Quest Journals
Journal of Software Engineering and Simulation
Volume 11 ~ Issue 2 (2025) pp: 16-29
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Research on Power Trading Decision Support System based on GMM and Light GBM

Guichang LI[1]

1(Department of Computer Science，North China Electric Power University Baoding 071003，Hebei Province;)

ABSTRACT:*With the continuous growth of renewable energy in the power system, accurate prediction of wind power becomes an important issue in the auxiliary decision-making system of power trading, but the uncertainty of wind power generation brings great challenges to the stable operation of the power system. Accurate wind power prediction is helpful for the efficient development of wind power energy. Aiming at the problem of insufficient data mining in the face of complex wind power data, an ultra-short-term wind power prediction algorithm based on Gaussian mixture model (GMM), ridge regression algorithm and light gradient lift machine (LightGBM) is proposed. The proposed algorithm is specifically designed for application in the auxiliary decision-making system for power trading. The method uses GMM to cluster the features, and then uses the ridge regression algorithm to predict each cluster after clustering. Considering the limitations of a single prediction algorithm in the face of highly volatile wind power data, the prediction value after the ridge regression algorithm and the data after feature importance screening together constitute the final data set and transmit it to the LightGBM algorithm to obtain the final prediction result. Compared with widely used prediction algorithms such as CNN, SVM, and LightGBM, this method has higher stability and prediction accuracy. By applying this methodology in the auxiliary decision-making system for power trading, the proposed algorithm enables effective short-term wind power generation prediction and provides reliable decision-making support for the electricity market. Its application contributes to the improvement of operational efficiency in power systems and facilitates the sustainable utilization of renewable energy resources.*

KEYWORDS: *gmm; ridge regression; Light GBM; wind power prediction*

R*eceived 12 Feb., 2024;* ***R****evised 22 Feb., 2025;* ***A****ccepted 24 Feb., 2025*
P*ublished with open access at* ***www.questjournas.org***

I. INTRODUCTION

At the United Nations General Assembly on September 22, 2020, China promised the world that it would strive to achieve carbon neutrality before 2060. Against this backdrop, vigorously developing clean and renewable energy has become a major topic in recent years. As a low - carbon clean energy, wind power has become the third - largest energy source after thermal power and hydropower. However, the uncertainty of wind power generation poses a great challenge to the stable operation of the power system. Therefore, accurate wind power forecasting is of great importance [1][3]. At the same time, with the rapid development of the power trading market, the management core of power generation enterprises participating in power trading has shifted from the "production - based sales" planned management to the real - time response to market demand of "sales - based production". Wind power forecasting can accurately provide power information, offer suggestions for spot quotations, and maximize the revenue of power generation enterprises in power trading.

In power trading decisions, accurately predicting wind power is crucial for ensuring the balance between power supply and demand, optimizing power dispatching, and formulating trading strategies [4]. Especially in the ultra - short - term range, that is, the prediction within a few minutes to a few hours, the rapid changes in wind speed and wind power need to be considered so as to adjust the power generation plan and market transactions in a timely manner [5].

Currently, the methods for ultra - short - term wind power forecasting can be divided into two categories: physical methods and AI - based forecasting methods [6]. Physical methods mainly refer to directly solving the thermodynamic and fluid - mechanical equations satisfied by wind speed through the weather forecast data and geomorphic information around the wind farm to obtain information such as wind speed and direction. Then, the output power is calculated according to the power curve of the unit. This type of method does not require historical data and mainly relies on weather data. However, the weather is highly uncontrollable, so the error of this method is relatively large. Moreover, the calculation amount of solving the fluid - dynamics and thermodynamic equations is extremely large, resulting in slow prediction updates, which is not suitable for ultra - short - term forecasting.

AI - based forecasting methods mainly use machine - learning approaches to model the input - output relationship of the system by learning various rules, and use this model for prediction. Early methods mainly include artificial neural networks (ANN) [7] and support vector machines (SVM) [8], etc. In recent years, AI - based forecasting methods have been more widely applied, giving rise to methods such as recurrent neural networks (RNN) [9], long - short - term memory networks (LSTM) [10], k - means clustering algorithm [11], random forest algorithm (Random Forecast, RF) [12], etc. However, due to the large volatility of wind power generation, a single model is sensitive to data, so the prediction effect is not satisfactory. Researchers have begun to pay more attention to composite models. For example, in reference [13], the CNN and LightGBM algorithms are combined to improve the prediction accuracy and robustness. The disadvantage is that the features of the dataset are not fully explored, and there is a lack of processing of abnormal data [13].

In order to enable the model to better cope with the highly volatile wind power data, further explore the information contained in wind power generation data, and improve the prediction accuracy, a composite algorithm combining the Gaussian Mixed Model (GMM), Ridge regression algorithm, and Light Gradient Boosting Machine (LightGBM) is proposed. The dataset is taken from the SCADA system of a wind farm in Inner Mongolia, China, containing the collected data of a single unit throughout 2019. The time resolution of the data is 15 minutes. Research on this dataset verifies the effectiveness of the algorithm.

II. ULTRA - SHORT - TERM WIND POWER PREDICTION METHOD COMBINING GMM, RIDGE REGRESSION AND LIGHTGBM

With the advancement of data monitoring and storage technologies, data in the wind power field has become increasingly complex and diverse. Through data processing, useful information can be extracted from the vast amount of wind power data. The processing method proposed in this paper mainly relies on the GMM (Gaussian Mixture Model) algorithm and the ridge regression algorithm. Combined with abnormal data processing, feature construction, feature selection, etc., experiments have verified that this method can effectively reveal the internal laws of wind turbine power output and improve the accuracy of wind power prediction. An overview of the method is shown in Figure 1 below:

