Energy Conversion and Management 90 (2015) 84–92

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting

Cheng Hao Jin [a], Gouchol Pok [b,e], Yongmi Lee [a], Hyun-Woo Park [a], Kwang Deuk Kim [c], Unil Yun [d], Keun Ho Ryu [a,*]

[a] Department of Computer Science, Chungbuk National University, Republic of Korea
[b] Department of Computer Science, Yanbian University of Science and Technology, China
[c] Department of Renewable Energy Research, Korea Institute of Energy Research, Republic of Korea
[d] Department of Computer Engineering, Sejong University, Seoul, Republic of Korea
[e] Division of Information and Communication, Baekseok University, Republic of Korea

ARTICLE INFO

Article history:
Received 22 July 2014
Accepted 4 November 2014
Available online 22 November 2014

Keywords:
Artificial neural network
Clustering
Self organizing feature map
Time series forecasting
Pattern sequence
Topology relations

ABSTRACT

In this paper, we propose a new day-ahead direct time series forecasting method for competitive electricity markets based on clustering and next symbol prediction. In the clustering step, pattern sequence and their topology relations are obtained from self organizing map time series clustering. In the next symbol prediction step, with each cluster label in the pattern sequence represented as a pair of its topologically identical coordinates, artificial neural network is used to predict the topological coordinates of next day by training the relationship between previous daily pattern sequence and its next day pattern. According to the obtained topology relations, the nearest nonzero hits pattern is assigned to next day so that the whole time series values can be directly forecasted from the assigned cluster pattern. The proposed method was evaluated on Spanish, Australian and New York electricity markets and compared with PSF and some of the most recently published forecasting methods. Experimental results show that the proposed method outperforms the best forecasting methods at least 3.64%.

1. Introduction

Short-term load and price time series forecasting has become an important issue in the competitive electricity markets. A great number of approaches have been proposed for electricity load and price forecasting and these approaches can be classified into two categories: traditional approaches and modern intelligent approaches. The traditional approaches such as regression models, autoregressive integrated moving average (ARIMA) [1], exponential soothing are mainly designed based on linear models. Although simple and easy to use, the traditional approaches are not suitable to address short-term load and price forecasting problem since the load and price series has the complex characteristics of nonlinearity, nonstationarity and high volatility. One of modern intelligent approaches, artificial neural network (ANN) is a common choice for such a numeric forecasting which has the capability of modeling the nonlinear function between input and output data.

The authors in [2] used evolving fuzzy neural network, ARIMA and artificial neural network to forecast electricity demand in the State of Victoria and concluded that evolving fuzzy neural network shows better prediction accuracy than that of the latter. The author in [3] introduced a neural network approach to forecast next-week prices. Several combinations were tested to find an optimal neural network architecture with different number of hidden layers, different number of neurons in each layer and different activation functions. Adaptive wavelet neural network (AWNN) [4] was proposed to forecast short-term electricity prices. The activation function used in the hidden layers was adapted by the wavelet basis function and experiment results show that AWNN outperforms wavelet-ARIMA, multi-layer perceptron (MLP) and radial basis function (RBF) neural networks. In order to more improve the forecasting accuracy, the most important features were selected as input for neural network [5–12]. Neural network combined with other techniques also can be found in [13–21]. However, most of these forecasting methods which uses ANN techniques predict each time series value at a time.

SOM (self organizing map) [22], known as Kohonen map, is also a special type of ANN for clustering and visualization. The author of method [23] proposed an adaptive hybrid method based on SOM and SVM. Through SOM clustering, input data was partitioned into

* Corresponding author. Tel.: +82 17 430 2254; fax: +82 43 275 2254.
E-mail addresses: kimsunghoyust@gmail.com (C.H. Jin), gcpok2000@gmail.com (G. Pok), forymlee@gmail.com (Y. Lee), hwpark@dblab.chungbuk.ac.kr (H.-W. Park), kdkim@kier.re.kr (K.D. Kim), yunei@sejong.ac.kr (U. Yun), khryu@dblab.chungbuk.ac.kr (K.H. Ryu).

http://dx.doi.org/10.1016/j.enconman.2014.11.010
0196-8904/

several clusters and then different SVMs were applied to each cluster to forecast the next day's load curve. An adaptive fuzzy combination model (AFCM) based on SOM and support vector regression (SVR) was proposed in [24] for electric load forecasting. The key idea of AFCM is to build a human-understandable knowledge base by constructing a fuzzy membership function for each homogeneous sub-population. AFCM was compared with three other comparison models backpropagation neural network trained by particle swarm optimization algorithm (PSO-BP), SVR model, and optimal selection approach of SVR parameters based on PSO algorithm (PSO-SVR) and the comparison results showed the effectiveness of the proposed AFCM. However, after SOM clustering, the topology relations between cluster patterns were not considered which could be an important information in forecasting.

