Daily Pattern Prediction Based Classification Modeling Approach for Day-Ahead Electricity Price Forecasting

Fei Wang[1,2,3 *], Kangping Li[2], Lidong Zhou[4], Hui Ren[2], Javier Contreras[5], Miadreza Shafie-khah[6], João P. S. Catalão[6,7,8]

1. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Baoding 071003, China
2. Department of Electrical Engineering, North China Electric Power University, Baoding 071003, China
3. Hebei Key Laboratory of Distributed Energy Storage and Micro-grid (North China Electric Power University), Baoding 071003, China
4. Weifang Power Company of State Grid Shandong Electric Power Company, Weifang 261000, China
5. E.T.S. de Ingenieros Industriales, University of Castilla-La Mancha, 13071 Ciudad Real, Spain
6. C-MAST, University of Beira Interior, Covilhã 6201-001, Portugal
7. INESC TEC and the Faculty of Engineering of the University of Porto, Porto 4200-465, Portugal
8. INESC-ID, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, Portugal

***Abstract*—Day-ahead electricity price forecasting (DAEPF) plays a very important role in the decision-making optimization of electricity market participants, the dispatch control of independent system operators (ISOs) and the strategy formulation of energy trading. Unified modeling that only fits a single mapping relation between the historical data and future data usually produces larger errors because the different fluctuation patterns in electricity price data show different mapping relations. A daily pattern prediction (DPP) based classification modeling approach for DAEPF is proposed to solve this problem. The basic idea is that first recognize the price pattern of the next day from the "rough" day-ahead forecasting results provided by conventional forecasting methods and then perform classification modeling to further improve the forecasting accuracy through building a specific forecasting model for each pattern. The proposed approach consists of four steps. First, K-means is utilized to group all the historical daily electricity price curves into several clusters in order to assign each daily curve a pattern label for the training of the following daily pattern recognition (DPR) model and classification modeling. Second, a DPP model is proposed to recognize the price pattern of the next day from the forecasting results provided by multiple conventional forecasting methods. A weighted voting mechanism (WVM) method is proposed in this step to combine multiple day-ahead pattern predictions to obtain a more accurate DPP result. Third, the classification forecasting model of each different daily pattern can be established according to the clustering results in step 1. Fourth, the credibility of DPP result is checked to eventually determine whether the proposed classification DAEPF modeling approach can be adopted or not. A case study using the real electricity price data from the PJM market indicates that the proposed approach presents a better performance than unified modeling for a certain daily pattern whose DPP results show high reliability and accuracy.**

Keywords—Electricity price; day-ahead forecasting; classification modeling; daily pattern prediction; weighted voting mechanism

Nomenclature

EPF	Electricity price forecasting	RBF	Radial basis function
DAEPF	Day-ahead electricity price forecasting	ELM	Extreme learning machine
ISO	Independent system operator	AR	Autoregression
DR	Demand response	MA	Moving average
DPP	Daily pattern prediction	ARIMA	Autoregressive integrated moving average
DPR	Daily pattern recognition	GARCH	Generalized autoregressive conditional heteroskedastic
WVM	Weight voting mechanism	MAPE	Mean absolute percentage error
ANN	Artifical neural network	RMSE	Root mean square error
SVM	Support vector machine	SDE	Standard deviation of errors
LSSVM	Least square support vector machine	ENN	Elman nerual network
SVC	Support vector classification	ED	Effective distances
SVR	Support vector regression	ER	Effective radius

1. Introduction

1.1. Background

The power industry market reform all over the world has made electricity prices become the core of market operation and electricity energy trading [1-4]. The time scales of electricity price forecasting (EPF) include long term (years), midterm (months), short term (days) and very short term (hours) [5]. Accurate EPF in all time scales is necessary for market participants and market operation [6-8]. In particular, day-ahead electricity price forecasting (DAEPF) has already been analysed due to its importance for the day-ahead market. For example, sellers and buyers need DAEPF to strategically submit their bids and offers before a certain closing time of the day-ahead energy market for the delivery of electricity during each hour of the next day [9] and independent system operators (ISOs) need DAEPF to foresee and deal with those possible abnormal conditions such as extreme high or low values and great fluctuations of electricity prices, so as to operate accordingly to guarantee the stability and orderly competition of the electricity market. Only with an accurate DAEPF can the maximum profit for all market participants be obtained as well as the minimum risk for the ISOs. However, the characteristics of electricity prices, such as non-stationarity, high volatility, seasonality and calendar effect [10, 11], make EPF more complex and challenging than electricity load or renewable energy power forecasting [12-16], particularly in the presence of demand response (DR), which brings uncertainty to electricity prices [17-21].

1.2. Literature review and motivation

The objective of forecasting is to fit the mapping relations between the future values of a forecasted variable (output) and the historical data of any related impact factors including the forecasted variable itself (input) as accurately as possible. Basically, the forecasting accuracy depends on the two following aspects. As shown in Fig. 1, the first aspect is the expression of the forecasting problem, i.e. how we formulate the problem logically and mathematically, which mainly determines how complex the forecasting will be and how difficult the modeling will be. The second aspect is the specific adopted forecasting method, which is composed of two sub-aspects including modeling theory, which is application dependent; and the model optimization to decide the optimal structure and parameters of the established model.

In terms of the second aspect, various DAEPF forecasting methods have been proposed. Time series methods such as autoregressive integrated moving average (ARIMA) [22, 23], generalized autoregressive conditional heteroskedastic (GARCH) [24]; machine learning methods such as artificial neural networks (ANN) [25,26] and support vector machine (SVM) [27,28]; hybrid methods such as two-stage hybrid model based on panel cointegration and particle filter [29], hybrid chaotic gravitational search algorithm-least square support vector machine (LSSVM) [30], hybrid models such as combining LSSVM and ARIMA, or combining ANN, an adaptive neuro-fuzzy inference system and ARMA [31, 32], and dynamic choice artificial neural network [33], etc. Although the above forecasting methods have achieved good forecasting results in some designed scenarios, it is difficult to further improve the forecasting accuracy if we only focus on the optimization of forecasting methods themselves. Because for a given forecasting issue, the expression of forecasting mainly determines the upper bound of forecasting accuracy while forecasting methods are just used to close this upper bound. Hence, more attention should be paid to more reasonable formulation of the forecasting problem to further improve the accuracy of DAEPF.

In terms of the first aspect (expression of the forecasting problem), there are two different modeling approaches: unified modeling and classification modeling. Unified modeling directly uses historical data to forecast future data and establishes only one single forecasting model without classification. Classification modeling first divides historical data into several clusters according to the daily fluctuation patterns and then build a forecasting model for each cluster [34]. Classification modeling usually outperforms the unified modeling. The reasons are illustrated as follows. Most of the forecasting methods usually divide all data into two subsets: training set and testing set. The data in training set is used to learn the mapping relation between the input (historical data) and output (future data) while the data in testing set is used to test the forecasting accuracy. For a certain forecasting method, the forecasting accuracy is mainly determined by whether the mapping relations between training set and testing set are close or not. Apparently, large errors will occur if the mapping relation between the input and output in testing set is different from it in training set. In other words, it should be ensured that the mapping relations between the training set and testing set are as close as possible in order to obtain more accurate forecasting results. Actually, there are several different fluctuation patterns in the daily electricity price data, which means there must exist different mapping relations between the historical data and future data. However, the unified modeling just fits a single mapping relation between the historical data and future data to perform the forecasting. Therefore, the cases, the mapping relations between the training set and testing set are totally different, will most likely happen in unified modeling, which will result in large forecasting errors. By contrast, classification modeling groups all historical data with different fluctuation patterns into different clusters and establish a model for each cluster. Thus, the mapping relations between the input and output in the same cluster are close, hence the classification modeling can produce more accurate forecasting results.

The key to classification modeling is the pattern prediction of next day. Because if we want to perform classification modeling, we first must know the pattern of next day so that we can choose the corresponding classification model to perform the forecasting. It should be noted that the pattern prediction must be accurate and reliable. Otherwise, the wrong classification model will be chosen and large forecasting errors will occur. Several studies concerning pattern prediction have been conducted. Martínez Álvarez et al. [35] presented an approach to forecast the behavior of time series based on the similarity of pattern sequences. First, clustering techniques were applied to group and label the samples of historical dataset, then the pattern sequence before the day to be predicted with a window size W was searched in the whole historical pattern sequence and the prediction was calculated by averaging all the samples after the matched sequence. Jin et al. [36] used ANN to conduct the pattern prediction by training the relationship between previous pattern sequences and the pattern for the next day, and the forecasting results were generated by averaging the historical data belonging to the predicted pattern. These two studies have indicated the feasibility of pattern prediction and obtained satisfactory results in some certain cases. However, they only focused on how to predict the future pattern while the pattern prediction results were not used to further perform the classification modeling.