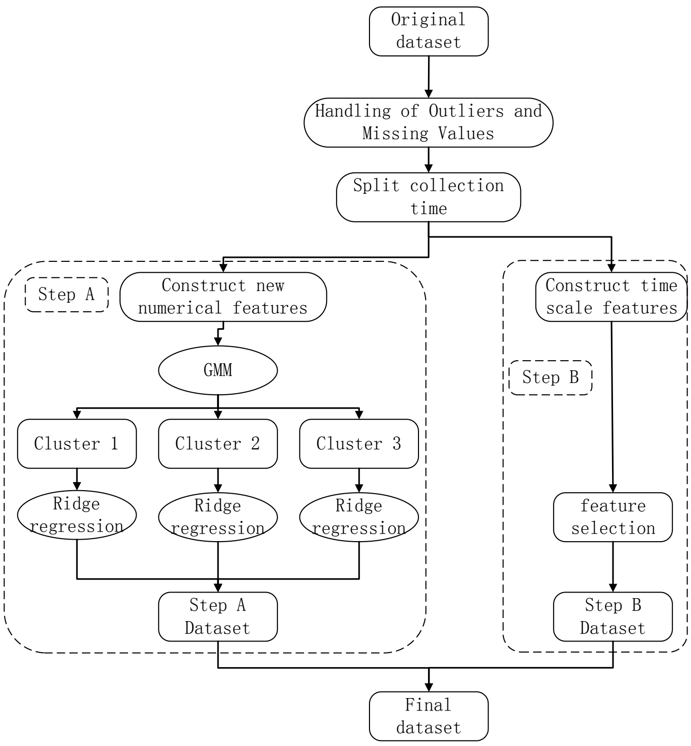

Figure 1 Dataset processing steps

(1) Traverse the dataset, and fill in the outliers and missing values by calculating the average of adjacent units.

(2) Split the collection time in the dataset into month, day, and hour, and then delete the collection time item. This is to facilitate the construction of new features for ultra - short - term prediction. After that, construct two parts of the dataset by taking Step A and Step B.

(3) In Step A, through new feature construction, the GMM algorithm performs clustering according to month, day, and hour, and the ridge regression algorithm is used to obtain 3 preliminary prediction results, which serve as part of the final dataset.

(4) Step B is to construct time - scale features, and combine the features of the original dataset with the time - scale features. Then, jointly perform feature importance screening, which serves as another part of the final dataset.

(5) Combine the results of Step A and Step B into the final dataset.

A. Gaussian Mixture Model (GMM)

1) data processing

The uncertainty of wind power generation is closely related to environmental changes. In the dataset, features such as wind speed, temperature, humidity, and pressure have varying degrees of impact on the actual power generation at different times. There are regular seasonal changes throughout the year, and there are also variations between high - wind and low - wind periods within a day. Therefore, considering clustering the data features according to three time periods: month, day, and hour [14], not only can the correlations between features be explored, but also outliers can be separated.

The data processing method before the GMM algorithm is shown as Step A in Figure 1. By constructing cross - terms and square terms of numerical features, the correlations between features can be further explored, and then the features are clustered. The original dataset has 27 features in total. The collection time in the dataset is split, and numerical features are separated from time features. After new feature construction, the dataset in Step A contains 256 features in total. The detailed description of the dataset processing method is as follows:

(1) After separating the time - type features and numerical - type features in the original dataset, the time features include 3 items: month, day, and hour; and there are 22 numerical features in total, such as wind speed, wind direction, temperature, etc.

(2) Add cross - terms and square - terms of numerical - type features (excluding time - type features). Constructing cross - terms and square - terms is a commonly used data - mining method. This can expand the feature base. By means of feature combination, the interaction effects between features can be explored, and sufficient variables can be prepared for the subsequent clustering algorithm. The formula for constructing the cross - term is shown in Equation (1), and the formula for the square - term is shown in Equation (2), where data represents the dataset, i represents any feature in the dataset, and j represents any feature in the dataset except i:

$$data[i*j] = data[i]*data[j] \quad (1)$$

$$data[i**] = data[i]*data[i] \quad (2)$$

After data processing, there are 3 time - related features, as shown in Table 1:

name	meaning
month	月
day	日
hour	时

Table 1 Time related Characteristics

name	meaning
ws10	10 meter wind speed
ws30	30 meter wind speed
dir10	10 meter wind direction
mslp	sea-level pressure
mslp*dir10	Cross term between sea level pressure and 10 meter wind direction
...	...
T2m	2-meter temperature

Table 2 Partial Numerical Characteristics

2) GMM algorithm

The Gaussian Mixed Model (GMM) is a widely used clustering algorithm that combines the advantages of parametric and non - parametric models [15]. The GMM algorithm estimates the probability density distribution of random variables through a linear combination of Gaussian components. When there are enough Gaussian components, the numerical characteristics of Gaussian components such as the expected value, variance, and covariance matrix can be solved to accurately represent the probability density distribution of feature values.

The probability distribution model satisfied by GMM is as follows:

$$\begin{cases} p(x) = \sum_{k=1}^{K} w_k N(x \mid \mu_k, \Sigma_k) \\ N(x \mid \mu_k, \Sigma_k) = \dfrac{1}{\sqrt{|\Sigma_k|(2\pi)^{d/2}}} \exp[-\dfrac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)] \\ \sum_{k=1}^{K} w_k = 1 \end{cases} \quad (3)$$

where $p(x)$ is the mixed probability density, $N(x \mid \mu_k, \Sigma_k)$ is the single - Gaussian probability density, and w_k, μ_k, Σ_k are the weight, mean, and covariance matrix of the k-th Gaussian model, respectively. The GMM algorithm fits k mixed Gaussian distributions by solving the Expectation - Maximization (EM). The specific methods are shown in Table 3 below:

Step 1: Initialize ϕ_k, μ_k, Σ_k and set the number of classes K

Step 2: Calculate the posterior probability $w_j^{(i)}$, of which

$w_j^{(i)} = p(z^{(i)} = j \mid x^{(i)}; \phi_k, \mu_k, \Sigma_k)$ 。

Step 3: Update parameters

$$\begin{cases} \phi_k = \dfrac{1}{m}\sum_{i=1}^{m} w_j^{(i)} \\ \mu_k = \dfrac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}} \\ \Sigma_k = \dfrac{\sum_{i=1}^{m} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}} \end{cases}$$

Step 4: Calculate the log - likelihood function:

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p(x^{(i)}; \phi, \mu, \Sigma)$$

Step 5: Determine whether the Gaussian parameters or the likelihood function has converged. If it has converged, end the iteration and assign the sample to the class with the maximum likelihood function. If it has not converged, return to Step 2 and continue the iteration.

