Deep neural network based load forecast

Wan He*

State Grid Energy Research Institute, State Grid Corporation of China

Received 1 March, 2014, www.tsi.lv

Abstract

Accurate electrical load forecast has great economic and social value. In this paper, we study deep neural networks based load forecast approaches. We first analyse the critical features related to load forecast. Then we present details of deep neural networks and pre-training technologies, including RBM pre-training and discriminative pre-training. We compare the performances of different neural network models and show the advantages of the proposed methods using a rather large data set of loads.

Keywords: Load Forecast, Deep Neural Networks, Pre-training, RBM

1 Introduction

In modern society, electrical power plays a key role in supporting national economy. Moreover, with the rapid development of China's economy and society, the demand for electricity is continuously growing. However, within a certain period, the electricity demand may be affected by a number of factors and varies greatly. Moreover, an important feature of electricity is that it is difficult to store it. Thus, when the electricity supply is greater than consumption, it will bring wastes and cause losses to power companies. Therefore, we need to forecast the load of electrical power, to guide electricity production and scheduling.

Because of the important economic and social significance, great attention has been paid on electrical load prediction research. The EUNITE load forecast completion held in 2001 attracted 56 registered competitors from 21 countries. In addition, 105 teams from well-known universities and companies, such as Cambridge University and Petrobras, joined the load forecast subtask in Global Energy Forecasting Competition 2012 (GEFCom2012).

From a broad perspective, load forecasting is a time series prediction problem, with many similar problems, such as stock price prediction, oil risk prediction, etc. According to different planning horizon lengths, load predictions can be roughly classified into three categories: one hour to one week for short-term forecast, a week to a year for medium forecasts, and longer than a year for long-term forecasts [1]. We focus on short-time prediction in this paper.

To improve the performance of load forecasts, various approaches have been proposed. Mbamalu and El-Hawary [2] used regression analysis and they represented load as a function of factors such as time and weather conditions, while Kiartzis and Bakirtzis [3] introduced fuzzy logic to predict daily peak load. Since the sequence of load data can be described as a time series, prediction methods for time series, which are widely used in economics can be applied. Taylor considered several recently developed exponentially weighted methods for load forecasting and compared their performances [4]. Huang and Shih [5] proposed an autoregressive moving average (ARMA) model including non-Gaussian process considerations and applied it on a practical power system.

Machine learning methods have also been widely used in load forecast. The most commonly used machine learning methods include Support Vector Machines [6], Gaussian process [7] and artificial neural networks [8]. Load forecast is a very complicated problem, which is highly nonlinear and has no simple analytical formulas. Neural networks are quite suitable for forecasting loads for they can also be highly nonlinear, and can approximate any complex function when they have enough nodes in hidden layer or have enough number of hidden layers. Since Peng et al. first proposed a neural network based approach to tackle the influence of holidays and drastic changes in weather patterns [8], a lot of researchers employed neural networks to handle various problems in load forecast [9, 10]. The method proposed in this paper is also based on neural network.

Although a neural network with only one hidden layer can represent arbitrarily complex functions when the number of hidden layer neurons is large enough, a network with multiple hidden layers not only has many theoretical advantages, but also brings practical benefits [11, 12]. Recently, deep neural networks have achieved great success in image processing and speech recognition, and led to the rise of deep learning and the "renaissance" of neural network research. In this paper, we propose a method that is based on deep neural networks and combines a rich set of electrical load related features to predict one day ahead hourly load. We demonstrate the advantages of our method through carefully designed experiments using load data in a city of north China.

* ***Corresponding author*** e-mail: streamlethe@126.com

The rest of this paper is organized as follows. In Section 2, we discuss the characteristics of electrical load series and various factors related to load forecast. After that, we detail our methods in Section 3. Then we provide experiments and evaluation results in Section 4. Finally, we conclude our work in Section 5.

2 Factors for Load Forecast

The actual electrical loads are influenced by a variety of factors. In this section, we analyse some of the most important factors. On the basis of these analyses, we consider extracting representative features which are used as input of our deep neural network model for load prediction. We focus on load periodicity, time dependency, holiday effect and weather influence in the following.

Periodicity

Short-term electrical load usually exhibits remarkable periodicity. Intuitively, electricity consumption is directly associated with daily work and rest patterns of people. Indeed, by plotting the hourly loads of 7 consecutive days in figure 1, it is rather obvious that hourly loads in adjacent days have similar patterns.