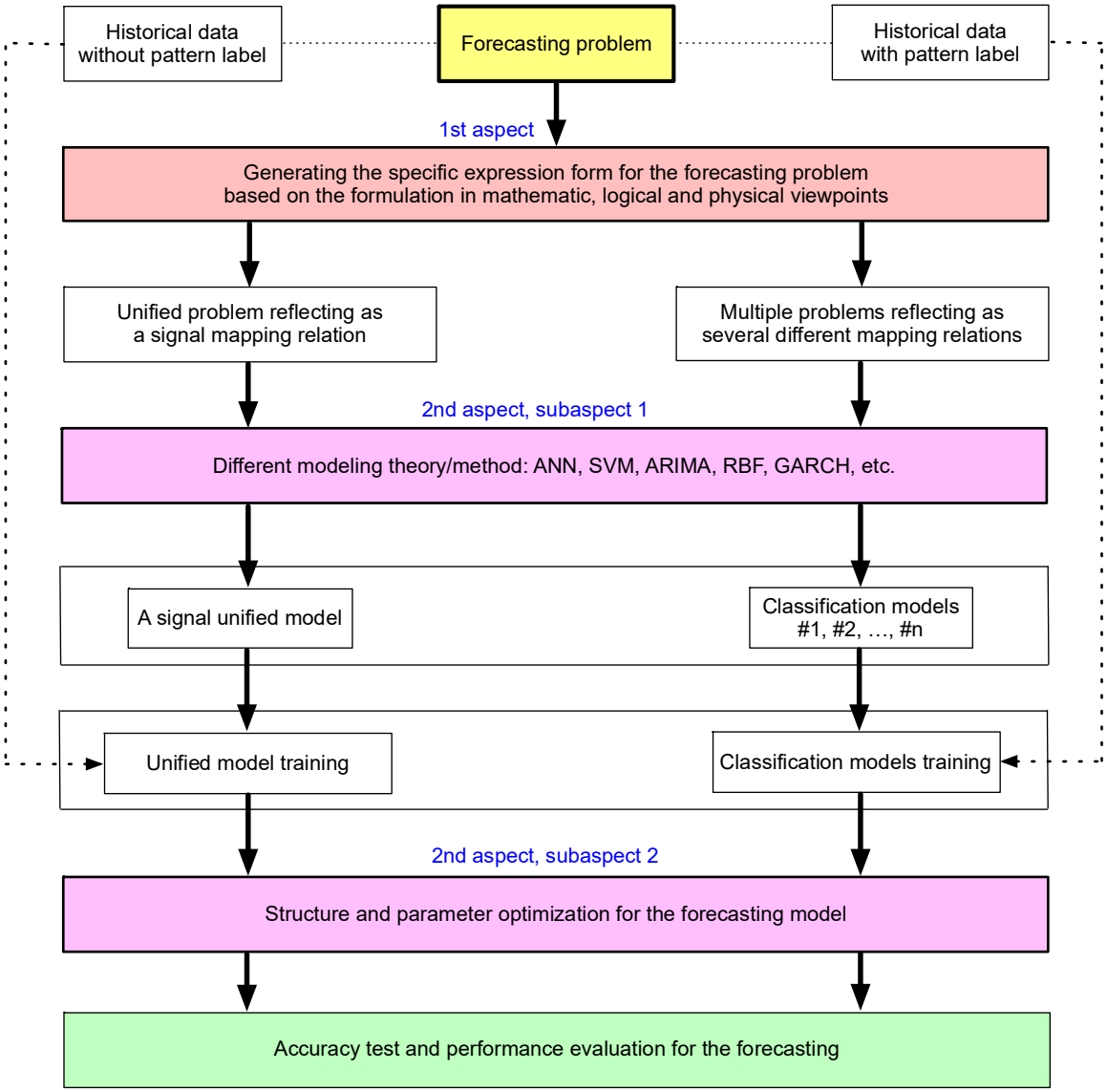

Fig. 1. Two aspects affecting the forecasting accuracy: Expression form and forecasting model

Accurate price pattern prediction is not easy to realize but possible. Although current forecasting methods cannot directly provide very accurate DAEPF results, at least the general trend could be obtained, which contains the pattern information of the next day. If we can accurately recognize the pattern from this "rough" forecasting result, classification modeling can then be applied to further improve the forecasting results. Even if we can only accurately predict one of the patterns, the forecasting results corresponding to

that pattern can be improved. In other words, as long as we can accurately predict a specific pattern, it is still of significance for improving the overall accuracy.

1.3. Contributions

The main contributions of this paper can be summarized as follows:

(1) The error generation mechanism of unified modeling and the superiority of classification modeling are revealed. Unified modeling that only fits a single mapping relation between the historical data and future data usually produces larger errors because the different fluctuation patterns in electricity price data show different mapping relations. Classification modeling establish a model for each different pattern so as to better fit the multiple mapping relations separately.

(2) A daily pattern prediction (DPP) based classification modeling approach for DAEPF is proposed in this paper. The basic idea is that first recognize the price pattern of the next day from the "rough" day-ahead forecasting results provided by conventional forecasting methods and then perform classification modeling to further improve the forecasting accuracy. Furthermore, the proposed approach can be applied into other time series forecasting issues and is not limited to any specific algorithm.

(3) A weighted voting mechanism (WVM) method is proposed in the step of price pattern prediction to combine multiple day-ahead pattern predictions to obtain more accurate pattern prediction results.

1.4. Structure of the paper

The remainder of the paper is organized as follows. Section 2 describes the DPP-based classification modeling approach for DAEPF. Section 3 introduces the algorithms/methods adopted in the proposed approach in detail. Section 4 presents the case study using a real dataset from PJM market. Finally, Section 5 concludes the paper.

2. Methodology

The detailed framework of the proposed approach consisting of 4 main steps is shown in Fig. 2. In step1, all the historical daily electricity price curves are grouped into several non-overlapping clusters and each historical day is assigned a pattern label. The price pattern of the next day is predicted in step2. In step3, classification modeling is performed to build a forecasting model for each pattern. The credibility of the pattern prediction result will be checked in step 4 to determine whether the classification model established in step 3 can be used or not. If the pattern prediction results don't pass the check, unified modeling will be directly used to produce the forecasting results.

2.1 Electricity prices data clustering

If we want to perform classification modeling, we need to know the specific pattern of each historical day. To this end, clustering is used to group the historical daily electricity price curves into n clusters based on the similarity of price pattern such that the electricity price in the same cluster show similar patterns while electricity price in different clusters exhibit distinct patterns. The purpose of this step is to mark a pattern label for each historical day indicating which cluster it belongs to. This label will be used as the output of the pattern recognition model in step 2.2. K-means, the most widely used clustering algorithms, is adopted to perform the clustering due to its attractive advantages such as fast computation speed and effective clustering results. Euclidean distance is used as the similarity metric. It should be noted that other clustering algorithms such as Fuzzy clustering also can be used to perform this step.

2.2 Daily price pattern prediction

We have obtained the pattern label by step1, the next step is to predict the pattern label of the next day so that we can perform the classification modeling. The daily price pattern prediction can be realized by the following three sub-steps.

2.2.1 Multiple DAEPF modeling

To recognize the price pattern of the next day, the "rough" DAEPF results for the next day are required. Hence, in this step, m different forecasting methods are selected to establish m DAEPF models. The reason for selecting multiple forecasting models instead of one single model is to make the pattern prediction more reliable. For each DAEPF model, the electricity price data of the previous three days is used as the input and the electricity price of the next day is used as the output of the forecasting models. Thus, m different DAEPF results can be obtained from these models. In this paper, five conventional forecasting methods including ARIMA, RBF neural network, support vector regression (SVR), Elman neural network and extreme learning machine (ELM) are adopted, which will be introduced in Section3. However, there are no constraints or preferences for the selection of the forecasting methods, which means any applicable method all could be used in this step for DAEPF modeling.

2.2.2 Daily pattern recognition (DPR)

After obtaining the "rough" DAEPF results, the next step is to recognize the daily price pattern of the next day. A daily pattern recognition (DPR) model is established using support vector classification (SVC). The input vector of this DPR model is composed of the features extracted from the daily electricity price data, and the output is the corresponding daily pattern indicating which cluster it belongs to. We select six features (i.e. maximum price, minimum price, average price, price variance, difference between maximum and minimum prices, and average price volatility) instead of the whole daily electricity curve as the input of the DPR model due to the high computation complexity caused by the high dimension of the whole daily electricity price curve. So we can use this DPR model to recognize the daily price pattern of the "rough" DAEPF results provided by the m different DAEPF models in step2.1, m different DPP results for the next day can be obtained.

