A Modified Pattern Sequence-Based Forecasting Method for Electricity Price

Huaizhi Qiu, Lingling Zhao and Xiaohong Su
School of Computer Science and Technology, Harbin Institute of Technology Harbin, China

Abstract—**Electricity price forecasting is a relevant yet hard task in the field of one step time series forecasting. A new approach called Pattern Sequence-based Forecasting (PSF) shows a remarkable improvement in the time series prediction. The PSF consists of four steps: clustering, extraction pattern subsequence, sliding window to find similar subsequence, predicting. However, it doesn't consider the characteristic of correlation which changes with the length of the time gap. In this paper, a modified machine learning method based on similarity of pattern sequences which is combined with the locally weighted linear regression, is proposed to forecast electricity prices. The main novelty is that the modified PSF consider the problem of correlation which changes the mean value into weighted mean value and the PSF is firstly used in the price forecasting track of the Global Energy Forecasting Competition 2014 which shows a significantly better performance.**

Keywords-locally weighted linear regression; pattern sequence-based forecasting (PSF); electricity price; time series forecasting

I. INTRODUCTION

Electricity price forecasting has attracted a lot of attention from industry and academics in the past few years. In the electrical power industry, it is of great importance for electricity providers to make a prediction about the future electricity price accurately as it is one of the most important entries for production planning and trading on the electricity markets. Furthermore, as for the electricity consumers, knowing the future electricity price plays an important role on the electricity plans, programs, and other important strategies.

Recently, the need for reliable forecasting electricity prices has led to plenty of research and lots of approaches. The state-of-the-art methods are being divided into two main categories since the electricity price can be seen as a time series or a multi-factors prediction problem. There are substantial machine learning methods to handle this problem. Gradient boosted regression trees model which ensemble multiple linear regressors to forecast day-ahead electricity price is used in [1]. Reference [2] proposed a new hybrid model which utilizes the advantages of the feedforward neural networks model and the Auto-Regressive Fractionally Integrated Moving Average (ARFIMA) model. Also, deep learning has been applied to electricity price forecasting[3], which exploits a recurrent neural network model for the day ahead deregulated electricity market price forecasting that could be implemented using the Elman network. Reference [4] proposed an algorithm which used a principal component analysis (PCA) network associated with a multi-layer feedforward (MLF) network for electricity prices predicting in a day-ahead market. Reference [5] do a

survey on the performance of traditional hybrid methods combines the basic model of supply stacks modeling with the econometric model using price drive factor data to predict the next day's base load price. There are some competitions about the energy forecasting have been held these years such as Global Energy Forecasting Competition 2014 [6], a probabilistic energy forecasting competition consisted of four tracks on load, price, wind and solar forecasting, which attracted 581 participants from 61 countries. Reference [7] puts an interpretation on the hardness of the state-of-the-art methods, their advantages and disadvantages and the challenges that the forecasting models may be faced.

In this paper, we focus on forecasting local electricity price using modified PSF. The PSF has already been applied on some time series datasets. But the PSF has still some procedure that can be modified. A new methodology called modified pattern sequence-based forecasting based on PSF and locally weighted linear regression is proposed and being applied to each task. The performance of the proposed method compares with the original PSF and Gradient Boosting Decision Tree [8] using two common metrics: Mean Relative Error and Mean Absolute Percentage Error.

The paper contains three sections:

1) Methodology shows the principle of the PSF, locally weighted linear regression, and the proposed method.

2) Application provides some descriptions about the dataset and summarizes the comparison between the proposed method and the other state-of-the-art method.

3) Conclusions are drawn at the end.

