ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Electric Load Forecasting Using Wavelet Transform and Extreme Learning Machine

Song Li[1], Peng Wang[1] and Lalit Goel[1]

1 -School of Electrical and Electronic Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798 -Singapore

Abstract. This paper proposes a novel method for load forecast, which integrates wavelet transform and extreme learning machine. In order to capture more internal features, wavelet transform is used to decompose the load series into a set of sub-components, which are more predictable. Then all the components are separately processed by extreme learning machine. Numerical testing shows that the proposed method is able to improve the forecast performance with much less computational cost compared with other benchmarking methods.

1 Introduction

Accurate models for load forecasting are essential for electric utilities to make crucial operational and planning decisions. Many conventional methods have been proposed, including linear regression [1], exponential smoothing [2] and time series methods [3]. Most of these methods are linear approaches and cannot appropriately represent the nonlinear load behavior. Hence, artificial intelligences such as neural networks (NN) and support vector machines (SVM) have been introduced [4, 5].

In this paper, we propose a hybrid model for load forecasting, which combines wavelet transform and extreme learning machine. Wavelet transform is an efficacious treatment to unfold the inner features of load series [6]. The load series is decomposed into a set of different frequency components. Each of them is then processed by an independent forecaster. Instead of the commonly used back propagation (BP) neural network, extreme learning machine (ELM) is adopted as the independent forecaster. ELM has no iterative fine-tuning of parameters (e.g. weights and biases) in learning process, which is completely different from that of the traditional iterative algorithms (e.g. BP and its variants) [7]. The proposed hybrid method is tested using real-world data from ISO New England, USA.

This paper is organized as follows. Section 2 describes the proposed approach, including the details of wavelet transform and extreme learning machine. Simulations are provided in Section 3. Section 4 outlines the conclusions.

2 Proposed Hybrid Method

2.1 Wavelet transform

A family of wavelet and scaling functions can be derived from the mother wavelet $\psi(t)$ and the scaling function $\varphi(t)$ [8] by

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right) \tag{1}$$

$$\varphi_{j,k}(t) = 2^{j/2}\varphi\left(2^j t - k\right) \tag{2}$$

where j and k are integer variables. The wavelet and scaling functions can be used for function expansion. A signal or function $f(t)$ can be expressed by

$$f(t) = \sum_k c_{j_0}(k) 2^{j_0/2}\varphi\left(2^{j_0} t - k\right) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2}\psi\left(2^j t - k\right) \tag{3}$$

where j_0 is the starting scale of interest, $c_{j0}(k)$ and $d_j(k)$ are the expansion coefficients.

A prototype of three-level decomposition can be given by

$$S = A_1 + D_1 = A_2 + D_2 + D_1 = A_3 + D_3 + D_2 + D_1. \tag{4}$$

The load series S is cut up into an approximation component A_3 with a set of detail components D_3, D_2 and D_1. The approximation A_3 reflects the general trend and offers a smooth version of the load series. The coefficients of D_1 are very small, which carry information about the noise conditions in the load series.

2.2 Extreme learning machine

ELM is a single-hidden layer feedforward network (SLFN) with a special learning mechanism. An SLFN is made up of three layers: input layer, hidden layer and output layer. Suppose the SLFN has n hidden nodes and nonlinear activation function $g(\boldsymbol{x})$. For N training samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$, where $\boldsymbol{x}_i$ is the ith input vector and $\boldsymbol{t}_i$ is the ith desired output, the SLFN can be modeled by

$$\sum_{j=1}^{n} \beta_j g\left(\boldsymbol{w}_j \cdot \boldsymbol{x}_i + b_j\right) = \boldsymbol{o}_i, \quad i = 1, \ldots, N \tag{5}$$

where $\boldsymbol{w}_j$ is the input weight vector linking the jth hidden node and the input nodes, b_j is the bias of the jth hidden node, β_j is the output weight vector linking the jth hidden node and the output nodes, $\boldsymbol{o}_i$ is the actual network output.

If ELM can approximate all the training samples $(\boldsymbol{x}_i, \boldsymbol{t}_i)$ with zero error, then we claim that there exist β_j, $\boldsymbol{w}_j$ and b_j such that

$$\sum_{j=1}^{n} \beta_j g\left(\boldsymbol{w}_j \cdot \boldsymbol{x}_i + b_j\right) = \boldsymbol{t}_i, \quad i = 1, \ldots, N. \tag{6}$$

The matrix form of (6) can be expressed as $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where $\mathbf{H}$ is called the hidden layer output matrix.

In the beginning of learning, the input weights $\boldsymbol{w}_j$ and the hidden layer biases b_j are randomly assigned. For given activation function $g(\boldsymbol{x})$, $\mathbf{H}$ can remain unchanged in the rest of learning. As the input weights and the hidden biases are fixed, the SLFN develops into a linear system. The output weight $\boldsymbol{\beta}$ is the only unknown parameter in the SLFN. The objective of learning is transformed to find a least squares solution $\boldsymbol{\beta}^*$ to satisfy $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$. Here a special solution is given by $\boldsymbol{\beta}^* = \mathbf{H}^{\dagger}\mathbf{T}$, where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of $\mathbf{H}$ [7].