For the feature clustering and grouping of the wind power dataset proposed in this paper, new features have been created by constructing feature cross - terms and square - terms before inputting into the GMM algorithm, and the item of collection time has been split into hour, day, and month. In order to perform feature clustering according to time scales such as hourly (hour), daily (day), and monthly (month), hour, day, and month are used as representations, that is

$$x^{(i)} = [hour, day, mouth]$$

Therefore, the input data of the GMM algorithm is a three - dimensional array with a length of \(m\). After clustering, the GMM outputs three clusters of eigenvalues that are closely related to hour, day, and month respectively. The three clustered eigenvalue clusters are then transmitted to the Ridge Regression algorithm, which fits a preliminary power prediction result for each cluster of features.

3) Ridge Regression Algorithm

After clustering by the GMM algorithm, there will be a strong correlation among the eigenvalues in each cluster, that is, there is multicollinearity. Moreover, the number of features in the three clusters is too large and needs to be further condensed. When there is multicollinearity among independent variables, the mean squared error will become very large, and the least - squares method cannot be used at this time. In such a case, the Ridge Regression algorithm is more suitable [16]. Essentially, the Ridge Regression algorithm is a modified least - squares method. When constructing a multiple linear regression model, the Ridge Regression algorithm abandons the unbiasedness of the least - squares method and sacrifices part of the information and accuracy as a price to make the regression coefficients more reliable. Especially when the correlation among regression variables is strong, the multiple linear regression equation will become extremely unstable, and the influence of some regression variables on the dependent variable will be concealed. The Ridge Regression algorithm can solve such problems well.

The regression coefficients of the multiple linear regression equation satisfy the following formula:

$$\beta = (X^T X)^{-1} X^T Y \quad (3)$$

where X is an n*m matrix of regression variables, X^T is the transpose matrix of X, Y is an n*1 vector of the dependent variable, and β is an m*1 vector of regression coefficients.

In real - world tasks, $X^T X$ is usually not a full - rank matrix, or the linear relationship between some columns is relatively strong. Therefore, there will be a large deviation in the solution of β. To solve this problem, a non - negative factor k is added to the main diagonal elements of $X^T X$, and the ridge regression equation satisfies the following formula:

$$\beta = (X^T X + kI_m)^{-1} X^T Y \quad (4)$$

where I_m is an m- order identity matrix, and k is called the ridge parameter. When the correlation between regression variables is strong, at least one of the eigenvalues of $X^T X$ will be very close to 0. However, by adding the parameter k, this deviation can be corrected, making the solution of β more reliable.

For the algorithm model proposed in this paper, ridge regression is mainly used to fit preliminary wind power prediction values for each cluster of eigenvalues after clustering by the GMM algorithm, and these three groups of prediction data are used as part of the final feature set. The essence of feature processing is to remove impurities and retain the essence. If all three clusters of features are input into the LightGBM algorithm without any processing, it will inevitably lead to a dimensional explosion due to the excessive number of features. By using ridge regression to condense the preliminary power prediction values from the three groups of predicted values respectively, it can effectively extract the feature correlations within each cluster and avoid drowning out important information caused by excessive features. At the same time, the results after preliminary prediction can better reveal the fluctuation characteristics of wind power, enhance the generalization ability of the dataset, and improve the prediction reliability.

B. LightGBM algorithm modeling

1) Build LightGBM feature set

The final prediction algorithm uses LightGBM. The dataset for LightGBM algorithm modeling is mainly divided into two parts. One part is the dataset from Step A obtained through the GMM and ridge regression algorithms, and the other part is the dataset from Step B after feature importance screening of the original dataset. The dataset from Step B is only used for LightGBM modeling and has nothing to do with GMM and ridge regression. Its function is to maintain the characteristics of the original data and prevent overfitting [17].

The construction method of this part of the feature set is described as follows:

(1) Conduct feature statistics of the actual power according to month, day, and hour. That is, calculate the maximum, minimum, median, sum, standard deviation, average, etc. of the actual power with month, day, and hour as units respectively, and add these data as new feature values to the original feature set. This is to add actual power labels to month, day, and hour. The new features after construction are as follows:

Monthly level	Daily level	Hour level
month_amount	day_amount	hour_amount
month_max	day_max	hour_max
month_median	day_median	hour_median
month_min	day_min	hour_min
month_sum	day_sum	hour_sum
month_std	day_std	hour_std

month_average	day_average	hour_average

Table 4: New Features Calculated by Month, Day, and Hour

(2) Calculate the feature importance between the new statistical features and the features in the original dataset, and the prediction target (actual power) together. Then, according to the Sequential Forward Selection (SFS) method, select the top 30 features with the highest importance as the new feature set. A total of 46 features participate in the screening. The selected feature set and the degree of feature importance are shown in Figure 2 below.

(3) Construct the final dataset by combining the Step B dataset with the top 30 important features and the Step A dataset obtained through the GMM algorithm and ridge regression.