II. METHODOLOGY

A. Pattern Sequence-Based Forecasting

Reference [9] presents a new approach to forecast the behavior of a time series based on similarity of pattern sequences. In this algorithm, k-means is applied to the hourly electricity price data in order to obtain the cluster pattern sequence. Secondly, make an extraction on the sequence prior to the predicting day to get the target subsequence which is used to find the matched subsequence in the historical data. Thirdly, a sliding window whose length is equal to the target subsequence's length slides on the cluster pattern sequence with the aim of finding matched subsequences. Finally, the result of the forecast is calculated by averaging all samples immediately after the matched subsequence. The details of this algorithm are introduced as follows:

Consider that the 24-hourly prices are needed to predict corresponding to day d which given the hourly prices recorded in the past, up to day d-1.

Let $X^{(i)} \in IR^{24}$ be a 24-dimensional vector. Each element $x_j^{(i)}$ of this vector represents the hour j of a certain day i.

$$X^{(i)} = \left[x_1^{(i)}, x_2^{(i)}, ..., x_{24}^{(i)} \right] \tag{1}$$

Let $L_i \in \{1, ..., K\}$ be the label of the 24-hourly prices of the day i which is acquired by the k-means algorithm. The maximum value of L_i is the number of clusters.

Let Seq_W^i be the sequence which is consisted of the labels of the prices. The length of the sequence is W that means the sequence contains W continuously days, from day $i - W + 1$ to day i .

$$Seq_W^i = \left[L_{i-W+1}, L_{i-W+2}, ..., L_{i-1}, L_i \right] \tag{2}$$

Noted that the length of the sequence W is a parameter to be determined.

The core concept of the PSF algorithm is to search for the subsequences of labels which are exactly identical with Seq_W^{d-1} in the historical data. So, the following equation defines a set that contains the equal subsequences.

$$ES_d = \left\{ j \middle| Seq_W^j = Seq_W^{d-1} \right\} \tag{3}$$

There is such a situation that the search process cannot find any subsequences in the historical data equal to Seq_W^{d-1} . Therefore, in order to avoid this phenomenon, the PSF algorithm will decrease the length of the target sequence step by step. This will ensure that the algorithm is able to find at least some subsequences whose length are equal to 1.

The predicting time series for the day d is forecasted by averaging the element in the set ES_d , the full process is introduced as follows:

$$\hat{X}(d) = \frac{1}{num(ES_d)} \times \sum_{j \in ES_d} X(j+1) \tag{4}$$

where $num(ES_d)$ is the number of elements contained in the set ES_d .

B. Modified Pattern Sequence-Based Forecasting

The normal PSF doesn't consider the characteristic of correlation which changes with the length of the time gap. This will influence the accuracy of the algorithm.

In this paper, we draw our inspiration from the locally weighted linear regression [10] due to that the locally weighted linear regression using a non-negative valued weight to measure the difference between different samples. This procedure can exactly solve the problem that faced in the normal PSF. In the original linear regression algorithm, the full procedure of predicting the value at a query point x (i.e., to evaluate $h(x)$) is as follows :

1) The original linear regression algorithm: The full procedure of predicting the value at a query point x is as follows.

a) Selection: Search on the solution space to find a θ that can minimize the following equation.

$$\sum_i \left(y^{(i)} - \theta^T x^{(i)} \right)^2 \tag{5}$$

b) Predicting: The $\theta^T x$ is the predictive value.

2) The locally weighted linear regression algorithm: How to predict the value at point x is as follows.

a) Selection: Search on the solution space to find a θ that can minimize the following equation.

$$\sum_i w^{(i)} \times \left(y^{(i)} - \theta^T x^{(i)} \right)^2 \tag{6}$$

b) Predicting: The $\theta^T x$ is the predictive value.

The difference between the original linear regression algorithm and the locally weighted linear regression algorithm is that the latter has a non-negative valued weight $w^{(i)}$. As the equation(6), the first step in the algorithm is obviously an optimization problem. Each component of (6) is called an error term. So the $w^{(i)}$ plays an important role in the optimization problem. In conclusion, the weight $w^{(i)}$ is used to take the difference between the samples in the training set into consideration.

