

# Electricity Price Forecasting Using a Clustering Approach

Kh. Sokhanvar\* and A. Karimpour\*\* and N. Pariz\*\*\*

\* Ferdowsi University of Mashhad, Mashhad, Iran. Kh\_so712@stu-mail.um.ac.ir

\*\* Ferdowsi University of Mashhad, Mashhad, Iran. karimpor@um.ac.ir

\*\*\* Ferdowsi University of Mashhad, Mashhad, Iran. n-pariz@um.ac.ir

**Abstract**—This paper presents a new method to forecast the short term electricity price as a kind of time series. A clustering based forecasting method is introduced. The proposed method contains input-output decomposition and using a simple clustering approach to classify them and then for a new input (a specified number of past time series values), these clusters are sorted according to the probabilities calculated by using the Bayes' formula. The prediction is then generated using the weighted average of the forecasted outputs of  $M$  clusters with highest probabilities.

**Keywords**—Bayes' rule; Clustering; Time Series Forecasting

## I. INTRODUCTION

Time series processing has important application in various domains such as medicine, ecology, industrial control and finance. Several methods have been proposed in the literature for the prediction of time series data [1]-[13]. The most popular approaches include the linear methods such as ARIMA and other regressions and nonlinear methods such as algorithms based on artificial neural network and fuzzy logic [3], [7], [8].

Electricity price is a special kind of time series with the following characteristics:

- High frequency,
- nonconstant mean and variance,
- high volatility,
- multiple seasonality,
- calendar effect,
- unusual prices.

Price forecasting is required by producers and consumers in a pool-based electric energy market, which may be one of the most unique as it is designed strictly following the economic theory. In this paper a new method is developed for time series prediction. The basic idea is to transform a time series values to input-output data set and using a simple clustering algorithm according to both input and output. Then the input features (i.e. input mean and covariance) corresponding to each cluster are obtained. A specified number of past values of the time series are given as a new input. Having this input, all clusters are sorted according to their probabilities calculated using the Bayes' rule. The prediction is then generated by using the weighted mean of the forecasted outputs of  $M$  clusters with highest probabilities.

This paper is organized as follows. In section II, the proposed method is described in detail. Section III contains some experimental results. Finally, section IV concludes the paper with summarizing the main contributions of the work.

## II. DESCRIPTION OF THE METHOD

### A. Input-Output Decomposition

According to the correlation coefficients, one considers a fixed number of past values of time series (i.e. time window of fixed size) as an input vector for the next value of time series which is an output. Then an input-output matrix is organized, which is used to be clustered.

For example, if  $\{x_i\}_{i=1}^n$  is the time series and  $k$  is the number of past effective values, that can be chosen through the analysis of correlation coefficients, then input-output matrix  $IOM$  is organized as follows.

$$IOM = \begin{bmatrix} x_1 & x_2 & \cdots & x_k & x_{k+1} \\ x_2 & x_3 & \cdots & x_{k+1} & x_{k+2} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{n-k} & x_{n-k+1} & \cdots & x_{n-1} & x_n \end{bmatrix} \quad (1)$$

### B. Clustering Algorithm

There are many clustering algorithms [14]. The simplest and most popular one is C-means, which is one of the hard partitioning methods. From an  $N \times M$  dimensional data set, C-means allocates each data point to one of  $C$  clusters,  $\{A_1, A_2, \dots, A_C\}$ , to minimize the within cluster sum of square

$$\sum_{c=1}^C \sum_{k \in A_c} \|X_k - V_c\|_2^2 \quad (2)$$

Where  $A_c$  is the set of objects (data points) in the  $c$ -th cluster and  $V_c$  is the mean for that points over cluster  $i$ , as in

$$V_c = \frac{\sum_{k=1}^{n_c} X_k}{n_c}, X_k \in A_c, c = 1, 2, \dots, C. \quad (3)$$

Where  $n_c$  is the number of objects in  $A_c$ . Each  $X_k$  contains both input and output here. Although other clustering algorithms can be used, in this paper the simple C-means is chosen. So the number of clusters must have been specified before. The number of clusters can be chosen by some effort and depends on the variance of the data.

