Engineering Applications of Artificial Intelligence 95 (2020) 103852

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A novel integrated price and load forecasting method in smart grid environment based on multi-level structure

Yang Zhang [a,b,*], Caibo Deng [b], Ran Zhao [c], Sebastian leto [d]

[a] School of mechanical and electrical engineering, Nanchang Institute of Technology, Nanchang 330099, China
[b] State Grid Electric Power Research Institute of Jiangxi Province, Nanchang 330099, China
[c] Zhongyuan- Peterburg Aviation College, Zhongyuan University of Technology, Zhengzhou 45000, China
[d] Department of Elec. & Energy, Sunlife Co., Baku, Azerbaijan

ARTICLE INFO

Keywords:
MIMO model
Smart grids
DWT
Feature selection
Load and price prediction
LSSVM

ABSTRACT

Prediction of load and price are two critical key in power system planning and operation. Most of the recent works in this area forecast the load and price signals separately but, a dynamic model in smart grid is evaluated while, the customers may have opportunity to react the proposed prices changing through shifting the electricity usages from expensive to cheaper hours. So, the load and price signals are coupled strongly which made the previous prediction models ineffective. In this research, a synthetic prediction approach has been proposed by considering the load and price signals, simultaneously. This method works as multi-input multi-output (MIMO) model based on least square support vector machine (LSSVM) forecast engine. Furthermore, a dyadic wavelet transform (DWT) is suggested in this approach to decompose the original signal into different small sub-signals. Beside of that, the modified mutual information (MMI) filter has been used to choose the best candidate input of forecast engine. The learning section is also coupled with novel modified optimization algorithm based on gravitational search algorithm (GSA) which called as modified GSA (MGSA). Finally, various forecasting errors have been considered as average mean absolute percentage error and error variance to get the comparison outcomes and performance of forecasting approaches. For this purpose, different markets have been considered as test case to show the efficiency of suggested approach.

1. Introduction

Increasing of electricity consumption in power network, cause to use renewable energy sources i.e., wind power, solar energy and etc., which moves the grids to concept of smart grids (Duan et al., 2018). So, it supplies required energy for economic very safety and efficiently to integrate smart grids with new notions, methods, and supplementary facilities for production, transmission, and distribution of products to the users with very advanced detecting and management tools (Oveis and Amjady, 2016). Moreover, smart grids give this capability to their product users to control their demand level based on cost rate which is changing constantly. It is well-defined as the usage of assessments in order to monitor and manage the power production, and as a result decrease the used power (Oveis et al., 2016). According to the notion of smart grids, Fig. 1 demonstrates a diagram of actual-marketplaces. Despite of the accessibility of predicted data for high-frequency green energy sources, the review of electricity price prediction based on the exogenous data is still moderately rare. So, in this work we focus on complete this break based on linear methods in multi and solo parameters researches through comparing the deterministic and probabilistic forecasting approaches (Duan et al., 2018; Oveis and Amjady, 2016; Oveis et al., 2016; Grossi and Nan, 2019; Muniain and Ziel, 2020; Oveis et al., 2015; Altan et al., 2019). This study aimed to presents an energy market for consumers in order to have a perfect and economical market. So, in order to increase generators pure benefit, we have presented a precise estimation of upcoming power cost rate for bidding in actual marketplaces.

Actually, the load and price forecasting models can be categorized into different applications of; very short term, short term, mid-term and long-term while, this work considered the day-ahead forecasting model.

Various forecasting approaches have been proposed by researchers in power market for load and price signals. The load signal is coupled with simulation method, statistical approach or mixed of mentioned models (Duan et al., 2018). Some of the mentioned models can be summarized as; linear regression, back propagation neural network (BPNN), co-evolutionary and synthetic approaches for estimating the loads, in which a large data set is demanded (Altan et al., 2019; Yang et al., 2020; Noradin et al., 2018; Wu and Shahidehpour, 2014). In Hafeez et al. (2020), the authors proposed an improved learning strategy method for short term load forecast (STLF) while, extreme learning machine has been applied to get the better generalization

* Corresponding author at: School of mechanical and electrical engineering, Nanchang Institute of Technology, Nanchang 330099, China.
E-mail address: rxzhangyang@foxmail.com (Y. Zhang).

https://doi.org/10.1016/j.engappai.2020.103852
Received 13 November 2019; Received in revised form 4 July 2020; Accepted 27 July 2020
Available online 19 August 2020