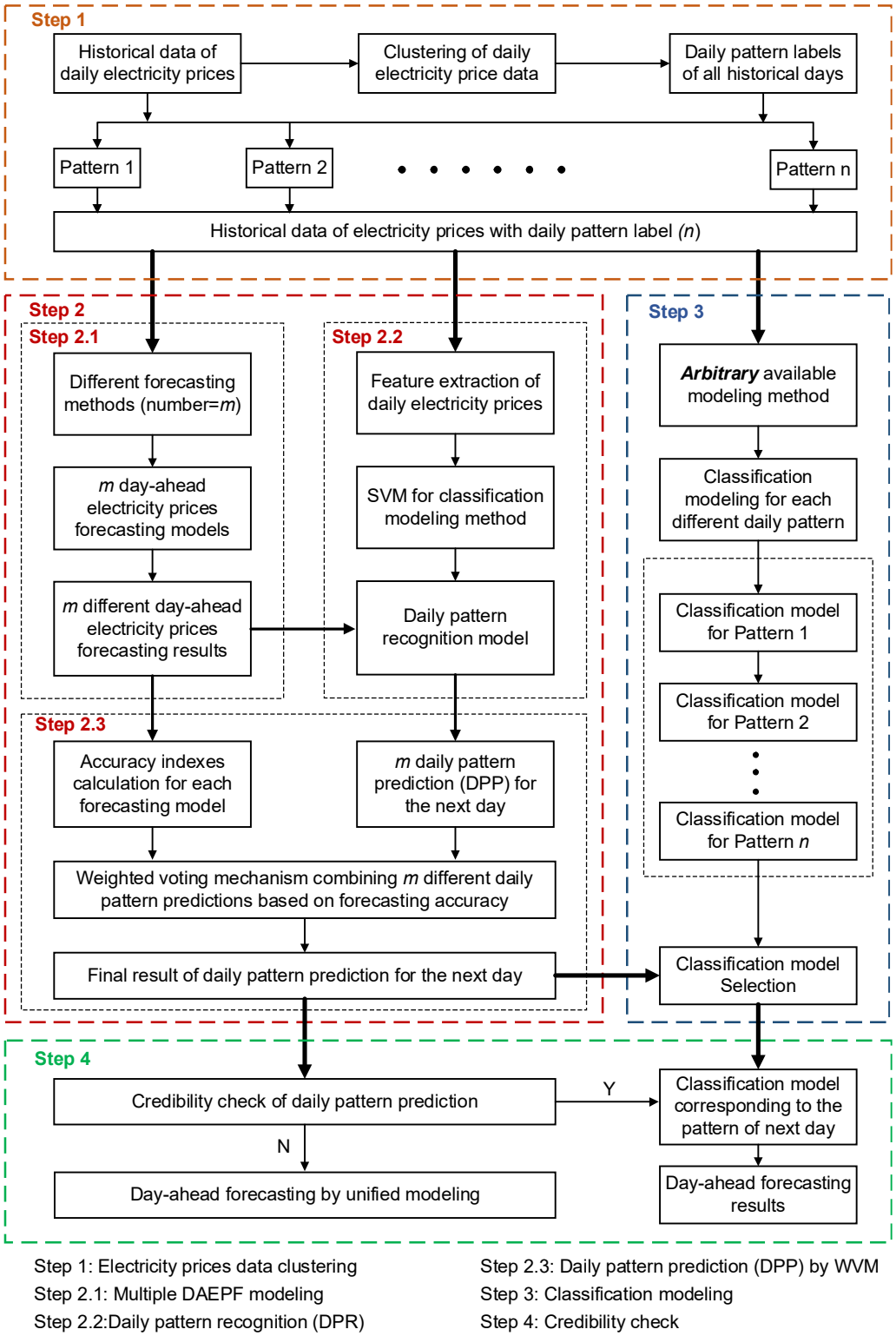

Fig. 2. Detailed framework of the proposed approach for DAEPF

2.2.3 DPP by weighted voting mechanism (WVM)

The purpose of this step is to combine the *m* pattern recognition results (*m* dimension) from step 2.3 together to obtain the final DPP result (1 dimension) for the next day. Usually, generating the final DPP result through a "simple majority" voting mechanism based on the pattern recognition results seems to be an easy way for this purpose. However, the *m* DAEPF models

exhibit different forecasting performances even under the same input vectors extracted from the same historical data, the prediction accuracy of these DAEPF models for a same day would be different. Therefore, the accuracy of each individual forecasting model should be considered when assigning the weight. To this end, a WVM method is proposed in this paper. The basic idea is that those individual models showing better performances should be assigned larger weight coefficients. The specific operative process of the proposed WVM is illustrated in Section 3.4.

2.3 Classification modeling

After obtaining the specific pattern of each historical day by step1 and the predicted pattern of the next day by step 2, the next step (step3) is to perform classification modeling to build n forecasting model for n clusters. For each cluster, the historical daily electricity price data in the same cluster will be used to train the classification model. Specifically, the electricity price data of the previous three days is used as the input and the electricity price of the next day is used as the output of the forecasting models. Then choose the corresponding model according to the DPP result to produce the final forecasting result. Note the same five forecasting methods adopted in step 2.1 will be used in this step in order to make a fair comparison between the proposed classification modeling and unified modeling.

2.4 Credibility check

Even if we have used multiple forecasting methods and WVM method to predict the daily price pattern of the next day, it still cannot be ensured that the DPP result is absolutely accurate. Hence, credibility check is conducted in this step to eventually determine whether the proposed classification DAEPF modeling approach for a certain pattern can work or not. If the final DPP result for a certain pattern shows high credibility, the classification DAEPF model selected in step 3 is accepted to forecast the electricity price for the next day. Otherwise, the proposed classification forecasting method cannot be used. Instead, the "rough" forecasting results from step 2.1 will be used as the final forecasting result. For more details about credibility check, please see Section 3.5.

3. Main algorithms

In this section, the specific algorithms/methods adopted in the proposed approach will be introduced. All of the algorithms/methods can be divided into five categories corresponding to the steps in Section 2: clustering algorithms, forecasting methods, classification methods, weight optimization methods and credibility check methods. It should be emphasized the proposed approach is not limited to specific algorithms/methods, other algorithms/methods that can realize the same function also can be used.

3.1. K-means clustering algorithm

As one of the unsupervised learning methods, K-means [37] divides the d-dimensional vector space of data point $\boldsymbol{x}_i$, $D = \{\boldsymbol{x}_i \mid i=1,2,\ldots,d\}$ into k categories by minimizing the objective function:

$$F=\sum_{i=1}^{n}\left(arg\ min\left\|\boldsymbol{x}_i-\boldsymbol{c}_j\right\|^2\right) \tag{1}$$

where $\boldsymbol{c}_j$ stands for the k-cluster centroids in center set $C = \{\boldsymbol{c}_j \mid j=1,2,\ldots,k\}$.

To minimize the objective function, K-means performs two-step procedures iteratively during the clustering. First, k-cluster centroids $\boldsymbol{c}_j$ ($j=1,2,\ldots,k$) are initialized randomly and all the d-dimensional data points are assigned to the closest centroid by implementing the Euclidean distance calculation. Second, new cluster centroids are calculated by the assigned data from the first step in each category. This iteration is repeated until $\boldsymbol{c}_j$ is stabilized.

3.2. Forecasting methods

3.2.1 Autoregressive integrated moving average (ARIMA)

ARIMA [22] is a forecasting procedure consisting of autoregression (AR), moving average (MA) and integration (I) to analyze and forecast time series. An ARIMA-based forecasting model with appropriate parameters could be used based on historical electricity prices to predict future them where the model should be updated throughout time in order to maintain a relative high accuracy using the closest electricity price data.

Assuming that the electricity prices time series are represented by P_t, the parts of P_t are described as:

$$P_t = f(t) + g(t) + X(t) \tag{2}$$

where $f(t)$ stands for the non-periodic increasing trend in P_t, $g(t)$ represents the periodic change trend in P_t, and $X(t)$ is a stationary stochastic process reflecting the random fluctuation of electricity prices.

These three parts should be treated in a forecasting mode to improve its performance. The ARIMA method provides a

difference operator, ∇, and a periodic difference operator, ∇_{24}, to eliminate the increasing and periodic variable trends in a nonstationary time series, respectively. The ARIMA-based model used for DAEPF is described as:

$$\varphi(B)\nabla^d\nabla_{24}^D P_t = \omega(B)\varepsilon_t \tag{3}$$

$$\varphi(B) = 1-\varphi_1 B-\varphi_2 B^2 -\cdots-\varphi_p B^p \tag{4}$$

$$\omega(B) = 1-\omega_1 B-\omega_2 B^2 -\cdots-\omega_q B^q \tag{5}$$

$$\nabla = 1-B \tag{6}$$

$$\nabla_{24} = 1-B^{24} \tag{7}$$

where B is the time shift operator, ∇ and ∇_{24} represent the difference operator and the periodic difference operator, p, q, d, D, φ_1, $\varphi_2 \ldots \varphi_p$ and $\omega_1, \omega_2 \ldots \omega_q$ are all the parameters of the model, and ε_t stands for a white noise interference whose mean value is zero, but whose variance is positive.