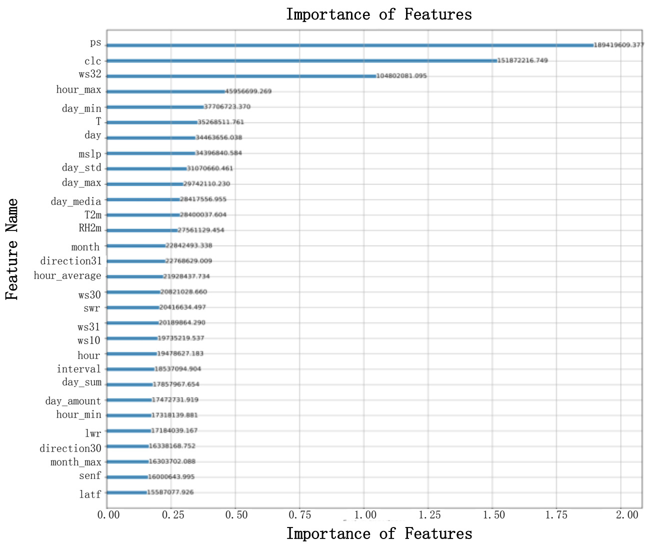

Figure 2. The top 30 features and their feature importance levels

2) LightGBM algorithm

The LightGBM algorithm (Light Gradient Boosting Machine) is a data model proposed based on the Gradient Boosting Decision Tree (GBDT) algorithm [18]. The LightGBM algorithm linearly combines \(M\) weak regression trees into a strong regression tree, achieving the goal of upgrading weak learners to strong learners, as shown in equation (7), where *F(x)* is the output and $f_m(x)$ is the output value of the m-th weak regression tree.

$$\mathrm{F}(x) = \sum_{m=1}^{M} f_m(x) \quad (5)$$

The traditional GBDT algorithm searches for the optimal splitting points for constructing decision trees through an exhaustive approach. This method is not only time - consuming but also requires a large amount of memory. In contrast, LightGBM adopts an improved histogram algorithm. Its basic idea is to discretize the continuous feature space into k intervals and search for the optimal splitting point within these k intervals. Therefore, the training speed and space efficiency of LightGBM are higher than those of GBDT [19].

At the same time, the LightGBM algorithm also employs a more efficient decision - tree growth strategy - the leaf - wise growth strategy with depth limitation. It searches for the leaf with the maximum gain among all current leaves and then splits it. This process repeats in a loop. As a result, the leaf - wise strategy has higher accuracy than the original decision - tree growth strategy. The depth limitation also ensures that the algorithm is efficient without overfitting.

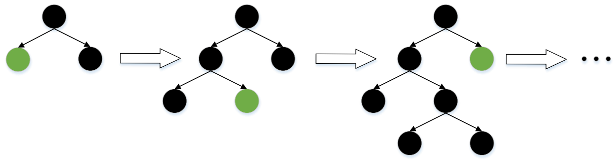

Figure 3 Leaf wise growth strategy

In terms of computing speed, the LightGBM algorithm uses histogram difference acceleration. That is, the histogram of a leaf node can be obtained by taking the difference between the histogram of its parent node and the histogram of its sibling node, which greatly improves the computing speed.

In the algorithm proposed in this paper, five - fold cross - validation is applied to the training set fed into LightGBM, and the Bayesian hyperparameter tuning algorithm is used to optimize the hyperparameters in LightGBM.

C. Algorithm Flow

A model combining the GMM, Ridge Regression, and LightGBM algorithms is established. The algorithm flow is shown in Figure 4.

(1) First, handle outliers and missing values in the original dataset, and split the collection time.

(2) Feed the processed dataset into the GMM clustering algorithm to cluster features according to time.

(3) Apply the Ridge Regression algorithm to each clustered group for a prediction.

(4) Conduct feature statistics of the actual power on a monthly, daily, and hourly basis. After merging with the original dataset, calculate the importance of feature values in the dataset, select the top 30 features with the highest importance, and combine them with the predicted values from Step A to form the final dataset.

(5) Input the final dataset into the LightGBM algorithm for prediction. Use five - fold cross - validation for the training set and optimize the hyperparameters using the Bayesian hyperparameter tuning algorithm.

(6) Output the prediction results.

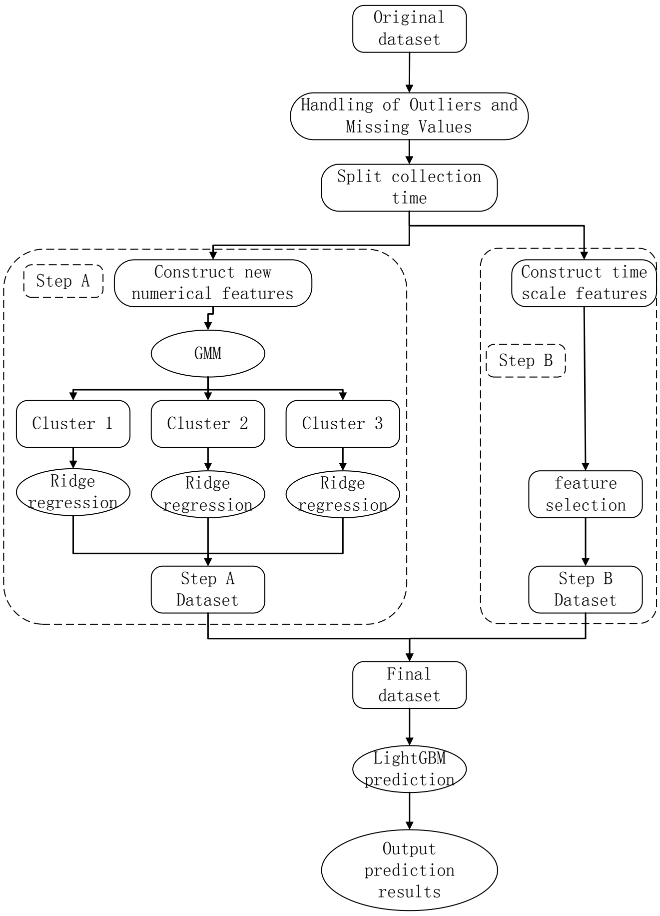

Figure 4 Algorithm flowchart

Diagram 1

III. Example Applications

A. data preparation

The dataset used is the fan data of a wind power plant in Inner Mongolia, China, collected throughout 2019 based on the SCADA system, with a sampling period of 15 minutes. The data in the initial dataset includes 27 features such as date, collection time, temperature, humidity, air pressure, precipitation, wind speed, wind direction, flux, actual power, and ID. The useless feature ID needs to be deleted.

