**Short-term load forecasting of**

**Australian National Electricity Market by**

**hierarchical extreme learning machine**

Jee Hyun Park1, Yan Xu2, Wan Su Na1

1SunKyunKwan University

2Nanyang Technological University

[Abstract]

Artificial Neural Network(ANN) is an effective approach for short-term load forecasting (STLF). While accurate load forecasting is pivotal for the economic and secure operation of the power system, there have been continuous efforts to achieve high load forecasting accuracy. There are two main ways to do this: the first is the improvement of the performance of the learning algorithm, and the second is how well the data features are extracted through the pre-processing process. In this paper, we design Hierarchical-Extreme Learning Machine (H-ELM) based model for forecasting the electricity load of Australian National Electricity Market (NEM) data. Owing to the very fast training/tuning speed of feed-forward neural network and multilayer concept, the H-ELM model, like the Extreme Learning Machine (ELM), is able to make fast and efficient predictions, while overcoming the instability of the forecast, which was a drawback of ELM. In addition, the H-ELM model shows better performance due to data pre-processing.

1. Introduction

블라블라

1. Hierarchical Extreme Learning Machine
   1. ELM theory

The ELM is a novel learning technology working for generalized single hidden layer feedforward networks (SLFN) [3] and it consists of three layers: input layer, hidden layer and output layer as shown in Fig 1.

Given a training data set with samples, the output function of the SLFN with hidden nodes and activation function . The activation function is expressed as following

ELM is completely different from traditional iterative learning algorithms as it randomly selects the input weights and biases for hidden nodes, w and b and analytically calculates the output weights, β, by finding least-square solution [3]. In doing so, it is proven that, the training error can still be minimized with even better generalization performance [3]. According to ELM theory, (1) can be rewritten into a compact format as follows [3]

For a training data set, given the activation function and hidden node number, the ELM learning can be summarized as the following major three steps:

Step 1. Randomly generate the input weights and , ;

Step 2. Calculate the hidden layer output matrix ; and

Step 3. Calculate output weights matrix ;

where is the Moor–Penrose (MP) generalized inverse of [3].

According to the paper [1], ELM is an effective solution for the single hidden layer feedforward networks (SLFNs), and has been demonstrated to have excellent learning accuracy/speed in various applications.

Unlike the other traditional ANN learning algorithms, ELM requires no iteratively adjustments of network parameters during the training. In other words, once ELM parameters of hidden layers of the ELM are generated randomly, it does not need to be tuned. Therefore, its training speed can be thousands of times faster [2].

And, as proved in [3], it can not only reach the minimized training error , but also the smallest norm of output weights . It is known that the feed-forward neural network achieves good generalize performance as the training error and the norm of weight become smaller [4]. Further, ELM can avoid difficulties such as stopping criteria, learning rate, learning epochs and local minima that can be commonly encountered by traditional algorithms.

Due to the characteristics of the SLFN, we can obtain very short training time, excellent efficiency and good generalize performance. However, ELM randomly selects the input weights and biases for hidden nodes, and it cause a crux in the stability of its outputs [2].

* 1. H-ELM framework

Over the past several years, Extreme Learning Machine (ELM) has been developed and applied to various fields. Due to the unique characteristics of ELM, we were able to gain many advantages such as extremely fast training, good generalization, and universal approximation/classification capability. However, it is considered that the instability of output is a disadvantage, and it is known that one hidden layer cannot perform learning of high-level features of input data. To solve these problems, H-ELM is developed, which is a feed-forward neural network with multiple layers concept.

According to [], unlike the traditional gradient-based framework, the HELM consists of two parts as shown in Fig 2: 1) unsupervised hierarchical feature representation and 2) supervised feature classification. For first part, ELM-based autoencoder is built in to extract multilayer sparse features of the input data. For second part, the original ELM-based regression is placed to conduct final decision making.

In 1) unsupervised feature learning phase, which is the first phase of H-ELM structure, it receives raw input data and projects it to the ELM random feature space, which plays an important role in extracting hidden feature information of input training data. Eventually, high-level sparse features can be obtained as output values through N-layer unsupervised learning, which is then passed to the input data of the second stage 2) supervised feature classification. In this step, the final decision is made through regression, which is the same process as the existing ELM. Since H-ELM receives input of high-level sparse features obtained from unsupervised feature learning, instead of raw input data, it is able to obtain more accurate and stable performance.

The experiments that were designed in the back will confirm that H-ELM shows better performance than ELM.

1. Data used in experiments
   1. Australian NEM

The NEM metrology procedure define a requirement for calculation of a Net System Load profiles (NSLP) for Victoria, New South Wales (NSW), Australian Capital Territory (ACT), South Australia (SA), Queensland (QLD) and Tasmania (TAS) and a CLP for each network area in NSW and SA and two Controlled Load profiles (CLP) for the Energex distribution area [5]. Both types of profiles have data recorded every 30 minutes (48 points per day). In this paper, we focused on NSLP and load profile of NSW region from 1 January 2006 to 31 December 2015 is used.

