

Chiller load prediction based on CEEMDAN-BiLSTM-Attention model for sufficient data and small sample data cases

Xu Zhang[†]

2200467

(of Affiliation)

Software Engineering and Artificial Intelligence program

2200467@student.uwtsd.ac.uk

1. INTRODUCTION

1.1. Project Overview and Importance

HVAC system chiller load forecasting is of great significance and importance in modern buildings and industries [1]. The main purpose of this forecasting is to help ensure efficient operation of the system, reduce energy wastage, improve energy utilization, reduce operating costs and minimize adverse impacts on the environment. The significance and importance of chiller load forecasting for HVAC systems is described in detail below.

Chiller load forecasting is a crucial task in Heating, Ventilation and Air Conditioning (HVAC) systems and it is of great importance. There are several reasons why this program is of interest and worth the investment of time and resources:

Energy conservation and energy management: Chillers typically account for a significant portion of energy consumption in buildings. Load forecasting can help decision makers understand future demand and thus better plan energy supply and use. By reasonably forecasting loads, unnecessary unit operation can be avoided, energy waste can be reduced, and energy costs can be lowered, thus achieving the goals of energy conservation and sustainable energy management [2].

Improve equipment life: Frequent starting and stopping of chiller units has an adverse effect on equipment life [3]. Load forecasting allows for better management of unit operation, reducing the number of starts and stops, extending equipment life, and reducing the cost of equipment maintenance and replacement.

Provide Comfort: In commercial and residential buildings, providing a comfortable indoor environment is critical to the productivity and quality of life of employees and residents. Accurate load forecasting ensures that proper indoor temperatures and humidity are maintained in all weather conditions [4], improving indoor comfort and providing a better indoor working and living environment.

Cost Savings: Chiller operation and maintenance costs are an important concern for building managers and owners [5]. Through load forecasting, it is possible to better plan the running time of the unit, avoid the increase of energy cost during peak period, optimize the running efficiency of the unit, reduce the maintenance cost, and achieve long-term economic savings.

Environmental protection: Energy consumption is closely related to environmental protection [6]. Through load forecasting, unnecessary energy consumption can be reduced, lowering the carbon footprint and helping to slow down climate change and environmental damage.

Why choose chiller load forecasting as a project?

Complexity: Chiller load forecasting is a complex project that requires a combination of factors to be considered, such as meteorological data, building characteristics, and people's activities. This comprehensive complexity makes the project a challenging one that attracts the interest of engineers and scientists.

Potential energy savings: Chiller energy consumption is a significant part of the building sector, and significant energy savings can be realized through load forecasting [7]. This is a compelling goal for both businesses and governments to reduce energy costs and improve competitiveness in the long term.

Technological innovation: chiller load forecasting requires the use of advanced technologies and methods [8], including data analytics, machine learning, and artificial intelligence. This project provides a broad field for technological innovation and promises to develop more efficient and accurate forecasting models.

Social impact: Improving the energy efficiency of buildings is part of a sustainable society [9]. Through chiller load forecasting, greenhouse gas emissions can be reduced, environmental pollution can be minimized, and society and the environment can be positively impacted.

Building management: chiller load forecasting can help building managers better manage and maintain equipment, provide a better indoor environment, reduce equipment damage and downtime, and improve building management efficiency [10].

To summarize, HVAC system chiller load forecasting is of great significance and importance in the building and industrial sectors. It not only helps to save energy and improve energy efficiency, but also ensures system stability and reliability, reduces costs, improves comfort and indoor air quality, protects the environment, and better manages

maintenance and spare parts. Load forecasting is therefore a key task not only for building owners and managers, but also has a positive impact on society and the environment as a whole.

1.2. Why chooses this program

Based on accurate load forecasting I would like to subsequently achieve the goals of energy allocation based on the correct chiller load, energy saving optimization, and ensuring sufficient cooling capacity, so finding an accurate load forecasting method is very important and a fundamental step.

These points will be expanded in detail below to explain the importance and application of this sentence.

Basic Concept of Load Forecasting: Load forecasting for chillers is the process of predicting the cooling load for a future period of time by analyzing information [11] such as historical data, meteorological conditions, building characteristics, and user demand. This load typically includes cooling and heating demand, as well as ventilation demand. Accurate load forecasting is the foundation of building energy management, and it provides an important basis for rational energy allocation and use.

Energy saving and energy allocation: Load forecasting allows building managers to know the future energy demand and thus optimize the chiller operation schedule [12]. This helps to avoid over-supply or under-supply of cooling energy and reduces unnecessary energy consumption, thus realizing the goals of energy saving and lower operating costs. In addition, based on accurate load forecasts, chiller starts and stops can be rationally arranged to reduce the energy waste caused by frequent starts and stops.

Environmental Sustainability: Optimizing energy distribution and use not only saves costs, it also helps reduce greenhouse gas emissions and has a positive impact on the environment [13]. Accurate load forecasting enables buildings to control cooling and heating more precisely, reducing unnecessary carbon emissions and supporting sustainable building and environmental goals.

Ensure adequate cooling capacity: Chillers are under tremendous stress during the hot season and high load periods [14]. Load forecasting ensures sufficient cooling capacity, prevents overloading of the units, reduces the risk of mechanical failure and extends equipment life, while ensuring a comfortable indoor environment.

Improve system reliability: Accurate load forecasting helps to improve the reliability of the chiller system [15]. Reasonable operation planning and maintenance scheduling can reduce the failure rate of the unit, minimize emergency repairs and downtime, and ensure continuous system operation without disrupting business and life.

User satisfaction: The comfort of the indoor environment directly affects user satisfaction. Through load forecasting, it is possible to maintain constant indoor temperature and humidity under different meteorological conditions [16], providing a better indoor environment that meets user needs and improves user satisfaction, whether in commercial, industrial or residential premises.

Data-driven decision-making: Load forecasting relies on extensive data analysis and modeling. This data-driven approach helps building managers make informed decisions about equipment investments [17], maintenance programs, energy purchases, and more. These decisions will impact the long-term performance and economics of the building.

Technological innovation: The field of load forecasting continues to evolve, with new technologies and methods emerging, including prediction models based on artificial intelligence and machine learning [18]. This provides researchers and engineers with opportunities for innovation that can improve forecast accuracy and efficiency.

Overall, accurate chiller load forecasting is a multi-benefit task that not only contributes to energy efficiency, cost reduction, and environmental protection, but also improves indoor environmental quality, enhances system reliability, increases occupant satisfaction, and drives technological innovation. Therefore, building managers and engineers should place a high priority on load forecasting and continually improve their methods to achieve more sustainable, efficient and comfortable building operations.

1.3. How the objectives will be achieved and what methods will be used:

I intend to use CEEMDAN-BiLSTM-Attention approach for load prediction, firstly, the original data is decomposed using CEEMDAN, each component of the decomposition is substituted into the BiLSTM prediction model, and weights are added to the model parameters through the Attention mechanism to increase the accuracy of the prediction.

The EMD algorithm [19] is a signal decomposition method used to decompose a complex signal into a set of Intrinsic Mode Functions (IMFs), each IMF represents an intrinsic feature of the signal. The basic idea of EMD is to obtain the IMFs by continuously extracting the local extremes and minima in the signal. Then, these IMFs can be used to analyze the frequency domain characteristics and time domain characteristics of the signal.

The EEMD algorithm [20] is an improvement of EMD, which mainly solves the stability problem of EMD when dealing with noisy signals. EEMD introduces randomness, and obtains multiple sets of IMFs by adding different random noises to the original signal several times, and then averages these IMFs in order to minimize the effect of random noises. This approach improves the performance of EMD on noisy signals.

The CEEMDAN algorithm is a further improvement of EEMD and is mainly used to solve the problems of EEMD in dealing with non-smooth and nonlinear signals [21]. CEEMDAN introduces adaptive noise and no longer relies on

predefined random noise. It generates noise based on the local characteristics of the signal and adaptively adjusts it in each iteration. This approach improves the performance of EEMD on nonlinear and nonsmooth signals.

Analyze and compare the above three decomposition algorithms:

EMD: The original EMD algorithm is prone to modal aliasing problems when dealing with noisy signals, resulting in inaccurate IMFs. It does not take into account the nonlinearity and nonsmoothness of the signal.

EEMD: EEMD improves EMD by introducing randomness, which improves the stability of processing noisy signals. However, it still cannot effectively handle nonlinear and nonsmooth signals.

CEEMDAN: CEEMDAN further improves the ability to handle nonlinear and nonsmooth signals by introducing adaptive noise and adjusting it to the local characteristics of the signal. This makes CEEMDAN a powerful tool for processing various types of signals.

Overall, EMD is a basic signal decomposition method but has some limitations when dealing with complex signals. EEMD improves the robustness of EMD by introducing stochasticity, while CEEMDAN further improves the performance in dealing with nonlinear and nonsmooth signals. The choice of method depends on the characteristics of the signal and the goal of the analysis. For highly nonlinear and nonsmooth signals, CEEMDAN is usually a better choice but requires more computational resources.