Recently, by clustering historical electricity time series data, a pattern sequence-based direct time series forecasting method (PSF) [25] was proposed. Based on PSF idea of pattern sequence-based time series forecasting and by using ANN techniques, we propose a new SOM Cluster Pattern Sequence-based Next Symbol Prediction method (SCPSNSP). SCPSNSP is a direct forecasting method and the topology relations between cluster patterns are also considered in the time series forecasting. It is applied to forecast electricity load and price of next day.

Referring to next symbol prediction, a lot of methods have been proposed such as compression model [26–31], Markov model [32–37], recurrent neural network model [38–44] and hybrid model [45]. However, it is important to note that the purpose of this paper is not to address any drawbacks in PSF or next symbol prediction methods, but rather to propose a novel cluster pattern sequence-based time series forecasting method.

The main novelty of SCPSNSP is that: with each cluster label of the cluster pattern sequence obtained from SOM time series clustering represented as its unique topological coordinates, ANN is used to predict the topological coordinates of next cluster label and according to the topology relations obtained from previous SOM time series clustering, next cluster label is predicted as that of the nearest nonzero hits pattern to the predicted coordinates. To the best of our knowledge, such a next symbol prediction method has not ever been proposed.

The proposed SCPSNSP could expect high forecasting accuracy since both time series vectors and topology relations between cluster patterns can be well captured in SOM clustering and ANN is also a powerful forecasting tool in addressing numeric regression problems.

The rest of the paper is organized as follows. Section 2 introduces the proposed method SCPSNSP in detail and the extensive experimental results tested on real-world electricity time series datasets are reported in Section 3. Finally, conclusions and future work are given in Section 4.

2. The SCPSNSP algorithm

The hourly electricity load/price time series vector of day i can be represented by $X(i) = [x_1, x_2, \ldots, x_{24}]$. D is the whole electricity time series dataset which consists of training dataset D_{train} and test dataset D_{test}. $[R, C]$ is SOM feature map size where R is the number of rows and C is the number of columns respectively. All the same feature maps are used in this section for illustrations (map size $[R, C] = [8, 5]$).

A cluster pattern sequence-based time series forecasting method was proposed in [25]. As shown in [25], since clustering method [46] can be used to transform the time series data into a symbolic sequence by assigning the cluster labels to pattern vectors, the time series forecasting problem also can be addressed through the next symbol prediction of the symbolic sequence obtained from time series clustering. Based on the framework of clustering and next symbol prediction, in this section, we propose a new time series forecasting method SCPSNSP. Then the forecasting problem can be formulated as: given the pattern sequence by clustering historical $X(i)$s, the goal is to predict the next cluster label.

First, in SCPSNSP, the cluster pattern sequence needs to be obtained from electricity time series clustering. In principle, any clustering algorithm can be used for this purpose, however, we choose SOM. SOM is one of the most widely used clustering methods. The main difference between SOM and other clustering methods is SOM's topology preserving property which means that close vectors in the input space are also mapped to close neurons in the output space of feature map. This property leads close neurons in the feature map to take much more similar weight vectors than those of further way neurons and the topology relations between neurons can be measured with the Euclidean distance between their topological coordinates. After time series clustering, such topology relations will be used in the next symbol prediction step so that among many clustering algorithms, only SOM can be applied.

Second, the next cluster label of the obtained cluster label sequence needs to be predicted. Thanks to SOM's topology preserving property, we can solve this symbolic sequence prediction problem with numeric regression problem by training the complex relationship between a certain length of pattern sequence and its next pattern at which each cluster label is represented as a pair of its topological coordinates. ANN will be applied for nonlinear approximation between input and output data. Then based on the previously obtained topology relations, the nearest nonzero hits pattern can be assigned for the next cluster label.

Fig. 1 shows the main framework and detail flow chart of SCPSNSP and Algorithm 1 shows the pseudo code of SCPSNSP. All the details are described in the following subsections.

Algorithm 1. The SCPSNSP Algorithm

Input: (1) Time series dataset D ($D = D_{train} + D_{test}$)
(2) Parameters related to SOM clustering
(3) Parameters related to ANN
Output: Time series values of each day $d \in D_{test}$
1: Data normalization to D
2: After various training, choose optimal feature map size $[R_{opt}, C_{opt}]$
3: SOM clustering to D with map size $[R_{opt}, C_{opt}]$
4: ANN trains the relationship between a window of W pattern sequence and its next pattern with training data D_{train}
5: ANN predicts the topology coordinates of each day $d \in D_{test}$
6: Assign the nearest nonzero hits pattern to d
7: Denormalization

2.1. Data normalization

PSF is the most related to the proposed SCPSNSP. In order to facilitate direct comparison with PSF, the same normalization is carried out as in PSF [25].

2.2. Feature map size optimization

Feature map size is important to detect the deviation of the data and it directly influences the SOM clustering results. Therefore, it is necessary to check the map quality. Several map quality measures are introduced in [47], however, there is no best measure for map quality of SOM. In this research, topographic product [48], which can represent the appropriateness of the feature map size for the