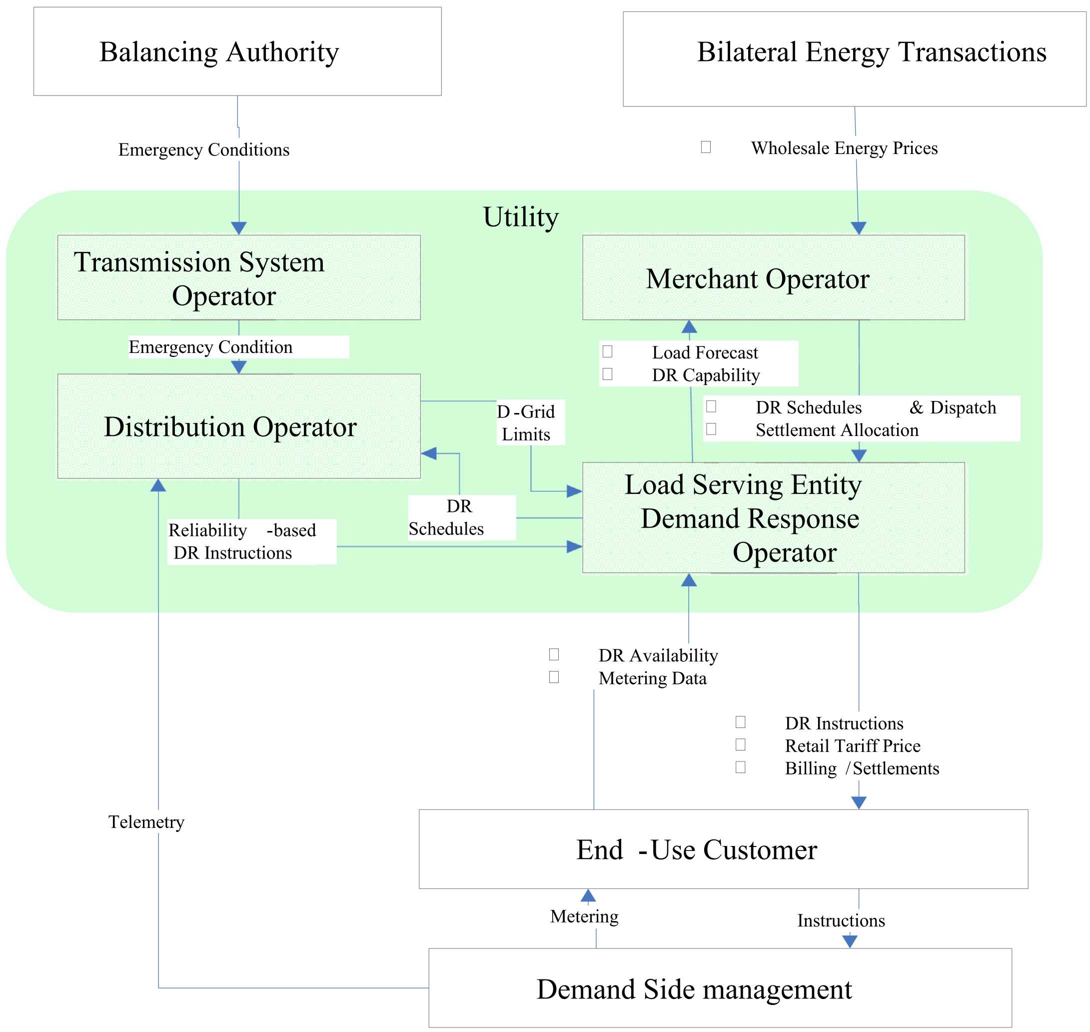

Fig. 1. Suggested notion of smart grids on the basis of relationship of cost and load prediction.

performance. In Han et al. (2012), hybrid model for very short term load forecast is presented by considering the recursive least square support vector machine as dynamic specifications of multi-node active demand distribution items and power factors, and the Takagi–Sugeno fuzzy control model has been improved for the feedback compensation. In the mentioned approaches, the error is between 105% (Oveis and Amjady, 2016; Oveis et al., 2016; Grossi and Nan, 2019; Muniain and Ziel, 2020; Oveis et al., 2015; Altan et al., 2019).

Beside of that, the short term price forecast (STPF) volatility is much more that load signal due to seasonal spikes and other factors which needs more accurate forecasting approach. Different approaches have been proposed by researchers for this signal as mathematical finance models, game theory models and regression models. In the game theory based model, all players are rational that are not always the case in electricity markets. Furthermore, the time consumption process to evaluate the Nash equilibriums is another backward of this approach. Beside of that, the regression and artificial intelligent based models are dealing with electricity price fluctuations in historical data and other coupled items such as load, weather factors and etc. Lin et al. (2020). The autoregressive integrated moving average (ARIMA) approach, combined wavelet and ARIMA, generalized autoregressive conditional heteroscedastic (GARCH), neural network (NN) as well as other synthetic approaches have been applied to the STPF (Wu and Shahidehpour, 2014; Hafeez et al., 2020; Han et al., 2012; Lin et al., 2020; Salehpour and Moghaddas Tafreshi, 2019). Such models are working simply and efficiently but, accurate details are demanded in fine tuning of the mentioned models.

In Elattar (2013), a new model based on kernel principal component analysis is presented based on local informative vector machine model for STPF while, this model extracts the features of input and kernel principal components. A modified relief algorithm based feature selection which is combined with neural network is suggested in Amjady et al. (2010). In Aggarwal et al. (2009), different comparison approaches have been presented, which proofs that there is no systematic evidence of out-performance of one model over other methods on a consistent basis. Such mentioned models could get suitable forecasting error between 5 to 20 percent (Yang et al., 2020; Noradin et al., 2018; Wu and Shahidehpour, 2014).

But, both of STLF and STPF are far from reaching maturity, particularly by the integration of price-sensitive loads in emerging power systems. In smart grids the price-based demand response delivers clients chances and incentives to conversion their usual consumption approaches against the price variations, which can assistance the markets set effective energy prices, mitigate the market, develop economic effectiveness and rise security by allocating scarce resources (Wu and Shahidehpour, 2014).

Actually, 9% of peak demand can be offset if the demand response extended to shield the whole country and extra number of price-responsive programs. Additionally, if the dynamic pricing started to all clients, 20% of peak load will be offset through the demand response program. But, DR integration presents price-related load changing and provides the demand profile stochastic and price-sensitive. Accordingly, the effects of price signal on power market and demand being more noticeable. So, prediction of load may needs more information related to the price while, considering single signal of load cannot be effective in traditional models. Consequently, the load and price are strongly coupled as exogenous data to have accurate forecasting approach.

In Amir et al. (2012) had presented the MIMO method in which had been created on the basis of artificial neural network (ANN) that is used for forecast aided state estimator to study on the relationship of energy cost and load indicators, whereas it does not use correctly the cost rate in all predictions. Additionally, the implemented exterior acquiring pattern is recognized as data association mining and Fuzzy in order to improve the preciseness of prediction. In Shayeghi et al. (2015) had presented a new compound algorithm for the cost forecasting and demand level that applies a range of efficient instruments in preprocessing section, predict the algorithm of engine and its tune. In order to show our assistance, we had divided the suggested forecasting algorithm into two basic sections. The first section applies a wavelet packet transform in order to dissect a sign into different phrases, and a novel characteristics choice approach that uses mutual information and related characteristics to choose the best data for input. The next part had composed of a new MIMO method which is shaped on the basis of nonlinear least square support vector machine to form the linear and nonlinear correlation between the energy cost and load in extra to levels. In Zhao et al. (2007), a new prediction process has been presented for price spike based on classification process. In Amjady (2007), two stage prediction model is suggested for high volatility and complexity of busload time series in price signal.