A random sample of the wind power generation power curve of this unit within 24 hours on a certain day in 2019 is shown in Figure 5. By analyzing the curve in the figure, it can be found that there are power fluctuation periods and power stable periods in wind power. This is due to the changes in weather factors such as wind speed, air pressure, and temperature in a short period. There are climate change laws throughout the four seasons of the year, and there are also high - wind periods and low - wind periods every day. It is precisely because the weather changes over time that the wind power changes. Therefore, it is necessary to construct the time - scale features shown in Method B (calculate the maximum, minimum, median, sum, standard deviation, average, etc. of the actual power with month, day, and hour as units).

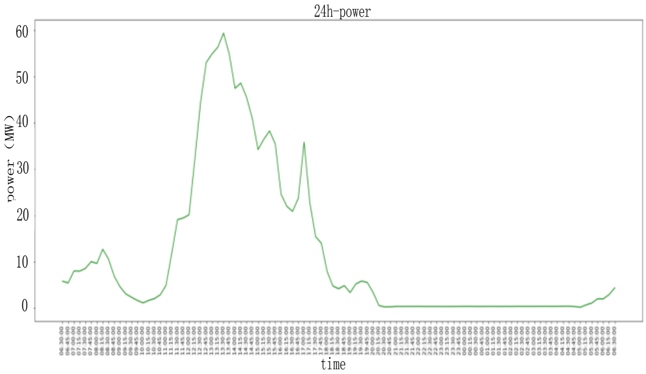

Figure 5 Power variation curve within 24 hours

A total of 256 features, which are obtained after constructing square terms and cross - terms of features, are input into the GMM algorithm for clustering. Then, each clustered group is used to make a preliminary prediction through the Ridge Regression algorithm. Sequentially, the first 1000 results after Ridge Regression prediction are selected and compared with the actual output power. The results are shown in Figure 6:

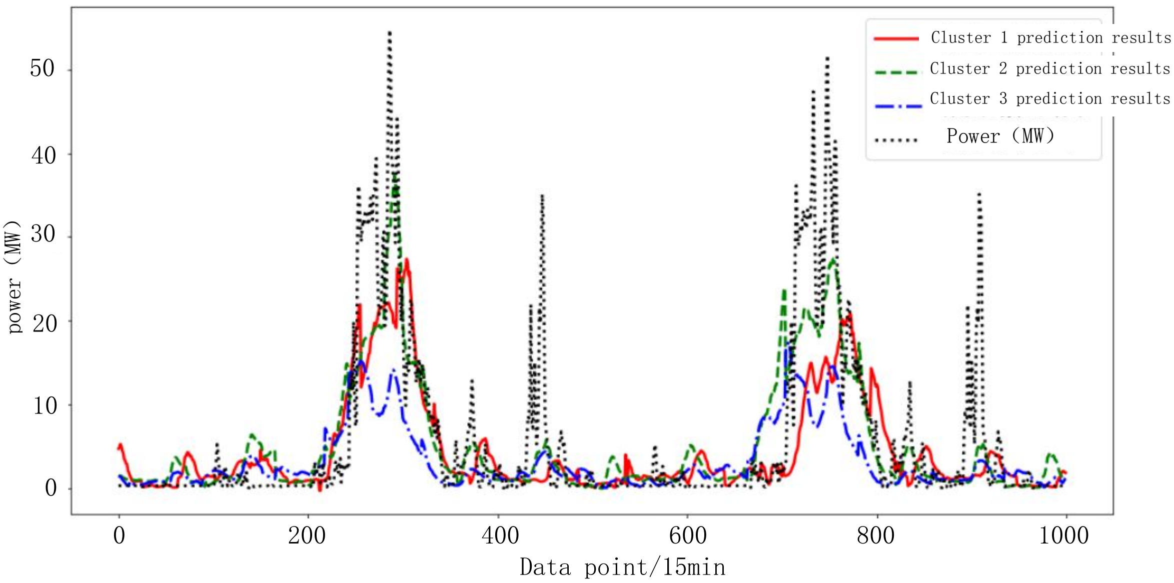

Figure 6 Comparison of Ridge Regression Prediction Results

As can be seen from the figure, after GMM clustering and preliminary prediction by ridge regression, the prediction results can already reflect the fluctuation characteristics of the actual power. However, the fitting degree of these three groups of prediction results with the actual power is not satisfactory in some time periods. This is because the datasets input into the ridge regression algorithm are three clustered datasets. The features within such datasets have a high correlation, but they lack the characteristics of the original dataset. Thus, they cannot fully reflect the characteristics of the real power, resulting in poor fitting with the actual power at peak values and in time periods with more intense fluctuations.

To correct the final prediction results and improve the generalization ability of the model, the data processing method of Step B shown in Figure 1 is adopted. The time - series distribution features of the actual power are constructed, and through the feature importance screening method, the top 30 features with the highest importance are selected. These features, together with the three groups of prediction results after clustering, jointly form the final dataset, which is then input into the LightGBM algorithm.

B. Model optimization

The final prediction is completed by the LightGBM algorithm. To obtain a more stable model, the Bayesian optimization algorithm is used for LightGBM to acquire the optimal hyperparameters, and five - fold cross - validation is adopted during training. Figure 7 shows the modeling flow chart of the LightGBM algorithm.

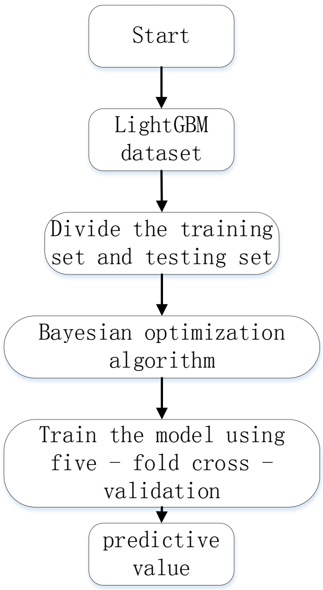

Figure 7: Flowchart of LightGBM Algorithm Modeling

Bayesian Optimization is an automatic hyperparameter - tuning algorithm [20]. It can utilize the prior knowledge of hyperparameters for intelligent optimization, quickly search for the optimal solution, and effectively avoid unnecessary hyperparameter sampling. The important hyperparameters of LightGBM after Bayesian optimization are set as shown in Table 5 below:

名称	含义	数值
max_depth	Maximum depth of the tree	8
num_leaves	Number of leaves	86
bagging_fraction	The proportion of data used in iteration	0.615
feature_fraction	The proportion of features selected when constructing the tree	0.6
min_child_samples	The minimum number of records a leaf node should have.	65
learning_rate	learning rate	0.01

Table 5 Optimal Hyperparameters of LightGBM after Bayesian Optimization

K - fold cross - validation is a commonly used model training method that can effectively evaluate the generalization performance of a model. In K - fold cross - validation, the dataset is evenly divided into K parts, with each part referred to as one fold. During training, K-1 folds are used for model training, and the remaining 1 fold serves as the validation set. Therefore, using K - fold cross - validation can prevent uneven data distribution during dataset partitioning, thereby improving the generalization ability of the model. The evaluation metric for model stability is the Root Mean Squared Error (RMSE), and the calculation formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (6)$$

In this experiment, five - fold cross - validation [21] was adopted. The learning curves of the model during the training process are shown in Figure 8. As can be observed from the figure, when the model approaches 6,000 iterations, the RMSE of both the training set and the validation set has tended to stabilize.

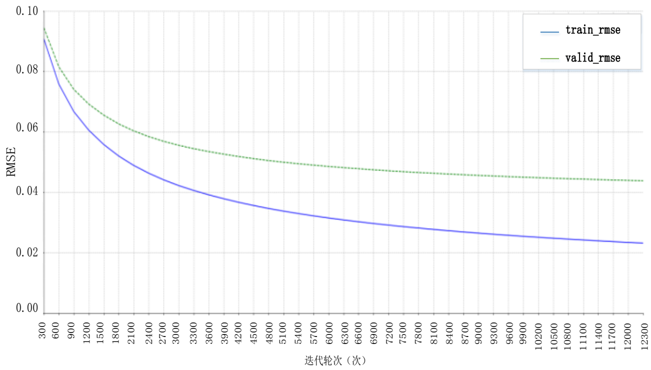

Figure 8 Model RMSE Curve

C. Result analysis

For the evaluation of the prediction accuracy of regression models, commonly used criteria currently include the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). To fully verify the prediction accuracy of the model proposed in this paper, MAE and MSE are adopted as the model evaluation indicators. The calculation formulas for MAE and MSE are shown in equations (9) and (10), where y_i represents the true value and $\hat{y}_i$ represents the predicted value of the model:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \quad (7)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (8)$$

The experiment adopts a comparative verification method. The algorithm proposed in this paper is simulated and compared with algorithms such as CNN, SVM, LightGBM, and CNN + LightGBM. For the LightGBM algorithm and CNN + LightGBM, Bayesian optimization and five - fold cross - validation are also used. The smaller the MAE and MSE values of the prediction results are, the better the prediction effect of the algorithm is. During the test, the same dataset is used, and the dataset is divided in a ratio of 8:1:1. That is, 80% of the data is used as the training set to train the model, 10% of the data is used as the validation set for model optimization, and the remaining 10% of the data is used as the test set to verify the model's performance. To avoid randomness, each algorithm is run five times, and the MSE and MAE values are statistically recorded respectively. Then, the average values of the five runs are calculated. The results are shown in Table 6 and Table 7.

MSE (10^{-3})					
	CNN	SVM	LightGBM	CNN+LightGBM	GMM+Ridge+LightGBM
1	1.912	2.441	1.880	1.759	1.511
2	1.913	2.446	1.872	1.776	1.530
3	1.907	2.432	1.862	1.774	1.477
4	1.911	2.445	1.877	1.741	1.500
5	1.904	2.439	1.871	1.788	1.483
avg	1.909	2.441	1.872	1.768	1.500

Table 6 MSE (10^{-3}) of Five Algorithms

MAE (10^{-3})					
	CNN	SVM	LightGBM	CNN+LightGBM	GMM+Ridge+LightGBM
1	23.415	35.635	23.150	22.972	22.326
2	23.423	35.332	22.916	22.983	21.517
3	24.881	34.932	22.938	22.911	21.353
4	25.192	35.535	23.235	23.115	22.158
5	23.911	34.793	23.384	22.921	21.106
avg	24.164	35.245	23.125	22.980	21.692

Table 7 MAE (10^{-3}) of Five Algorithms

Analysis of Table 5 shows that when the Mean Squared Error MSE (10^{-3}) is used as the evaluation criterion, the algorithm proposed in this paper has improvements of 21.42%, 38.55%, and 19.87% compared to CNN, SVM, and LightGBM respectively. Compared with the current CNN + LightGBM algorithm, there is also a 15.16% improvement. Analysis of Table 6 shows that when the Mean Absolute Error MAE (10^{-3}) is used as the evaluation criterion, the algorithm proposed in this paper has improvements of 10.23%, 38.45%, and 6.20% compared to CNN, SVM, and LightGBM respectively. Compared with the current CNN + LightGBM algorithm, there is a 5.60% improvement.

To more intuitively reflect the prediction effects of the five algorithms, according to the standards of ultra - short - term wind power prediction, the prediction results for four hours starting from the initial time are taken, with a sampling period of 15 minutes. The curves of the four algorithms and the actual power are shown in Figure 9.

It can be observed that when facing a relatively large and unevenly distributed dataset, LightGBM has a better prediction effect compared to CNN and SVM. This is because the LightGBM algorithm with Bayesian parameter tuning and five - fold cross - validation has stronger stability and higher accuracy, and will not fall into a local optimal solution state when facing unevenly distributed data. However, the single LightGBM algorithm cannot achieve the mining of dataset features. For the current CNN + LightGBM algorithm, although a convolutional neural network CNN is added on the basis of LightGBM to extract feature vectors, which improves the accuracy and stability of the model to a certain extent, this method focuses on studying the correlation and non - linear features of data at adjacent time points while ignoring the overall fluctuation characteristics of the data. At the same time, it lacks abnormal data processing, and the data mining by a single CNN network is still not ideal.