RNN (Recurrent Neural Network) [22] is an important algorithm in the field of deep learning, but it has some limitations, such as difficulty in capturing long-range dependencies and the gradient vanishing problem. To overcome these limitations, researchers have proposed more complex models such as LSTM (Long Short-Term Memory Network) and BiLSTM (Bidirectional Long Short-Term Memory Network). In the following article, we will analyze these algorithms step by step and explain why the BiLSTM algorithm was chosen to cope with the low repetition rate problem.

RNN is a sequence model which is very useful when dealing with time-series data. The basic idea of RNN is to capture sequence information by passing the hidden state of the previous time step as input to the current time step. However, RNN has one major problem, which is the difficulty in capturing long distance dependencies. This is because the gradient tends to vanish or explode in backpropagation, resulting in a model that is unable to effectively learn information in long sequences.

To solve the problem of RNN, LSTM introduces gating mechanisms [23], these gating units are responsible for controlling the flow of information. LSTM consists of forgetting gates, input gates, and output gates, which learn which information should be forgotten, and which information should be stored and passed on. This structure allows LSTMs to better capture long-distance dependencies and is less susceptible to the gradient vanishing problem.

While LSTM performs well on certain tasks, it still has the limitation that it can only see past information. In some applications, we need to consider future information, and this is where BiLSTM is introduced [24]. BiLSTM captures both past and future information by running two LSTM networks simultaneously at each time step, one from the front to the back and the other from the back to the front. This allows BiLSTM to provide a more comprehensive understanding of the sequence data and helps to improve task performance.

The main reason for choosing BiLSTM algorithm is to cope with the low repetition rate problem. In many tasks, such as sentiment analysis, named entity recognition, load forecasting, etc., bi-directional information extraction from data is a challenge. Bidirectional information extraction captures more patterns and regularities in time series data.

BiLSTM can better handle the work of bi-directional data feature extraction because it can consider the information of the front and back time series at the same time, and even if there are overlooked areas of the forward information, they can be captured by the backward information. In addition, BiLSTM can learn both local and global information in time series data, which helps to perform tasks such as classification and prediction more accurately.

In conclusion, RNNs, although fundamental to sequence data processing, perform poorly when faced with the problems of long distance dependence and low repetition rate. LSTM introduces a gating mechanism to address these problems, while BiLSTM further enhances the model's capabilities, especially for applications dealing with bidirectional extraction of data information. The main reason for choosing the BiLSTM algorithm is its comprehensiveness and robustness to better accommodate diverse textual data, and the same can be applied innovatively for load prediction.

Bidirectional Long Short-Term Memory Network (BiLSTM) is a deep learning model for processing sequential data that has the potential for a wide range of applications in natural language processing and other domains. When combining the Attention mechanism with BiLSTM, some significant benefits can be realized.

Bi-directional information extraction capability: BiLSTM itself already has some bi-directional information extraction capability ability to process sequence data by considering both past and future information. However, the addition of the Attention mechanism can further improve the model's context-awareness capabilities. The Attention mechanism allows the model to selectively focus on different parts of the input sequence at each time step, thus better capturing important information about the time-series data. This helps to improve the model's ability to understand and model the input data.

Processing Long Sequences: When processing long sequential data, BiLSTM may encounter problems such as gradient vanishing or gradient explosion, resulting in performance degradation. The Attention mechanism can help to solve these problems as it allows the model to focus on a portion of the sequence that is relevant to the task at hand.

This allows BiLSTM to focus on the information that is most important for the current time step without having to process the entire sequence at the same time.

Reducing Information Loss: Traditional pooling operations (e.g., average pooling or maximum pooling) often result in information loss because they discard some information from the input sequence. The Attention mechanism allows the model to dynamically weight different parts of the sequence before pooling, thus reducing information loss and allowing the model to better retain critical information from the input sequence.

Noise Suppression: When there is noise or irrelevant information in the input sequence, the Attention mechanism can selectively suppress this information without letting it negatively affect the model's output. This helps to improve the robustness and performance of the model.

Adaptability: the combination of BiLSTM and Attention mechanism makes the model more adaptable. It can adapt to different prediction task requirements by dynamically allocating attention based on different parts of the input data. This adaptability makes the BiLSTM-Attention model perform well in load prediction tasks.

In summary, combining the Attention mechanism with BiLSTM can significantly improve the model's performance in sequence data processing tasks. This combination makes the model more context-aware, better able to process long sequences, align information, learn semantic representations, reduce information loss, suppress noise, and provide greater adaptability. As a result, BiLSTM-Attention has become a powerful tool in the field of natural language processing, widely used in tasks such as text categorization, named entity recognition, machine translation, load prediction, etc., and has helped researchers and practitioners make significant progress.

The attention mechanism is an important technique in deep learning that improves the performance of models in areas such as processing sequential data, natural language processing, and computer vision by allowing them to dynamically attend to different parts of the input. The core idea is to determine the relative importance of input data by calculating weights that can be dynamically adjusted according to task and context. The attention mechanism has become a standard tool in the field of deep learning, providing powerful modeling capabilities for solving a variety of complex tasks, and is one of the key factors driving the continuous development of the AI field.

Meanwhile, for small-sample data, the neural network freezing layer approach has been used in the past to construct a migration learning model to improve the accuracy of prediction for small-sample data. Due to the CEEMDAN-BiLSTM-Attention method's decomposition of data and the enhancement of neural networks, the prediction effect of using the migration learning model is compared with the CEEMDAN-BiLSTM-Attention method to observe whether the CEEMDAN-BiLSTM-Attention method also has a superiority when it comes to small sample data. has superiority.

1.4. What I've learned

I learned the CEEMDAN-BiLSTM-Attention method, which combines the three key techniques of cyclic empirical modal decomposition (CEEMDAN), bi-directional long and short-term memory network (BiLSTM), and Attention mechanism (Attention), and meanwhile learned the following important information and capabilities in the process of applying the above techniques:

Temporal feature extraction and denoising: CEEMDAN, as the first step, can effectively extract temporal features in sequence data. With CEEMDAN, the model can learn the local patterns, trends and periodicity of the signal and separate the useful signal components from the noise, thus improving the data quality.

Contextual dependencies: BiLSTM introduces the modeling of contextual dependencies, where the model is able to learn the dependencies between different time steps in the sequence data. This allows the model to better understand long-range dependencies in sequences and thus predict future data points or events more accurately.

Bidirectional information fusion: BiLSTM learns the ability to fuse bidirectional information by processing both forward and reverse sequence information. This means that the model can more fully consider information from different directions in the sequence data, improving contextual understanding and modeling.

Importance weighting: the Attention mechanism allows the model to dynamically focus on different parts of the sequence data as it is processed, learning to weight the importance of different elements in the sequence. This allows the model to focus more on key information in the sequence, improving performance in prediction and classification tasks.

Multi-scale feature extraction: CEEMDAN can decompose the signal into components of different frequencies, and the BiLSTM and Attention mechanisms can handle features of different scales simultaneously. This allows the model to understand data at multiple levels, thus better adapting to changes in different data scales.

TIME-SCHEDULED DATA MODELING: The CEEMDAN-BiLSTM-Attention approach can better model time-series data when integrating these techniques. It is able to capture the time-varying, periodic, and anisotropic nature of the data, and is suitable for a variety of tasks that require analysis, prediction, or classification of time-series data.

Removal of redundant information: Thanks to the introduction of the Attention mechanism, the model can dynamically select the parts to focus on, thus reducing the reliance on redundant information. This helps to improve the generalization performance of the model and reduce the risk of overfitting.

Adaptive tuning: the CEEMDAN-BiLSTM-Attention method is adaptive and is able to dynamically adjust its internal parameters and weights according to different features and contexts of the input data. This allows the model to adapt to different types of sequence data without the need to manually adjust the hyperparameters.

In summary, the power of the CEEMDAN-BiLSTM-Attention approach lies in its ability to integrate different techniques to better understand, extract, and utilize information in sequence data. This approach is applicable to a variety of application areas, including time series prediction, signal processing, natural language processing, computer vision, etc., providing a powerful tool for data analysis and prediction tasks. By learning capabilities such as temporal features, context dependency, and importance weighting, the CEEMDAN-BiLSTM-Attention method can improve the performance and efficiency of the model in sequence data processing tasks.

2. LITERATURE REVIEW

Chiller load forecasting is one of the key issues in modern buildings and industries. With the increase in energy efficiency and sustainability requirements, as well as the increase in complexity of power systems, load forecasting methods have been improved and evolved. In this process, a variety of methods have emerged, including CEEMDAN, BiLSTM, and Attention mechanisms. Chiller load forecasting is a key research area that involves several fields such as energy management, environmental protection, and equipment operation and maintenance. Researchers have addressed this problem through different methods and techniques, and the following are some common research methods and research directions that demonstrate how people have studied the process of chiller load forecasting.