3.2.2 RBF neural network (RBFNN)

The RBFNN [38] is a feed-forward and fully connected network with three layers: input, hidden and output layer. The basic principle of the RBF neural network is to transform the low-dimensional input data into a high-dimensional space by the hidden layer composed of some RBF-based nodes. As a result, some linearly non-separable problems in low-dimensional spaces could be solved in this way.

The output layer adjusts the linear weights with a linear optimization strategy and the learning speed is relatively fast. However, the hidden layer is responsible for adjusting the parameters of the activation function with a nonlinear strategy whose speed is slow. The Gaussian function is usually used as the RBF function and the activation function $R(x)$ is described as:

$$R\left(x_p - c_i\right) = exp\left(-\frac{1}{2\sigma^2}\left\|x_p - c_i\right\|^2\right) \tag{8}$$

where $\|x_p - c_i\|$ is the Euclidean norm, c_i and σ represent the centre and variance of the Gaussian function, respectively and $x_p = \{x_1^p, x_2^p, \ldots, x_m^p\}$ is the p-th input vector.

Fig. 3 shows the structure of the RBF neural network where the final output can be expressed by:

$$y_j = \sum_{i=1}^{h} w_{ij}\, exp\left(-\frac{1}{2\sigma^2}\left\|x_p - c_i\right\|^2\right) \quad j = 1, 2, ..., n \tag{9}$$

where p=1, 2, …, P, P being the sample number, w_{ij} represents the connected weight between the hidden layer and the output layer, i=1, 2, …, h is the node number of hidden layer, and y_j stands for the output of the j-th node corresponding to the input data.

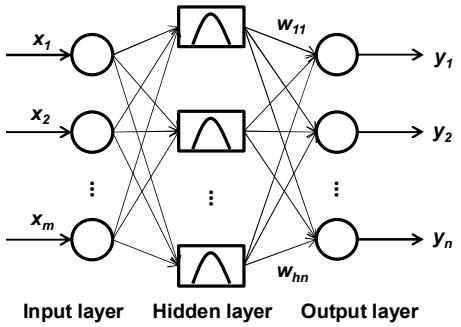

Fig. 3. Structure of RBF neural network

3.2.3 Support vector regression (SVR)

SVM [39] is a machine learning method based on statistical learning theory adopting the structural risk minimization criterion. According to different functions, SVM can be divided into SVC (for classification) and SVR (for regression).

Fig. 4 shows the schematic diagrams for SVR. The difference is that we can get $\boldsymbol{\alpha}^* = [\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*]$ and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ from the dual-form model. The regressive function can be expressed as:

$$f\left(\boldsymbol{x}\right) = \sum_{i=1}^{m}\left(\alpha_i - \alpha_i^*\right) K\left(\boldsymbol{x}_i, \boldsymbol{x}\right) + b^* \tag{10}$$

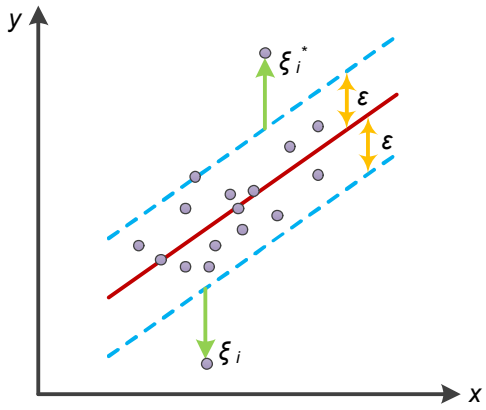

Fig. 4. Schematic diagrams for SVR

3.2.4 Elman neural network (ENN)

ENN [40] is a kind of feedback neural network that adds a context-sensitive layer after the hidden layer compared with the traditional three-layer structure. The context-sensitive layer acts as a delay operator to achieve memory function, making the system adaptable to time-variant characteristics. The nonlinear state space expressions of ENN are:

$$\boldsymbol{y}(k) = g\left(\omega^3 \boldsymbol{x}(k)\right) \tag{11}$$

$$\boldsymbol{x}(k) = f\left(\omega^1 \boldsymbol{x}_c(k) + \omega^2 \left(\boldsymbol{u}(k-1)\right)\right) \tag{12}$$

$$\boldsymbol{x}_c(k) = \boldsymbol{x}(k-1) \tag{13}$$

where k represents the sample number, $\boldsymbol{u}(k)$ and $\boldsymbol{y}(k)$ are the k-th input and output vectors, and $\boldsymbol{x}(k)$ and $\boldsymbol{x}_c(k)$ stand for the output vector of the hidden layer and the feedback state vector, respectively, ω^1, ω^2 and ω^3 are the connected weights between the context-sensitive and hidden layers, the input and hidden layers, and the hidden and output layers, respectively, and $g(*)$ and $f(*)$ are the neural transfer functions of the output and hidden layers.

3.2.5 Extreme learning machine (ELM)

ELM [41] is a kind of fully connected single-hidden layer feed-forward neural network whose structure is similar to RBF. The output vector $\boldsymbol{T}$ can be described as:

$$\boldsymbol{T} = \sum_{i=1}^{l} \beta_{ik} g(\boldsymbol{w}_i \boldsymbol{x}_j + b_i) \tag{14}$$

where l is the node number of the hidden layer, β_{ik} represents the connected weight between the i-th node in the hidden layer and the k-th node in the output layer, $g(*)$ is the activation function for a neuron in the hidden layer, $\boldsymbol{w}_i$ stands for the i-th weight row vector between the input and hidden layers, $\boldsymbol{x}_j$ is the j-th input column vector, and b_i is the threshold of the i-th neuron in the hidden layer.

3.3 Support vector classificaiton(SVC)

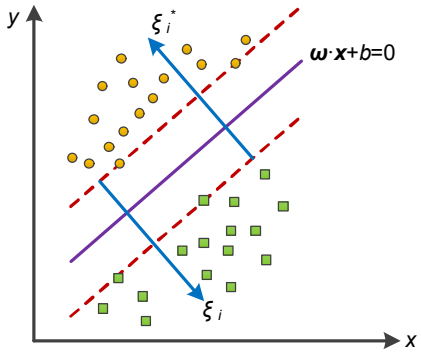

Fig. 5. Schematic diagrams for SVC

A variety of classification methods have been proposed in the past few decades. In this paper, SVC is chosen due to its excellent performance in many applications [42]. The target of SVC is to find an optimal hyperplane splitting samples into two categories.

Fig. 5 shows the linear separability of SVC with hyperplane $\boldsymbol{\omega}\cdot\boldsymbol{x}+b=0$, where $\boldsymbol{x}$ is the input vector of training samples, $\boldsymbol{\omega}$ (vector) and b (scalar) are both parameters that define the hyperplane. By introducing slack variables ξ_i and ξ_i^*, the problem to find $\boldsymbol{\omega}$ and b can be described as an optimization model. The model can be transformed to its dual form using Lagangian function. Assuming that $\boldsymbol{\alpha}^*=[\alpha_1^*, \alpha_2^*,\ldots,\alpha_m^*]$ is the solution of the dual-form model, then the optimal separable function for linear separability of SVC can be defined as:

$$f(\boldsymbol{x}) = sgn\left[\sum_{i=1}^{m}\alpha_i^* y_i(\boldsymbol{x}\boldsymbol{x}_i)+b^*\right] \tag{15}$$

where y_i represents the output of the training samples, m is the dimension of the input vector, and b^* stands for the optimal value obtained by $\boldsymbol{\alpha}^*$.

Correspondingly, for nonlinear separability of SVC, the dot product operation is replaced with $K(\boldsymbol{x}_i,\boldsymbol{x}_j)$ and the optimal separable function can be described as:

$$f(\boldsymbol{x}) = sgn\left[\sum_{i=1}^{m}\alpha_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b^*\right] \tag{16}$$

3.4 Optimization algorithm for WVM

An optimization procedure is conducted in WVM to determine the weight of each individual DAEPF model, which is illustrated in Fig. 6.