* 1. Australian Bureau of Meteorology

The Australian Government Bureau of Meteorology (BOM) is the national meteorological authority for Australia. It provides weather forecasts, warnings and observations for all states and territories of Australia. There is also information about climate, hydrology and other weather services such as weather charts, radar images, satellite images and marine weather [6]. In this paper, we choose to use the maximum temperature, minimum temperature, rainfall, and solar exposer data by date (4 points per day) among the weather data provided in the BOM. The data used are for NSW region from 1 January 2006 to 31 December 2015.

1. Implementation structure
   1. Data pre-processing strategies

Data-preprocessing is an important process for STLF training and is a process for determining how to configure the feature input data. Sufficient and well-defined feature input data is required to obtain valid predictions in the forecasting experiment. If the weather information consisting of the maximum temperature, minimum temperature, rainfall, and solar exposer data obtained from the BOM was time series data consisting of 48 points recorded every 30 minutes like the load data provided by the NEM, the weather data could be directly used as the feature input data and the load data as the label input data. However, all the weather data that can be obtained from the BOM is time series data, only one value is recorded per day. Using this data directly as feature input data will result in inaccurate and invalid predictions.

To overcome this problem, we will use data-preprocessing to include the historical load data from the NEM in the feature input data consisting of the weather data of the BOM. Which load data to include in the feature input data will be described in the following correlation study. In short, the forecasting method is based on historical load data and takes weather forecast as a hint to predict future load data. It will be confirmed from the following experiment that the prediction obtained through data-preprocessing which intentionally increases the feature input data in this way can allow more accurate and valid results.

* 1. Correlation study

In this paper, we use the correlation between data given as a measure to select the most suitable input data for model training, and the correlation analysis based on Pearson’s correlation coefficient (PCC). Pearson's correlation coefficient is the covariance of the two variables, which will be the electricity load variables of each half an hour in our studies, divided by the product of their standard deviations, which can be calculated as follows

사용할 전력량 데이터를 고를 때에 PCC 이용할 것

pcc보면 그 전날과 일주일 전이 가장 좋은 점수를 보임

case00, case01, case02, case03…

* 1. Architecture

The architecture of the STLF model for NEM is shown in Fig 3. The learning model part described in Fig 3 will consist of ELM or H-ELM. In the following, we will check performance differences between ELM and H-ELM. This architecture is divided into two phases: a) training and b) testing. In the a) training phase, the learning model receives the feature vector obtained through the data pre-processing process and performs model training. In the b) testing phase, forecasting is performed based on the input feature using the model obtained through the training process. For reference, the input features of the training phase and the testing phase are of different contents extracted from training data and testing data, respectively. Performance evaluation for forecasting will be performed using mean absolute squared error (MAPE) and mean absolute error (MAE), which are calculated as follows

where and are, respectively, the real and forecasted value of electricity load, is the total number of the points being forecasted.

좀 더 쓸까??

* 1. Input
  2. Output

1. Test results

In this experiment, NEM and BOM data from 2006 to 2015 were used. The data were divided into training data and testing data at a ratio of 8: 2. The entire process of this experiment can be found in [7]. All the simulations are conducted in MATLAB R2017a platform running on an ordinary PC with 2.4 GHZ CPU.

* 1. Optimal number of hidden nodes

필요성

방법

결과

* 1. Single ELM against H-ELM

결과 나열

* 1. Concluding remarks

결과 해석

1. Conclusion

ELM is a learning algorithm that can compensate for the drawbacks of conventional gradient-descent. However, due to

the characteristics of ELM, the output value of the algorithm is unstable, which degrades the accuracy of prediction. By constructing the prediction model through HELM, it is possible to solve the instability of the output value which is pointed out as a disadvantage of ELM and to make more accurate prediction.

1. References
2. Jiexiong Tang, Chenwei Deng, Guang-Bin Huang, Extreme Learning Machine for Multilayer Perceptron, IEEE transactions on neural Network and learning systems, Vol. 27, No. 4, April 2016
3. Rui Zhang, Zhao Yang Dong, Yan Xu, Ke Meng, Kit Po Wong, Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine, December 2012
4. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: ‘Extreme learning machine: theory and applications’, Neurocomputing, 2006, 70, pp. 489–501
5. Bartlett, P.L.: ‘The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network’, IEEE Trans. Inf. Theory, 1998, 44, (2), pp. 525–536
6. Australian Energy Market Operator [Online]. Available at <http://www.aemo.com.au/>
7. Bureau of Meteorology of Australian Government [Online]. Available at <http://www.bom.gov.au/index.shtml>
8. Experiment repository on Github. <https://github.com/jhyun0919/PowerLoad_Forecast>