2.1. Others' research on chiller load forecasting

Load forecasting based on empirical methods: early load forecasting methods were mainly based on rules of thumb and rules. These methods relied on historical data and manual rules, e.g., estimating future loads based on the previous year's data for the same dates. Although these methods are simple, they lack accuracy and flexibility.

Time series analysis methods: The accuracy of load forecasting has improved with the rise of time series analysis methods such as moving average, exponential smoothing and ARIMA (Autoregressive Integrated Moving Average Model). These methods can capture seasonality and trends in the data, but have limited ability to model non-linear relationships.

Introduction of machine learning methods: The development of machine learning methods has opened up new opportunities for load forecasting. Methods such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are beginning to be applied to load forecasting. These methods are better able to capture nonlinear relationships, but have a high dependence on feature engineering.

Rise of Deep Learning Methods: load forecasting has made significant progress with the rise of deep learning methods such as Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM). These methods are able to process sequential data and capture temporal dependencies, but there are challenges in modeling long sequences.

Studies based on statistical methods: early chiller load forecasting studies mainly used statistical methods, such as time series analysis, regression analysis, and cluster analysis. These methods rely on the statistical properties of historical data for load forecasting. For example, seasonality and trends in loads can be captured using time series methods such as ARIMA (Autoregressive Integrated Moving Average Model).

Feature engineering and data preprocessing: regardless of which method is used, good feature engineering and data preprocessing are key to success. Researchers need to select appropriate features and deal with missing data and outliers to ensure the quality of the input data.

Multi-model fusion: some studies use multi-model fusion to combine different predictive models to improve the overall predictive performance. This can be achieved through techniques such as voting, weighted averaging or stacking.

Data-driven and real-time: with advances in sensor technology, more and more real-time data is available for load forecasting. Researchers are focusing on how big data and real-time data can be utilized to improve forecasting models to more accurately reflect current load conditions.

Interpretability and visualization: Load forecasting requires not only high accuracy but also model interpretability. Researchers are concerned with how to interpret the model's prediction results and help users understand load forecasting through visualization tools.

The CEEMDAN-LSTM hybrid model [25] is a much-anticipated study in the field of time series analysis that combines the features of Empirical Mode Decomposition (EMD) and Long Short-Term Memory (LSTM) networks for processing and forecasting time series data. This hybrid model has attracted a lot of attention. This hybrid model has attracted extensive research interest and has been thoroughly studied and explored by researchers to explore its applications and potential advantages in different fields. The following is a detailed description of a study on CEEMDAN-LSTM hybrid model by others.

The study of CEEMDAN-LSTM hybrid model is a cutting-edge work in the field of time series analysis, which combines two important techniques, i.e., empirical modal decomposition (EMD) and long-short-term memory network (LSTM). EMD is a data-driven signal decomposition method that can decompose a time series into a set of intrinsic modal functions (IMFs) that have different frequency characteristics. LSTM is a deep learning model for sequence data

with powerful sequence modeling capabilities. Combining them together, the CEEMDAN-LSTM hybrid model aims to capture the complex features and nonlinear dynamics of time series data.

Time series analysis is a central issue in several fields, including finance, meteorology, energy management, healthcare, and so on. Traditional time series analysis methods often struggle to deal with nonlinearity and nonsmoothness, and the CEEMDAN-LSTM hybrid model attempts to address these challenges. As a result, researchers have developed a strong interest in its performance and application potential.

Researchers have conducted extensive studies on CEEMDAN-LSTM hybrid models and have achieved a number of important research results covering different areas and applications. Some of the major research results are summarized below:

1. Load Forecasting

In power system management, researchers use CEEMDAN-LSTM hybrid models to forecast load demand. By applying the model to historical load data, they are able to predict future load demand more accurately, which is critical for planning and optimizing power supply. In addition, the model is able to handle nonlinearities and non-stationarity in the load data, making it excellent for practical applications.

2. Financial Time Series Analysis

In the field of finance, researchers use a hybrid CEEMDAN-LSTM model to analyze stock prices, exchange rates, and other financial time series data. Such models are better able to capture the complexity and nonlinear characteristics of markets, thus improving the accuracy of financial forecasts. Some studies have also explored combining sentiment analysis with this model to improve the modeling and prediction of market sentiment.

3. Healthcare

In the field of healthcare, the CEEMDAN-LSTM hybrid model is used to analyze biomedical signals such as electrocardiograms (ECGs) and electroencephalograms (EEGs). The model can help physicians better understand and diagnose patients' conditions, especially for those suffering from heart disease or neurological disorders. In addition, some studies have used the model to monitor patients' physiological conditions to detect and intervene in health problems in advance.

4. Environmental Monitoring

In environmental science, the CEEMDAN-LSTM hybrid model is used to process meteorological and climatic data in order to better predict natural disasters such as hurricanes, floods, and droughts. The ability of the model to capture seasonal and long-term trends in meteorological data provides a powerful tool for risk management and emergency response.

5. Strengths and potential applications

Research on the CEEMDAN-LSTM hybrid model has shown that it has the following advantages and potential applications:

Nonlinear modeling: the model is able to handle nonlinear relationships in time series data, which is critical in many practical problems.

Time-series modeling: The LSTM part allows the model to capture the dynamics of time-series data, which is suitable for problems that need to take into account temporal dependencies.

Prediction performance: Studies have shown that the CEEMDAN-LSTM hybrid model achieves good prediction performance in several domains, outperforming traditional methods.

Wide range of applications: The model has a wide range of potential applications in different domains, including power, finance, healthcare and environmental monitoring.

Data-driven: the model adopts a data-driven approach, which can adapt to different types of time series data without the need to force assumptions on the data beforehand.

Therefore, the CEEMDAN-LSTM hybrid model is a compelling approach to time series analysis that has achieved remarkable research results in several fields. It combines the advantages of empirical mode decomposition and deep learning to better capture the nonlinear and dynamic features of time series data. In the future, researchers can further explore the improvement and application of the model to solve more complex practical problems, improve forecasting accuracy, and advance the field of time series analysis.

2.2. Superiority of the CEEMDAN-BiLSTM-Attention method

CEEMDAN development process:

The evolution of Cyclic Empirical Modal Decomposition with Adaptive Noise (CEEMDAN) can be traced back to Empirical Modal Decomposition (EMD) and Improved Empirical Modal Decomposition (EEMD). The following is the evolution of CEEMDAN from EMD evolution to EEMD to CEEMDAN:

1. Empirical Modal Decomposition (EMD):

EMD is a method for signal processing and data decomposition, which was first proposed by Huang Jianzhong in 1998. Its main idea is to decompose a signal into a series of components called Intrinsic Modal Functions (IMFs), where each IMF represents a different frequency component of the signal. The process of EMD consists of the following steps:

Connect the extreme maximum and minimum points of the signal to form the upper and lower envelopes.

Calculate the average of the signal with the upper and lower envelopes to obtain a set of local extreme points.

Interpolate the signals between the local extreme points to obtain the first round of IMF.

Repeat the above steps until an IMF satisfying a certain stopping condition is obtained.

However, the traditional EMD method has some problems, such as unstable number of modal functions and noise sensitivity.

2. Improved Empirical Modal Decomposition (EEMD):

In order to overcome some of the problems of EMD, the improved empirical modal decomposition (EEMD) was introduced by Wu and Huang in 2005. The main improvement of the EEMD is the introduction of white noise, which stabilizes the number of modal functions by adding different random noises several times. The core idea of the EEMD is to introduce different random noises in each decomposition, which mitigates the problem of unstable number of modal functions. problem.

3. Cyclic Empirical Modal Decomposition with Adaptive Noise (CEEMDAN):

CEEMDAN is a further improvement of EEMD and was proposed by Rato et al. in 2008. It introduces adaptive noise so that the energy of the noise decreases as the decomposition proceeds. This allows better handling of the noise component of the signal and improves the quality and stability of the decomposition.

The steps of CEEMDAN include:

Summing the signal with the gradually decreasing adaptive noise.

Decompose the noise-added signal using the EEMD method to obtain the IMF.

Repeat the above steps to gradually reduce the energy of the noise until the stopping condition is satisfied.

The introduction of CEEMDAN makes the signal decomposition more stable and reliable, especially in the presence of noise. This makes CEEMDAN widely used in the fields of signal processing, data analysis and feature extraction, providing a powerful tool for higher-level data processing and analysis.

In summary, the development process of CEEMDAN from EMD to EEMD to CEEMDAN is a continuous improvement and optimization of empirical modal decomposition methods to make them more stable and reliable in practical applications. The introduction of CEEMDAN has had a far-reaching impact in the field of signal processing and data analysis, and has provided better solutions for handling complex signals and data.

