Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)
IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7

A Comparative Study of the Performance of Machine Learning based Load Forecasting Methods

[1] P.PRASHANTHI
Research Scholar, Department of CSE,
Vels Institute of Science Technology and Advanced Studies.
Tamilnadu, India
Mail id: prashanthi.peram@gmail.com

[2] Priyadarsini .K
Dept of CSE,
Vels Institute of Science Technology and Advanced Studies,
Tamilnadu, India
Mail id: priyadarsini.se@velsuniv.ac.in

2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9395834

***Abstract*— A significant aspect of power system control is short-term power load forecasting. It is the basis for the preparation of network systems, the exchange of energy and load scheduling. The accuracy of power load forecasting is directly linked to the security, stability and economic activity of the power system. In this paper, short-term load forecasting methods based on machine learning are studied and evaluated, and the efficiency of the load forecasting methods based on machine learning is compared to a conventional load forecasting approach widely used.**

Keywords— Time series classification, Short term load forecasting, Machine learning, deep learning, SVM.

I. INTRODUCTION

The power industry serves as an important part of society, which has great significance to the security of the whole country, social stability and people's lives. Electric energy is difficult to store, which sets higher requirements for power generation, transmission and sale. Power energy should not be supplied in excess of demand, which results in wasting of energy resources. It should not be in short supply either, since it may cause out age in some districts. Therefore, power load forecasting is very important to maintain the balance of power supply and demand [1]. Environmental factors and historical data are used to predict loads in the future, which is beneficial to make plans for power generation and transmission. The power load forecasting began in the 1980s, but the original forecasting work did not use complex methods and it is mainly done through manual calculation by experienced people, the forecasting results are quite different from the actual situation. With the development of society, people's demand for electricity is getting higher and higher, which puts forward higher requirements for the accuracy of load forecasting. What we can do is to do everything possible to ensure a real-time balance between electricity supply and demand. Unfortunately, it is impossible to achieve a complete supply demand balance via a forecasting method because of the existence of emergencies and the influence of various factors. Therefore, power grids have some capacities to achieve a dynamic balance of power generation and demand. Accurate forecasting of power load can reduce the reserve capacity of power grid and contribute to a better utilization of electricity.

In terms of short-term electric load forecasting methods, no matter it is based onstatisticallearning-basedforecastingmethodsorartificialintelligence-based forecasting methods, their basic idea is to explore the potential value of historical electricity loads, and establish a mathematical model to make use of the historical data. Then, mathematical models are used to predict the power load at a certain time in the future. In the past ten years, smart grid technology has developed vigorously. Smart meters have replaced traditional meters. The existence of a large number of sensors in smart meters has dramatically improved the observability of power grid. After years of development, the electric power industry has accumulated a large amount of historical load data, which laid a solid foundation for the application of various forecasting models. With the development of smart grid and information technology, computers have undergone many updates. The computing performance of computers has experienced explosive growth. At the same time, the extensive application of graphics processors also provides powerful computing ability for deep neural networks.

In recent years, with the rapid development of machine learning, artificial intelligence has made great breakthroughs in many fields. At present, many artificial intelligence algorithms have played an important role in improving the accuracy of prediction, which make it possible for high-precision short-term load forecasting.

II. BACKGROUND

This chapter introduces the classification of power load forecasting problems and summarizes the short-term power load forecasting methods.

A. Power load forecasting

In respect to different time scales of prediction, power load forecasting can be roughly divided into long-term forecasting and short-term forecasting. There are also some differences in definitionsoflong-termandshort-termindifferent application scenarios. According to different time scale, power load prediction can be further divided in to into four classes [2]: Forecasting ultra-short-term loads, forecasting short-term loads, forecasting medium-term loads and forecasting long-term loads. Minutes, hours, months and years are the associated time scales. The classification of power load forecasts based on a different time scale is shown in Figure 1.

