**International Conference on Applied Energy 2019**
**Aug 12-15, 2019, Västerås, Sweden**
**Paper ID: 0060**

# SHORT-TERM LOAD FORECASTING AIMING AT CHINESE FESTIVALS USING TREND-BIAS PREDICTION AND VIRTUAL LOAD REPLACEMENT

Y. Yu[1], M. S. Li [1*], Z. G. Li[1], Q. H. Wu[1]
1 School of Electric Power Engineering, South China University of Technology, Guangzhou, 510641, China

## ABSTRACT

This paper proposes a novel load forecasting algorithm aiming at improving the accuracy of prediction near the holiday.

The improvement can be divided into two aspects: during the holiday, and a period after the holiday. Trend-Bias Prediction (TBP) algorithm is applied for predicting the load during the holiday, while Virtual Load Replacement (VLR) algorithm is used after a period of the holiday. The data in this experiment is from an industrial part in China. Comparing with the benchmark, both proposed TBP and VLR are efficient and have better performances.

**Keywords:** short-term load forecast, local predictor, Chinese festivals

## NONMENCLATURE

| | |
|---|---|
| *Abbreviations* | |
| TBP | Trend-Bias Prediction |
| VLR | Virtual Load Replacement |
| *Symbols* | |
| Dow | $\mathrm{Dow} \in (1,7)$, the day of week ( from Sunday to Monday) |
| lag2 | the same time as present moment in the two days ago |
| lag7 | the same time as present moment in the one week ago |
| isHoliday | boolean value, represents whether the forecast day is a festival or not |
| isHoliday2 | boolean value, represents two days before the forecast day is a festival or not |
| isHoliday7 | boolean value, represents one week before the forecast day is a festival or not |
| $\mathrm{Y_{pred}}$ | predictive value during the holiday |
| $data_t$ | *load of holiday in the* $\mathrm{t^{th}}$ *year.* |

## 1. INTRODUCTION

For optimal power system operation, electrical generation must follow electrical load demand. Generation utilities use electrical load forecasting techniques to schedule their generation resources to meet the future load demand. Transmission utilities use electric load forecasting techniques to optimize the power flow on the transmission network to reduce congestion and overloads [1].

The Short-term Load Forecasting (STLF) plays a key role in the formulation of economic, reliable, and secure operating strategies for the power system, they can be classified into three categories:

1) Statistical technique
2) Artificial intelligent technique
3) Hybrid techniques

Statistical approaches require an explicit mathematical model which gives the relationship between load and several input factors. Multivariate adaptive regression splines (MARS) is a kind of statistical technique for regression analysis [4].

Back-propagation algorithm (BP) is a commonly used Deep Learning algorithm. Performing back-propagation directly on a freshly initialized network may get stuck in poor local minima [5]. Stacked Autoencoder (SAE) are trained in an unsupervised, greedy, layer-wise fashion. It begins training with just the first layer of the network and then add new layers gradually. And Deep Belief Network is a probabilistic generative model created by stacking multiple restricted Boltzmann machines (RBMs) on top of each other. It has already been applied successfully to solve many problems [6]. Support Vector Machine (SVM) is also a popular algorithm [7]. It applies the structural risk minimization principle which can avoid being trapped into the local optimal solution. However, it may be disappointment if the parameters are not tuned properly.

In the real world, we are unable to obtain data in time because of some realistic reasons. In this paper, we have to use yesterday's data to forecast tomorrows load, which is equal to a 24-steps to 48-steps forecasting problem. Due to the lag of getting data, most of the models perform not well during the holiday period. Thus, we proposed Trend-bias Algorithm (TBA) and Virtual Load Replacement (VLR). The results indicate that the accuracy of the proposed Trend-bias and Virtual Load Replacement algorithm outperforms other algorithms with stronger reliability.

The rest of the paper is organized as follows. Section 2 provides a brief description of basic principles about MARS, BPNN, SAE and DBN. Also, TBA and VLR are be introduced. Section 3 contains the specific experiments, which indicate that the algorithm proposed is effective in forecasting holidays load. Section 4 provides the conclusion.