In this work, three stage forecasting approach is proposed while, the first stage filters the input signal to select the best candidate inputs as feature selection process. The second stage the proposed pattern is extracted or rule in the deviations of input signal predicts from their corresponding observed values. And finally in the third stage, the obtained patterns to the input signal forecasts generated by new optimization algorithm on second stage for the considered forecasting horizon. Consequently, the contribution of this paper can be presented as follows:

(i) Decomposition and reconstruction the input signals to various sub-signals in which the higher and lower pass filters had been changed in order to evade the lacks of WT in lower pass filtering.

(ii) Propose a new feature selection method based on mutual information and relevancy and redundancy terms to extract the valuable load and price candidate signals.

(iii) Present a new MIMO predicator based on modified LSSVM bundled with fuzzy kernel to efficiency find the signal patterns.

(iv) Application of improved gravitational search algorithm (GSA) to enhance the local and global search of standard GSA.

(v) Evaluate the proposed forecasting method on some well-known and real electric market based on error-base indices.

2. Description of proposed tools in forecasting

2.1. The proposed wavelet transforms

In this part, a novel changeable wavelet transform (WT), which is named DWT had been presented, which informs that WT could not master resolution with lower common in oscillatory signs like cost and load. For further study in the field of WT see reference Muthukrishnan et al. (2019). We had selected a comprehensive (mother) wavelet in order to calculate a dyadic wavelet change, next they expanded with double powers. In order to generate a set of performances which are named dyadic wavelet transform we had used wavelet with dyadic dilates simultaneously in basic performance. If $\psi(x) \in L^2(\mathbb{R})$ are considered as wavelet function, it should be satisfying:

$$\int_{-\infty}^{+\infty} \psi(x)dx = 0. \tag{1}$$

Let $\psi_s(x)$ represent the expansion by a factor s:

$$\psi_s(x) = \frac{1}{s}\psi(\frac{x}{s}). \tag{2}$$

By considering (1) and (2), The function of f(x) in scale s and position x, could be determined as follow:

$$W_s f(x) = f * \psi_s(x). \tag{3}$$

The regularity and size of signal features is symbolized by s factor in wavelet transform. For small scale factor, the sustenance of $\psi_s(x)$ reduces, so the fine details signals of $W_s f(x)$ could be achieved. For high value of s, the $W_s f(x)$ eliminates fine specifics, so that large value of signals can be sensed. s and x factor are very important in wavelet transform theory. In some class of this transform, s factor can be sampled using dyadic sequence 2^j ($j \in \mathbb{Z}$). In this case the above equation could be rewritten as;

$$W_{2^j} f(x) = f * \psi_{2^j}(x). \tag{4}$$

By considering $\theta(x)$ as smoothing function, the two function of ψ^1 and ψ^2 is determined by:

$$\psi^1(x) = \frac{d\theta(x)}{dx}, \psi^2(x) = \frac{d^2\theta(x)}{dx^2}. \tag{5}$$

Therefor, the two of wavelets can be rewritten as:

$$\begin{aligned} W^1_{2^j} f(x) &= f * (2^j \frac{d\theta(x)}{dx})(x) = 2^j \frac{d}{dx}(f * \theta_{2^j})(x) \\ W^2_{2^j} f(x) &= f * (2^j \frac{d^2\theta(x)}{dx^2})(x) = 2^{j2} \frac{d^2}{d^2x}(f * \theta_{2^j})(x) \end{aligned}. \tag{6}$$

where, $W^1_{2^j} f(x)$ and $W^2_{2^j} f(x)$ are first and second smoothed signals of f(x). In this relations, the extremum value of $W^1_{2^j} f(x)$ function is related to zero-crossings of $W^2_{2^j} f(x)$ and to the variation points of $f^*\theta_{2^j}(x)$. The other derivatives of smoothing function $\theta(x)$ are not considered.

2.2. Multi-input multi-outputLSSVM

Generally, the support vector machine (SVM) is a learning method that successfully was implemented in many fields of real word (dos Santos et al., 2012). Using optimization method has critical role in finding solution (response) for SVM. Indeed, SVM rooted in the minimizing structural risks. The rule of great limitation in this model limits the complexity of class function and expresses the truth that for some of functions is related to the great limitation with smaller diminution. Based on the different needs in the training speed, the memory limitation and the preciseness of optimization variables enable using different methods. As it said before, this section aims to model multi-input multi-output systems, so other information about the support vector network is provided in reference dos Santos et al. (2012). This paper introduces a novel simplification of SVM model to prediction of multi-variables problems. A MIMO model is defined for learning of relation between sending and received signals of system. By considering multi-dimension regression function the noise signals effects is reduced. Also, the prediction performance against scarce data and nonlinear kinds of noise signals are raised by operation of all channels together. As example, suppose $x_i = [x_{\ ij} :: :x_{id}]$; $i = 1 + 1, \ldots,$ n is the input signal of load price with output y_i. the simple regression prediction model will be find the relation between input vector $x \in \mathbb{R}^d$ and output $y \in \mathbb{R}$ using the given data. The classic SVM model will be solved this problem based on minimizing of $\|W\|^2/2 + C\sum_{i=1}^{n} L_v(y_i - (\phi^T(x_i)w + b))$. The SVR model needs to specify a kernel function $k(x_i, x_j) = \phi^T(x_i)\phi(x_j)$ for finding the map of input and observed output data. Hence, based on decreasing training error and weight function, we had:

$$L_P \min_{(W,b,e)} = \frac{1}{2}\sum_{j=1}^{B}\|w^j\|^2 + \frac{1}{2}\gamma\sum_{i=1}^{n} L(\|e_i\|) \tag{7}$$

In which

$$W = [w^1, \ldots, w^B], b = \begin{bmatrix} b^1 \\ \vdots \\ b^B \end{bmatrix}, \|e_i\| = \sqrt{e_i e_i^T} \tag{8}$$

$$e_i^T = y_i^T - \phi^T(x_i)W - b^T,$$
$$L(\|e\|) = \begin{cases} 0 & , \quad \|e\| < \varepsilon \\ \|e\|^2 - 2\|e\|\varepsilon + \varepsilon^2, & \|e\| > \varepsilon \end{cases} \tag{9}$$