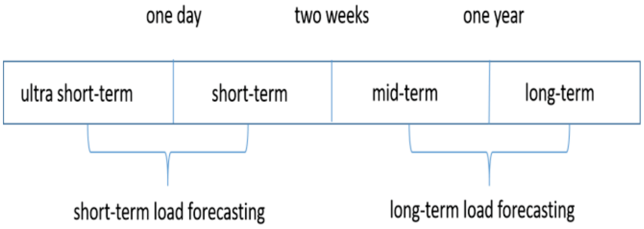

Figure 1: Classification of power load forecasting

Ultra-short term load forecasting aims to predict the power load in the next few minutes, which is mainly used to monitor the operation of the power grid [3]. Short-term load forecasting is to forecast the power load in the next few hours, which mainly provides data for the optimal dispatch of power plants [4]. Medium-term load forecasting is to forecast the load in the next few months, which is used to make maintenance plans [5]. Long-term power load forecasting is to forecast the load in the next few years, which is used to guide the transformation of power grid [5]. Each kind of power load forecasting has its own application scenario. This work mainly studies short-term power load forecasting, which is also the basis for power companies to dynamically adjust their generation and transaction plans in the market.

B. Traditional methods for short-term load forecasting

In the 1990s, computers gradually entered all walks of life. However, their computing ability is very limited. Most people mainly used statistical methods to predict short-term load [6]. Traditional methods mainly include the following:

Time Series Model:

Time series model uses the characteristics such as auto-correlation, trend and seasonal variation to predict short-term load [7]. Time series models have been studied for decades. The most popular prediction methods are ARIMA and ARMA [8]. The time series model assumes a linear relationship between future loads and historical loads over the past few hours and stochastic error functions. Both ARMA model and ARIMA have achieved good prediction results.

Multiple Linear Regressions

Multivariate regression uses a linear function to map the relationship between the output y and multiple independent variables x1,x2,...,xk. The purpose of multiple linear regression is to find a function to describe the relationship between the output and variables, it tries to make predictions according to those independent variables. Power load can be affected by various of environmental features such as temperature, humidity and precipitation. The mathematical formula of multivariate linear regression prediction model can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

Where y is the short-term load to be predicted, xi is the feature that affects power load, βi is the regression parameter of is the random error. Multiple linear regression has been widely used for smooth time series prediction. However, for time series with strong fluctuations, the accuracy of multiple linear regression is very low.

Grey Model

Grey model (GM) uses a small amount of historical data to build differential equations to predict short-term loads. Firstly, the historical data are accumulated to generate new sequences to weaken the randomness of the original data. Secondly, differential equations are established by using the generated sequence. For example, the GM (1, 1) model represents differential equations of 1-order and one variable. Other traditional methods need a lot of historical data to train the model, while the grey model needs relatively few data. Grey model is suitable for load curve that grows exponentially. For the stationary power load curve, its prediction accuracy is limited.

Expert system

Expert system is a computer program that has the ability to analyze the current circumstance and expand the knowledge base following the emergence of new information. The basic structure of the expert system is shown in figure 2, where the arrow shows the direction of data flow. Expert systems also simulates the decision-making process of human and tries to provide the optimal solution at current circumstance.

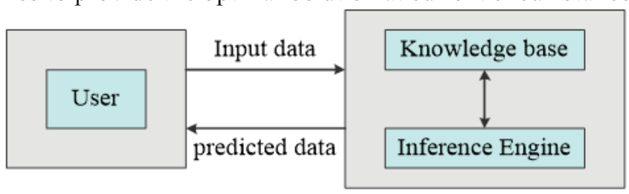

Figure 2: Correlation matrix of Pearson similarity method

Firstly, users input features such as historical data and environmental factors. Then, the inference engine matches the input features with the circumstances stored in the knowledge base, the most similar circumstance will be found. Finally, the expert system makes the prediction based on the stored conclusion. Expert system is easy to maintain. However, the process of acquiring knowledge is the main obstacle of expert system.

C. Machine Learning Methods for forecasting

SVM(Support Vector Machine)

Support vector machine(SVM)buildsahyper-planeinhighdimensionalspace, which is used to classification or regression [10]. If the sample is linearly inseparable, the feature can be mapped to high-dimensional space by using the kernel function, and then a linear classifier is established. In terms of regression, the normal vector of the hyper-plane contains a function that makes the objective and the estimation as close as possible. This hyper-plane should be able to accurately predict the distribution of data. Compared with traditional methods, SVM does not make a prior assumption on data, so it can deal with both stationary and non-stationary sequences. However, if the training set has a large number of samples, the training speed of SVM will be very slow [11].