The development process of BiLSTM:

The development of Bidirectional Long and Short-Term Memory Networks (BiLSTM) can be traced back to Recurrent Neural Networks (RNN) and Long and Short-Term Memory Networks (LSTM), and the following is the evolution of BiLSTM from RNN to LSTM to BiLSTM:

1. Recurrent Neural Network (RNN):

RNN is one of the earliest neural networks used for sequence data processing, and its basic idea is to introduce recurrent connections to make the network memorized and able to deal with time-dependent relationships in sequence data. However, traditional RNNs suffer from problems such as gradient vanishing and gradient explosion, which lead to difficulties in capturing dependencies in long sequences, limiting their application in practical tasks.

2. Long Short-Term Memory Network (LSTM):

In order to overcome the problems of RNNs, Long Short-Term Memory Networks (LSTMs) were introduced. LSTMs are a variant of RNNs with a gating mechanism, which was proposed by Hochreiter and Schmidhuber in 1997. LSTMs introduce three kinds of gating units: forgetting gates, input gates, and output gates, in order to control the flow of information and the preservation of memories. These gating mechanisms enable LSTM to better capture dependencies in long sequences while effectively preventing the gradient problem.

3. Bidirectional Long Short-Term Memory Network (BiLSTM):

BiLSTM is developed on the basis of LSTM and its main feature is the introduction of bidirectional structure. While traditional LSTMs consider only past historical information at each time step, BiLSTM considers both past and future information. BiLSTM consists of two independent LSTM networks, one for forward propagation and the other for back propagation. The outputs of these two networks are concatenated together at each time step, providing more comprehensive contextual information.

Advantages of BiLSTM include:

More comprehensive contextual understanding: BiLSTM is able to consider both past and future information, giving the model a more comprehensive understanding of the context of sequence data, which is particularly suitable for natural language processing tasks and sequence modeling.

Solving the long-term dependency problem: With the introduction of the bi-directional structure, BiLSTM is able to better capture dependencies in long sequences, which helps to improve the performance of the model.

Better adaptability: BiLSTM is not only suitable for fixed-length sequences, but can also handle variable-length sequences, as it generates an output at each time step and finally aggregates these outputs into a sequence representation.

In summary, the development of BiLSTM is a natural extension of RNN and LSTM, providing a more powerful tool for processing sequence data. Its introduction of a bidirectional structure enables the model to better capture contextual information and long-term dependencies, and has been widely used in the fields of natural language processing, speech recognition, and time series analysis, providing a powerful solution for modeling and predicting sequence data.

Add the superiority of the Attention mechanism:

The introduction of Attention mechanism into BiLSTM (Bidirectional Long Short-Term Memory Network) neural network has significant superiority in load forecasting tasks. This combination can help to improve model performance and achieve more accurate load forecasting, and it is useful for handling highly dynamic and complex load data. This advantage is explained in more detail below, including the role and benefits of the Attention mechanism.

Capturing key information in the time series: The Attention mechanism allows the BiLSTM network to dynamically assign different weights to individual time points in the input sequence at each time step. This means that the network is able to focus on the time periods that are most important for load forecasting, rather than treating the entire historical data equally. This is critical for capturing key information and trends in the load data, especially if the load data has low repeatability, as the data may exhibit large variations across time.

Adaptive modeling: Load data may be affected by a variety of factors, such as seasonal variations, weather conditions, holidays, etc. The introduction of the Attention mechanism allows the BiLSTM network to adaptively adjust its focus to better accommodate changes in these external factors. This allows the model to better capture the changing patterns of load data over time and improves the model's ability to generalize, especially when faced with low repetitive data.

Reduced information redundancy: Load data often contains a great deal of redundant information, especially over long time spans. The Attention mechanism helps the model identify and focus on those points in time that are most valuable for load forecasting, thus reducing the burden on the network to process redundant information. This improves the computational efficiency of the model and reduces the risk of overfitting.

Handling anomalies: In load forecasting, anomalies such as sudden load increases or decreases can have a significant impact on accuracy. The Attention mechanism helps the model to be more flexible in adjusting the weights under these anomalies, thus better capturing these sudden events and improving the robustness of the model.

Interpretability: Introducing the Attention mechanism increases the interpretability of the model as we can analyze the attention weights assigned by the model on different time steps. This helps to understand why the model makes specific decisions in load forecasting and provides more insight into decision making.

Overall, incorporating the Attention mechanism into the BiLSTM neural network improves the accuracy, adaptability, and generalization of load forecasting, especially when dealing with load data with very low repetition rates. This combination allows the model to better understand important patterns and trends in time-series data, which helps power system managers to more accurately plan and manage power supply, thereby improving energy efficiency and reliability.

Migration methods for small sample data:

In the field of load forecasting, in the face of small sample data, a migration learning strategy using frozen layers is a powerful approach [26] that can effectively improve model performance and reduce the risk of overfitting. In this strategy, we utilize a pre-trained model trained on a large-scale load dataset and then freeze some of the model layers to fit the small sample data. The advantages and implementation steps of this strategy are explained in detail below.

1. Background of Load Forecasting

Load forecasting is an important task in power system management and involves predicting the demand for electricity over a period of time in the future for better planning of generation and transmission. Load data typically consists of time-series information, such as hourly or minute-by-minute electricity usage. In some cases, the available historical data may be very limited, which requires us to use migration learning methods to solve the problem of small sample data.

2. Migration Learning and Freezing Layer Strategy

Migration learning is a machine learning approach in which a model that has been trained on one task (called the source model) is adapted to perform better on a new task. In the case of small sample data, traditional deep learning models may be prone to overfitting because they require a large amount of data to tune a large number of parameters. Therefore, migration learning strategies are particularly useful.

The frozen layer strategy is a migration learning technique that centers on the idea of freezing (or partially freezing) layers of an already trained model on top of the source model so that they are no longer updated. The reason for this is that the underlying layers of the source model usually learn generic features, whereas in a new task we are more concerned with task-specific features of the data. By freezing the underlying layers, we can avoid destroying the generic features already learned by the source model and thus better adapt to small sample data.

3. Implementation steps

The following are the specific steps for using the migration learning strategy of freezing layers for load prediction models with small sample data:

Step 1: Select a source model

Select a source model that has been trained on a large-scale load dataset, typically a deep neural network such as a convolutional neural network (CNN) or recurrent neural network (RNN). This source model should have some performance on a relevant domain (e.g., time series prediction).

Step 2: Model restructuring

Build a new target model based on the source model according to the requirements of the new task. Typically, you will need to add some additional layers that will be fine-tuned in the new task to capture specific patterns in the load data. These additional layers can include fully connected layers, LSTM layers, or other appropriate hierarchies.

Step 3: Freeze Source Model Layers

In the target model, freeze a portion of the source model's layers, usually the underlying hierarchy. The weights of these layers will remain unchanged during training and will not be updated. The frozen layers should include generic features learned from the source model so that they remain constant, while the top layers of the model can be fine-tuned for new tasks.

Step 4: Training and fine-tuning

The entire target model is trained using a small sample of data. Since the bottom level hierarchy has been frozen, the model will focus on learning specific features and relationships that apply to the new task. This will help reduce the risk of overfitting because the model will not be overly dependent on a limited amount of small sample data.

Step 5: Evaluate and Adapt

Evaluate the performance of the target model on the test data and fine-tune and optimize the model as needed. This iterative process can be repeated as many times as necessary to achieve optimal performance.

4. Superiority and benefits

The migration learning strategy using a frozen layer has the following superiority and benefits for load prediction on small sample data:

Leveraging source model knowledge: the strategy leverages the knowledge of source models trained on large-scale load data. This means that the model is able to derive valuable generalized features from the source domain rather than learning from scratch.

Reduces the risk of overfitting: By freezing the underlying layers, the model is easier to generalize to small sample data, reducing the overfitting problem that occurs with small samples.

Accelerates training: freezing layers reduces training time because only some of the layers need to have their weights updated. This is especially important for small sample data as it reduces computational costs.

Improve performance: By fine-tuning the top layer hierarchy on the target task, the model is able to adapt to the specific requirements of the new task, thus improving the performance and accuracy of load forecasting.

Adaptability: This strategy allows the model to remain adaptive to the source domain while facing small samples of data. This is useful for load forecasting in power system management, where the model needs to adapt to changing power demand.

So, freeze Layer Migration Learning is a commonly used technique for training deep neural network models in which some of the model layers are left unchanged while only some of the specific layers are trained or fine-tuned. In this process, the weight parameters of the frozen layers are not updated, while the parameters of only some of the layers are tuned to the new task. This approach is commonly used to apply a model trained on one task to a related task to accelerate model training and improve performance. By freezing the layers, the learning from the previous task can be preserved while adapting to the specific requirements of the new task, leading to more efficient transfer learning. A classic example is for example the CNN-LSTM migration model [26], which can be used to solve the sample lack problem. However, freezing layer migration learning requires a large number of raw datasets from other related buildings for model training, which may lead to poor prediction of the migration model when the correlation between the raw dataset and the target dataset is not strong or when there is a lack of raw datasets. Therefore, it is also important to find a prediction method that is directly applicable to small sample datasets.