## 2. LOAD FORECASTING AIMING AT CHINESE FESTIVALS PAPER STRUCTURE

### *2.1 Feature Selection*

The load data we analyzed is obtained from an industrial park in China. It is a complete micro-grid system, which including distribution generation, control system, energy storage device and so on. The load has obvious diurnal change and cycle changes. It is shown in Fig. 1. Thus, we choose the same time as the predicted hour in the two days ago as the feature "lag2", and the same time as the predicted hour in one week ago as the feature "lag7". "lag2" is the latest and correlative load we can obtained, which is also represent for the diurnal change. "lag7" stands for the cycle change.

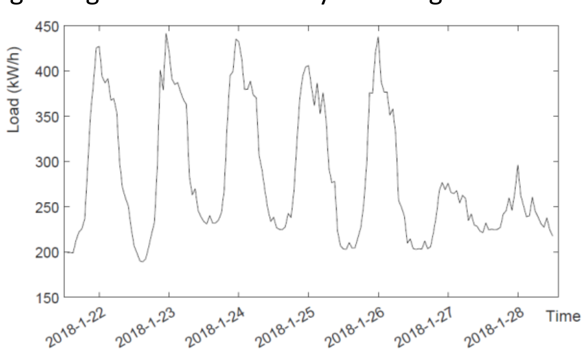


Fig.1 Typical load trends in this industrial park

During data exploration, we found that the load in festivals is much lower than an ordinary day. Thus, the day is a festival or not will influence the load definitely. In conclusion, the features we choose are as follow:

$$[\mathrm{DoW}, \mathrm{lag2}, \mathrm{lag7}, \mathrm{isHoliday}, \mathrm{isHoliday2}, \mathrm{isHoliday7}] \tag{1}$$

### *2.2 Improved Algorithm aiming at Chinese Festival*

#### *2.2.1 Universal Forecasting Algorithms*

*MARS is a flexible regression modeling of statistical technique. The model takes the form of an expansion in product spline basis functions, where the number of basis functions as well as the parameters associated with each one is automatically determined by the data. BPANN, DBN and SAE all belong to artificial intelligent technique. SVM applies the structural risk minimization principle which can avoid being trapped into the local optimal solution [9].*

*For fairness and validation in comparison, all neural net- works have the same network structure. The network includes four layers, consisting of input layer (6 input neurons), 1st hidden layer (24 hidden neurons), 2nd hidden layer (12 hidden neurons) and the output layer (one output neuron).*

#### *2.2.2 Trend-Bias Prediction Algorithm*

There are several holidays in China, including the Spring Festival, Tomb-sweeping Day, Labor Day and so on. Most of them are three-day-holiday (sometimes, it will be four-day), but the Spring Festival is slightly different. It will last ten days.

Each holiday has similar pattern in different years. Moreover, the load will slightly increase with the development of industrial park every year. Thus, we construct the formula to calculate the holidays load as follow:

$$\mathrm{Y_{pred}} = \mathrm{trend} + \mathrm{meanValue} \times \mathrm{ratio}$$

$$\mathrm{trend} = a_{t+1} \times data_{t+1} + \cdots + a_{t+n} \times data_{t+n}$$

$$\mathrm{ratio} = a_{t+1} \times rate_{t+1} + \cdots + a_{t+n} \times rata_{t+n}$$

$$rate_t = \frac{data_t}{data_{t-1}} \tag{2}$$

subject to:

$$\mathrm{a_{t+1}} + \mathrm{a_{t+2}} + \cdots + \mathrm{a_{t+n}} = 1$$

The variable of trend is calculated by load of holiday in many years. $a_t$ ∈(0, 1), which will be set to the most appropriate value according to the experiments. n means that we have n years load data. Restrictive condition makes sure sum of the proportion will be 1.

The variable trend represents the loads performance during the holiday. The variable ratio is the representative of loads growth. The forecasting load will be the most relevant to the load of the same festival in the last year. Considering actual status, we can get the real load during the holiday in this year. Then we will use the mean value of real load to instead the value of trend we calculate.

*2.2.3 Virtual Load Replacement Algorithm*

The load during the holiday is lower than usual. If features of lag2 or lag7 are in the holiday, the predictive value will be much lower. Considering that the load has weekly cycle, if lag2 or lag7 are in the holiday, we will replace them with the load in 2/7 days before. And if 2/7-days load before is still in the holiday, we will keep finding the load before them until we find the appropriate load. These loads will be considered as the virtual load. The flow graph is illustrated in Fig. 5.