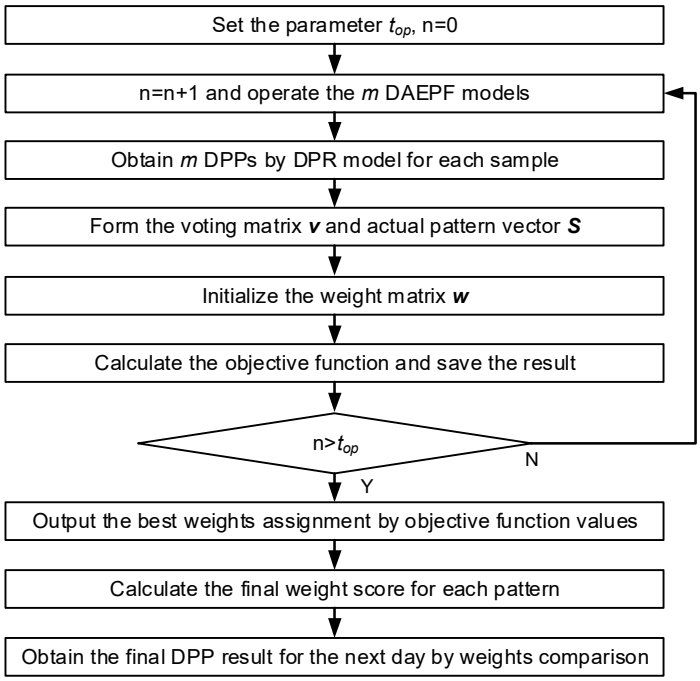

Fig. 6. Optimization procedures for WVM

The objective is to find the best weight allocation according to all training samples. And the optimization model is as follows:

$$Minimize\sum_{i=1}^{D}\left(max\text{-}row\left(diag\left(\boldsymbol{w}\boldsymbol{v}^i\right)\right)-S_i\right)^2 \tag{17}$$

subject to:

$$\sum_{k=1}^{m} w_{j,k} = 1, \qquad \forall j \in Tn \tag{18}$$

$$0 < w_{j,k} < 1, \qquad \forall j \in Tn;\ \forall k \in Tm \tag{19}$$

$$\sum_{t=1}^{n} v_{u,t}^{i} = 1, \qquad \forall u \in Tm \tag{20}$$

$$v_{u,t}^{i} = 0 \text{ or } 1, \qquad \forall u \in Tm, \forall t \in Tn \tag{21}$$

$$S_i = p, \qquad \forall p \in Tn \tag{22}$$

Matrix $\boldsymbol{w}$ and $\boldsymbol{v}^i$ can be expressed by:

$$\boldsymbol{w} = \begin{bmatrix} w_{1,1} & w_{1,2} & \mathrm{L} & w_{1,m} \\ w_{2,1} & w_{2,2} & \mathrm{L} & w_{2,m} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ w_{n,1} & w_{n,2} & \mathrm{L} & w_{n,m} \end{bmatrix} \tag{23}$$

$$\boldsymbol{v}^i = \begin{bmatrix} v_{1,1}^i & v_{1,2}^i & \mathrm{L} & v_{1,n}^i \\ v_{2,1}^i & v_{2,2}^i & \mathrm{L} & v_{2,n}^i \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ v_{m,1}^i & v_{m,2}^i & \mathrm{L} & v_{m,n}^i \end{bmatrix} \tag{24}$$

where D is the number of training samples, $diag(\boldsymbol{w}\boldsymbol{v}^i)$ returns a row vector which is formed by the diagonal value of matrix $\boldsymbol{w}\boldsymbol{v}^i$. *max-row*(∗) returns the column number of ∗'s maximum. Matrix $\boldsymbol{w}$ given by (23) is the weight matrix and $\boldsymbol{v}^i$ given by (24) is the DPP result of m normal DAEPF models for sample i. $w_{j,k}$ stands for the weight assigned to pattern j for DAEPF model k. $v^i_{u,t}$ represents the voting state of DAEPF model u for pattern t of sample i. If the DAEPF model u votes for pattern t, $v^i_{u,t}$ =1; otherwise, $v^i_{u,t}$=0. S_i is the actual pattern ranging from 1~n for sample i; m, n are the number of individual DAEPF models and patterns, respectively. Tm, Tn are two sets which are composed of the positive integers that are less than or equal to m and n. In this paper, the initial weights are set according to the DPP accuracy:

$$W_{X,Y} = \frac{1}{D} \sum_{i=1}^{D} \frac{C_{X,Y,i}}{T_{X,i}} \tag{25}$$

where $W_{X,Y}$ represents the weight assigned to pattern X for the DAEPF model Y, D is the operation number, $C_{X,Y,i}$ stands for the correct number of samples of the DAEPF model Y for pattern X in the i-th operation for DPP, and $T_{X,i}$ is the total correct number of samples of all the DAEPF models for pattern X in the i-th operation for DPP.

Assuming that the operation number of this model is t_{op}, the target for the model is to find the best weight allocation among t_{op} operations. All the weights can be updated if new historical data is introduced into the model. When m DPP results for the next day are exported from the DPR model, every price pattern's voting score is calculated, acting as the evaluation criteria for the final DPP result. In other words, the price pattern whose voting score is the maximum will be the final DPP result. This is exactly the function of *max-row*(∗).

For example, assuming that we have extracted two price patterns, A and B, in step 1, and forecasting methods P, Q, R are selected to establish three DAEPF models in step 2.1. Based on all the training samples, a reasonable weight allocation is obtained through solving the optimization model above. The solution is represented by matrix $\boldsymbol{w}$ as equation (26) shows.

$$\boldsymbol{w} = \begin{bmatrix} w_{A,P} & w_{A,Q} & w_{A,R} \\ w_{B,P} & w_{B,Q} & w_{B,R} \end{bmatrix} \tag{26}$$

Assuming that Table 1 is the DPP results of every model for each price pattern, then the $\boldsymbol{v}^i$ for sample i is expressed as equation (27):

Table 1. DPP results statistics for a certain sample

Algorithm	P	Q	R
Pattern A		✓	✓
Pattern B	✓		

$$\boldsymbol{v}^i = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \tag{27}$$

So, the final voting score for patterns A and B is:

$$\begin{aligned}\left[S_A, S_B\right] &= diag\left(\boldsymbol{w}\boldsymbol{v}^i\right) \\ &= diag\begin{bmatrix} W_{A,Q}+W_{A,R} & W_{A,P} \\ W_{B,Q}+W_{B,R} & W_{B,P}\end{bmatrix} \\ &= \left[W_{A,Q}+W_{A,R}, W_{B,P}\right]\end{aligned} \tag{28}$$

The final DPP result for the next day of the above sample will be given by *max-row*(∗):

$$\begin{cases} S_A > S_B \Rightarrow A \\ S_A < S_B \Rightarrow B \end{cases} \tag{29}$$

3.5. Method for credibility check

If we obtain a DPP result (for example, A) of the next day, the problem is how to determine whether the actual pattern is A or not. To ensure a completely accurate DPP result, a method based on distance calculation is presented to deal with this problem. For the sample whose DPP result is A, the DAEPF results obtained in step 2.1 are selected and, then, the Euclidean distance between the DAEPF results and the clustering center of pattern A are calculated. The mean value of the Euclidean distances for the same sample, defined as effective distances (ED) is calculated as:

$$ED = \frac{\sum_{i=1}^{m} \lambda_i T_A\left(r_i\right)}{\sum_{i=1}^{m} \lambda_i} \tag{30}$$

where λ_i represents the computed state of the DAEPF model i, if the DPP of the DAEPF model i is A, λ_i=1, otherwise, λ_i=0. $T_A(r_i)$ is the Euclidean distance calculation between the forecasting result, r_i, and the clustering center of pattern A. r_i is the DAEPF result of model i. Numbers 1~m represent m forecasting methods, respectively. For example, a sample whose final DPP result is A, the initial DPPs from DPR in step 2.2 are shown in Table 2.

Table 2. DPP results by DPR model for a certain sample

Models	**1**	**2**	**3**	**…**	**m**
Pattern A	✓		✓	✓	✓
Pattern B		✓			

Assuming the total number among m forecasting models for pattern A is p, the ED for this sample is:

$$ED = \frac{1}{p}\left(\lambda_1 T_A\left(r_1\right) + \lambda_3 T_A\left(r_3\right) + \mathrm{L} + \lambda_m T_A\left(r_m\right)\right) \tag{31}$$

The minimum ED of all the incorrect predictions is chosen as the bound for the correct and incorrect results, defined as the effective radius (ER). Furthermore, a margin value range from 0 to 1 can be set to obtain credible results. As a result, the ER can be expressed as:

$$ER = \mu \cdot min\left(ED_{wrong}\right) \tag{32}$$

where μ is the margin value between 0~1, and $min(ED_{incorrect})$ represents the minimum ED of the incorrect samples.

Classification modeling is conducted only when the ED is less than the ER, since only these samples show high credibility of DPP.

4 Metrics for accuracy assessment

Three metrics are used to evaluate the performance of the proposed approach: mean absolute percentage error (MAPE) and the root mean square error (RMSE) and standard deviation of errors (SDE). MAPE and RMSE are used to evaluate the forecasting accuracy and SDE is used to assess the robustness of different DAEPF models.

4.1. Root mean square error (RMSE)

The computational formula of RMSE is expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y(t)^* - y(t)\right)^2} \tag{33}$$

where N is the number of testing samples, $y(t)$ and $y(t)^*$ are the actual and predicted electricity price values at time t, respectively.