Eq. (9) describes the Vapnik ε-insensitive loss function. In the above equation, when $\varepsilon = 0$, the kernel function is defined distinctively for each signal but when $\varepsilon > 0$ there will be created a connection between the price and load signal in the forecasting model. By substitution of Eqs. (8)–(9) in Eq. (7) we will have a non-linear function that we used IRWLS method to solve it. Due to the non-linear nature of constructed equation it could not be solved directly (Shayeghi et al., 2015). Now we will consider Taylor series expansion:

$$\begin{aligned} L_P'(W,b) &= \frac{1}{2}\sum_{j=1}^{B}(\|W^j\|^2 \\ &+\frac{1}{2}C\sum_{i=1}^{n}L(v_i^l + \frac{dL(v)\cdot(e_i^l)^T}{du\cdot v_i^l}|_{v_i^l}[e_i - e_i^l])) \end{aligned} \tag{10}$$

In which

$$\begin{cases} v_i^l = \|e_i^l\| = \sqrt{(e_i^l)^T e_i^l} \\ (e_i^l)^T = y_i^T - \phi^T(x_i)W^l - (b^l)^T \\ \forall W \in R^H \times R^B, b \in R^B \ when \ W = W^l \ and \ b = b^l \\ then, L_P'(W^l, b^l) = L_P(W^l, b^l) \\ \nabla L_P'(W^l, b^l) = \nabla L_P(W^l, b^l) L_P'(W^l, b^l) \le L_P(W, b) \end{cases} \tag{11}$$

Now with expanding the second class equation, will obtain:

$$\begin{aligned} L_P''(W,b) &= \frac{1}{2}\sum_{j=1}^{B}(\|W^j\|^2 \\ &+\frac{C}{2}(\sum_{i=1}^{n}L(v_i^l + \frac{dL(v)}{dv}|_{v_i^l}\frac{v_i^2 - (v_i^l)^2}{2v_i^l})) \\ &= \frac{1}{2}\sum_{j=1}^{B}\|W^j\|^2 + \frac{1}{2}\sum_{i=1}^{n}\zeta_i v_i^2 + SC \end{aligned} \tag{12}$$

To solve the above equation, the following steps are provided:

First step: Determining primary quantities for $W^1 = 0$, $B^1 = 0$ and $1 = 0$ consequently to compute primary quantities of $u^1{}_i$ and a_i.

Second step: arrangement of weights and obtained bias from Eq. (13) as a vector:

$$P^l = \begin{bmatrix} W^s - W^l \\ (b^s - b^l)^T \end{bmatrix} \tag{13}$$

Third step: computing next answers:

$$\begin{bmatrix} W^{l+1} \\ (b^{l+1})^T \end{bmatrix} = \begin{bmatrix} W^l \\ (b^l)^T \end{bmatrix} + \gamma^l \tag{14}$$

where, γ^l corresponding is achieved with program iteration and at first equals with 1.

Fourth step: Computing the primary quantities of u_i^{l+1} and a_i and adding training repetitions of $j = j+1$, if the ending condition which is the quantity of final responses, take place otherwise go back to the second step. As it determined in the second step obtaining W^s and B^s is unavoidable and for simplification can use gradient equivalence with zero as follow:

$$\nabla_{W^j} L_P'' = W^j - \sum_i \Phi(x_i)\alpha_i(y_{ii} - \Phi^T(x_i)W^j - b^j) = 0$$
$$\nabla_{b^j} L_P'' = -\sum_i \alpha_i(y_{ii} - \Phi^T(x_i)W^j - b^j) = 0 \tag{15}$$
$$j = 1, 2, \dots, B$$

By making linear of answers as Matrix we have:

$$\begin{bmatrix} I + \Phi^T\Omega_a\Phi & \Phi^T a \\ a^T\Phi & I^T a \end{bmatrix}\begin{bmatrix} W^j \\ b^j \end{bmatrix} = \begin{bmatrix} \Phi^T\Omega_a y^i \\ a^T y^i \end{bmatrix} \tag{16}$$

While, $\phi = [\phi(x_1), \dots, \phi(x_n)]^T$, $a = [a_1, \dots, a_n]^T$, $(\Omega_a)_{ij} = a_i\delta(i-j)$, and $y^j = [y_{1j}, \dots, y_{nj}]^T$. Each column of W^s are composed with the solutions of Eq. (16) for every j.

2.3. Fuzzy feature selection algorithm

2.3.1. Approaching degree

The approaching degree is applied to assess the similarity between two given fuzzy sets in a particular universe of discourse X (Sharmin et al., 2019). Let assume $S(X)$ be a power set of normal fuzzy sets with $A_k^j \ne 0$ and $A_k^j \ne X$. Let A, B be two normal fuzzy sets where A, $B \in S(X)$. The similarity degree between two fuzzy sets A and B is assessed in:

$$SM(A,B) = \min((A\cdot B), \overline{(A \oplus B)}) \tag{17}$$

where $(A \cdot B)$ the inner product is defined by Eq. (2): and $(A \oplus B)$ is the outer product defined by Eq. (3)::

$$(A\cdot B) = \max(\min[\mu_A(x), \mu_B(x)]), x \in X \tag{18}$$

$$(A \oplus B) = \min(\max[\mu_A(x), \mu_B(x)]), x \in X \tag{19}$$

Particularly, when the value of SM(A, B) approaches a value 1, this represents that the two fuzzy sets A and B are "more closely similar". When SM(A, B) approaches a value 0, the two fuzzy sets are "more dissimilar". Since each fuzzy set represent a cluster we represent similarity between two clusters same to similarity between two fuzzy sets.

The evaluation mechanism used to evaluate the best feature subset is given in the following definitions:

Definition 1. Similarity between two clusters for given m-dimensional features:

Let C_i be a fuzzy cluster given by $C_i = \{C_{i1}, C_{i2}, .., C_{im}\}$ where M is the number of features in a particular feature subset, C_{i1} is the membership function for the ith cluster in the feature 1, C_{i2} is the membership function for cluster i in the feature 2 and so forth. Then the similarity between two clusters C_x and C_y is given by: $SM\left(C_x, C_y\right) = \max_{i=1}^{M}\left(W_i.SM_{F_i}\left(C_{xi}, C_{yi}\right)\right)$, $x \ne y$. $\forall$k $= 1, 2 \dots$ n.$\sum_{i=1}^{M} W_i = 1$, W_l is normalized weighting factor representing the importance of some features among others.