SVM should not only be extended to classification problems, but also to regression. Like the classification method, there is an incentive to look for and refine the generalisation limits for regression.

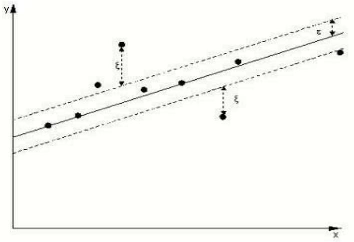

Figure 3: Classification of SVM

Gradient boosting decision tree

In the field of machine learning, boosting is a method to build a strong learner by combining a group of weak learners with low complexity, low training cost, and not easy to overfit. In each classification process, the weight of misclassified data is increased a bit for the next training procedure, and finally obtain a strong learner[12]. Gradient descent is a boosting method that commonly used in classification and regression problems. It generates models in the form of a set of weak learners (usually decision tree). The main idea is to optimize the loss function using gradient descent [13]. XGBoost and LightGBM are the most popular gradient boosting frameworks for regression and classification. Very similar to regression trees except that the terminal node leaves, predict the outcome using a linear model as shown in figure 4.

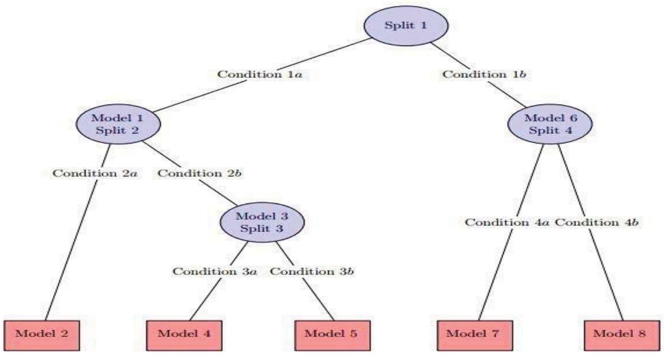

Figure 4: gradient boosting frameworks for regression and classification

Deep learning methods

Deep learning is one of the machine learning techniques that has neural networks with several hidden layer. With a more complex model architecture, it usually has a better performance than shallow learning when solving complicated problems, the advantages of deep learning can be summarized as follows [14]:

- With more hidden layers in the architecture, deep neural network is easier to capture non-linear data characteristics, and it has stronger ability to analyze the internal correlation between input vector and corresponding output.
- Deep neural networks are easier to learn features of samples. Feature information is transformed from one layer to another to form a new feature space, which makes the model easier to learn.
- Deep learning uses large data set to train the model, which is capable of digging more intrinsic information from training data.

The typical deep learning models include convolution neural network (CNN), stacked auto-encoder network, deep belief network (DBN) and recurrent neural network (RNN) [15].

III. ANALYSIS

A. Analysis of power load data characteristic

Power load data preprocessing

Before the power load forecasting, the most important step is the selection of historic information. Historic data is the base of the future load forecasting, the quality of historic power load data directly affect the prediction accuracy. Nowadays, power system data are mainly collected from SCADA system. However, due to some reasons, data collected from SCADA system are incomplete, and even have some errors. Thus, how to distinguish and correct those bad data is of great importance for short-term power load forecasting.

Correction of abnormal data

There are two correction methods for abnormal data:

(1) processing method

Normally, power load change within a day should be smooth and continuous, which means the difference between the load value at a certain moment and the load before and after it will not be too large. If large difference is observed within short time, it means that deviation may occur due to equipment records or human factors. In this case, we can handle the outliers using horizontal processing method, which can be described as:

$$|Y(d,t) - Y(d,t-1)| > \alpha(t)$$

$$|Y(d,t) - Y(d,t+1)| > \beta(t)$$

$$Y(d,t) = \frac{Y(d,t-1) + Y(d,t+1)}{2}$$

(3.5) Where $\alpha(t)$,$\beta(t)$ is the difference threshold, $Y(d,t)$ is the power load on day d at moment t, $Y(d,t-1)$ is the power load on day d at moment $t-1$, $Y(d,t+1)$ is the power load on day d at moment $t+1$.