4.2. Mean absolute percentage error (MAPE)

The formula of MAPE can be expressed as:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y(t)^{*} - y(t)}{y(t)}\right| \times 100\% \tag{34}$$

4.3. Standard deviation of errors (SDE)

The proposed SDE can be calculated by:

$$SDE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_i - \bar{E}\right)^2} \tag{35}$$

where N is the number of testing samples, $\bar{E}$ is the average value of errors, and E_i represents the i-th error. In this paper, we use RMSE as the error index.

5 Case study

5.1. Data

In this paper, a case study is conducted using the real electricity price data from PJM (Pennsylvania, New Jersey, Maryland in the US) market [43]. The time range of the historical data consists of 365 days from March 1st, 2016 to February 28th, 2017 with 24 observation values every day. That's to say, the daily electricity price series is a 24-dimension column vector. As a result, all the historical data composes a 24×365 matrix. All the simulations are carried out in MATLAB and run on an Intel Core i3-2310M CPU @ 2.10 GHz computer with a 4 GB RAM. The average run time for all the steps of this approach is about 1100s.

5.2. Results

5.2.1 Results of electricity price clustering

The number of clusters K is chosen as 2 in this paper. Table 3 shows the clustering results of 365 days of historical data of electricity prices by K-means.

Table 3. Results of K-means clustering

Type	Pattern A	Pattern B	Total
Number of days/d	311	54	365
Percentage/%	85.21	14.79	100

In Table 3 we can see that patterns A and B contain 311 and 54 samples respectively. That's to say, the historical data includes more samples belonging to daily fluctuation pattern A. Then, six features are extracted to represent the high-dimensional electricity price series (For convenience, they are represented by Max, Min, Average, Variance, Max-Min and Average volatility in the figure). The input of DPR model is a 6-dimension feature vector. In order to verify whether the features could cover the typical characteristics of these patterns, so as to distinguish diverse patterns, the normalized features data of samples for pattern A and B are shown in Fig. 7. The numbers in the figure represent the normalized values of different features from -1 to 1.

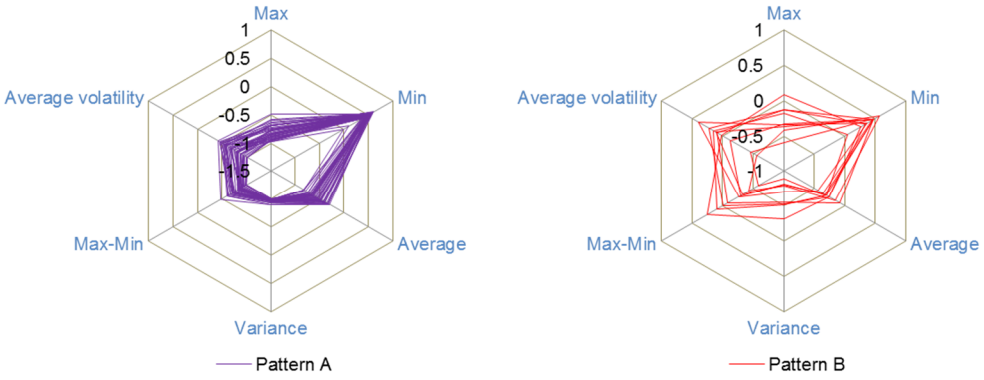

Fig. 7. Radar plot diagram of six features for patterns A and B

From fig.7 we can see that the extracted six features are able to distinguish patterns A and B effectively.

5.2.2 Results of daily price pattern prediction

1) Results of multiple DAEPF modeling

ARIMA, RBFNN, SVR, ENN and ELM are then used to establish 5 DAEPF models independently. Since the input of each DAEPF model is the price data of previous three days, thus 362 sample couples can be obtained from the 365 historical data of electricity prices. In order to cover the electricity prices in different period of one year, training samples are selected randomly form 362 sample couples. And the reminder sample couples act as testing set. All the DAEPF models are established based on the same training set. The ratio of number of training samples to testing samples is approximately 5:1. The detailed partition is shown in Fig. 8. It can be seen that the training set contains 255 samples of pattern A and 45 samples of pattern B, while the testing set includes 53 samples of pattern A and 9 samples of pattern B.

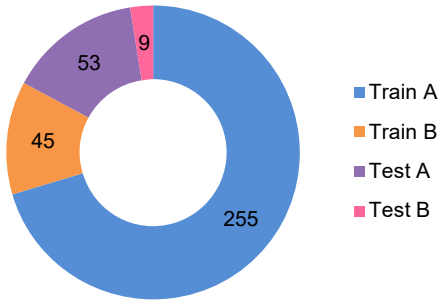

Fig. 8. The partition of training and testing samples

2) Results of daily pattern recognition

The DPR model is established using SVC after obtaining the "rough" price forecasting results. The extracted six features are used as the input of the DPR model. The pattern labels are used as the output. Table 4 shows the recognition results of DPR model.

Table 4. The results of SVC-based DPR model

Type	Pattern A	Pattern B	Total
Number of training samples	255	45	300
Correct number of training samples	252	38	290
Training accuracy/%	98.82	84.44	96.67
Number of testing samples	53	9	62
Correct number of testing samples	49	7	56
Testing accuracy/%	92.47	77.78	90.32

It is shown in Table 4 that the recognition accuracy for pattern A is better than pattern B. Table 5 shows the DPP results of the five DAEPF models in terms of the 62 testing samples (53 samples of pattern A and 9 samples of pattern B) for the first operation.

Table 5. DPP results for different DAEPF models in the first operation

Models	ARIMA	RBFNN	SVR	ENN	ELM
No. of correct A[a]	42	52	53	53	52
No. of correct B	6	4	1	4	3
Total	48	56	54	57	55

a: number of pattern A that is correctly predicted

3) Results of WVM

The proposed WVM method is then conducted and t_{op} is set to be 100. As a result, different weights are assigned to the 5 DAEPF models for each pattern based on the optimization of DPP's accuracy in the t_{op} operations. The final weights of the five DAEPF models for patterns A and B are listed in Table 6.

Table 6. Weights of five DAEPF models for patterns A and B

Models	ARIMA	RBFNN	SVR	ENN	ELM
Pattern A	0.1721	0.1903	0.2168	0.2082	0.2126
Pattern B	0.3918	0.2060	0.0686	0.1750	0.1586

Table 6 indicates that different forecasting methods present similar accuracy for predicting pattern A because the weight assigned to each model is similar. By contrast, for pattern B, the weight of each DAEPF model is different from each other. Specifically, ARIMA is assigned with the maximum weight of 0.3918 while SVR is assigned with the minimum weight of 0.0686.

5.2.3 Results of credibility check

The final DPP results for the 62 testing samples are listed in Table 7.

Table 7. The final DPP results of 62 testing samples

Type	Pattern A	Pattern B	Total
Total number	53	9	62
Correct number	53	3	56
Accuracy/%	100	33.33	90.32

It can be seen from Table 7 that the DPP of pattern A presents a good performance but for pattern B it is not ideal. In this case, the DPP of pattern B cannot be accepted because its forecasting accuracy is only 33.33%. Hence, classification modeling can be used only when the DPP result is pattern A.

From Table 7 we can see that there are 59 samples that are predicted as A among the 62 testing samples, including 53 correct DPP results and 6 incorrect DPP results. The EDs of these 59 samples are calculated and shown in Fig. 9. EDs are displayed in a circle whose radius is the maximum of them and the angle for every point is generated randomly from 0 to 2π.

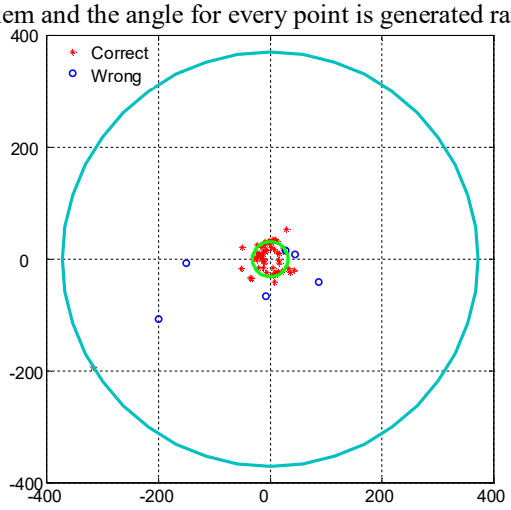

Fig. 9. Scatter diagram of 59 EDs

The ED of these 59 samples are then calculated. μ is set to be 1 in this paper. It should be pointed that electricity prices from different electricity markets perform various characteristics. Consequently, parameter μ can be flexible based on real simulation results of historical data. Only the sample whose ER is larger than its EDs can be used for classification modeling. Finally, 39 samples pass this credibility check.