Definition 2. Overall similarity between all clusters in a feature subset S_i is given as:

$E_{s_i} = \max_{j=1}^{N}\left(SM\left(C_x, C_y\right)\right)$ where N represents number of possible combination between pair of clusters.

Definition 3. Predictive feature subset D_{best} is the one with minimum E_{si}.

2.3.2. The FFSS algorithm

The goal of our algorithm is to assess the similarity degree between two given fuzzy clusters and consequently the similarity between all clusters for all selected features. We presume that as long as the similarity degrees between all pairs of clusters for selected features is as small as it can be, then the clustered data are well distributed. Hence, they present predictive features. The similarity between all cluster pairs for a feature subset is given by:

$SM(C_1,C_2) = max[W_{FA}.SM_{FA}(C_1,C_2),\ W_{FB}.SM_{FB}(C_1,C_2),\ W_{FC}.SM_{FC}(C_1,C_2)]$

$SM(C_1,C_3) = max[W_{FA}.SM_{FA}(C_1,C_3),\ W_{FB}.SM_{FB}(C_1,C_3),\ W_{FC}.SM_{FC}(C_1,C_3)]$

$SM(C_2,C_3) = max[W_{FA}.SM_{FA}(C_2,C_3),\ W_{FB}.SM_{FB}(C_2,C_3),\ W_{FC}.SM_{FC}(C_2,C_3)]$

Then overall similarity: $E_{si} = min[SM(C_1,C_2), SM(C_1,C_3), SM(C_2,C_3)]$

After calculating E_{si} for all possible feature subsets, the best feature subset is one with minimum E_{si}

Our proposed algorithm is described by five main steps:

Step1: select a particular feature subset from available feature space (e.g. $S_1 = \{F_1, F_3, F_4\}$) and weight each feature in this subset by finding out value of R-squared that indicates how much the selected feature can explain the output which in our case is the total development effort. R-squared is the fraction of the variation in the value of the observed value that is explained by the regression (Amir et al., 2012). The weighing parameter is crucial task in the similarity measures because it takes into account the relevancy of particular features for effort estimation.

Step 2: Fuzzify each feature subset using fuzzy identification model based on FCM clustering. To fuzzify clustered dataset, there are two main ways: first is the expert knowledge (Pablo et al., 2009) which is formed in if-then-rules where parameters and memberships are tuned using input and output data. The second is no prior knowledge about the system (Pablo et al., 2009), so the model is built based on particular algorithms.

Step 3: Assess similarity degree between all pairs of clusters in each feature subset using Definition 1.

Step 4: Assess overall similarity degree in each feature subset using Definition 2.

Step 5. Evaluation: after calculating similarity degrees for all features subsets, we extract the subset that has minimum similarity degree E_{si} according to Definition 3. The output of the algorithm is the best feature subset.

Input:
D(F_1, F_2, … F_M) //input dataset (NxM)
Out //Output Dataset(Nx1)
Output:
D_{best} //feature subset of high predictive features.
begin
Do:
Step1: Select feature subset to be searched S_i.
Step2: Fuzzify the feature subset (S_i).
Step3: For feature subset S_i, assess similarity degree between all pairs of clusters (i.e. fuzzy sets) in all features in S_i.
Until all feature subsets are searched.
Step4: Evaluate each feature subset S_i using E_{si}
Step5: Evaluation: best feature subset is one with minimum $E_{s_i}.\min_{i=1}^{k}\left(E_{s_i}\right)$
End;

2.4. Proposed modified GSA

2.4.1. The standard GSA

In this section the standard (single-objective) GSA is briefly reviewed. Interested readers are referred to Giladi and Sintov (2020) for more details. To describe the original GSA, consider a vector with s agents (x_i^d is ith mass in the dth dimension) and the position of ith mass is given by:

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n), i = 1, 2, \dots, s \tag{20}$$

After that mass of ith agent is calculated as follows:

$$M_i(t) = \frac{q_i(t)}{\sum_{j=1}^{s} q_i(t)} \tag{21}$$

where, $M_i(t)$ and $q_i(t)$ are the mass value of the ith agent at time t and the gravitational mass, one gets:

$$q_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{22}$$

where, $\text{fit}_i(t)$, worst(t) and best(t) are the fitness value of the ith agent at time t, worst and best fitness are given by:

$$\begin{cases} worst(t) = \underset{j\in\{1,\dots,s\}}{Max} fit_i(t) \\ best(t) = \underset{j\in\{1,\dots,s\}}{Min} fit_i(t) \end{cases} \tag{23}$$

To obtain the acceleration of the agent i, $a_i^d(t)$, total forces $F_i^d(t)$ from a set of heavier masses should be considered based on the law of gravity by:

$$\begin{cases} F_i^d(t) = \sum_{j=kbest, j\neq i} rand_i G(t) \frac{M_j(t) M_i(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \\ a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j=kbest, j\neq i} rand_j G(t) \frac{M_j(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \end{cases} \tag{24}$$

Thenceforth, next velocity of the agent is computed as:

$$V_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{25}$$

Then, the agent position can be updated by:

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{26}$$

where, $rand_i$ and $rand_j$ are two uniformly distributed random numbers in the interval [0, 1], $R_{ij}(t)$ is the Euclidean distance between two agents i and j, $R_{ij}(t) = \|X_i(t), X_j(t)\|_2$, ε is a small value, k_{best} is the set of first K agents with the best fitness value and biggest mass, which become a function of time with the initial value K_0 at the beginning and decreasing with time. Here, K_0 is set to the total number of agent's s and is eliminated linearly to 1. G shows the gravitational constant, initialized to G_0, and updated with time:

$$G(t) = G(G_0, t) \tag{27}$$

2.4.2. Modified GSA

To enhance the local and global ability of standard GSA, two modifications in the initial population and cognitive/social coefficients and extra recover operator to expand the GSA search ability is proposed. These modifications are:

(i) Initial population

How generate the initial population in a population-based algorithm is very important, since it may leads to least computational burden time and escapable option from local solutions. Generally, there is no prior knowledge of the optimal solution, therefore, the initial population (or solution) generates randomly. Of course the successful model among them, one that could provide better distribution in the search space. Let f be the objective function in feasible search space S with variables x_i, i = 1, 2, …, n. In a small neighborhood of x_i, if $f(x_i)<f(x)$ then have a local minimum while the global minimum is lowest value for $f(x^*)$ throughout the feasible space. Let $I^n \subset R^n$ be the closed n-dimensional unit cube, P denotes a set of $x^1, \dots, x^n \subset I^n$. For an arbitrary subset B and its characteristic (c_B), one gets:

$$A(B; P) = \sum_{n=1}^{N} c_B(x_n) \tag{28}$$

If B is a nonempty family of Lebesgue-measurable subsets of I^n (Niederreiter, 1992), then discrepancy can be calculated by:

$$D_N(b; P) = \sup_{B\in b} \left| \frac{A(B; P)}{N} - \lambda_S(B) \right| \tag{29}$$

Note that $0 < D_N(b; P) < 1$. By appropriate specializations of the family b, get two most vital concepts of discrepancy. We put $I^n = [0, l)^n$.

Concept (i) The star discrepancy $D_N{}^*(P) = D_N{}^*(x_1,\dots,x_N)$ of the point set P is defined by $D_N{}^*(P) = D_N(b^*;P)$, where b^* is the family of all subintervals of I^n of the form $\prod_{i=1}^{n}[0, u_i)$.

Concept (ii) The extreme discrepancy $D_N{}^*(P) = D_N{}^*(x_1,\dots,x_N)$ of the point set P is defined by $D_N^*(P) = D_N(b^*;P)$, where b is the family of all subintervals of I^n of the form $\prod_{i=1}^{n}[u_i, v_i)$.

Based on the above description, $D_N^*(P) \le D_N(P) \le 2^n D_N^*(P)$. In order to have best relation between star discrepancy and dispersion ($d_N(P;X) = \sup_{x\in X} \min_{1\le k\le N} d\left(x, x^k\right)$), and have least error in population distribution, one gets:

$$w(f;t) = \sup_{\substack{x,y\in X \\ d(x,y)\le t}} |f(x) - f(y)|, t \ge 0 \tag{30}$$

Which the related error bound shows low-dispersion sequence for random search method by:

$$\min_{1\le k\le N} f(x^k) - f(x^*) \le w(f; d_N(P;X)) \tag{31}$$

(ii) Chaotic operator

Recently, chaos theory shows popular performance in the optimization algorithm (Yang et al., 2007). It makes an optimization algorithm more powerful in local search when faced with a large number of variables. One of well-known chaos model is Lorenz system which comes from a natural behavior of the earth's atmospheric convection flow heated from below and cooled from above:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{32}$$

where, x, y and z come from system state, t is time. σ, ρ and β denote the system parameters. In this paper, the chaos theory has following steps:

Step1: Set control parameters of chaos and algorithm

Step 2: Generate initial population between lower and upper bounds of variables.

Step 3: Map the decision variables.

Step 4: Convert the chaotic variables with decision variables.

Step 5: Calculate the new solution with decision variables.

3. Proposed price and load forecasting

In the proposed MIMO prediction model, an incorporating analysis between load and price prediction in single framework will be done. High accurate, adaptive and effectiveness are the main advantages of the proposed model. In first step, based on day D-1 historical available data, the price and load initial value are predicted for day D. A correlation analysis of this signals by initial price and load signals prediction in previous step, will be done in second step. All of day D-1 price and load signals plus day D initial price prediction value are as input of second step. Based on correlation analysis, day D signals will be calculated and replace with previous values. This process will be done continuously until minimum difference between predicted signals is observed. In this part, we explained the suggested simultaneous cost and load forecasting in smart grids in ordered phrases. To put it in another word, we demonstrated that how the provided tools in prior parts cooperate together in the suggested forecasting model. We deliberated and classified the mentioned phases:

Phase 1: the previous cost and load vectors are determined in place of input signals. Hence, $X_p = \{X_{p1}, X_{p2}, \ldots, X_{pm}\}$ and $X_l = \{X_{I1}, X_{I2}, \ldots, X_{Im}\}$ with the cooperation of m- variables change the price (subscript p) and load (subscript I) signals. Moreover, in order to achieve optimal scattering or a comprehensive range of X_p and X_l, we had applied $X_{np} = (X_p - \min(X_P)\ (max\ (X_p) - min\ (X_p))$ and $X\,nl = (X_l - min\ (X_l)\ (max\ (X_l) - min\ (X_l))$ as standard to obtain $X_{np} = \{X_{np1}, X_{np2}, \ldots, X_{npm}\}$ and $X_{nl} = \{X_{nl1}, X_{nl2}, \ldots, X_{nlm}\}$. Reminder that the act of standardization alters the feature section function, it cuts back the unreliability and redundancy. At this time these experimented vectors are prepared for next phase to cushion.

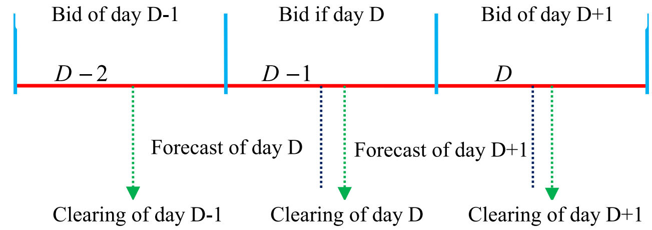

Fig. 2. Suggested time model of 24-ahead price forecast.

Phase 2: this phase uses suggested DWT to dissect the X_{np} and X_{np} in two four distinct sub parts, a_h, b_h, and d_h, the two sets of X_{np} and X_{np} are achieved by appraisal and detail coefficients which is then is illustrated in cA_p (or cA_L) and cA_p (or (cD_L).