A fairly standard choice for the weights is as follows.

$$w^{(i)} = \exp\left(-\frac{\left(x^{(i)} - x \right)^2}{2\tau^2} \right) \tag{7}$$

Moreover, θ is chosen giving a much higher "weight" to the (errors on) training examples close to the query point x.(Note also that while the formula for the weights takes a form that is much likely to the density of a Gaussian distribution, the $w^{(i)}$ do not directly have anything to do with Gaussians, and in particular the $w^{(i)}$ are not random variables, normally distributed or otherwise.) The parameter τ controls how fast the

weight of a training example is to be pushed down with distance of it's $x^{(i)}$ from the query point x ; τ is called the bandwidth parameter.

As for the method of this paper, the equation (7) is used to deal with the correlation which changes with the length of the time gap.

Let x be the day which we want to predict the 24-hourly prices, $x^{(i)} \in ES_d$,then, $\left| x^{(i)} - x \right|$ is the length of the time gap.

So $\hat{X}(d)$ can be changed below:

$$\hat{X}(d) = \sum_{j \in ES_d} \frac{w^{(j)}}{\sum_{j \in ES_d} w^{(j)}} \times X^{(j+1)}$$

(8)

We will call this model the MPSF (Modified Pattern Sequence-based Forecasting) model in the following and Figure 1 shows the pseudo-code of the MPSF.

III. APPLICATION

In this section, we apply the algorithm of Section II to real electricity data. The idea of probabilistic forecasting is abandoned as this is a fairly new approach and our goal was to gain comparable results with conventional forecasting algorithms. So we first describe the dataset and then show the results of the experiment. In the last part of this section is devoted to the analysis of one task.

A. Data Description

The dataset is from GEFCom2014-P. The forecast horizon was 24 h. About 3 years of hourly data were provided (from January 1st, 2011, to December 17th, 2013), including the locational marginal price, zonal load forecasts and system load forecasts (see table I for details). The dataset is separated to 15 tasks totally, as listed in table II. Every task uses the historical data given right before each forecasted period. The price series with a length of 360 is intercepted to show the characteristics of the dataset (see Figure 2 for more details).

B. Overall Forecast Results

There are two algorithms being applied in this subsection to compare with the MPSF. And we measure the forecasting performances of these algorithms on every subset (from task 1 to task 15) with two metrics : Mean Absolute Percentage Error (MAPE) and Mean Relative Error(MRE). These measures are very classical tools in time series and the formula of these two metrics are in (9) and (10).

$$MAPE = \frac{1}{length(X^{(i)})} \sum_{j=1}^{length(X^{(i)})} \frac{\left| \hat{X}_j^{(i)} - X_j^{(i)} \right|}{X_j^{(i)}} \times 100$$

(9)

$$MRE = \frac{1}{length(X^{(i)})} \sum_{j=1}^{length(X^{(i)})} \frac{\left| \hat{X}_j^{(i)} - X_j^{(i)} \right|}{Avg\left(X^{(i)} \right)} \times 100$$

(10)

Where $length\left(X^{(i)} \right)$ is the dimension of the vector $X^{(i)}$ which in this paper is 24.

Firstly, the PSF and MPSF are applied to the dataset above. The forecasting results are presented in detail in table III. Through the two metrics which display in the table III,

the performance of MPSF is better than that of the PSF in most of the tasks, except task8, task11 and task15. Then, the GBDT is applied to the dataset above. The predicting results which are compared with the result of the MPSF are presented in detail in table III. As the results showed, the change of the PSF certainly improves the performance of the original method.

Moreover, the proposed method MPSF get better results than GBDT which is the most commonly used algorithm in the field of prediction.

C. Focus on One Substation

The overall experimental results are shown in Table III and this subsection focuses on the detail of the results in one task.

