Based on Multi-Dimensional Features and BP Neural Network," Energy Reports, vol. 8, no. 4, pp. 523-531, July 2022.

[4] K. Li, W. Huang, G. Hu, and J. Li, "Ultra-Short Term Power Load Forecasting Based on CEEMDAN-SE and LSTM Neural Network," Energy and Buildings, vol. 279, article no. 112666, January 2023.

[5] S. Wang, Y. Lei, B. Yang, X. Li, Y. Shu, and N. Lu, "A Graph Neural Network-Based Data Cleaning Method to Prevent Intelligent Fault Diagnosis from Data Contamination," Engineering Applications of Artificial Intelligence, vol. 126, no. C, article no. 107071, November 2023.

[6] Q. Li, L. Zhang, G. Zhang, H. Ouyang, and M. Bai, "Simultaneous Detection for Multiple Anomaly Data in Internet of Energy Based on Random Forest," Applied Soft Computing, vol. 134, article no. 109993, February 2023.

[7] R. Morrison, X. Liu, and Z. Lin, "Anomaly Detection in Wind Turbine SCADA Data for Power Curve Cleaning," Renewable Energy, vol. 184, pp. 473-486, January 2022.

[8] X. Gao, J. Yu, S. Zha, S. Fu, B. Xue, P. Ye, et al., "An Ensemble-Based Outlier Detection Method for Clustered and Local Outliers with Differential Potential Spread Loss," Knowledge-Based Systems, vol. 258, article no. 110003, December 2022.

[9] Y. Mei, Y. Li, W. Zhou, Y. Guo, W. Deng, and X. Qiao, "Dynamic Data Cleaning Method of Abnormal and Missing Data in a Distribution Networkbased on Machine Learning," Power System Protection and Control, vol. 51, no. 07, pp. 158-169, 2023. (In Chinese)

[10] Y. Zhang, Q. Ai, H. Wang, Z. Li, and X. Zhou, "Energy Theft Detection in an Edge Data Center Using Threshold-Based Abnormality Detector," International Journal of Electrical Power & Energy Systems, vol. 121, article no. 106162, October 2020.

[11] Y. Yao, X. Zhang, and W. Cui, "A LOF-IDW Based Data Cleaning Method for Quality Assessment in Intelligent Compaction of Soils," Transportation Geotechnics, vol. 42, article no. 101101, September 2023.

[12] X. Li, M. Liu, K. Wang, Z. Liu, and G. Li, "Data Cleaning Method for the Process of Acid Production with Flue Gas Based on Improved Random Forest," Chinese Journal of Chemical Engineering, vol. 59, pp. 72-84, July 2023.

[13] L. Xie, H. Li, Y. Yuan, and H. Zhou, "Calculation of Line Loss in Transformer District Based on K-Means Clustering Algorithm and Improved MRVM," Journal of Electrical Engineering, vol. 16, no. 1, pp. 62-69, March 2021. (In Chinese)

[14] K. S. Sambaiah and T. Jayabarathi, "Loss Minimization Techniques for Optimal Operation and Planning of Distribution Systems: A Review of Different Methodologies," International Transactions on Electrical Energy Systems, vol. 30, no. 2, article no. e12230, February 2020.

[15] L. Huang, G. Zhou, J. Zhang, Y. Zeng, and L. Li, "Calculation Method of Theoretical Line Loss in Low-Voltage Grids Based on Improved Random Forest Algorithm," Energies, vol. 16, no. 7, article no. 2971, April 2023.

[16] S. J. Arand, A. Akbari, and M. Ardebili, "Investigation into the Thermal Behavior and Loadability Characteristic of a YASA-AFPM Generator via an Improved 3-D Coupled Electromagnetic-Thermal Approach," International Journal of Engineering and Technology Innovation, vol. 11, no. 2, pp. 88-102, April 2021.

[17] S. Klaiber, F. Bauer, and P. Bretschneider, "Improvement of Grid Loss Forecast for Energy Transmission Grids," at - Automatisierungstechnik, vol. 68, no. 9, pp. 738-749, August 2020. (In German)

[18] X. Zhong, J. Chen, M. Jiang, and X. H. Zheng, "A Line Loss Analysis Method Based on Deep Learning Technique for Transformer District," Power System Technology, vol. 44, no. 2, pp. 769-774, 2020. (In Chinese)

[19] N. N. Bon and L. V. Dai, "Fault Identification, Classification, and Location on Transmission Lines Using Combined Machine Learning Methods," International Journal of Engineering and Technology Innovation, vol. 12, no. 2, pp. 91-109, February 2022.

[20] M. X. Zhang, X. R. Liu, Y. Shang, L. Y. Kang, Q. T. Wu, and W. Y. Cheng, "Research on Comprehensive Diagnosis Model of Anti-Stealing Electricity Based on Big Data Technology," Energy Reports, vol. 8, no. 4, pp. 916-925, July 2022.

[21] A. Farhadi, A. Akbari, A. Zakerian, and M. Tavakoli Bina, "An Improved Model Predictive Control Method to Drive an Induction Motor Fed by Three-Level Diode-Clamped Indirect Matrix Converter," International Journal of Engineering and Technology Innovation, vol. 10, no. 4, pp. 265-279, September 2020.

[22] W. Zhou, Y. Li, Y. Guo, X. Qiao, Y. Mei, and W. Deng, "Daily Line Loss Rate Forecasting of a Distribution Network Based on DAE-LSTM," Power System Protection and Control, vol. 49, no. 17, pp. 48-56, 2021. (In Chinese)

[23] L. Li, B. Lu, W. X. Xu, Z. H. Gu, Y. S. Yang, and D. P. Tan, "Mechanism of Multiphase Coupling Transport Evolution of Free Sink Vortex," Acta Physica Sinica, vol. 72, no. 3, article no. 034702, February 2023. (In Chinese)
[24] C. P. Lu, Y. L. Huang, and P. J. Lai, "Development of the Abnormal Tension Pattern Recognition Module for Twisted Yarn Based on Deep Learning Edge Computing," International Journal of Engineering and Technology Innovation, vol. 13, no. 4, pp. 284-295, October 2023.
[25] U. M. Kaliyaperumal, M. S. B. Somasundaram, and N. Savarimuthu, "Partitioning-Based Data Sharing Approach for Data Integrity Verification in Distributed Fog Computing," International Journal of Engineering and Technology Innovation, vol. 13, no. 2, pp. 160-174, April 2023.
[26] C. Chen, Q. Zhang, B. Yu, Z. Yu, P. J. Lawrence, Q. Ma, et al., "Improving Protein-Protein Interactions Prediction Accuracy Using XGBoost Feature Selection and Stacked Ensemble Classifier," Computers in Biology and Medicine, vol. 123, article no. 103899, August 2020.
[27] W. Cao, Y. Liu, H. Mei, H. Shang, and Y. Yu, "Short-Term District Power Load Self-Prediction Based on Improved XGBoost Model," Engineering Applications of Artificial Intelligence, vol. 126, no. A, article no. 106826, November 2023.
[28] D. S. K. Karunasingha, "Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well," Information Sciences, vol. 585, pp. 609-629, March 2022.
[29] S. M. Robeson and C. J. Willmott, "Decomposition of the Mean Absolute Error (MAE) into Systematic and Unsystematic Components," PLoS One, vol. 18, no. 2, article no. 0279774, February 2023.