In the algorithm proposed in this paper, after GMM clustering, feature correlations can be mined, and outliers can be separated. Then, the ridge regression algorithm makes a preliminary prediction for each cluster. The predicted data and the data screened from the original dataset jointly form the final dataset, which is input into the LightGBM algorithm for the final prediction. In this way, feature information can be deeply mined, greatly improving the prediction accuracy.

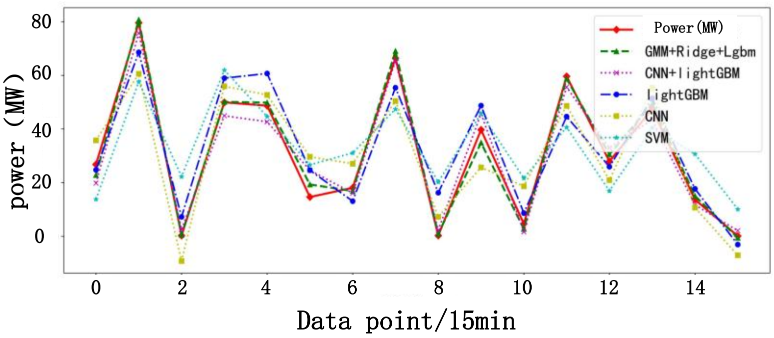

Figure 9 Prediction Curves of Five Methods

To visually reflect the accuracy of the prediction results, Figure 10 shows a comparison between the actual power (MW) data of a wind power plant in Inner Mongolia and the predicted power (MW) data. It can be observed that due to the uncertainty of weather changes, the actual power generation can fluctuate significantly in a short period. However, the model proposed in this paper still demonstrates a high prediction accuracy for the highly volatile wind power generation data.

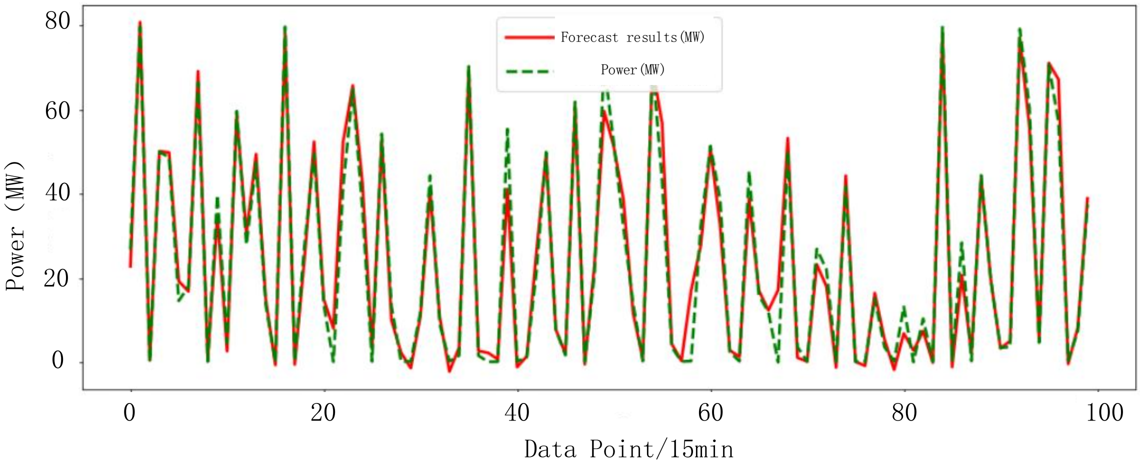

Figure 10 Prediction Curve and Actual Power

IV. RESULT

Firstly, this paper proposes a composite model based on GMM, Ridge Regression, and LightGBM. When dealing with highly volatile wind power data, it still exhibits good accuracy, addressing the problem of poor prediction accuracy of traditional AI prediction methods for non - linear data. Secondly, the algorithm proposed in this paper combines the GMM, Ridge Regression, and LightGBM algorithms. The GMM algorithm clusters and separates outliers, and then the Ridge Regression algorithm makes a preliminary prediction to reflect data characteristics. The prediction results are used as part of the dataset. Finally, the LightGBM algorithm is used for prediction. This approach can not only deeply explore the characteristics of the dataset but also solve the instability problem of a single model, improving the model's robustness. It is beneficial for improving the operation of the electricity market, reducing market costs, enhancing the stability of power grid operation, promoting the consumption of renewable energy, and improving the efficiency of power system planning and operation. These advantages can help achieve sustainable energy development goals, promote the application of clean energy, and the development of the electricity market. From the final results, although the algorithm proposed in this paper has good prediction accuracy and efficiency, it has a relatively long calculation time and requires a large amount of computing space. Therefore, improving the model's prediction speed and saving computing space should be the focus of future research.

This paper proposes a composite model based on GMM, Ridge Regression, and LightGBM. Through experimental verification, the following conclusions can be drawn:

(1) This paper proposes a data processing method of constructing numerical feature cross - terms and square - terms, and then using GMM clustering and Ridge Regression to fit the preliminary prediction curve. This method can establish associations between different features by multiplying numerical features pairwise, expanding the feature base. Then, the GMM algorithm clusters according to different time scales to reflect the relationship between various data and time. Finally, the Ridge Regression algorithm condenses the preliminary predicted values. This method can deeply explore the information of the dataset and avoid the dimensional explosion caused by inputting a large number of features into the model.

(2) To maintain the characteristics of the original dataset, this paper proposes another dataset construction method of merging new features statistically calculated in units of months, days, and hours with the original dataset and then performing feature importance screening.

REFERENCES

[1]. Wang Shen, Lv Lianhong, Zhang Chuai, etc Low cost carbon peak and carbon neutrality pathways in China based on multi-objective models [J]. Environmental Science Research, 2021,34 (09): 2044-2055. DOI: 10.13198/j.issn.1001-6929.2021.06.18

[2]. E Zhijun, Wang Guilin, Li Zhenbin, etc Optimization Control Method for Hybrid Energy Storage System to Improve the Consumption Level of New Energy Grid [J]. Journal of Electric Power System and Automation, 2021, 33 (03): 132-137. DOI: 10.19635/j.cnki.csu-epsa.000681

[3]. Ganley, W., Simple pendulum approximation. American Journal of Physics, 1985. **53**(1): p. 73-76.