5.2.4 Comparison

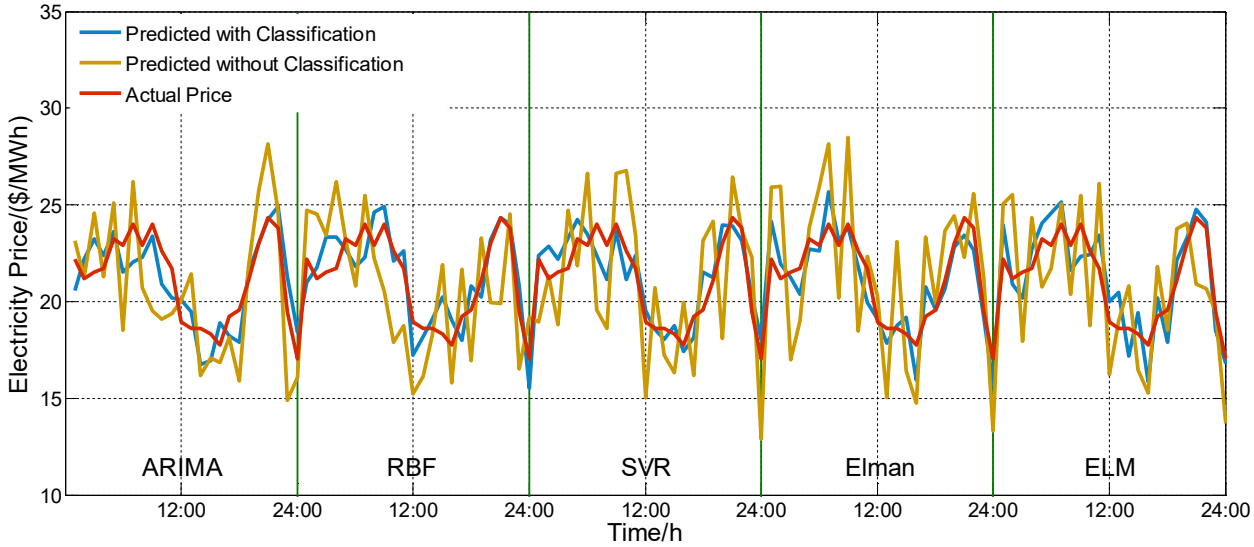

Fig. 10. DAEPF results before and after classification modeling for a certain day

For these 39 examples, we still use five different forecasting methods to establish the classification model and unified model (without classification), respectively to verify the superiority of classification modeling. Taking one of the 39 samples as an example, the DAEPF results produced by five forecasting method before and after classification modeling are shown in Fig. 10. It is shown in Fig.10 that the classification modeling presents more accurate forecasting results than unified modeling for all the forecasting methods. In other words, the classification modeling shows higher accuracy than unified modeling whatever forecasting method is used.

We combine the five forecasting models to obtain the final prices forecasting results by averaging the results of five different models. The three evaluation indexes, RMSE, MAPE and SDE, are calculated to evaluate the forecasting accuracy before and after classification, as shown in Table 8.

Table 8. Comparison of evaluation indexes before and after classification modeling

Algorithms	RMSE			MAPE			SDE		
	Before	After	Improvement %	Before	After	Improvement %	Before	After	Improvement %
ARIMA	18.92	10.43	44.87	36.69	26.10	28.86	572.45	94.42	83.51
RBFNN	16.83	7.68	54.37	29.84	20.79	30.33	937.54	38.99	95.84
SVR	11.18	6.00	46.33	18.85	13.95	25.99	227.80	36.06	84.17
ENN	16.63	6.19	62.78	27.66	16.93	38.79	932.47	35.87	96.15
ELM	11.97	6.24	47.87	25.42	16.79	33.95	250.77	36.76	85.34
Final Prices		5.90			12.14			30.79	

From Table 8 we can see that all the evaluation indexes are improved after using the DPP based classification modeling approach, indicating the effectiveness of the proposed methodology. Among the five forecasting methods, ENN forecasting methods presents the maximum improvements in accuracy after adopting the proposed approach. Comparison of RMSE between the results of these five single classification models and the final DAEPF result for 39 samples is shown in Fig. 11. We can see that the final DAEPF result show better performance than each single model in most cases.

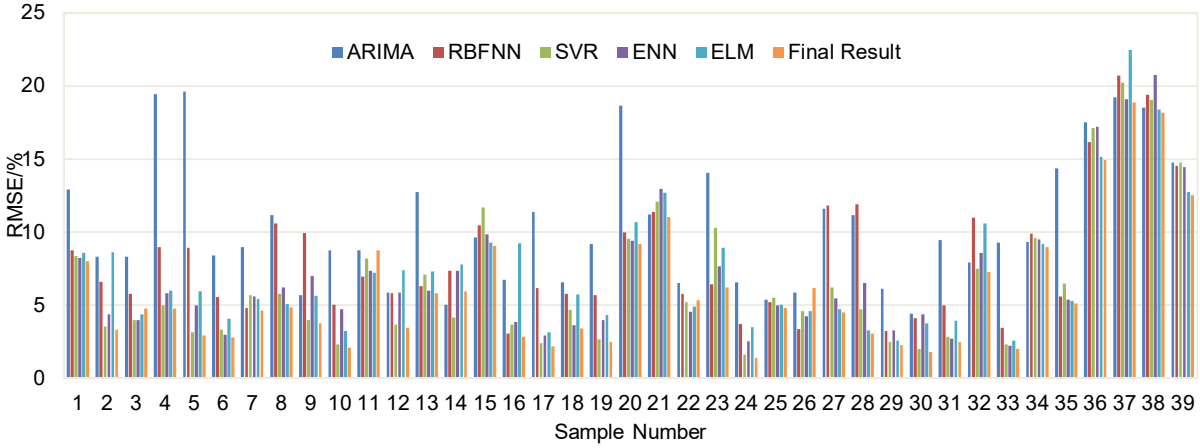

Fig. 11. RMSE comparison among the five DAEPF results with classification and the final DAEPF result

5. Conclusions

In this paper, a DPP based classification modeling approach is proposed to improve the forecasting accuracy of DAEPF. The proposed approach consists of four main steps: electricity price pattern clustering, daily price pattern prediction, classification modeling and credibility check. A case study using the real electricity price data from the PJM market shows that the proposed approach has a better forecasting accuracy than the unified modeling approach for a certain daily pattern of electricity prices, which verifies the effectiveness and feasibility of the proposed approach. The future works of this paper are listed as follows:

(1) The electricity prices from different markets may exhibit different characteristics which will affect the forecasting results. In future work, data from other electricity markets will be used to further verify the effectiveness of the proposed approach.

(2) The DAEPF is regarded as a time series forecasting problem in this paper. Actually, there are many impact factors affecting the electricity price. Correlation analysis methods can be used to quantify the specific impact of these factors and feature selection can then be applied to further improve the forecasting accuracy [44, 45].

(3) This paper proposes a very complex method containing multiple steps to predict the pattern for the next day. The final

accuracy of pattern prediction is limited by the accuracy of each step. We will try to use some state prediction method such as Markov chain to directly predict the pattern of the next day.

Acknowledgment

This work was supported by the National Key R&D Program of China (grant No. 2018YFB0904200), the National Natural Science Foundation of China (grant No. 51577067), the Beijing Natural Science Foundation of China (grant No. 3162033), the Hebei Natural Science Foundation of China (grant No. E2015502060), the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (grant Nos. LAPS18008), the Headquarters Science and Technology Project of State Grid Corporation of China (SGCC), the Open Fund of State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems (China Electric Power Research Institute) (No. 5242001600FB), the Fundamental Research Funds for the Central Universities (No. 2018QN077). This work was also supported by FEDER funds through COMPETE 2020, by Portuguese funds through FCT, under Projects SAICT-PAC/0004/2015 - POCI-01-0145-FEDER-016434, POCI-01-0145-FEDER-006961, UID/EEA/50014/2013, UID/CEC/50021/2013, UID/EMS/00151/2013, 02/SAICT/2017 - POCI-01-0145-FEDER-029803, and also by the EU 7th Framework Programme FP7/2007-2013 under grant agreement no. 309048.