## 3. EXPERICAL STUDY

*3.1 Experimental Settings*

The load data used for this paper is collected from an industrial park in China. The data was sampled every hour from January 1, 2016 to December 31, 2018. We choose the quality load data before January 1, 2018 for training, and data between January 1, 2018 and May 8, 2018 for testing. We also need the schedule of holidays between 2016 and 2018. These data will be transformed into the format like Equ 1, consisting of 6 features. In the training process, training data will be picked at random.

Usually, we use Mean Absolute Percentage Error (MAPE) to evaluate the accurate of forecasting results. However, the average of each points may mask some problems during the holiday's predictions. And the capability to predict the upper limits of load is quite important for Power Grid company for transmission capacity. In order to reflect reality of prediction from a multitude of perspectives, we choose the MAPE of maximum value $(\mathrm{MAPE}_{max})$ and average value of Root Mean Square Error (RMSE) as evaluating standard system in the predictions near festivals. The formulas are shown as following:

$$\mathrm{MAPE_{max}} = |\frac{\mathrm{Y_{max}} - \overline{\mathrm{Y}}_{\mathrm{max}}}{\mathrm{Y_{max}}}|$$

Where $Y_{max}$ is the maximum true value in the day. $\bar{Y}_{max}$ is the predictive value at the same time. Smaller $\mathrm{MAPE}_{max}$ indicates that the predictive value is more accurate.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N_s}}\sum_{\mathrm{i=1}}^{\mathrm{N_s}}(\mathrm{Y_i} - \overline{\mathrm{Y}}_{\mathrm{i}})^2}$$

where $N$ is the number of samples, $Y_i$ is the forecast output of $i_{th}$ sample and $\bar{Y}_i$ is the actual measure of $i_{th}$ sample. Smaller RMSE shows that the trend of predictive values is more similar to true values.

*3.2 Experimental Results*

*3.2.1 Predictions of Universal Algorithms*

The prediction results of universal algorithms mentioned above is shown in Table I.

*Table I Prediction results of universal algorithms*

| *Algorithm* | *MAPE* | *Rank* |
|---|---|---|
| *MARS* | *0.1648* | *5* |
| *SVM* | *0.1427* | *3* |
| *BP* | *0.1341* | *2* |
| *DBN* | *0.1436* | *4* |
| *SAE* | *0.1331* | *1* |

In general, artificial intelligent technique outperforms other algorithms when there are many input neurons. It can find out pattern hidden in the dataset. However, Table I indicates that it still performs a little better than statistical algorithms, although input are just six neurons. Prediction accuracy of BPANN, SAE and DBN are very similar. Taking into account these situations, we choose the forecasting results of SAE algorithm as our benchmark.

*3.2.2 Implementation of Trend-Bias Prediction Algorithm*

Considering that we just get the load data between January 1, 2016 and May 9,2018, the n of Equation 2 is set to 2. And $a_1$ is set to 0.2, $a_2$ is set to 0.8. Table II shows the RMSE and $\mathrm{MAPE}_{max}$ before and after processed by the TBP.

While in some points which have been predicted very worse, the proposed algorithm has achieved remarkable effects. For instance, RMSE and $\mathrm{MAPE}_{max}$ both decrease more than 300% on April 29, 2018 and February 14, 2018. In general, the proposed algorithm is effective.