Considering the daily periodicity of load sequences, three types of features can be extracted. Firstly, the average load of the previous day is a significant feature since it is often close to that of current day. Secondly, loads from the same hour of the previous days are good indicators for the current hour's load. Finally, loads from the previous hours of yesterday (might extend to hours of even earlier days) are also good indicators for predicting the current hour's load. As an example, loads at 3:00 and 4:00 AM on yesterday can help forecast load at 5:00 AM today.

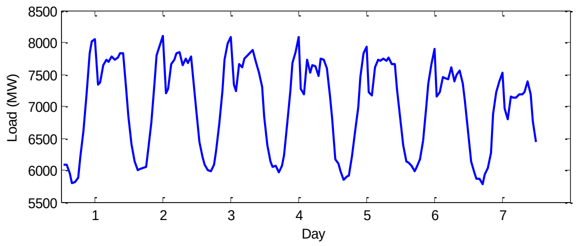

FIGURE 1 Daily periodicity of loads

In addition, load data in adjacent weeks show strong positive correlations. Figure 2 shows how load sequences of three adjacent weeks match one another. The data used in figure 2 come from hourly loads from April 4th to April 24th 2011 in our data set. So load from the same hour of same day of the previous week is used a type of feature.

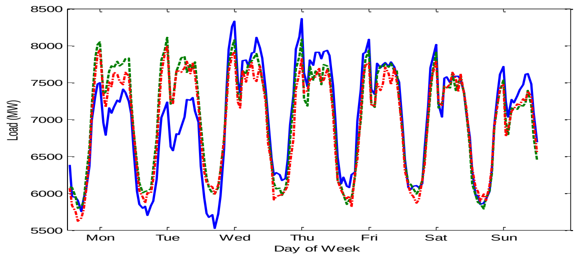

FIGURE 2 Weekly periodicity of loads

Time dependency

Figure 1 and figure 2 demonstrate how the load values change over time. So load can be expressed as a stochastic function of time. Based on this, hour of day and day of week are two features in our methods.

Holiday effect

Earlier works have pointed out that holiday is a type of factor for load. It is easy to find out that load on holidays is lower than usual. For example, in our data the average load of May 1st 2011 is 12.6% lower than that of April 25th. We add a feature to indicate whether the current day is holiday or not.

Weather influence

It is common sense that weather factors have a large impact on electrical load. On one hand, some production works are arranged according to weather conditions. On the other hand, people consume electricity to regulate temperature if it is too high or too low. Thus, daily maximum temperature and minimum temperature are used as features. Hourly temperature may be more informative, but unfortunately we cannot get them. Figure 3 show the correlation between daily average load and maximum temperatures in our 2011 data.

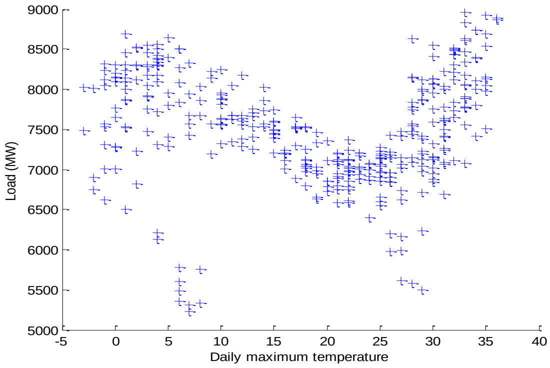

FIGURE 3 Correlation between daily average load and temperature

3 Deep Neural Network Based Load Forecast

Neural networks have the ability to learn complex nonlinear relationship between input patterns and the target to predict. The method proposed in this paper is based on deep neural networks (DNN). In general, a DNN is a hybrid model that combines traditional multilayer perceptron and recently developed pre-training technologies. Although many papers on load forecast employ neural networks, the research of neural networks became much less hot in 1990s [12]. The situation continued until Hinton et al. proposed Deep Belief Network (DBN) and greedy layer-wise pre-training methods using Restricted Botzmann Machine (RBM) in 2006 [11]. Deep neural networks have received extensive attention and achieved remarkable success in industry since then. In the following, we will first briefly introduce neural network (NN) and then introduce the two pre-training methods used in this paper: RBM pre-training and discriminative pre-training.

