Adaboost-Based Power System Load Forecasting

WuNeng Ling¹, Pan Zhang¹, QiuWen Li¹*, Dong Mo¹, CuiYun Luo¹, YiMing Li¹ and ZhenCheng Liang¹

¹Guangxi Power Grid Corporation, Dispatch and Control Center, Nanning, Guangxi, China

* Corresponding author: liujiawen@orait.cn

Abstract. This study presented a penetrating insight into the basic principle of ensemble learning and the ensemble technique Boosting, and deduced the theoretical model and learning principle of the adaptive ensemble learning. Besides, it proposed a Adaboost-base power system load forecasting method, and validated the effectiveness of this method through the empirical forecasting of a provincial medium and long-term load. The calculation example in this paper proves that high-accuracy of medium and long-term load forecasting can be achieved by using Adaboost-based power system load forecasting method.

1. Introduction

Based on previous and current known economic situations, social development, power supply market conditions, and analysis and investigation of previous data of power indicators and relevant factors, the medium and long-term load forecasting is performed to accurately forecast the future power demand in accordance with the forecasting of economic situations and social development during the planning period. In this regard, the medium and long-term load forecasting is fundamental to the safety of power grid planning and operation [1,2,3]. As China’s economy entered a “new normal” over these years, there are increased uncertainties on load development, leading to the restricted accuracy of conventional forecasting methods applicable to the stage of a steadily increasing load. Additionally, the small sample size of the load data in the economic new normal results in increased difficulties in the modeling of medium and long-term load forecasting.

Since medium and long-term load forecasting is subject to multiple uncertain factors and insufficient previous data are available, it is challenging to perform the accurate quantitative analysis. As a result, the development of medium and long-term load forecasting is lagging behind. The models and methods regarding medium and long-term load forecasting at home and abroad primarily include traditional electricity consumption per unit output method, trend extrapolation, elastic coefficient method, time series method, regression analysis method, neural network method based on modern forecasting techniques, fuzzy forecasting method, and grey forecasting method [4]. In the literature [5], a medium- and long-term cooling load forecasting is proposed based on the support vector machine (SVM) regression combination model, and such forecasting can be extensively implemented. Grey forecasting can forecast the system with uncertain factors through the integration and analysis of raw data. However, conventional grey forecasting models require a continuous increase in the power load in strict accordance with the exponential law, and are unable to ensure a high-accuracy forecasting in case of sudden changes in power load. However, the most optimized grey model can transform the raw data sequence with fluctuations into a sequence with an enhanced regularity [6]. Literature [7] suggests that corresponding models can be adaptively selected for forecasting based on the classification results of
raw data, thus improving the forecasting accuracy. However, such method requires massive data, and features the complex screening and processing of data. In the literature [8], a combined forecasting model is put forward through the interval division using Markov chain. In the literature [9,10], a multidimensional power consumption characteristic model is constructed based on the power consumption information of 40 million users, so as to investigate the medium-and long-term power load forecasting. In summary, most models can only provide single forecasting results or help to obtain result intervals through fluctuations, and the medium and long-term load forecasting features high levels of uncertainties, leading to high risks of single forecasting results as well as difficulties in achieving a high-accuracy forecasting.

In view of this, this study investigates the basic principle of ensemble learning and Boosting based on the aforementioned requirements, and deduces the theoretical model and learning principle of the adaptive ensemble learning. It also proposes a Adaboost-based power system load forecasting method, and validates the effectiveness of this method through the empirical forecasting of a provincial medium and long-term load.

2. Adaboost-based power system load forecasting

2.1. Algorithm principle

2.1.1. Ensemble learning

Currently, multiple machine learning algorithms (e.g., SVM, decision tree, perceptron) are available. In practical implementation, ensemble learning is extensively used. Specifically, almost all teams with high scores adopt ensemble learning during matches. The philosophy of ensemble learning refers to “three stooges top one wise”. Based on multiple learners (e.g., same algorithms with different parameters or different algorithms), ensemble learning, generally, can perform more satisfactorily than any other single learners, especially single “weak learners”. A weak learner refers to a learner with a poor performance (e.g., a binary classifier with an accuracy slightly exceeding 50%).

Consider a binary classification problem \( y \in \{-1, +1\} \), a real function \( f \), and an individual learner (or base learner) \( h_i \) whose \( m \) (odd number) error probabilities are independent and \( y \), we use a voting approach for ensemble learning, namely, over half of the results of base learners are taken as the classification results:

\[
H(x) = \text{sign}(\sum_{i=1}^{M} h_i(x))
\]  

(1)

According to Hoeffding’s inequality, the error (i.e., more than half of base learners make errors) probability after ensemble learning satisfies

\[
P(H(x) \neq f(x)) \leq \exp\left(-\frac{1}{2} M(1 - 2e)^2\right)
\]

(2)

In case of a large number \( M \) of base learners with an independent error probability, the error probability is close to 0 after ensemble learning, which is in line with the intuitive idea: The probability that most people simultaneously make errors is relatively low. All the aforementioned inferences are based on the fact that base learners independently make errors. However, these learners cannot be independent of each other in practical applications. In this regard, how to make base learners “relatively independent” (i.e., an increase in the diversity of base learners) remains the major issue to be tackled for ensemble learning.