References

[1] G. Aneiros, J. Vilar and P. Raña, Short-term forecast of daily curves of electricity demand and price, *International Journal of Electrical Power & Energy Systems* Vol. 80, pp. 96-108, Sep 2016.
[2] M. Shafie-khah and J. P. S. Catalão, A stochastic multi-layer agent-based model to study electricity market participants behavior, *IEEE Trans. Power Systems*, Vol. 30, no. 2, pp. 867-881, March 2015.
[3] M. Ortiz, O. Ukar, F. Azevedo and A. Múgic, Price forecasting and validation in the Spanish electricity market using forecasts as input data, *International Journal of Electrical Power & Energy Systems*, Vol. 77, pp. 123-127, May. 2016.
[4] F. Wang, L. Zhou, H. Ren, X. Liu, S. Talari, M. Shafie-khah and J. P. S. Catalão, Multi-objective Optimization Model of Source-Load-Storage Synergetic Dispatch for Building Energy System Based on TOU Price Demand Response, *IEEE Trans. on Industry Applications*, vol. 54, no. 2, pp. 1017-1028. March/April, 2017.
[5] L. Wu and M. Shahidehpour, A hybrid model for day-ahead price forecasting, *IEEE Trans. Power Systems*, Vol. 25, no. 3, pp. 1519-1530, Aug. 2010.
[6] J. Vilar, G. Aneiros and P. Raña, Prediction intervals for electricity demand and price using functional data, *International Journal of Electrical Power & Energy Systems* Vol. 96, pp. 457-472, Mar 2018.
[7] V. Guerrero-Mestre, A. A. Sánchez de la Nieta, J. Contreras and J. P. S. Catalão, Optimal bidding of a group of wind farms in day-ahead markets through an external agent, *IEEE Trans. Power Systems*, Vol. 31(4), pp. 2688-2700, July 2016.
[8] H.M.I. Pousinho, V.M.F. Mendes and J.P.S.Catalão, Short-term electricity prices forecasting in a competitive market by a hybrid PSO–ANFIS approach, *International Journal of Electrical Power & Energy Systems* Vol. 39, pp. 29-35, Jul 2012.
[9] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong and K. P. Wong, A hybrid approach for probabilistic forecasting of electricity price, *IEEE Trans. Smart Grid*, Vol. 5, no. 1, pp. 463-470, Jan. 2014.
[10] P. Kou, D. Liang, L. Gao and J. Lou, Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning, *Energy Conversion and Management*, Vol. 89, pp. 298-308, 2015.
[11] A. Khosravi, S. Nahavandi and D. Creighton, Quantifying uncertainties of neural network-based electricity price forecasts, *Applied Energy*, Vol. 112, pp. 120-129, 2013.
[12] J. P. S. Catalão, H. M. I. Pousinho and V. M. F. Mendes, Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting, *IEEE Trans. Power Systems*, Vol. 26, no. 1, pp. 137-144, Feb. 2011.
[13] N. A. Shrivastava and B. K. Panigrahi, A hybrid wavelet-ELM based short term price forecasting for electricity markets, *International Journal of Electrical Power & Energy Systems*, Vol. 55, pp. 41-50, Feb 2014.
[14] F. Wang, Z. Mi, S. Su and H. Zhao, Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters, *Energies*, Vol. 5, pp. 1355-1370, 2012.
[15] F. Wang, Z. Zhen, Z. Mi, H. Sun, S. Su and G. Yang, Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting, *Energy and Buildings*, Vol. 86, pp. 427-438, 2015.
[16] Wang F, Zhen Z, Liu C, Mi Z, Hodge BM, Shafie-khah M, et al. Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting. *Energy Convers Manag* Vol.157, pp.123–135, 2018.
[17] F. Wang, H. Xu, T. Xu, K. Li, M. Shafie-khah and J. P.S. Catalão, The values of market-based demand response on improving power system reliability under extreme circumstances, *Applied Energy*, Vol. 193, pp. 220-231, 2017.

[18] Q. Chen, F. Wang, B.M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. S. Catalão, Dynamic Price Vector Formation Model-Based Automatic Demand Response Strategy for PV-Assisted EV Charging Stations, *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2903–2915, 2017.
[19] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-khah, and J. P. S. Catalão, Synchronous Pattern Matching Principle Based Residential Demand Response Baseline Estimation: Mechanism Analysis and Approach Description, *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2018.
[20] S. Talari, M. Shafie-khah, F. Wang, J. Aghaei and J. P. S. Catalão, Optimal scheduling of demand response in pre-emptive markets based on stochastic bilevel programming method, *IEEE Trans Ind Electron* 2017; pp:1-1.
[21] F. Wang, L. Liu, Y. Y, G. L, J. Li, M. Shafie-khah and J. P. S. Catalão, Impact Analysis of Customized Feedback Interventions on Residential Electricity Load Consumption Behavior for Demand Response, Energies, vol. 11, no. 4, p. 770, 2018.
[22] J. Contreras, R. Espínola, F. J. Nogales and A. J. Conejo, ARIMA models to predict next-day electricity prices, *IEEE Trans. Power Systems*, Vol. 18, no. 3, pp. 1014-1020, Aug. 2003.
[23] T. Kristiansen, A time series spot price forecast model for the Nord Pool market, *International Journal of Electrical Power & Energy Systems*, Vol. 61, pp. 20-26, Oct 2014.
[24] R. C. Garcia, J. Contreras, M. van Akkeren and J. B. C. Garcia, A GARCH forecasting model to predict day-ahead electricity prices, *IEEE Trans. Power Systems*, Vol. 20, no. 2, pp. 867-874, May 2005.
[25] D. Singhal and K.S. Swarup, Electricity price forecasting using artificial neural networks, *International Journal of Electrical Power & Energy Systems*, Vol. 33, pp. 550-555, Mar 2011.
[26] I. P. Panapakidis and A. S. Dagoumas, Day-ahead electricity price forecasting via the application of artificial neural network based models, *Applied Energy*, Vol. 172, pp. 132-151, 2016.
[27] N. A. Shrivastava, A. Khosravi and B. K. Panigrahi, Prediction interval estimation of electricity prices using PSO-tuned support vector machines, *IEEE Trans. Industrial Informatics*, Vol. 11, no. 2, pp. 322-331, April 2015.
[28] X. Yan and N. A. Chowdhury, Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach, *International Journal of Electrical Power & Energy Systems*, Vol. 53, pp. 20-26, Dec 2013.
[29] X. R. Li, C. W. Yu, S. Y. Ren, C. H. Chiu and K. Meng, Day-ahead electricity price forecasting based on panel cointegration and particle filter, *Electric Power Systems Research*, Vol. 95, pp. 66-76, 2013.
[30] H. Shayeghi and A. Ghasemi, Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme, *Energy Conversion and Management*, Vol. 74, pp. 482-491, 2013.
[31] A. Ghasemi, H. Shayeghi, M. Moradzadeh and M. Nooshyar, A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management, *Applied Energy*, Vol. 177, pp. 40-59, 2016.
[32] A. Darudi, M. Bashari and M. H. Javidi, Electricity price forecasting using a new data fusion algorithm, *IET Generation, Transmission & Distribution*, Vol. 9, no. 12, pp. 1382-1390, 2015.
[33] J. Wang, F. Liu, Y. Song and J. Zhao, A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system, *Applied Soft Computing*, Volume 48, Pages 281-297, 2016.
[34] Wang F, Zhen Z, Liu C, Mi Z, Shafie-Khah M, Catalão JPS. Time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization. Energies 2018; 11:184.
[35] F. Martínez Álvarez, A. Troncoso, J. C. Riquelme and J. S. Aguilar Ruiz, Energy time series forecasting based on pattern sequence similarity, *IEEE Trans. Knowledge and Data Engineering*, Vol. 23, no. 8, pp. 1230-1243, Aug. 2011.
[36] C. H. Jin, G. Pok, Y. Lee, H.-W. Park, K. D. Kim, U. Yun and K. H. Ryu, A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting, *Energy Conversion and Management*, Vol. 90, pp. 84-92, 2015.
[37] J. Ghosh and A. Liu, K-means, *The Top Ten Algorithms in Data Mining*. Boca Raton, FL, USA: Chapman & Hall, chapter 2, pp. 21-35, 2009.
[38] F. Bonanno, G. Capizzi, G. Graditi, C. Napoli and G.M. Tina, A radial basis function neural network-based approach for the electrical characteristics estimation of a photovoltaic module, *Applied Energy*, Vol. 97, pp. 956-961, 2012.
[39] H. Xue, Q. Yang and S. Chen, SVM: support vector machine, *The Top Ten Algorithm in Data Mining*. Boca Raton, FL, USA: Chapman & Hall, chapter 3, pp. 37-59, 2009.
[40] H. Liu, H.-Q. Tian, X.-F. Liang and Y.-F. Li, Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks, *Applied Energy*, Vol. 157, pp. 183-194, 2015.
[41] S. Li, L. Goel and P. Wang, An ensemble approach for short-term load forecasting by extreme learning machine, *Applied Energy*, Vol. 170, pp. 22-29, 2016.
[42] Wang F, Zhen Z, Wang B, Mi Z. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. Appl Sci 2017; 8:28.
[43] http://www.pjm.com/markets-and-operations.aspx

[44] Y. Sun, F. Wang, B. Wang, Q. Chen, N. A. Engerer, and Z. Mi, Correlation Feature Selection and Mutual Information Theory Based Quantitative Research on Meteorological Impact Factors of Module Temperature for Solar Photovoltaic Systems, *Energies*, vol. 10, no. 1, 2017.
[45] Wang F, Li K, Duić N, Mi Z, Hodge BM, Shafie-khah M, et al, Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns, *Energy Convers Manag*, Vol.171, pp. 839–854, 2018.