The neural network in this paper is standard feed forward multilayer perceptron structured with one input layer, one or multiple hidden layers and one output layer. The input to the network is a vector of various features extracted from the load sequence and meta-information such as weather and holiday. Each hidden layer receives a vector of input from the previous layer and converts it to its output vector through a linear transformation followed by a nonlinear activation.

Assume the set of patterns and corresponding target values are $D=\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),...,(x^{(T)},y^{(T)})\}$. For each pair $(x^{(p)},y^{(p)})$, $x^{(p)}$ is the input to the network. Since the input layer in our network is an identity mapping, its output is also $x^{(p)}$. For neuron j in hidden layer l, supposing the inputs from the previous layer are $\mathbf{a}^{(l-1)}=(a_1^{(l-1)},a_2^{(l-1)},...,a_n^{(l-1)})^T$, it outputs

$$a_j^{(l)}=f\left(\sum_{i=1}^{n}w_{ji}^{(l)}a_i^{(l-1)}+b_j^{(l)}\right), \tag{1}$$

where $w_{ji}^{(l)}$ and $b_j^{(l)}$ are parameters that represent weights and bias in layer l, and f is the activation function which is the sigmoid function in this paper

$$f(z)=\sigma(z)=\frac{1}{1+e^{-z}}. \tag{2}$$

Note that in our work the output layer has a single neuron because the target to predict is hourly load, which is a scalar value. Moreover, the output layer applies only the linear transformation and uses no nonlinear activation function.

For each input feature vector $x^{(p)}$, let $o^{(p)}$ denote the corresponding output of the neural network. The expected load value $o^{(p)}$ should be close to the actual load $y^{(p)}$. The network adjusts its parameters (weights and biases) to learn good estimation of $y^{(p)}$ by minimizing the square error function $E=1/2\cdot\sum_{t=1}^{T}(o^{(t)}-y^{(t)})^2$. In the optimization process, Back Propagation (BP) algorithm is used to compute the derivatives of parameters.

The standard BP algorithm is slow and tends to get trapped in local minima. In order to find better parameters and accelerate the optimization process, we use Conjugate Gradient method, which is a kind of second order optimization algorithm [13]. We refer to the implementation of the Polack-Ribiere conjugate gradient method as in [11].

For neural networks with multiple hidden layers, the BP training with random initialization of parameters frequently gets trapped in bad parameters. In addition, in such cases, the performance of the network is usually worse than networks with a single hidden layer. To solve this problem, Hinton et al. first introduced a greedy layer-wise pre-training method [11, 12] by use of Restricted Boltzmann Machines (RBM), to find better initialization of neural network parameters. Their method has achieved huge success and lead to the rise of "deep learning".

Restricted Boltzmann Machine is a special type of energy-based models whose energy function is bilinear [11, 12]. The energy function of an RBM is defined as:

$$E(\mathbf{v},\mathbf{h})=-\mathbf{b}^T\mathbf{v}-\mathbf{c}^T\mathbf{h}-\mathbf{h}^T\mathbf{W}\mathbf{x}, \tag{3}$$

where $\mathbf{v}$ and $\mathbf{h}$ are visible and hidden variables respectively, and $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{W}$ are parameters. Because of the particular structure of RBMs, visible and hidden units are conditionally independent given one another. According to this, the conditional probabilities $p(\mathbf{h}\,|\,\mathbf{v})$ and $p(\mathbf{v}\,|\,\mathbf{h})$ can be written as

$$\begin{aligned}p(\mathbf{h}\,|\,\mathbf{v})&=\prod_i p(h_i\,|\,\mathbf{v})\\ p(\mathbf{v}\,|\,\mathbf{h})&=\prod_j p(v_j\,|\,\mathbf{h})\end{aligned}. \tag{4}$$

If the values of $\mathbf{v}$ and $\mathbf{h}$ are further limited to set $\{0,1\}$, the conditional probabilities of variables can be expressed as,

$$\begin{aligned}p(h_i=1\,|\,\mathbf{v})&=\sigma(c_i+\mathbf{W}_i\mathbf{v})\\ p(v_j=1\,|\,\mathbf{h})&=\sigma(b_j+\mathbf{W}_j^T\mathbf{h})\end{aligned}, \tag{5}$$

where $\mathbf{W}_i$ and $\mathbf{W}_j$ are the *i*-th and *j*-th row of $\mathbf{W}$, respectively. Based on formula 5, an RBM can be trained effectively using 1-step Contrastive Divergence. For more details on RBM and Contrastive Divergence, please refer to [12, 11].