### C. Calculating Probability

After finishing clustering, the cluster input mean ( $\mu_{in}^c$ ) and covariance ( $\Sigma_{in}^c$ ) and cluster prior probability ( $P_c$ ) are calculated using the following formulas for each cluster.

$$\mu_{in}^c = \frac{1}{n_c} \sum_{i=1}^{n_c} (X_{in}^c)_i \quad (4)$$

$$\Sigma_{in}^c = \frac{1}{n_c} \sum_{i=1}^{n_c} ((X_{in}^c)_i - \mu_{in}^c)((X_{in}^c)_i - \mu_{in}^c)^T \quad (5)$$

$$P_c = \frac{n_c}{N} \quad (6)$$

where  $(X_{in}^c)_i$  is the input vector of  $i$ -th point in the  $c$ -th cluster,  $n_c$  is the number of points in the  $c$ -th cluster and  $N$  is the number of all points. The dimension of input, ( $d$ ) is one less than length of each input-output vector.

After obtaining the clustering features, a Gaussian distribution function is used to describe the cluster membership function, i.e.

$$P(X_{in}|c) = (2\pi)^{-d/2} \left| \Sigma_{in}^c \right|^{-1/2} e^{-\frac{(X_{in} - \mu_{in}^c)^T (\Sigma_{in}^c)^{-1} (X_{in} - \mu_{in}^c)}{2}} \quad (7)$$

### D. Prediction

Given the new input (i.e. the last  $k$  values of time series), the posterior probability that the input belongs to cluster  $c$  can be approximated by using the Bayes' rule.

$$P(c|X_{in}) = \frac{P(X_{in}|c)P_c}{\sum_{c=1}^C P(X_{in}|c)P_c} \quad (8)$$

By sorting the clusters in a descending order of posterior probabilities, the cluster with the maximum posterior probability is picked as the predicted cluster and the predicted output ( $\hat{x}_{n+1}$ ) is then obtained by calculating the mean of output values in the selected cluster.

### E. Improvement of the method

To improve the prediction we can choose the average of output of  $M$  top clusters with highest posterior probabilities as the prediction, i.e.

$$\hat{x}_{n+1} = \frac{\sum_{c=1}^M P(c|X_{in}) \hat{x}_{n+1}^c}{\sum_{c=1}^M P(c|X_{in})}. \quad (9)$$

Another improvement of the results is held by choosing threshold ( $\beta$ ) between zero and one and calculating (9) with the smallest  $M$  (instead of the fixed one), which is tend to

$$\sum_{c=1}^M P(c|X_{in}) \geq \beta. \quad (10)$$

To extend over the data which have not been occurred before, the latest data is checked. If it was grater or smaller than the means of outputs of all clusters, an extra cluster for that point is added and continue.

## III. EXPERIMENTAL RESULTS

The proposed forecasting method has been applied to predict the electricity prices of California PX. The data which have been used was the average price of each hour of each weekday of a month. April and November 2000 have been chosen to forecast daily. The data used to forecast are from beginning of 1999 until the prediction period. It has not need to separate weekdays and weekends here, because clustering has done it as well and this is one of the advantages of using this method to predict electricity price and other time series which are depended on date and calendar effect. The performance of the estimators is measured by the normalized mean squared error (NMSE) of the 24 hour of each day of the week on every test set, i.e.

$$e = \frac{\sum_{k=1}^{n_{test}} (y_k - \hat{y}_k)^2}{\sum_{k=1}^{n_{test}} (y_k - \bar{y})^2}, \quad (11)$$

$$\bar{y} = \sum_{k=1}^{n_{test}} y_k, \quad (12)$$

where  $y_k$  and  $\bar{y}_k$  are, respectively, the test set measurements and their predictions.

Fig. 1 corresponds to the April's week real and forecast prices and table I consists of seven daily normalized mean squared errors.

Fig. 2 corresponds to the November's week real and forecast prices and table II consists of seven daily normalized mean squared errors.

TABLE I. DAILY NMSE OF APRIL 2000

Days	1	2	3	4	5	6	7
NMSE	3.5%	2.09%	1.97%	1.62%	1.71%	3.75%	5.8%

TABLE II. DAILY NMSE OF NOVEMBER 2000

Days	1	2	3	4	5	6	7
NMSE	3.25%	1.39%	2.69%	2.93%	1.84%	2.95%	3.47%

The results are acceptable comparing with other techniques applied on these datasets [15]-[16].

#### IV. CONCLUSION

In this paper, a new method for predicting time series was presented. It is based on clustering. The proposed method contains input-output decomposition and using a simple clustering algorithm to classify the data points. For a new input, which is a specified number of last time

series values, these clusters are sorted according to the probabilities calculated by using the Bayes' formula. The prediction is then generated by using the weighted mean of forecasted outputs of some clusters with highest probabilities. The results obtained through the Californian data shows the applicability of the proposed method.