*Table* II RMSE and $MAPE_{max}$ before and after processed by TBA

| *Festival* | *Date* | *RMSE_before* | *RMSE_after* | *MAPE_before* | *MAPE_after* |
|---|---|---|---|---|---|
| *Spring Festival* | *Feb 13,2018* | *23.4560* | *45.5201* | *0.1384* | *0.1439* |
| | *Feb 14,2018* | *153.5591* | *42.0736* | *1.4853* | *0.2398* |
| | *Feb 15,2018* | *42.6909* | *3.4394* | *0.5082* | *0.0084* |
| | *Feb 16,2018* | *48.0380* | *3.1608* | *0.5930* | *0.0201* |
| | *Feb 17,2018* | *67.2563* | *3.4153* | *0.8341* | *0.0263* |
| | *Feb 18,2018* | *52.5798* | *3.4279* | *0.2065* | *0.0093* |
| | *Feb 19,2018* | *29.6343* | *4.0682* | *0.1227* | *0.0153* |
| | *Feb 20,2018* | *38.0847* | *20.6565* | *0.2412* | *0.3237* |
| | *Average* | *54.4588* | *14.3796* | *0.4687* | *0.1002* |
| *Tomb-sweeping Day* | *Apr 5,2018* | *27.7601* | *28.4349* | *0.2241* | *0.1602* |
| | *Apr 6,2018* | *26.0564* | *25.1003* | *0.3208* | *0.1018* |
| | *Apr 7,2018* | *28.0103* | *28.2830* | *0.2420* | *0.3024* |
| | *Average* | *27.2756* | *27.2727* | *0.2623* | *0.1887* |
| *Labor Day* | *Apr 29,2018* | *89.4735* | *28.7833* | *1.2399* | *0.3879* |
| | *Apr 30,2018* | *34.2698* | *15.1119* | *0.5478* | *0.1703* |
| | *May 1,2018* | *22.0691* | *15.9948* | *0.1681* | *0.0043* |
| | *May 2,2018* | *16.4843* | *12.5391* | *0.2902* | *0.1424* |
| | *Average* | *40.5742* | *18.1073* | *0.5615* | *0.1762* |

Fig.2 shows the true values, and predictions before and after processed by the proposed algorithms during the holidays. We can reach the same conclusions as the Table II.

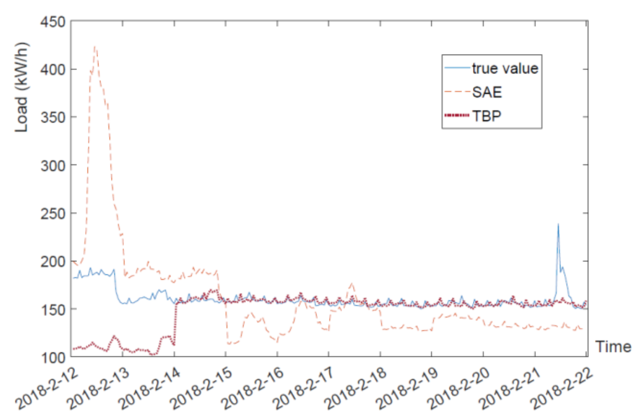
### *3.2.3 Implementation of Virtual Load Replacement*

Table III shows the similar results to the predictions during the holiday. The RMSE or $\mathrm{MAPE}_{max}$ will be slightly increase in some time, but the general effect of prediction is very pleasing.

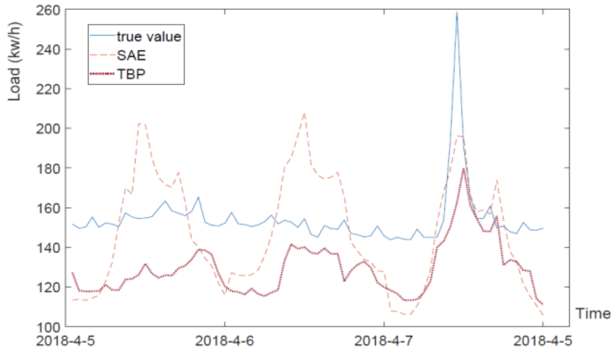
The true values and predictions before and after processed by VLR is shown in Fig. 3. It indicates that the predictions improve a lot after processed.

## *4. Conclusion*

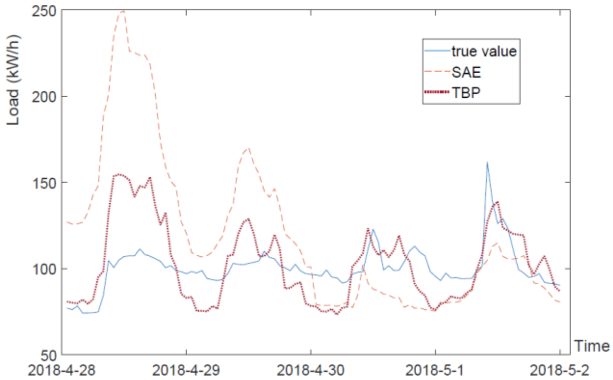
This paper proposes a novel load forecasting algorithm aiming at improving the prediction near the holiday. The improvement can be divided into two aspects: during the holiday, and a period after the holiday. During the holiday, we predict the load through trend and ratio which are calculated by load in the recent years. In the prediction of a period after the holiday, we construct virtual load as feature for prediction. Then we get the more accurate results through these virtual load values in the next week after the holiday. The proposed algorithm is a universal improvement method, which is applicable for any prediction algorithms.