ELM has been shown to offer an extremely fast learning speed with competitive generalization performance in many applications. Based on its distinctive installation, ELM can avoid some common problems in conventional learning algorithms such as

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

learning rate, stop criteria and local minima. Moreover, ELM can be easily employed. Only two parameters are left for consideration: the number of hidden nodes and the activation function, which are essential in all NN-based forecasters.

2.3 Proposed forecast method

The proposed method can be roughly divided into three stages: data preprocessing, independent forecasters and data post-processing, which are presented in Figure 1.

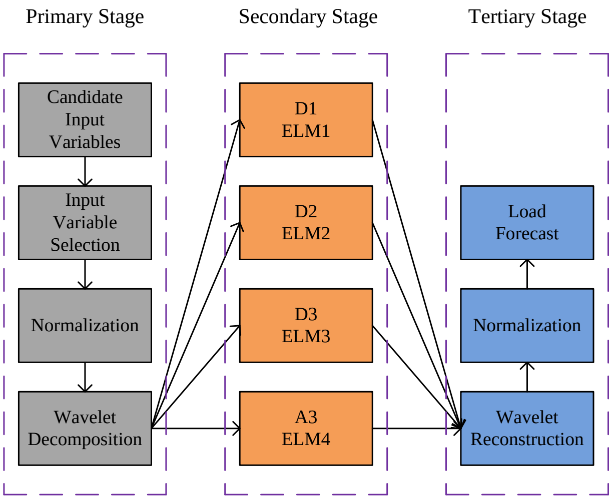

Fig. 1: Schematic diagram of proposed method.

The primary stage aims to make the subsequent forecasters more efficient and manageable. In this stage, data are collected and prepared for the forecasters. Two issues must be considered: input variable selection and wavelet decomposition.

In this work, three kinds of inputs are studied: historical load, temperature and day of the week. The lagged values of previous load are selected using autocorrelation function. Temperature and day of the week are involved as the explanatory variables, which are determined by heuristic and experience. The binary numbers are applied to represent the day of the week, which are defined as follows: 1000000 for Monday, 0100000 for Tuesday, 0010000 for Wednesday, 0001000 for Thursday, 0000100 for Friday, 0000010 for Saturday and 0000001 for Sunday.

For the wavelet transform, the fourth-order *coiflet* with three-level is selected to conduct wavelet decomposition and reconstruction. Hence, load series is decomposed into four sub-components, i.e., an approximation component A_3 associated with the low frequency and three detail components D_3, D_2 and D_1 associated with the high frequency.

The secondary stage employs ELMs as the independent forecasters to estimate the future load profile. In the tertiary stage, the outputs of all forecasters are collected to produce the overall forecast result.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

3 Simulations

The proposed method was compared with other benchmarking methods such as linear regression and back propagation network. The case of one hour ahead forecasting was studied using actual load and temperature data from ISO New England, USA. Mean absolute percentage error (MAPE) is the measure to evaluate the performance, which is given by

$$MAPE = \frac{1}{M}\sum_{i=1}^{M}\left|\frac{A_i - F_i}{A_i}\right| \times 100\% \qquad (7)$$

where M is the number of data, A_i is the actual value and F_i is the forecast value.

3.1 Dataset

The data from Dec. 21, 2009 to Nov. 21, 2010 were used to run simulations. The maximum load (occurred on Jul. 6, 2010) is 27102 MW, which is almost three times the minimum load of 9155 MW (occurred on Apr. 25, 2010). In order to observe the performance upon seasonal changes, four testing weeks were selected from four seasons: Feb. 8-14 (winter), May 17-23 (spring), Aug. 9-15 (summer) and Nov. 15-21 (fall), respectively. For each season, the six weeks before the testing week were used to build the forecast model.

3.2 Forecast models

The forecast models are listed as follows:

1) Linear regression (LR) is the first forecaster. The future load is expressed as a linear combination of its previous values and some explanatory variables.

2) Back propagation (BP) neural network is the second forecaster. BP has 10 hidden nodes and one output node. In fact, the proper number of hidden nodes is determined based on cross validation method, which can reach a compromise between the training and testing performance.

3) ELM is the third forecaster. Due to its distinctive learning mechanism, ELM needs more hidden nodes to obtain good forecast performance. In this work, ELM has 80 hidden nodes.

4) The fourth forecaster combines wavelet transform and BP network.

5) The proposed method is the fifth forecaster.

The input variables for LR, BP and ELM are shown in Table 1. It can be seen that the load L_h is highly correlated with its previous values, such as L_{h-1} and L_{h-24}.

Input	Variable	Lagged values	Output
1-4	Temperature	0, 1, 2, 24	Load(h)
5-11	Day of the week	Not Available	
12-22	Historical load	1, 2, 23, 24, 25, 48, 72, 96, 120, 144, 168	

Table 1: Input variables for LR, BP and ELM.