#### REFERENCES

- [1] William W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 1<sup>st</sup> ed., Addison-Wesley, 1990.
- [2] J. M. Mendel, *Lessons in Digital Estimation Theory*, Prentice-Hall, Inc. 1987, pp. 26-70.
- [3] C. H. Ong, J. J. Huang, G. H. Tzeng, "Model identification of ARIMA family using genetic algorithm," *Applied Mathematics and Computation*, 164, 2005, pp. 885-912.
- [4] O. Renaud, J. L. Strack, F. Murtagh, "Wavelet-based forecasting of short and long memory time series," 2002. Available: <http://www.unige.ch/ses/metri>
- [5] E. Ni, P. B. Luh, "Forecasting power market clearing prices and its discrete PDF using a bayesian-based classification method," *IEEE, Power Engineering Society Winter Meeting*, vol. 3, 2001, pp. 1518-1523.
- [6] J. J. Guo, P. B. Luh, "Selection input factors for clusters of gaussian radial basis function networks to improve market clearing price prediction," *IEEE Trans. Power System*, vol. 18, no. 2, 2003, pp. 665-672.
- [7] I. Rojas, H. Pomares, "Soft-computing techniques for time series forecasting," *ESSAN'2004 proceedings- European Symposium on Artificial Neural Network*, 2004, pp. 93-102.
- [8] R. Bone, M. Crucianu, J. Pierre, "An algorithm for the addition of time-delayed connection to recurrent neural networks," *ESSAN'2000 proceedings- European Symposium on Artificial Neural Network*, 2000, pp. 293-298.
- [9] S. Soltani, "On the use of the wavelet decomposition for time series prediction," *Elsevier, Neurocomputing* 48, 2002, pp. 367-377.
- [10] H. Tong, K. S. Lim, "Threshold autoregression, limit cycle and cyclical data," *Journal of the Royal Statistical Society*, B42, 1980, pp. 245-292.
- [11] M. B. Priestly, *Spectral Analysis and Time Series*, Academic press, 1983.
- [12] Weigend A. S., B. A. Huberman, D. E. Rumelhart, "Predicting the future: a connectionist approach," *International journal of Neural Systems*, 1(3), 1990, pp. 193-209.
- [13] A. M. Jade, V. K. Jayaraman, B. D. Kulkarni, "Improved time series prediction with a new method for selection of model parameters," *J. Phys.A:Math.*, Gen. 39, 2006, pp. L483-491.
- [14] T. Warren Liao, "Clustering of time series data- a survey," *Pattern Recognition* 38, 2005, pp. 1857-1874.
- [15] F. J. Nogales, J. Contreras, A. J. Conejo, R. Spinola, "Forecasting next day electricity prices by time series models," *IEEE Trans. Power Systems*, vol. 17, no. 2, May 2002.

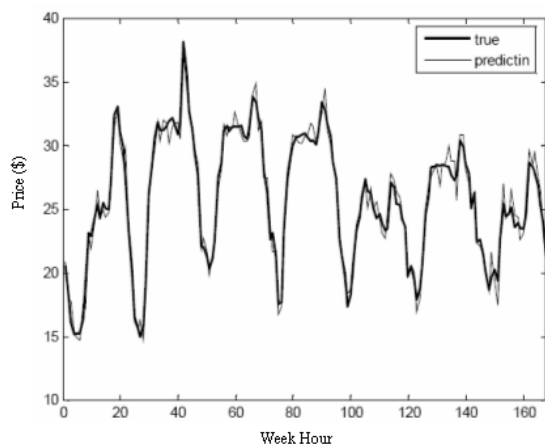


Figure 1. Real and forecasted average price of each hour of each weekday of November 2000.

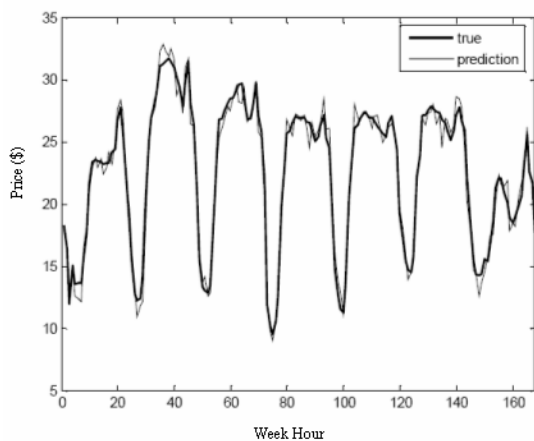


Figure 2. Real and forecasted average price of each hour of each weekday of April 2000.

- [16] R. Weron, A. Misiorek, "Forecasting spot electricity prices with time series models," International Conf. The European Electricity Market EEM-05, 2005.