TABLE I. DESCRIPTIVE ATTRIBUTIONS IN RAW DATA

Field Name	Description
ZONEID	The number of one zone.
Timestamp	A time that is accurate to a fraction of an hour.
Forecasted Total Load	The electricity total load in one zone.
Forecasted Zonal Load	The electricity zonal load in one zone.
Zonal Price	The hourly price in one zone.

TABLE II. DAYS TO BE FORECASTED IN EACH TASK

Task	Forecasted day	Task	Forecasted day	Task	Forecasted day
1	2013-06-16	6	2013-07-13	11	2013-07-24
2	2013-06-17	7	2013-07-16	12	2013-07-25
3	2013-06-24	8	2013-07-18	13	2013-12-07
4	2013-07-04	9	2013-07-19	14	2013-12-08
5	2013-07-09	10	2013-07-20	15	2013-12-17

Algorithm 1 MPSF

Input: the electricity price data D, number of clusters K, length of the window W, Test Set T, and bandwidth parameter τ .

Output: Forecasts $\hat{X}(d)$ for all days of T.

MPSF()

$ES_d = \{\ \}$

$\hat{X}(d) = 0$

$D = Normalization(D)$

$SubTimeSeries = \{\ \}$

$[L_1, L_2, ...L_{d-1}] = Clustering(D, K)$

For each day d for T do:

$Seq_W^{d-1} = [L_{d-W}, L_{d-W+1}, ...L_{d-1}]$

   For $j = 1 \rightarrow d - 2$ do:

      if $\left(Seq_W^j = Seq_W^{d-1}\right)$

        $ES_d = ES_d \bigcup j$

$SubTimeSeries = SubTimeSeries \bigcup X(j+1)$

   End for

      For $j = 1 \rightarrow size(ES_d)$

        $w(j) = Weight(ES_d(j), d, \tau)$

        $SumWeight = SumWeight + w(j)$

      End for

For $j = 1 \rightarrow size(ES_d)$

$\hat{X}(d) = \hat{X}(d) + \dfrac{w(j)}{SumWeight} \times SubTimeSeries(j)$

    End for

    $D = D \triangleright \hat{X}(d)$

    $[L_1, L_1, ..., L_{d-1}, L_d] = Clustering(D, K)$

    $d = d + 1$

End for

Return $\hat{X}(d)$ for all days of T

FIGURE I. THE PSEUDO-CODE OF THE MPSF
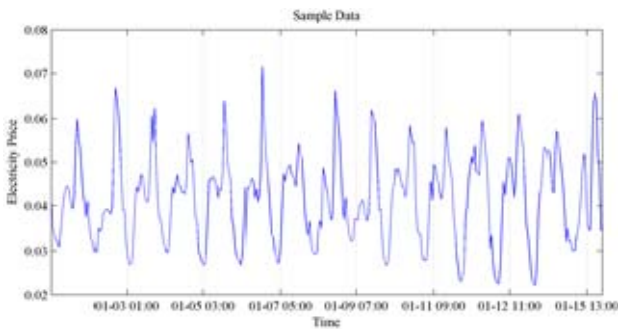

FIGURE II. SAMPLE DATA

TABLE III. PERFORMANCE OF THE MPSF, PSF AND GBDT

Task	PSF		MPSF		GBDT	
	MRE	*MAPE*	*MRE*	*MAPE*	*MRE*	*MAPE*
1	6.658	6.510	5.497	5.676	6.580	6.619
2	2.848	3.061	2.848	3.061	12.127	12.853
3	8.206	7.852	6.286	7.006	10.386	10.257
4	6.562	7.052	5.945	6.390	7.088	7.141
5	6.523	6.169	4.812	4.967	8.936	8.911
6	5.355	6.060	5.355	6.060	5.117	6.214
7	6.089	6.412	5.522	5.464	21.791	24.162
8	10.053	13.115	18.705	22.982	11.393	9.764
9	11.162	14.005	10.129	12.242	13.031	12.044
10	10.010	10.150	10.007	10.150	8.996	9.342
11	4.709	4.616	5.313	5.198	4.737	4.785
12	3.783	3.936	3.745	3.927	6.807	7.096
13	3.716	3.577	3.646	3.584	5.886	5.812
14	5.065	4.924	4.378	4.333	5.298	5.271
15	6.474	6.782	6.491	6.813	11.062	11.486

Figure 3 makes a comparison between the three algorithms' outputs and the norm price data for task 3. The MPSF fits the curve of the real electricity price data of task 3 better than other methods.