Assemble learning can be divided into two categories based on whether there is a dependency between base learners:

- There is a strong dependency between base learners, a series of base learners should be generated serially, and Boosting is the representative algorithm;
There is no strong dependency between base learners, a series of base learners can be generated in parallel, and Bagging and Random Forest are the representative algorithms. Boosting is an ensemble method that is used to boost a weak base learner into a strong learner. The steps are described as follows:
1. First, an initial base learner is trained using the training set with an equal weight of each sample;
2. The weight of samples in the training set is adjusted based on the performance of the training set forecast using the learner obtained in the previous round of training (e.g., an increase in the weight of misclassified samples causes the samples to receive increasing attention in the next round of training), and a new base learner is trained on this basis;
3. Repeat Step 2 until m base learners are obtained, and the final ensemble result is the combination of m base learners.

Therefore, Boosting is a serial process. AdaBoost is the most popular boosting algorithm among the Boosting clustering algorithms.

2.1.2. AdaBoost principle
For the aforementioned steps regarding Boosting, two questions should be answered:
- How to adjust the weight of samples in each training set?
- How to combine m learners obtained into the final learner?
For AdaBoost (Adaptive Boosting), the following methods are adopted:
The weight of the misclassified samples in the last round of training is increased, and the weight of the correctly classified samples is reduced; the linear weight sum method is used. Based learners with a small error rate have a larger weight, while base learners with a large error rate have a smaller weight.

The following figure shows the Adaboost structure.

![Adaboost structure](image)

**Figure 1** Adaboost structure

2.2. Algorithm steps
Consider the following binary classification training dataset (the standard AdaBoost is applicable to the binary classification only):

\[(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\]

Where, \(x_i\) is a column vector with \(d\) elements, i.e., \(x_i \in \mathbb{R}^d\); \(y_i\) is a scalar, \(y \in \{+1, -1\}\).

The steps of AdaBoost are described as follows:
1. The weight of initialized samples
D1= (w11, w12, ...w1N), w1i=1N, i=1,2...N

2. For m=1, 2, …M, repeat the following steps to obtain M base learners:
   (1) Using the training data with the sample weight distribution Dm, the m\(^{th}\) base learner Gm(x) is obtained:
   \[ Gm(x) : X \rightarrow \{-1, +1\} \]
   (2) The classification error rate of Gm(x) for the weighted training dataset is calculated:
   \[ e_m = \sum_{i=1}^{N} P(G_m(x_i) \neq y_i) = \sum_{i=1}^{N} w_m i(G_m(x_i) \neq y_i) \]  \( (3) \)
   In the aforementioned formula, \( I() \) is the indicator function, and AdaBoost should determine whether the basic conditions (e.g., whether the generated base learner performs better than random guessing) are met at this step. If not, the current base learner should be abandoned, and the learning process should be terminated earlier than scheduled.
   (3) The coefficient of Gm(x) (i.e. the weight of the base learner used for the final ensemble) is calculated:
   \[ \alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \]  \( (4) \)
   (4) The weight of training samples is updated
   \[ D_{m+1} = (w_{m+1,1}, w_{m+1,2}, \ldots w_{m+1,N}) \]  \( (5) \)
   \[ w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i)), i = 1, 2, \ldots N \]  \( (6) \)
   Where, \( Z_m \) is a normalization factor, which is used to make the sum of all elements of \( D_{m+1} \) equal to 1. i.e.
   \[ Z_m = \sum_{i=1}^{N} w_{mi} \exp(-\alpha_m y_i G_m(x_i)) \]  \( (7) \)
   The final linear combination of classifiers is constructed
   \[ G(x) = sign(f(x)) = sign(\sum_{i=1}^{M} \alpha_m G_m(x)) \]  \( (8) \)
   The final classifier obtained is
   \[ G(x) = sign(f(x)) = sign(\sum_{i=1}^{M} \alpha_m G_m(x)) \]  \( (9) \)
   Therefore, when the error rate of the base learner Gm(x) is \( e_m \leq 0.5 \), and \( \alpha_m \) increases with the fall in \( e_m \), namely, a smaller classification error rate indicates a larger ratio of the base learner in the final ensemble. In other words, AdaBoost can adapt to the training error rate of each weak classifier, which is cause contributing to the origin of its name “Adaptive”.
   The weight of samples misclassified by the base learner Gm(x) is increased, while the weight of correctly classified samples is reduced.
   It should be noted that the sum of all \( \alpha_m \) dose not equal to 1 (because no softmax operation is conducted), and the sign of f(x) determines the forecast class, and its absolute value represents the certainty of classification.
3. Case analysis
The power consumption data of a province from January 2004 to December 2018 is used as the input data, the trained model is used to forecast the power consumption from January to December 2019, which is shown below:

![Forecast results of a provincial power consumption from January to December 2019](image_url)

Figure 2 Forecast results of a provincial power consumption from January to December 2019 (ensemble learning, Adaboost)

The following table shows the comparison between forecast data and actual data:

| Time  | Forecast value (10k kWe·h) | Actual value (10k kWe·h) | Forecast accuracy (%) |
|-------|--------------------------|-------------------------|-----------------------|
| 2019-12 | 1897958                  | 2160191                 | 87.86                 |
| 2019-11 | 1891427.2                | 1950864                 | 96.95                 |
| 2019-10 | 1891427.2                | 1781873                 | 93.85                 |
| 2019-09 | 1891427.2                | 1728784                 | 90.59                 |
| 2019-08 | 1891427.2                | 1874086                 | 99.07                 |
| 2019-07 | 1891427.2                | 1972117                 | 95.91                 |
| 2019-06 | 1840575.11               | 1865217                 | 98.68                 |
| 2019-05 | 1759816.67               | 1848608                 | 95.2                  |
| 2019-04 | 1897958                  | 1800463                 | 94.59                 |
| 2019-03 | 1840575.11               | 1840647                 | 99.99                 |
| 2019-02 | 1897958                  | 1715390                 | 89.36                 |
| 2019-01 | 1897958                  | 2080795                 | 91.21                 |
| Mean   |                          |                         | 94.44                 |
| Cumulative value | 22489934.89           | 22619035               | 99.43                 |

According to the aforementioned data, the average monthly forecast accuracy from January to December 2019 is 94.40%, demonstrating that a satisfactory monthly forecast accuracy is obtained. Additionally, the annual power consumption forecast accuracy in 2019 is as high as 99.43%, suggesting excellent application results. The forecast accuracy is relatively low in January, February, September, October and December, which is primarily attributable to the flexible adjustment of power consumption because of Spring Festival in January and February, and the great interval from the actual time to year-ending months.
4. Conclusion
Medium and long-term load forecasting is crucial to the safety of power grid planning and operation. With a view to exploring methods to address challenges regarding difficulties in modeling of the medium and long-term load forecasting due to increasing uncertainties of load development and small sample size of load data in the economic new normal, this study delved into the basic principle of ensemble learning and Boosting, deduced the theoretical model and the learning principle of adaptive ensemble learning, proposed Adaboost-based power system load forecasting, and validated the effectiveness of this method through the empirical forecasting of a provincial medium and long-term load, thereby achieving a high-accuracy medium and long-term load forecasting.

Acknowledgements
This work was supported by Guangxi Power Grid Dispatch and Control Center (Grant No.: 046000KK52200025).

Reference
[1] Zhong Q, Sun W, Yu NH, Liu CF, Wang F, Zhang X. Load and Power Forecasting in Active Distribution Network Planning. Proceedings of the CSEE 34, 3050-56 (2014)
[2] Zhao T, Wang LT, Zhang Y, Tian SM. Relation Factor Identification of Electricity Consumption Behavior of Users and Electricity Demand Forecasting Based on Mutual Information and Random Forests. Proceedings of the CSEE 36, 604-14 (2016)
[3] Y. M, Z. M, X. H, W. D. Research on mid-long-term load forecasting based on combination forecasting mode, 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) pp, 1-05 (2015)
[4] Kang CQ, Xia Q, Zhang BM. REVIEW OF POWER SYSTEM LOAD FORECASTING AND ITS DEVELOPMENT. Automation of Electric Power Systems 2004, 1-11 (2004)
[5] Wang N, Xie M, Deng JL, Liu MB, Li JL, Wang Y. Mid-long term temperature-lowering load forecasting based on combination of support vector machine and multiple regression. Power System Protection and Control 44, 92-97 (2016)
[6] Sun LJ, Liu D. An investigation of medium and long-term power load combination forecasting based on grey theory and regression analysis. Journal of North China Electric Power University (Social Science Edition) 2011, 132-34 (2011)
[7] Zhu MJ, Wang CB, Wang TZ, Liu ZW, Du MQ, Li Y. Adaptive Power Load Forecasting Method Based on Multi-Source Data. Water Resources and Power 35, 200-03 (2017)
[8] Zhang DL, Yan J, Li XB, Ren XD, Zhang JZ, Zhang FL. Mid-long-term load forecasting based on Markov chain screening combination forecasting models. Power System Protection and Control 44, 63-67 (2016)
[9] Xu WT, Li T, Liu YB, Yan XQ, Liu Y, Zhu M. Long term load forecasting method considering the influence of the power market transaction. Electrical Measurement & Instrumentation 54, 100-04 (2017)
[10] Wang S, Xue YD, Tong ZH, Song HM, Wang GQ, Liu HY. Variable weight combination method for mid-long term load forecasting based on golden section algorithm. Electrical Measurement & Instrumentation 53, 85-92 (2016)