RMBs are building blocks for DBNs because they share parameters with neural networks. The parameters $\mathbf{W}$ and $\mathbf{c}$ are also weights and biases in the corresponding layer of the neural network. The initial values of weights and biases of each hidden layer can be obtained by layer-wise learning of RBMs consists of the current hidden layer and the previous layer. In this situation, the current hidden layer corresponds to the hidden variable of the RBM, while the previous layer corresponds to visible layer. Please refer to [12] and [11] for details on the construction of DBN using RBM and greedy layer-wise pre-training methods.

Besides RBM, we employ another type of pre-training, discriminative pre-training. Hinton et al. have pointed out that discriminative pre-training may perform better [14]. The RBM does not require the target value of each sample (pattern). Therefore, it is good for unsupervised learning. However, in our load data, all samples have the corresponding targeting load values. Therefore, we can pre-train the neural network discriminatively. We do this by starting from a network with a single hidden layer and increasing hidden layers gradually. The following pseudo-code illustrates this process.

Input: training data set D, hidden layer size vector L;
for l from 1 to length(L) – 1
construct a neural network with l hidden layers according to L;
if $l > 1$ then
use $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$ from the last run to initialize $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$, k=1, …, l-1;
end
randomly initialize other parameters;
train the network using data set D;
save parameters $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$, k=1, …, l;
end
Output: values of parameters $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$, l=1, …, length(L) – 1;

4 Experimental Results

In this section, we describe the evaluation of the proposed DNN based load forecast methods. First, we briefly introduce how we create the data set. Then we present our experimental results with quantitative evaluation of our methods on this data set and some qualitative discussions.

We prepared a data set by collecting hourly load data of a city in north China. Then we collected weather information about this city on the Internet. The final data set contains load and weather from February 10st 2000 to November 30th 2012. Hourly temperature is of great significance for predicting load of that hour. However, we can only get the highest and lowest daily temperature. The data in November 2012 are used as testing data, while data in October 2012 are used as validation data. All other data are used as training data for the neural networks. Thus, we have 23136 training samples (hourly load to predict), 744 validation samples and 720 testing samples.

Our task is to provide one day ahead prediction of hourly loads. In other words, we are to forecast the load at an hour using weather information of the targeting hour and historical load information of 24 hours before. For example, to predict the load at 12:00 on July 7th, we can only use historical loads up to 12:00 on July 6th.

In section 2, we have described the types of features used in our system. The final feature representation of each sample is a vector of 10 dimensions: 3 dimensions for history load at the same hour in previous days, 1 for average load of previous day, 1 for load at the same hour of the same day in previous week, 2 for daily highest and lowest temperatures respectively, 1 indicates day of week and 1 indicates hour of day, and 1 indicates whether it is a holiday or not. So the input layers of the neural networks have 10 neurons. We have tried different number of hidden layers and different size for each hidden layer. The best performed network has three hidden layers and each hidden layer has 30 neurons.

The common way to train a DNN contains two steps, an optional pre-training step and a final tuning step. For pre-training, we have tried RBM pre-training and discriminative pre-training, and a network with no pre-training. The performances of different methods are presented in the following table, where performance of a very strong baseline with only one hidden and 100 hidden neurons is also provided.

TABLE 1 Performance of different methods

Method	MAPE (%)	MAE (MW)
NN single hidden layer	2.07	179.0
DNN no pre-train	2.08	178.1
DNN RBM pre-train	1.98	169.7
DNN disc pre-train	1.90	164.2

The performance metrics are mean absolute percentage error (MAPE) and mean average error (MAE), which are computed according to formula (6) and formula (7), respectively,

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y^{(i)} - o^{(i)}}{y^{(i)}}\right| \times 100\,, \tag{6}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y^{(i)} - o^{(i)}\right|, \tag{7}$$

where $y^{(i)}$ is the actual value of load, $o^{(i)}$ is the prediction given by neural network, and N is the number of testing samples. We can see from table 1 that either the DNN using RBM pre-training or the one using discriminative pre-training outperforms other methods. In addition, the DNN using discriminative pre-training performs the best. Moreover, even without temperature of each hour, the DNN based methods can produce very promising prediction results.