Based on the above research, this paper combines CEEMDAN, BiLSTM and Attention mechanism does have some feasibility to increase the load forecasting accuracy, the advantages of this combination can be reflected in the following two aspects:

Multi-level feature extraction: CEEMDAN provides time-frequency features, BiLSTM modeling extracts bi-directional feature learning, and the Attention mechanism focuses on the important parts, allowing the model to extract and utilize features at multiple levels.

Adaptive modeling: the combination of these three approaches makes the model more adaptive and able to handle a wide range of types and complexity of sequence data.

High performance: this combination improves the performance of load forecasting, enabling more accurate prediction of future loads and improved energy utilization efficiency.

In summary, the CEEMDAN-BiLSTM-Attention method represents the latest development trend in the field of load forecasting, which combines the key techniques of feature extraction, context modeling, and importance weighting, enabling load forecasting to be more accurate and flexible in dealing with different challenges. Meanwhile, experiments show that the CEEMDAN-BiLSTM-Attention method also works well for migration learning using frozen layers when facing small sample data. This method has important application potentials in improving energy efficiency, optimizing resource allocation and reducing energy costs, which makes an important contribution to the future smart energy system and sustainable development.

3. RESEARCH METHODS

3.1. CEEMDAN

CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) is a method for signal processing and data analysis. It is an adaptive signal decomposition technique used to decompose a complex signal into a set of intrinsic mode functions (IMFs) and residual terms to better understand the components of the signal.

The steps to run CEEMDAN include:

Preparing the data: first, the signal data to be decomposed needs to be prepared. This can be time series data, images, audio or any continuous data.

Modal Function Decomposition: CEEMDAN goes through a series of steps to decompose the signal into IMFs, each of which is a component of the signal with a different frequency and amplitude, typically used to represent a localized characteristic of the signal.

Noise Modal Function Decomposition: In CEEMDAN, the IMF of the noise is also estimated and extracted. this helps to better understand the noise component of the signal and to separate it from the actual components of the signal.

Residual term extraction: after the first two steps, the remaining part is usually the residual term, which contains the components of the signal that are not represented by the previous IMF and noise modal function.

Repeat Steps: CEEMDAN is usually an iterative process by repeatedly applying steps 2 through 4 in order to better decompose the signal until a satisfactory decomposition is achieved.

Analysis and application: Once the signal is successfully decomposed into IMFs and residual terms, these components can be further analyzed or applied. For example, the IMF can be utilized for tasks such as feature extraction, time-frequency analysis, or signal recovery.

In conclusion, CEEMDAN is a method that helps to decompose complex signals into their essential components, which can be used in a variety of application areas such as signal processing, data analysis and feature extraction. Its adaptive nature enables it to efficiently handle different types of signals and provide deeper insights.

Figure 1 shows the CEEMDAN decomposition:

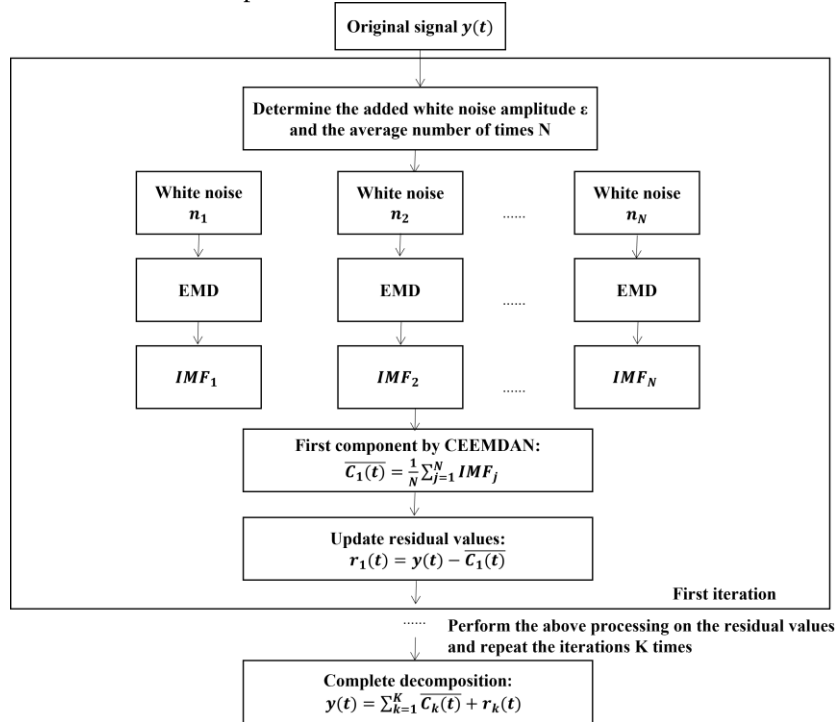


Figure 1

3.2. BiLSTM

Bidirectional Long Short-Term Memory Network (BiLSTM) is a deep learning neural network model for processing sequence data. It consists of two directional long short-term memory networks (LSTMs), one starting at the beginning of the sequence and the other starting at the end of the sequence in order to capture both past and future information. This bi-directional structure allows BiLSTM to better understand and model the contextual relationships of sequence data and is suitable for many natural language processing and temporal data analysis tasks such as text categorization, named entity recognition, and load forecasting. Its structure is shown in Figure 2.

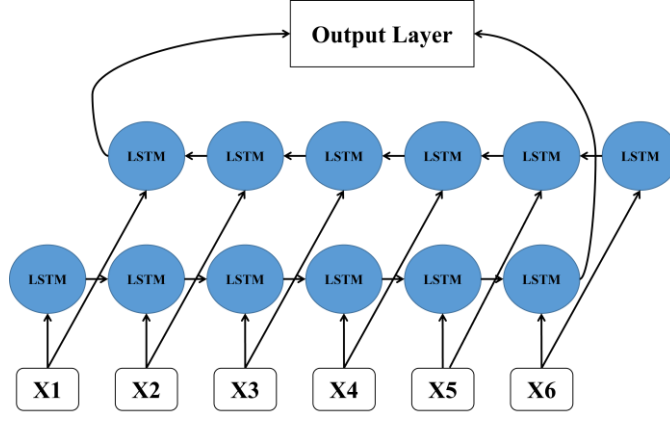


Figure 2

3.3. Attention mechanism

The Attention method is a technique used to improve the processing of sequence data by neural networks. It allows the model to assign different weights to different parts of the input sequence as they are processed so that the model can better focus on the information relevant to the task at hand. What makes this approach unique is that instead of simply taking all inputs into account equally, it dynamically adjusts attention to the needs of the current context to improve the model's performance. This can help the model better capture long-distance dependencies and semantic information in sequence data, leading to better results in tasks such as natural language processing.

The calculation formula is as follows:

$$Attention(Q, K, V) = \sum_{i=1}^n \alpha_i * V_i$$

$$\alpha_i = softmax(f(Q, K_i))$$

$$f(Q, K_i) = Q^T K_i$$

where Q is the query and K is the key, which are feature vectors used to compute the attention weights. α_i is the correlation between the query and the key.

3.4. Migration modeling

In order to form a comparison experiment, this paper establishes two migration models, CNN-LSTM-TL and CNN-BiLSTM-TL, for small sample data, and the CNN-BiLSTM-TL migration model is explained below.

In this paper, the CNN model is combined with BiLSTM model as shown in Figure 3. It uses the superposition of four convolutional layers and four pooling layers as a CNN layer to extract data features, which are then fed into the BiLSTM neural network for chiller load prediction.

For the implementation of the CNN-BiLSTM-TL transfer model, firstly, the prediction model is pre-trained using buildings with sufficient data. Then, the method of freezing some neural network layers and retraining the unfrozen network weight parameters on new buildings with small data samples is used to realize the CNN-BiLSTM-TL transfer model construction. The process of transfer model construction is shown in Figure 4.