IV. CONCLUSION

In this paper, we propose a new method called MPSF to predict electricity price. As a previous step to the prediction, a clustering technique to label 24-dimensional time series has been used and uses the labels obtained by the clustering method to forecast the future behavior of the time series. The main contribution lies on the predicting step. The proposed method uses the weighted averaging instead of the normal averaging and combines the PSF with the locally weighted linear regression which can give MPSF a method to get the weight. These changes consider the characteristic of correlation which changes with the length of the time gap.
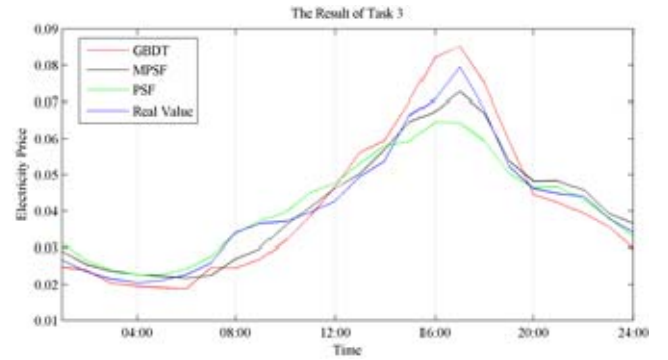

FIGURE III. THE RESULT OF TASK 3

Through the experimental results, the MPSF solves the problem that the characteristic of correlation which changes with the length of the time gap and outperforms PSF and GBDT in almost every task.

The other contribution is that our proposed method has been successfully applied in GEFCom2014-P for the first time and provides very competitive results. The performance of MPSF is better than PSF in almost all tasks. Though the GBDT

plays an important role in the state-of-the-art forecasting methods, the MPSF get more competitive results than GBDT too.

REFERENCES

[1] Barta G, Nagy G B G, Kazi S, et al. GEFCOM 2014—probabilistic electricity price forecasting[M]//Intelligent Decision Technologies. Springer International Publishing, 2015: 67-76.

[2] Chaâbane N. A hybrid ARFIMA and neural network model for electricity price prediction[J]. International Journal of Electrical Power & Energy Systems, 2014, 55: 187-194.

[3] Anbazhagan S, Kumarappan N. Day-ahead deregulated electricity market price forecasting using recurrent neural network[J]. IEEE systems journal, 2013, 7(4): 866-872.

[4] Hong Y Y, Wu C P. Day-ahead electricity price forecasting using a hybrid principal component analysis network[J]. Energies, 2012, 5(11): 4711-4725.

[5] Gonzalez V, Contreras J, Bunn D W. Forecasting power prices using a hybrid fundamental-econometric model[J]. IEEE Transactions on Power Systems, 2012, 27(1): 363-372.

[6] Hong T, Pinson P, Fan S, et al. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond[J]. 2016.

[7] Weron R. Electricity price forecasting: A review of the state-of-the-art with a look into the future[J]. International journal of forecasting, 2014, 30(4): 1030-1081.

[8] Friedman J H. Stochastic gradient boosting[J]. Computational Statistics & Data Analysis, 2002, 38(4): 367-378.

[9] Alvarez F M, Troncoso A, Riquelme J C, et al. Energy time series forecasting based on pattern sequence similarity[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(8): 1230-1243.

[10] Cleveland W S, Devlin S J. Locally weighted regression: an approach to regression analysis by local fitting[J]. Journal of the American statistical association,1988,83(403):596-610.