In the following figure, we presented the actual load values and the forecasted values for the testing data, so as to get an intuitive vision. We can see that the forecasted load is very close to the actual value in most of the cases. By further analysing the data, we find that our methods perform much worse for weekend loads. The MAPE for is 2.93% for weekends in the testing set. This suggests that we need to put more efforts on improving weekend load forecast in the future.

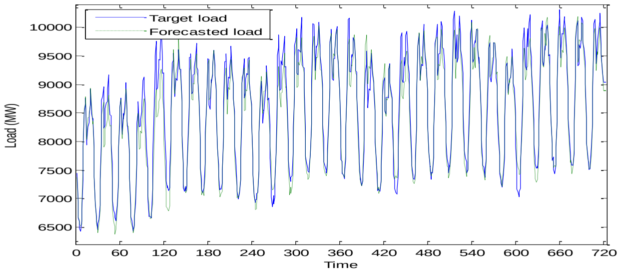

FIGURE 4 Actual load versus forecasted load

5 Conclusions

In this paper, we proposed to forecast electrical load using deep neural networks, which have attracted much attention in research and achieved great success in industry recently. We first introduced the various features we used to forecast load. Then we presented the details of deep neural networks as well as RBM pre-training and discriminative pre-training technologies. Finally, we evaluated the performances of deep neural networks in load forecast using a large data set, which contains nearly three years loads. The experimental results showed the advantages of deep neural networks in load prediction.

Load forecast is a very complex problem and can be improved in a number of ways. First, hourly temperature is critical for hourly load, and we need to improve the performance of our system with this type of data. Second, loads in weekends show different patterns and are more difficult to predict. We will pay more attention to improving the accuracy in weekend load prediction.

References

[1] Feinberg E A, Genethliou D 2005 *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence* Spinger 269-85

[2] Mbamalu G A N, El-Hawary M E 1993 Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation *IEEE Transactions on Power Systems* 8(1) 343-8

[3] Kiartzis S J, Bakirtzis A G 2000 A fuzzy expert system for peak load forecasting: application to the Greek power system *Proceedings of the 10th Mediterranean Electrotechnical Conference* **3** 1097–100

[4] Taylor J W 2012 Short-term load forecasting with exponentially weighted methods *IEEE Transactions on Power Systems* 27(1) 458-64

[5] Huang S J, Shih K R 2003 Short-term load forecasting via ARMA model identification including non-Gaussian process considerations *IEEE Transactions on Power Systems* **18**(2) 673-9

[6] Chang M W, Chen B J, Lin C J 2001 EUNITE network competition: electricity load forecasting

[7] Lloyd J R 2012 Hierarchical load forecasting using gradient boosting machines and Gaussian processes *GEFCom2012*

[8] Peng T M, Hubele N F, Karady G G 1992 Advancement in the application of neural networks for short-term load forecasting *IEEE Transactions on Power Systems* **7**(1) 250-7

[9] Zhu Y, Fang G 2012 Short-term load forecasting based on dynamic adaptive artificial neural network *Power System Protection and Control* **40**(1)

[10] Shi H, Niu X, Lu Y 2012 The short-term load forecasting model based on bayesian neural network *Chinese Journal of Management Science* **20**(4) 118-24

[11] Hinton G E, Osindero S, Teh Y 2006 A fast learning algorithm for deep belief nets *Neural Computation* **18** 1527-54

[12] Bengio Y 2009 *Learning Deep Architectures For AI: Foundations and Trends in Machine Learning* 1-127

[13] Nocedal J, Wright S J 2006 *Numerical Optimization* Second Edition

[14] Mohamed A, Dahl G E, Hinton G E 2012 Acoustic modeling using deep belief networks *IEEE Transactions. on Audio, Speech, and Language Processing* **20** 14-22

Authors

Wan He, born in December 2, 1981, Beijing, China

Current position, grades: PhD of economics, senior researcher in State Grid Energy Research Institute of State Grid Corporation of China
University studies: Quantitative Economics in Renmin University of China
Scientific interest: Electricity Markets and Power System Economics, Country Risk Analysis
Publications: 10 papers
Experience: She is a senior researcher in State Grid Energy Research Institute of State Grid Corporation of China. She got her doctor's degree from Renmin University of China in 2010. Her research interests include energy economics, power system economics and country risk control and analysis.