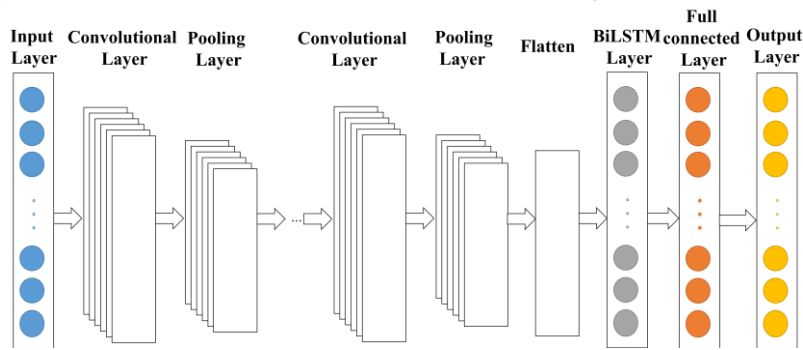


Figure 3

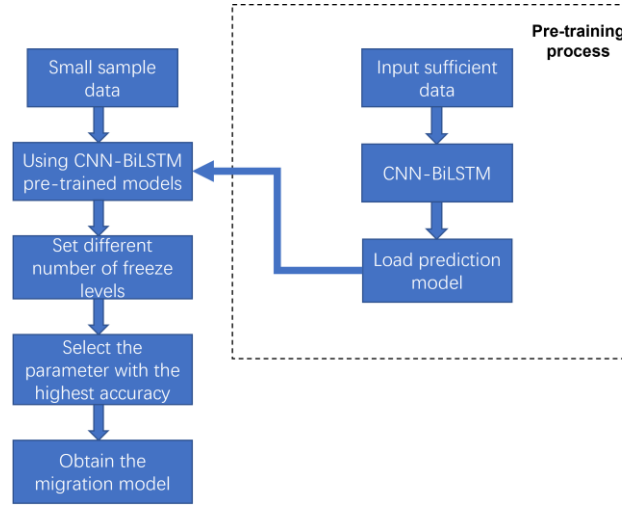


Figure 4

4. EXPERIMENTAL DESIGN PROCESS

4.1. Experimental data analysis

The data in this paper comes from the data of HVAC system operation in a building, the data includes parameters such as load factor, chilled water flow rate, cooling water flow rate and so on. Since the parameters such as chiller cooling outlet temperature, chiller freezing inlet temperature, chiller cooling inlet temperature and chiller freezing outlet temperature have a strong correlation with the operation status of the chiller. And the load factor, as a kind of time-varying characteristic data, is usually affected by the data of the previous moment. Therefore, these four parameters and the load rate at the current moment are selected as input variables, and the load rate after 1 hour is taken as output variable. For the large sample data as well as the small sample data, the variation of our target data (load rate) is shown in Figure 5 and Figure 6.

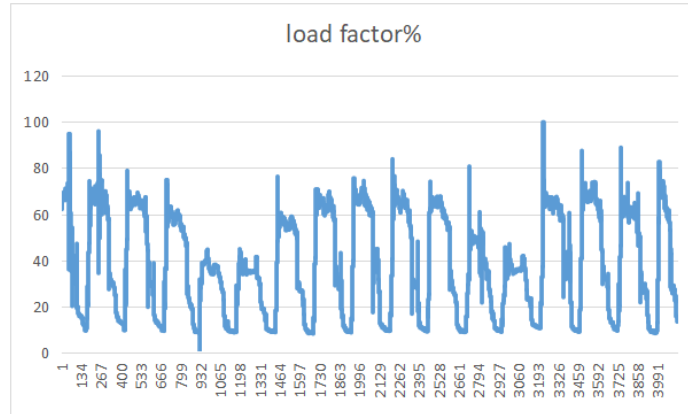


Figure 5

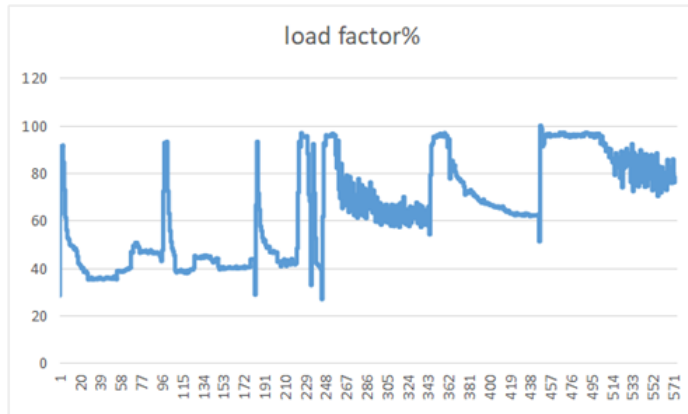


Figure 6

4.2. Experimental parameter setting

Load factor prediction after one hour is the subject of this paper. For the CNN-BiLSTM migration model, after comparing the experimental results, it was decided to choose the number of CNN layers $m=4$ and the number of BiLSTM layers $n=2$. In the BiLSTM structure, there are 64 hidden neurons in the first layer and 32 hidden neurons in the second layer. The number of neurons in the fully connected layer is $n_{FC}=32$. As for the CEEMDAN-BiLSTM-Attention model, its structure is similar to that of the above BiLSTM, and the number of nodes in the Attention mechanism is 32. As for the choice of optimizer, after comparing it with several optimizers, such as adam, radam, and sgd, the result shows that the adam optimizer is the best. As for the value of batch size, according to the generalized value of batch size which is generally a power of 2, 16 is used as the batch size. Finally, the raw data was divided into 90% for the training set and 10% for the test set. The programming environments used are python 3.6, Tensorflow 2.1, and keras 2.9.0. The model parameters are set as shown in Table 1.

Table 1

Prediction model	Parameters setting
CNN	n_{input} : 5, m : 4, kernel size: (3*3), n_{FC} : 32, n_{batch} : 16, activation: Relu, lr: 0.0002, loss: MSE, optimizer: Adam;
LSTM	n_{input} : 5, n : 2, n_{L1} : 64, n_{L2} : 32, n_{FC} : 32, n_{batch} : 16, activation: Relu, lr: 0.0002, loss: MSE, optimizer: Adam;
CNN-LSTM	n_{input} : 5, m : 4, n : 2, kernel size: (3*3), n_{L1} : 64, n_{L2} : 32, n_{FC} : 32, n_{batch} : 16, activation: Relu, lr: 0.0002, loss: MSE, optimizer: Adam;
CNN-BiLSTM	n_{input} : 5, m : 4, n : 2, kernel size: (3*3), n_{L1} : 64, n_{L2} : 32, n_{FC} : 32, n_{batch} : 16, activation: Relu, lr: 0.0002, loss: MSE, optimizer: Adam;
CEEMDAN-BiLSTM-Attention	n_{input} : 5, m : 4, n : 2, kernel size: (3*3), n_{L1} : 64, n_{L2} : 32, n_{nodes} : 32, n_{FC} : 32, n_{batch} : 16, activation: Relu, lr: 0.0002, loss: MSE, optimizer: Adam;

4.3. Evaluation indicators

When used to evaluate the performance of a predictive model, three common metrics are generally used, MAPE, RMSE and MAE, a brief description of which is stated below:

MAPE (Mean Absolute Percentage Error): MAPE is a metric expressed as a percentage that measures the prediction error of a model. It calculates the difference between the prediction error and the actual value for each observation and then averages these differences. The lower the MAPE, the more accurate the model's prediction.

RMSE (Root Mean Square Error): RMSE is a measure of prediction error that calculates the square of the prediction error for each observation and takes the square root. RMSE penalizes large errors because it sums the squares of the differences. Unlike MAE, a lower RMSE indicates that the model is more sensitive to outliers.

MAE (Mean Absolute Error): MAE is a measure of a model's prediction error that calculates the absolute value of the prediction error for each observation and averages these absolute values. MAE measures the average error of the model, regardless of the direction of the error. Lower values of MAE indicate that the model's average error is smaller.

In short, MAPE focuses on percentage error, RMSE considers the square of the error and is more sensitive to large errors, and MAE considers only the absolute value of the error. All three metrics can be used to assess the predictive accuracy of a model.

5. COMPARISON OF EXPERIMENTAL RESULTS

5.1. Comparison of Adequate Data Results

CEEMDAN was used to decompose the raw data to reduce its complexity, and the decomposition results are shown in Figure 7. Then the BiLSTM-Attention model is used to predict the load factor after one hour, and the prediction curve is shown in Figure 8.