Phase 3: use the suggested feature selection on the finest subparts of wavelet tree (the outputs are determined as signals and candidate features). Regarding the suggested feature selection, in the X_{np}^{best} and X_{nl}^{best}, the finest section with highest information will be chosen. This procedure is used to decrease the selection of unpleasant information that interrupts training process.

Phase 4: according to the existing correlation of cost and load signals, this phase presents a precise short-term forecasting. The elaborateness of simultaneous cost and load forecasting resulted to implementing of MIMO LSSVM. The suggested approach provides suitable performance for linear/nonlinear terms as indicated in Fig. 2. The prediction price as day D is generated based on previous day. The price signal at day D i.e., 24 h, is broadcasted through Independent System Operator (ISO) for *D-2*. Accordingly, the real prediction of price signal for day D, can provides among the clearing hour for D-1 of D-2 as well as the biding hour of D of D-1.

Phase 5: now, use combination procedure of DWT to achieve the end forecasting outcomes of cost and load such as; $W^{-1}(\{a_h^{est}, b_h^{est}, c_h^{est}, d_h^{est};\ h = T+1, \ldots, T+No\}) = P_h^{W,est}, No = \{24, 168\}$. Regarding to the previous step the objective performance will be:

$$Obj = \frac{1}{No}\left(\sum_{t=1}^{No}(\alpha_1 \frac{X_{a_t}^P - X_{f_t}^P}{X_{a_t}^P} + \alpha_2 \frac{X_{a_t}^L - X_{f_t}^L}{X_{a_t}^L})\right) \tag{33}$$

where, the factors of a_t and f_t stands for the real and predicted quantities. And the factors of P and L show the related information to cost and load. And at the last, stands for the factors with fixed quantities.

Phase 6: this phase aimed to achieve the finest training procedure with the lowest error rate, so we had suggested MGSA algorithm in order adjust LSSVM restrictions. It the factor of shopping has the satisfaction rate in the forecasting outcomes, in other case go back to phase 3. The proposed forecasting flowchart is shown in Fig. 3.

4. Simulation and discussion

4.1. Original data policy

The power cost and demand information from the New York independent system operator (NYISO) (NYISO, 0000) in 2014, of New South Wales (NSW) region in Australia marketplaces which is referred in NSW (0000) in 2010 and PJM (PJM, 0000) of 2013 are implemented in the process of simulation. All utilized data are achieved from available resources. The suggested hybrid algorithm for forecasting is conducted in the MATLAB. The statistical simulation for the study is achieved from a test within four weeks particularly in the weeks of Mar, Jun, Sep, and Dec. put it in another words, we conducted the train process in the two first months of season and first weeks of 3th month to estimate the preciseness of prediction. In order to assess the function of suggested algorithm we had applied a few of indexes lower smaller than the values that stands for the optimal forecasting preciseness.

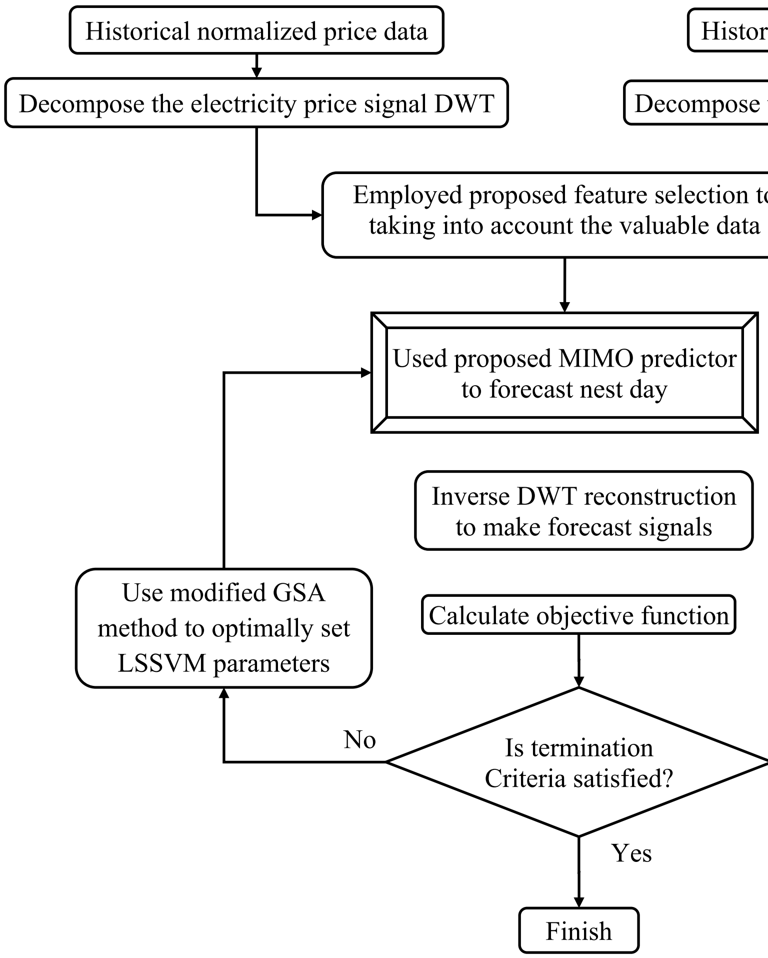

Fig. 3. Proposed forecasting flowchart.

4.2. Statistical indices

The mean absolute percentage error (MAPE) will be achieved by using following equations:

$$MAPE_{day/week} = \frac{1}{No}\sum_{i=1}^{No}\frac{|X_{iACT} - X_{iFOR}|}{X_{AVE-ACT}} \tag{34}$$

$$X_{AVE-ACT} = \frac{1}{No}\sum_{i=1}^{No} X_{iACT} \tag{35}$$

The forecast mean square error (FMSE) is obtained:

$$FMSE_{day/week} = \sqrt{\frac{1}{No}\sum_{i=1}^{No}(X_{iACT} - X_{iFOR})^2} \tag{36}$$

Finally, the following equation is used to estimate the root mean square (FMSE):

$$RMSE_{day/week} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i^{act} - P_i^{for})^2} \tag{37}$$

In the above equations, X_{iFOR} and X_{iACT} are predicted and real quantities that factor X recognized them as cost or load signals, and $X_{AVE-ACT}$ which are the mean quantities of X_{iACT}. X_{NO} shows the mean load and cost of each provided no = {24, 168}.