The input variables for the wavelet-based forecasters are given in Table 2. As the approximation component A_3 is a smoother version of the original load series,

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

their lagged values are the same. However, the lagged values of detail components are quite different from those of the approximation component.

Net	Input	Variable	Lagged values	Output
A_3	1-4	Temperature	0, 1, 2, 24	A_3 (h)
	5-11	Day of the week	Not Available	
	12-22	Historical load	1, 2, 23, 24, 25, 48, 72, 96, 120, 144, 168	
D_3	1-9	Historical load	1, 12, 24, 48, 72, 96, 120, 144, 168	D_3 (h)
D_2	1-7	Historical load	24, 48, 72, 96, 120, 144, 168	D_2 (h)
D_1	1-7	Historical load	24, 48, 72, 96, 120, 144, 168	D_1 (h)

Table 2: Input variables for wavelet-based forecasters.

3.3 **Case study**

In our study, 50 independent trials have been repeated for all network models. For the case of one step (hour ahead) load forecasting, the average results are reported. A comparison of the proposed hybrid method with other standard methods (LR and BP) is performed. The average results in MAPE are presented in Table 3.

	Winter	Spring	Summer	Fall	Mean
LR	1.0021	0.9744	0.8681	0.9017	0.9366
BP	0.8135	0.7264	1.0086	0.7502	0.8247
ELM	0.7310	0.6319	0.9651	0.6208	0.7372
BP+WT	0.7041	0.6147	0.6640	0.6014	0.6461
ELM+WT	0.6363	0.5698	0.7528	0.5289	0.6220

Table 3: MAPEs for forecast models.

It can be seen that ELM exhibits its own merit in load forecasting. ELM is able to generate better results in most weeks than LR and BP. Moreover, the forecast accuracy of neural networks is greatly improved when wavelet is integrated. Taking BP as an example, the forecast accuracy experiences an increase of 13.4% in winter, 15.4% in spring, 34.2% in summer and 19.8% in fall.

The results of Table 3 also show that the proposed method presents much better performance in most cases, as compared to other benchmarking methods. The forecast errors of the proposed method are 0.6363, 0.5698, 0.7528 and 0.5289 for winter, spring, summer and fall, respectively. Furthermore, taking spring as an instance, the enhancements of the proposed method compared to the previous methods are 41.5%, 21.6%, 9.8% and 7.3%, respectively.

In order to directly observe the performance of the proposed method, Figure 2 shows the forecast result of the fall testing week, which has the lowest forecast error in all testing weeks.

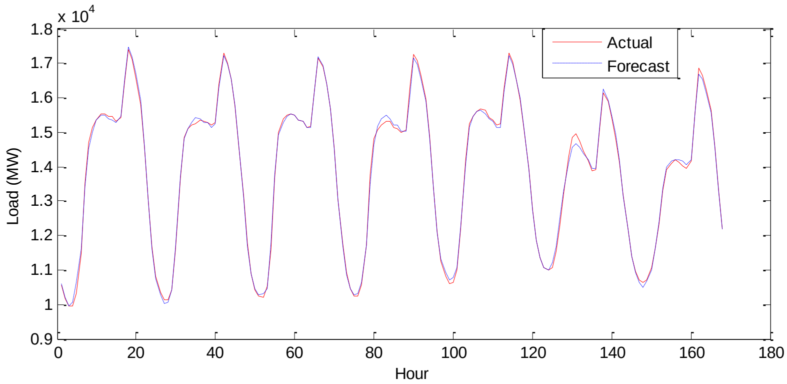

Fig. 2: Forecast result of fall week.

4 Conclusions

In this paper, we propose a hybrid model for one hour ahead load forecasting. In this model, wavelet transform is applied to decompose the load series into a series of different frequency components, which are more predictable. ELM is employed as the independent forecaster to estimate the future load profile. Compared to the popular gradient-based learning algorithms, ELM exhibits some important properties, which can improve the forecasting performance and relieve the computational burden. The simulation results reveal that the proposed method can produce excellent forecasting outcome beyond other benchmarking methods.

References

[1] T. Hong, P. Wang, and H. L. Willis, "A naive multiple linear regression benchmark for short term load forecasting," in *Power and Energy Society General Meeting, 2011 IEEE*, 2011, pp. 1-6.

[2] J. W. Taylor, "Short-Term Load Forecasting With Exponentially Weighted Methods," *Power Systems, IEEE Transactions on,* vol. 27, pp. 458-464, 2012.

[3] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *Power Systems, IEEE Transactions on,* vol. 16, pp. 798-805, 2001.

[4] Y. Wang, Q. Xia, and C. Kang, "Secondary Forecasting Based on Deviation Analysis for Short-Term Load Forecasting," *Power Systems, IEEE Transactions on,* vol. 26, pp. 500-507, 2011.

[5] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *Power Systems, IEEE Transactions on,* vol. 17, pp. 626-632, 2002.

[6] Z. A. Bashir and M. E. El-Hawary, "Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks," *Power Systems, IEEE Transactions on,* vol. 24, pp. 20-27, 2009.

[7] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, pp. 985-990 vol.2.

[8] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.