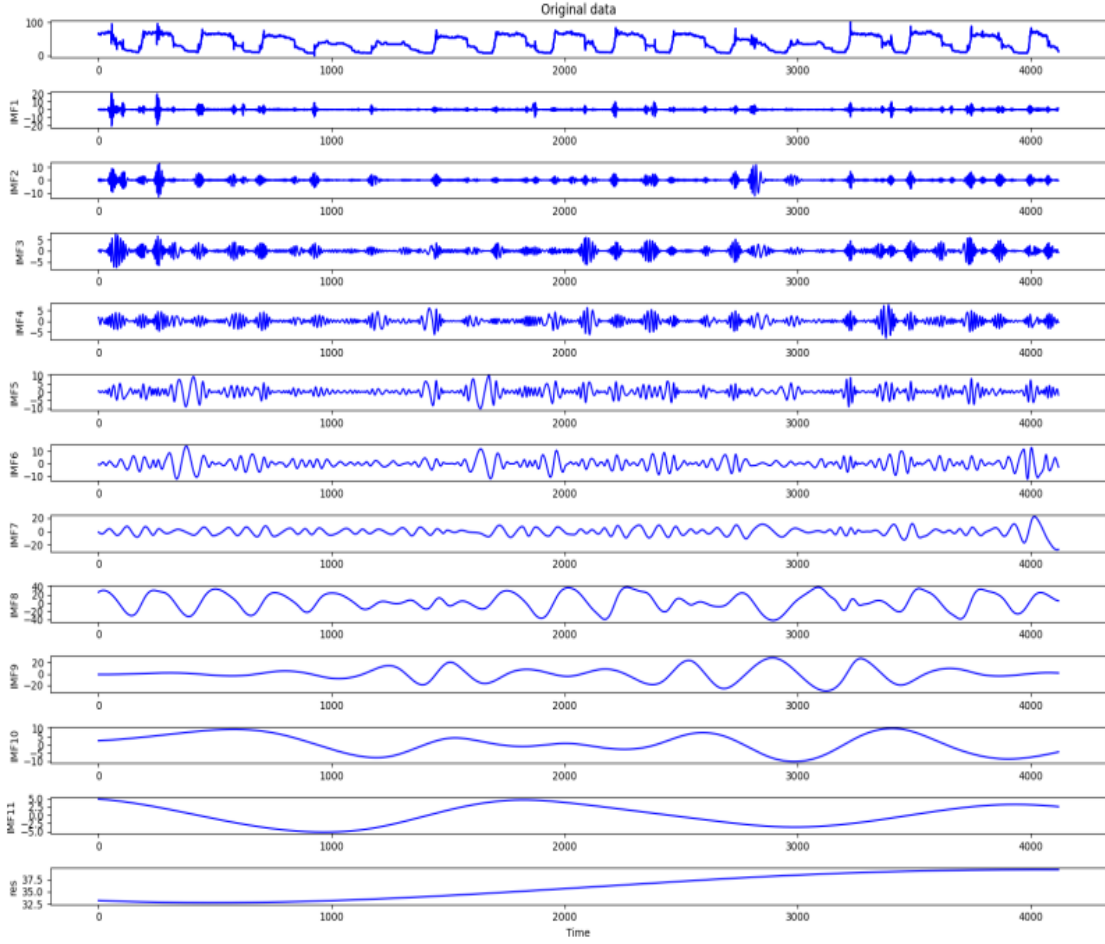


Figure 7

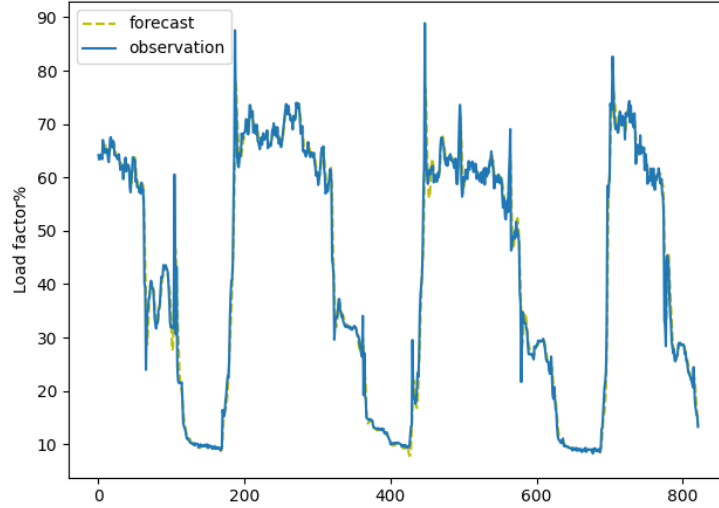


Figure 8

From Figure 7, it can be seen that after CEEMDAN decomposition, the original complex data, gradually become regular. From Figure 8, it can be seen that the prediction curve of CEEMDAN-BiLSTM-Attention stays in the same trend with the actual curve, and the prediction effect is very good. In order to reflect the accuracy of the proposed model in this paper, it is compared with the classical prediction models, which include CNN, LSTM, CNN-LSTM, CNN-BiLSTM, CEEMDAN-LSTM, and CEEMDAN-BiLSTM. MSE, MAE, and MAPE evaluation indexes are used, and the comparison results are shown in Table 2.

Table 2

models\evaluation merits	RMSE	MAE	MAPE
CNN	5.41	4.06	9.89
LSTM	4.56	3.42	9.38
CNN-LSTM	4.02	2.92	7.46
CNN-BiLSTM	3.70	2.57	6.51
CEEMADAN-LSTM	2.68	1.80	4.10
CEEMADAN-BiLSTM	2.53	1.50	3.92
CEEMADAN-BiLSTM-Attention	2.27	1.35	3.65

5.2. Comparison of experimental results for small sample data

From the above analysis of the data, it can be learned that the amount of data in the case of small samples is not much, so the IMF component obtained by CEEMDAN decomposition will be less than that of the decomposition of the large sample data, and the convergence of the algorithm is achieved faster. The results of decomposing the small sample data are shown in Figure 9. The results predicted by the CEEMDAN-BiLSTM-Attention model are shown in Figure 10. Meanwhile, in order to compare with the migration model, the prediction results of the CNN-LSTM-TL, and the CNN-BiLSTM-TL are given as shown in Figure 11 and Figure12.

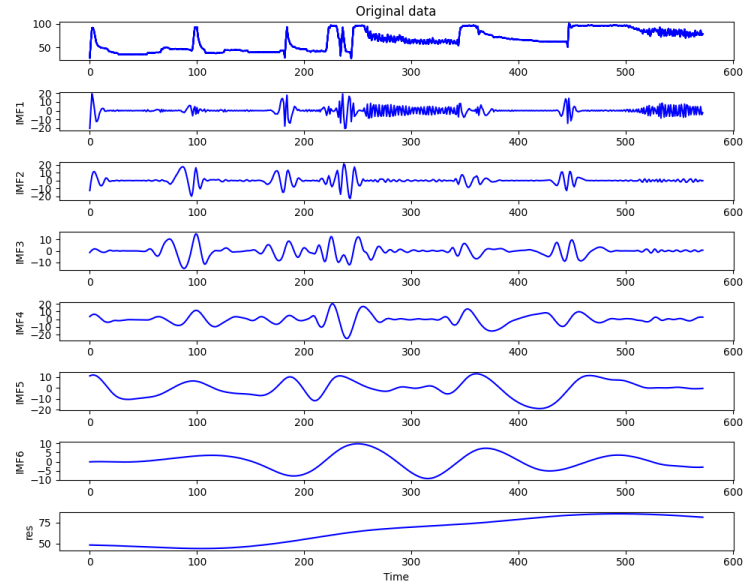


Figure 9

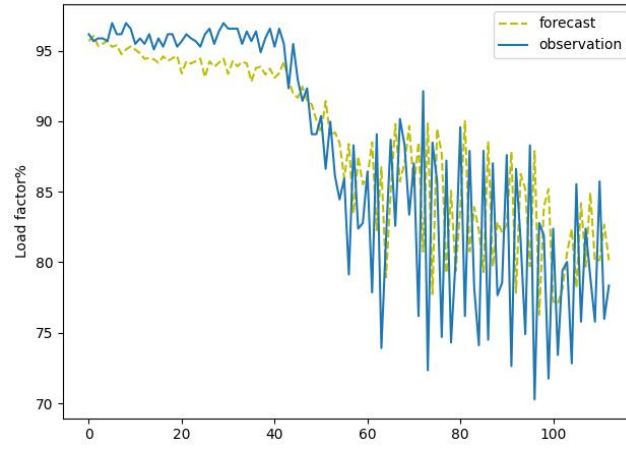


Figure 10

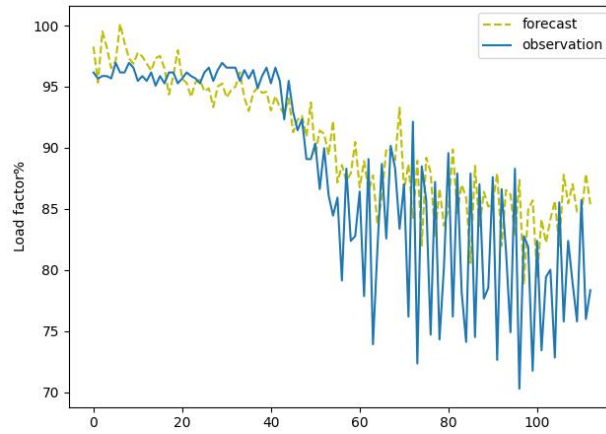


Figure 11

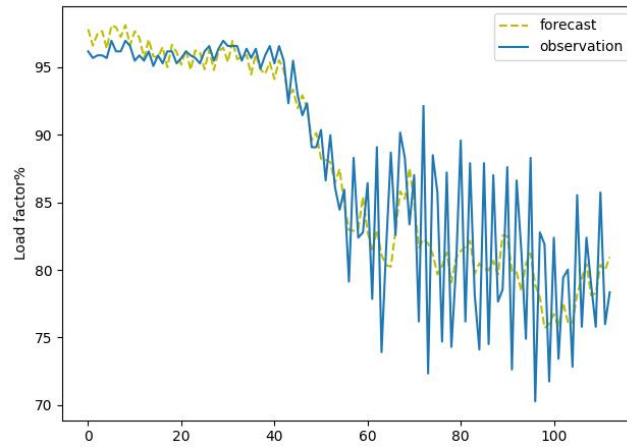


Figure 12

As can be seen from the figure, the IMF component quickly takes on a monotonic character due to the small amount of decomposed data. From the load factor prediction curves, it can be seen that the first half of the comparison migration model CNN-BiLSTM-TL model fits closer to the actual results. However, for the second half of the prediction curve, the CEEMDAN-BiLSTM-Attention model works better.