[4]. Shinbrot, T., et al., Chaos in a double pendulum. American Journal of Physics, 1992. **60**(6): p. 491-499.

[5]. von Herrath, F. and S. Mandell, The Double Pendulum Problem. 2000.

[6]. Nunna, R. and A. Barnett, Numerical Analysis of the Dynamics of a Double Pendulum. 2009.

[7]. Callen, H.B., Thermodynamics and an Introduction to Thermostatistics. 1998, AAPT.
[8]. The Mathworks, 1994-2014 The MathWorks, Inc
[9]. http://www.mathworks.com/help/physmod/sm/ug/model-double-pendulum.html.
[10]. Wang Shen, Lv Lianhong, Zhang Chuai, etc Low cost carbon peak and carbon neutrality pathways in China based on multi-objective models [J]. Environmental Science Research, 2021,34 (09): 2044-2055. DOI: 10.13198/j.issn.1001-6929.2021.06.18
[11]. E Zhijun, Wang Guilin, Li Zhenbin, etc Optimization Control Method for Hybrid Energy Storage System to Improve the Consumption Level of New Energy Grid [J]. Journal of Electric Power System and Automation, 2021, 33 (03): 132-137. DOI: 10.19635/j.cnki.csu-epsa.000681
[12]. Ye Lin, Lu Peng, Zhao Yongning, etc Overview of Active Power Model Predictive Control Methods for Wind Power Systems [J]. Chinese Journal of Electrical Engineering, 2021, 41 (18): 6181-6198. DOI: 10.13334/j.0258-8013.pcsee.202599
[13]. Han Zifen, Jing Qianming, Zhang Yankai, etc Overview of Wind Power Forecasting Methods and New Trends [J]. Power System Protection and Control, 2019, 47 (24): 178-187. DOI: 10.19783/j.cnki. pspc. 2012
[14]. Ren Shuang, Yao University, Liu Junxiang, etc Short term wind power prediction based on combined mode decomposition and IGWO BiLSTM [J/OL]. Computer Measurement and Control, 1-9 [2020-02-18] http://kns.cnki.net/kcms/detail/11.4762.tp.20250212.1428.039.html.
[15]. Sun Rongfu, Zhang Tao, He Qing, etc Overview of Key Technologies and Applications for Wind Power Prediction [J]. High Voltage Technology, 2021, 47 (04): 1129-1143. DOI: 10.1336/j.1003-6520.hve.20201780
[16]. Xu Longbo, Wang Wei, Zhang Tao, etc Ultra short term wind power prediction based on average influence value of neural network [J]. Power System Automation, 2017, 41 (21): 40-45
[17]. Ravi Kumar Pandit,David Infield. Comparative analysis of Gaussian Process power curve models based on different stationary covariance functions for the purpose of improving model accuracy[J]. Renewable Energy,2019,140:
[18]. Zhao Pei, Dai Yeming SVM power load forecasting based on real-time electricity prices and weighted grey relational projection [J]. Grid Technology, 2020,44 (04): 1325-1332. DOI: 10.13335/j.1000-3673.pst.2019.0942
[19]. M. -S. Ko, K. Lee, J. -K. Kim, C. W. Hong, Z. Y. Dong and K. Hur, "Deep Concatenated Residual Network With Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting," in IEEE Transactions on Sustainable Energy, vol. 12, no. 2, pp. 1321-1335, April 2021.
[20]. Guo Min, Zhao Qiao'e, Wang Xianjun, etc Application of Optimization Particle Swarm K-means Clustering Algorithm in Wind Power Prediction [J]. Automation Technology and Application, 2017, 36 (08): 24-26+39
[21]. FENG C, CUI M, HODGE B M, et al. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting[J]. Applied Energy, 2017, 190: 1245-1257.
[22]. Yun Ju,Guangyu Sun,Quanhe Chen,Min Zhang,Huixian Zhu,Mujeeb Ur Rehman. A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting.[J]. IEEE Access,2019,7:
[23]. Wang Yimei, Liu Hui, Song Peng, etc Short term power prediction method for wind farms based on Gaussian mixture model clustering [J]. Power System Automation, 2021, 45 (07): 37-43
[24]. Y. Li et al., "GMM-HMM-Based Medium- and Long-Term Multi-Wind Farm Correlated Power Output Time Series Generation Method," in IEEE Access, vol. 9, pp. 90255-90267,2021.
[25]. M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi and F. S. Wesleti, "A Hybrid Bayesian Ridge Regression-CWT-Catboost Model For PV Power Forecasting," 2020 IEEE Kansas Power and Energy Conference (KPEC), 2020, pp. 1-5.
[26]. Wang Zhao, Wang Weisheng, Liu Chun, etc Feature nearest neighbor search probability prediction method for wind power [J]. Grid Technology, 2022, 46 (03): 880-887. DOI: 10.13335/j.1000-3673.pst.2021.0764
[27]. Huang Nantian, Zhao Wenguang, Cai Guowei, etc Efficient identification of LightGBM power quality disturbances constrained by data transmission rate in the Internet of Things [J]. Chinese Journal of Electrical Engineering, 2021, 41 (15): 5189-5201. DOI: 10.13334/j.0258-8013. pcsee. 200374
[28]. Li Xiaozhu, Wang Weiqing Application of Bayesian theory based distribution robust optimization in energy storage configuration [J]. Grid Technology, 2022, 46 (10): 4001-4011. DOI: 10.13335/j.1000-3673.pst.2021.1978
[29]. Peng Shurong, Zheng Guodong, Huang Shijun, etc Based on XGBoost algorithm fusion of multi feature short-term photovoltaic power generation prediction [J]. Electric Measurement and Instrumentation, 2020,57 (24): 76-83. DOI: 10.19753/j.issn1001-1390.2020.24.010