*(a) Spring Festival*

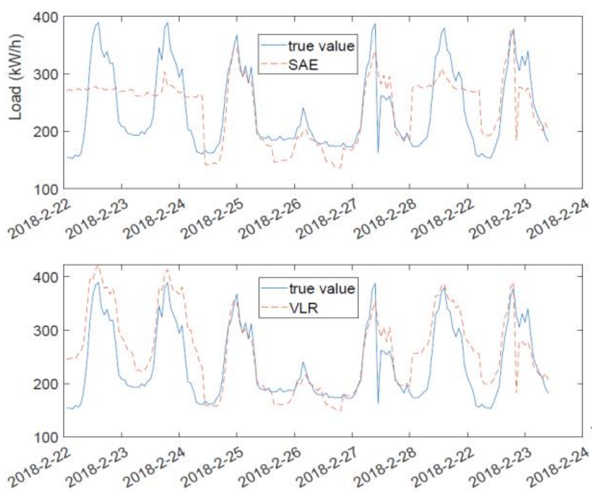



*(b) Tomb-sweeping Day*

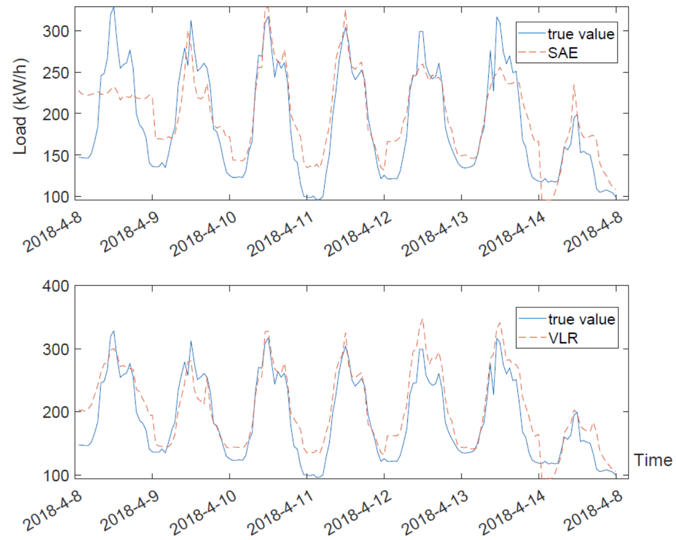



*(c) Labor Day*

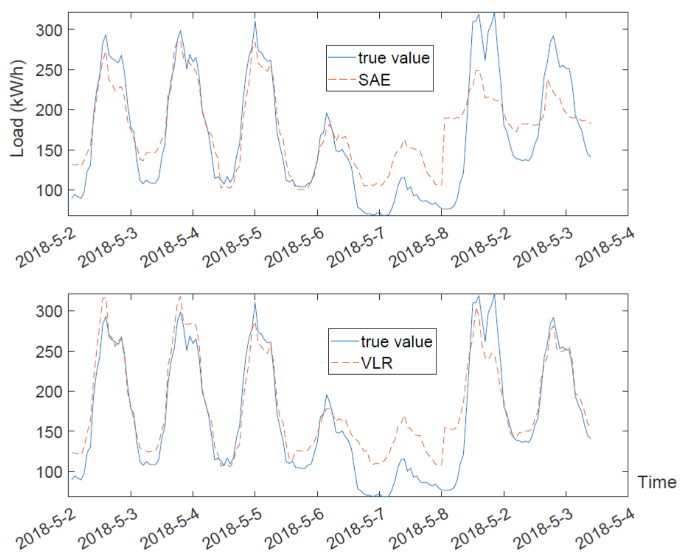
*Fig.2True values and predictions during the festivals by TBP*