As a whole, the CEEMDAN-BiLSTM-Attention model is able to respond better to the trend of the load factor. Meanwhile, the results of the evaluation index shown in Table 3 show that the model proposed in this paper is better than other classical models.

Table 3

models\evaluation merits	RMSE	MAE	MAPE
CNN	9.53	7.88	12.10
LSTM	9.42	8.27	10.48
CNN-LSTM	9.25	9.57	8.64
CNN-BiLSTM	8.55	6.76	7.99
CNN-LSTM-TL	6.53	4.54	5.72
CNN-BiLSTM-TL	6.11	4.82	5.47
CEEMADAN-LSTM	4.24	3.37	4.28
CEEMADAN-BiLSTM	3.97	2.98	3.63
CEEMADAN-BiLSTM-Attention	3.72	2.85	3.40

5.3. Analysis of model predictions

1. As can be seen from Table 2, the values of all three evaluation indexes of CEEMDAN-BiLSTM-Attention are optimal, and the prediction accuracy of hybrid prediction models (CNN-LSTM, CEEMDAN-LSTM, etc.) is a little bit higher compared to that of single models (CNN, LSTM, etc.), which may be due to the fact that the hybrid model has gone through several times of feature extraction of the data, which makes the time series information obtain more perfect, and thus achieves more accurate prediction. Moreover, the hybrid model of CEEMDAN gets better prediction results than the hybrid model of CNN, which indicates that the data decomposition of CEEMDAN can obtain more time series data information to a certain extent, thus supporting the accuracy of load prediction.

2. From Fig. 10, it can be seen that the CEEMDAN-BiLSTM-Attention prediction model for small-sample data, in the previous part of the prediction appeared to be a large deviation from the actual value, but compared with the traditional migration learning method, the method proposed in this paper is more in line with the trend of the actual load curve. And it can be found from the evaluation indexes in Table 3 that the migration learning model using the freezing layer still has certain advantages over some prediction models that directly carry out load prediction, but in terms of the overall indexes, it is still the CEEMDAN-BiLSTM-Attention prediction model that is in the optimal index value. Therefore, the prediction model proposed in this paper still has certain research and feasibility in the case of small sample data.

3. During the training process of the prediction model, it is found that CEEMDAN takes more time to decompose the large amount of historical data, so this may be a disadvantage of this method. However, it may be possible to improve the training speed of the whole model to a certain extent by improving the hardware facilities of the equipment, such as adding GPU for model training.

6. CONCLUSIONS

Load forecasting of chiller units is of great significance for subsequent building energy optimization, energy allocation management, etc. Therefore, this paper explores the CEEMDAN-BiLSTM-Attention algorithm and applies it to chiller load forecasting, which combines the decomposition of complex data and the use of the attention mechanism to enhance the prediction effect of BiLSTM neural network. For large-sample data, the CEEMDAN-BiLSTM-Attention algorithm demonstrates its superiority by comparing with other classical prediction models, and all evaluation indexes are better than the comparison models. Meanwhile, for small sample data, two migration models (CNN-LSTM-TL, CNN-BiLSTM-TL) using the freeze layer strategy are added for comparison, and it is found that the model proposed in this paper still has good results on small sample data, and the prediction accuracy is better than some migration models. For future work, it can be considered to establish an online learning system that can continuously update the model from new data to adapt to the evolution of the system performance; develop visualization tools and explanatory models to help users understand the prediction results and make better decisions; and integrate the load prediction with the energy management system in order to optimize the energy utilization of the chiller unit and reduce the energy cost.

REFERENCES:

- [1] Fan C, Ding Y. Cooling load prediction and optimal operation of HVAC systems using a multiple nonlinear regression model[J]. *Energy and Buildings*, 2019, 197: 7-17.
- [2] Jaruwongwittaya T, Chen G. A review: renewable energy with absorption chillers in Thailand[J]. *Renewable and Sustainable Energy Reviews*, 2010, 14(5): 1437-1444.
- [3] Patnaik D, Marwah M, Sharma R, et al. Sustainable operation and management of data center chillers using temporal data mining[C]//*Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009: 1305-1314.
- [4] Fan C, Ding Y. Cooling load prediction and optimal operation of HVAC systems using a multiple nonlinear regression model[J]. *Energy and Buildings*, 2019, 197: 7-17.
- [5] Piper J E. Operations and maintenance manual for energy management[M]. ME Sharpe, 1999.
- [6] Varbanov P S. The intertwining issues of energy security, environmental protection, and societal development[J]. *Clean Technologies and Environmental Policy*, 2022, 24(5): 1319-1320.
- [7] Mak B L, Chan W W, Li D, et al. Power consumption modeling and energy saving practices of hotel chillers[J]. *International Journal of Hospitality Management*, 2013, 33: 1-5.
- [8] Lu Y Y, Chen J, Liu T C, et al. Using cooling load forecast as the optimal operation scheme for a large multi-chiller system[J]. *International journal of refrigeration*, 2011, 34(8): 2050-2062.
- [9] Orehounig K, Mavromatidis G, Evins R, et al. Towards an energy sustainable community: An energy system analysis for a village in Switzerland[J]. *Energy and Buildings*, 2014, 84: 277-286.
- [10] Wang Z, Hong T, Piette M A. Building thermal load prediction through shallow machine learning and deep learning[J]. *Applied Energy*, 2020, 263: 114683.
- [11] Koschwitz D, Frisch J, Van Treeck C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale[J]. *Energy*, 2018, 165: 134-142.
- [12] Kang X, Wang X, An J, et al. A novel approach of day-ahead cooling load prediction and optimal control for ice-based thermal energy storage (TES) system in commercial buildings[J]. *Energy and Buildings*, 2022, 275: 112478.
- [13] Chiang W C, Li Y, Shang J, et al. Impact of drone delivery on sustainability and cost: Realizing the UAV potential through vehicle routing optimization[J]. *Applied energy*, 2019, 242: 1164-1175.
- [14] Yadav B, Pandey V, Yadav S, et al. Effect of misting and wallowing cooling systems on milk yield, blood and physiological variables during heat stress in lactating Murrah buffalo[J]. *Journal of animal science and technology*, 2016, 58(1): 1-10.
- [15] Powell K M, Sriprasad A, Cole W J, et al. Heating, cooling, and electrical load forecasting for a large-scale district energy system[J]. *Energy*, 2014, 74: 877-885.
- [16] Huang P, Huang G, Wang Y. HVAC system design under peak load prediction uncertainty using multiple-criterion decision making technique[J]. *Energy and Buildings*, 2015, 91: 26-36.
- [17] Hashem D T N. The reality of internet of things (IoT) in creating a data-driven marketing opportunity: mediating role of customer relationship management (CRM)[J]. *J. Theor. Appl. Inf. Technol*, 2021, 99(2).
- [18] Antonopoulos I, Robu V, Couraud B, et al. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review[J]. *Renewable and Sustainable Energy Reviews*, 2020, 130: 109899.
- [19] Damerval C, Meignen S, Perrier V. A fast algorithm for bidimensional EMD[J]. *IEEE signal processing letters*, 2005, 12(10): 701-704.
- [20] Yang D, Zhu Z, Liang B. Vital sign signal extraction method based on permutation entropy and EEMD algorithm for ultra-wideband radar[J]. *IEEE Access*, 2019, 7: 178879-178890.
- [21] Wang Y, Gu J, Zhou Z, et al. Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion[J]. *Knowledge-Based Systems*, 2015, 88: 12-23.
- [22] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network[J]. *Physica D: Nonlinear Phenomena*, 2020, 404: 132306.
- [23] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. *Neural computation*, 2019, 31(7): 1235-1270.
- [24] Siami-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series[C]//*2019 IEEE International conference on big data (Big Data)*. IEEE, 2019: 3285-3292.
- [25] Chen L, Liu X, Zeng C, et al. Temperature prediction of seasonal frozen subgrades based on CEEMDAN-LSTM hybrid model[J]. *Sensors*, 2022, 22(15): 5742.
- [26] Reddy G, Celani A, Sejnowski T J, et al. Learning to soar in turbulent environments[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(33): E4877-E4884.