(2) Vertical processing method

Load characteristics shows that it has daily periodicity. That is, it can be considered that the load data at the same time on adjacent days are similar, and the difference between the two should be maintained within a certain range. If it exceeds this range, it can also be regarded as bad data. In this

case, outliers can be handled by vertical processing method, which is described as follows:
If

$$|Y(d,t) - m(t)| > r(t)$$

Then

$$Y(d,t) = \begin{cases} m(t) + r(t), & Y(d,t) > m(t) \\ m(t) - r(t), & Y(d,t) < m(t) \end{cases}$$

Where $r(t)$ is the threshold, $m(t)$ is the mean power load at moment t in recent days, $Y(d,t)$ is the power load on day d at moment t.

Filling Missing Data

Based on the periodicity of the load, for the data lost due to some factors, we can use the data of the previous day or the next day at this moment instead.

B. Gradient Boosting Decision Tree

.4.1 XGBoost Algorithm

XGBoost is a decision tree based algorithm using gradient boosting framework proposed by Tianqi Chen and Carlo Guestrin in 2016. Like many other gradient boosting methods, XGBoost follows the idea of ensemble weak learners using gradient descent architecture. However, XGBoost distinguish itself in the following ways:

1) Better regularization capability: XGBoost penalizes the parameters of complex models through L1 and L2 regularization to avoid overfitting.
2) Sparsity Awareness: automatically handling missing value depending on training loss.
3) Cross-validation: XGBoost automatically conduct cross validation at each iteration process.
4) Parallelization: XGBoost uses paralleled implementation to handle the process of sequential tree building.

LightGBM Algorithm

Light gradient boosting machine (LightGBM) is a decision tree based gradient boosting framework. Unlike other decision tree algorithms, LightGBM employs a novel method called gradient based one side sampling to find the most suitable split for data samples. As shown in figure 5, for LightGBM, the decision tree grows leaf-wise while other boosting algorithms have a level-wise or depth-wise decision tree growth. In this way, more loss can be reduced and hence results in better performance.

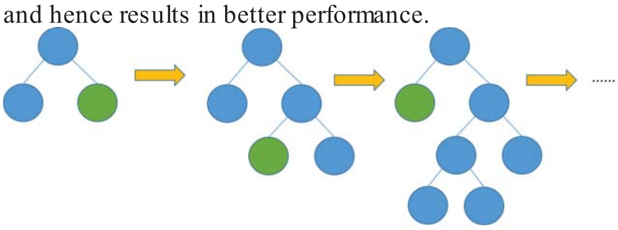

Figure 5: Leaf-wise growth for LightGBM

Advantages of LightGBM algorithm can be summarized as follows:

1) Shorter training time: LightGBM converts continuous feature values into discrete values which results in faster training process.
2) Lower memory usage: discrete values require less memory than continuous values.
3) High accuracy: decision tree built by leaf-wise growth approach is more complex than that of lever-wise growth approach. However, it may result in overfitting sometimes, this can be solved by tuning the hyperparameters in the model.
4) Compatibility with large data sets: even with large data set, LightGBM can perform equally good with shorter training time.

IV. Conclusion

Compared to the traditional ARIMA model forecasting, all machine learning based methods have an overall better forecasting performance in terms of MAE and MAPE. One reason may be that ARIMA model can only use the historical time series to make the prediction and it does not have the capability of analyzing environmental factors. In this case, among the four machine learning methods, LightGBM yields for the best performance. Another gradient boosting decision tree-based algorithm XGBoost also performs well, but the performance of GRU network is slightly better than that of XGBoost. Thus, it is hard to determine whether GBM algorithm or RNN stands out for this forecasting problem. Before we make the make the prediction, it is difficult for us to determine the most suitable method in the first place. Therefore, we should test several forecasting methods and pick the one with the best performance among them.

Features electing result suggest that temperature, dew point, uv-index are highly correlated to the power load consumption. In the feature extraction process, the Pearson similarity analysis of holiday index is not included. However, we add this feature to train the model. Feature importance analysis of XGBoost indicates that holiday impact does not contribute much to the forecasting result in this case, the reason may be that the power load consumption of the analyzed district is not largely influenced by holiday.