*(a) Spring Festival*




*(a) Tomb-sweeping Day*




*(b) Labor Day*

*Fig.3 True values and predictions after the festival by VLR*

*Table* III RMSE and $MAPE_{max}$ before and after processed by VLR

| *Festival* | *Date* | *RMSE_before* | *RMSE_after* | *MAPE_before* | *MAPE_after* |
|---|---|---|---|---|---|
| *Spring Festival* | *Feb 22,2018* | *83.1227* | *74.0599* | *0.2839* | *0.0865* |
| | *Feb 23,2018* | *66.8016* | *59.8589* | *0.2185* | *0.0635* |
| | *Feb 24,2018* | *20.5911* | *13.6996* | *0.0597* | *0.0330* |
| | *Feb 25,2018* | *29.1397* | *18.4895* | *0.1557* | *0.1014* |

| | | | | | |
|---|---|---|---|---|---|
| | *Feb 26,2018* | *34.9929* | *35.9047* | *0.1233* | *0.0857* |
| | *Feb 27,2018* | *72.3072* | *63.9528* | *0.1850* | *0.0195* |
| | *Feb 28,2018* | *43.2922* | *45.1337* | *0.0037* | *0.0323* |
| | *Average* | *50.0353* | *44.4427* | *0.1471* | *0.0603* |
| *Tomb-sweeping Day* | *Apr 8,2018* | *54.9279* | *43.6723* | *0.3211* | *0.2455* |
| | *Apr 9,2018* | *23.1238* | *20.3275* | *0.0156* | *0.0450* |
| | *Apr 10,2018* | *23.1663* | *23.0498* | *0.0117* | *0.0181* |
| | *Apr 11,2018* | *27.4760* | *28.8008* | *0.0330* | *0.0299* |
| | *Apr 12,2018* | *22.5847* | *23.7335* | *0.0839* | *0.0874* |
| | *Apr 13,2018* | *23.9361* | *21.8961* | *0.0514* | *0.0382* |
| | *Apr 14,2018* | *26.9093* | *23.0001* | *0.1429* | *0.0308* |
| | *Average* | *28.8749* | *26.3543* | *0.0941* | *0.0707* |
| *Labor Day* | *May 2,2018* | *28.3619* | *22.2275* | *0.0688* | *0.0808* |
| | *May 3,2018* | *22.0142* | *16.9557* | *0.0415* | *0.0654* |
| | *May 4,2018* | *18.1701* | *16.0673* | *0.0844* | *0.0761* |
| | *May 5,2018* | *20.1294* | *28.3967* | *0.0717* | *0.0902* |
| | *May 6,2018* | *44.0407* | *45.3493* | *0.4144* | *0.4743* |
| | *May 7,2018* | *73.4555* | *49.1047* | *0.2247* | *0.0511* |
| | *May 8,2018* | *41.6655* | *11.1627* | *0.1843* | *0.0353* |
| | *Average* | *35.4053* | *27.0377* | *0.1557* | *0.1247* |

**REFERENCE**

[1] Taylor and E. Lynn, "Short-term electrical load forecasting for an institutional/industrial power system using an artificial neural network," 2013.
[2] C. Tian, C. Tian, H. Yan, and H. Yan, "A novel nonlinear combined forecasting system for short-term load forecasting," Energies, vol. 11, 2018.
[3] G. Yi, H. Jiang, J. J. Zhang, Y. Zhang, E. Muljadi, and F. J. Solis, "Load forecasting based distribution system network reconfiguration-a distributed data-driven approach," 2018.
[4] A. Martin, V. Gayathri, G. Saranya, P. Gayathri, and P. Venkatesan, "A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means and mars," International Journal on Soft Computing, vol. 2, no. 1, pp. 12–24, 2011.
[5] Y. U. Feng and X. U. Xiaozhong, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved bp neural network," Applied Energy, vol. 134, no. 134, pp. 102–113, 2014.
[6] X. Zhang, W. Rui, Z. Tao, W. Ling, Y. Liu, and Y. Zha, "Short-term load forecasting based on rbm and narx neural network," 2018.
[7] H. Xie, J. P. Wei, and H. L. Liu, "Parameter selection and optimization method of svm model for short-term load forecasting," Proceedings of the Csee, vol. 26, no. 22, pp. 17–22, 2006.
[8] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of macedonian case," Energy, vol. 115, pp. 1688–1700, 2016.
[9] J. Huo, T. Shi, and J. Chang, "Comparison of random forest and svm for electrical short-term load forecast with different data sources," in 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Aug 2016, pp. 1077 – 1080.
[10] X. Chen, J. Qiu, and Z. Y. Dong, "An improved load forecast model using factor analysis: An australian case study," in Information and Automation (ICIA), 2017 IEEE International Conference on. IEEE, 2017, pp. 903–908.