References

[1] F. Hobbs et al. "Analysis of the value for unit commitment of improved load forecasts". In: IEEE Transactions on Power Systems 14.4 (Nov. 1999), pp. 1342–1348. issn: 1558-0679. doi: 10.1109/59.801894.
[2] Tulasi Radhika Patnala, Jayanthi D, Shylu D.S, Kavitha K, Prathyusha Chowdary, "Maximal length test pattern generation for the cryptography applications" meterialstoday proceedings, Aug 2020
[3] Viet Cuong Ngo et al. "Ultra-short-term load forecasting using robust exponentially weighted method in distribution networks". In: 2015 IEEE Power Energy Society General Meeting. July 2015, pp. 1–5. doi: 10.1109/PESGM.2015.7286602.
[4] Tulasi Radhika Patnala, Jayanthi D, Sankararao Majji, Manohar Valleti, Srilekha Kothapalli, Santhosh Chandra Rao Karanam, "Modernistic way for KEY Generation for Highly Secure Data Transfer in ASIC Design

Flow" https://ieeexplore.ieee.org/document/9074200, Published in IEEE digital Xplore, Electronic ISSN: 2575-7288, available from 23.04.2020.
[5] L. Duan, D. Niu, and Z. Gu. "Long and Medium Term Power Load Forecasting with Multi-Level Recursive Regression Analysis". In: 2008 Second International Symposium on Intelligent Information Technology Application. Vol. 1. Dec. 2008, pp. 514–518. doi: 10.1109/IITA. 2008.397.
[6] Moghram and S. Rahman. "Analysis and evaluation of five short-term load forecasting techniques". In: IEEE Transactions on Power Systems 4.4 (Nov. 1989), pp. 1484–1491. issn: 1558-0679. doi: 10.1109/ 59.41700.
[7] N. Amjady. "Short-term hourly load forecasting using time-series modeling with peak load estimation capability". In: IEEE Transactions on Power Systems 16.4 (Nov. 2001), pp. 798–805. issn: 1558-0679. doi: 10.1109/59.962429.
[8] Tulasi Radhika Patnala, Sankarrao Majji, Gopala Krishna Pasumarthi, "Optimization of CSA for Low Power and High Speed using MTCMOS and GDI Techniques", https://www.ijeat.org/wp-content/uploads/papers/v8i5S3/E10620785S319.pdf , International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-5S3, July, 2019.
[9] A. Almalaq and G. Edwards. "A Review of Deep Learning Methods Applied on Load Forecasting". In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). Dec. 2017, pp. 511–516. doi: 10.1109/ICMLA.2017.0-110.
[10] Sankararao Majji, Tulasi Radhika Patnala, Manohar Valleti, Srilekha Kothapalli, Santhosh Chandra Rao Karanam, "A Study on the Comprehensive Analysis of Electro Migration for the Nano technology trends", Published in IEEE digital Xplore, Electronic ISSN: 2575-7288, available from 23.04.2020.
[11] N. I. Sapankevych and R. Sankar. "Time Series Prediction Using Support Vector Machines: A Survey". In: IEEE Computational Intelligence Magazine4.2(May2009),pp.24–38.issn:1556-6048.doi:10.1109/MCI.2009.932254.
[12] Sajja Krishna Kishore;Tulasi Radhika Patnala;Arun S Tigadi;Aatif Jamshed "An On-chip Analysis of the VLSI designs under Process Variations" published in IEEE digital Xplore, 2020 International Conference on Smart Electronics and Communication (ICOSEC).
[13] JeromeH.Friedman "Greedy function approximation : Agradient boosting machine." In: Ann. Statist. 29.5 (Oct. 2001), pp. 1189–1232.
[14] Baraneetharan, E. "Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey." Journal of Information Technology 2, no. 03 (2020): 161-173.
[15] Smys, S., Abul Basar, and Haoxiang Wang. "CNN based Flood Management System with IoT Sensors and Cloud Data." Journal of Artificial Intelligence 2, no. 04 (2020): 194-200.