4.3. NYISO electric market

Marketplaces of NYISO presents a comprehensive chance among power markets due to the feature it includes three cost — responsive load plans that permits to the product users and offers them load rate in wholesale marketplaces. The NYISO synchronizes the supplements of direct arrangements and ancillary services that could not supply themselves. From the point that this markets presents great services, both in price rate and load quality they have more customers. The estimation of power cost, load and their connection in the range of 1/1/2014−1/3/2014 is illustrated in Fig. 4. In the suggested forecasting model, the obtained data of 600 participate inputs were investigated. To put it simply, the forecasting engine and the gathered data set are kept stable along with efficiency of feature selection model and it could be compared with presented model in Shayeghi et al. (2015). Additionally, the predicted and real signals of price and load which is obtained from sample test week for example week 1 (the most conducted awful test) are illustrated by Figs. 5 and 6 to present a total diagram view of forecasting preciseness which is obtained by suggested hybrid forecasting algorithm. As it observed from Figs. 5 and 6, the most awful test duration are investigated such as forecasting curve which is achieved by suggested algorithm based on actual curve shape of Figs. 5 and 6. By implementing the suggested indexes of Section 4.2, the comparison of suggested hybrid algorithm and two previously used methods are illustrated by Fig. 7.

The horizontal axis demonstrates the plotting indexes whereas the vertical axis demonstrates the percentage indexes. In comparison to other investigated methods, the suggested method has better function.

Considering this indexes, the second suggested method revealed same outcomes for weekly and daily MAPE, FMSE, and RMSE for the suggested model. Whereas, the proposed forecasting method obtained better results in all conducted test weeks compare to other available methods, particularly in the weekly indices of power price.

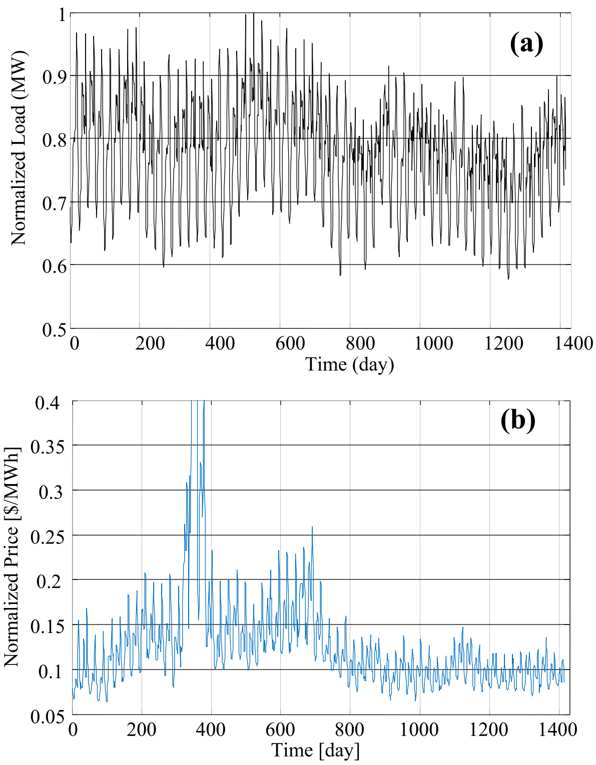

Fig. 4. (a) Standardized load, (b) standardized power cost.

4.4. PJM electric market

We had presented PJM of United States power marketplaces for the second data test. In order to model the trait of energy plants in the power marketplace, it is postulated that the bid cost of energy plants will be the same as actual markets like PJM. From the point that the form of price duration curve (PDC) has been formed as a result of bid cost, which monitors the marketplaces, the obtained PDC from the PJM is implemented to generalize the offer cost of production sections. Fig. 8 illustrates the hourly power cost, load and their relationship from date 01/01/2013 to the 31/12/2013.

PJM had used of same strategy like NYISO marketplace. Figs. 9 and 10 demonstrates the graphical view of power cost and load in the duration of day and week.

In order to raise the efficiency of suggested hybrid forecasting algorithm, we analyzed the suggested model by more indexes. Fig. 11 shows the numerical analysis of forecasting error. It is clear that, the suggested model has optimal function in the forecasting preciseness for all implemented indexes.

The proposed integrated strategy of load and price prediction results is presented in Figs. 9 and 10. As can be seen from of these figures, the obtained values for load and price are follow the actual values in most hours and more closely to their main signals. Because of effectiveness of the proposed model, it can be used for optimal bidding strategy in day-ahead markets for all of system operators. The evening peak value of price and load and earlier morning off-peak using the proposed model is highlighted in Figs. 9 and 10, too.

4.5. NSW electric market

For the ultimate test marketplace, we investigated NSW in order to compare the function of suggested hybrid algorithm by ANN-based MIMO network (Amir et al., 2012). The NSW power system had composed of each distinct area from Australian national market (NEM) and is attached to the areas of Queensland (QLD) AND Victoria (VIC) through interconnected lines. And transmits power in the areas and forms a unique pool market throughout the Australian NEM (which id determined for all transactions). Australian energy market operator (AEMO) manages the unique pool marketplace. There are two demand response program kinds in NSW market. From these two types, the first type is based on economic. It is developed in scheme of time-to-use (TOU) pricing, and is used for users that are equipped with smart meters. Weekdays contain three rates of off-peak, shoulder-peak and peak rates. The off-peak rate is in the period of 10 pm to 7 am; shoulder-peak is in periods of 8 pm to 10 pm and 7 am to 2 pm; the peak rate is in the period of 2 pm to 8 pm. The weekend and holidays also contain two rates of off-peak and shoulder-peak rates. These two rates are in the periods of 10 pm to 7 am and 7 am to 10 pm, respectively. So, users who equipped with smart meters will be charged the allocated rates for every half-an hour time range, and these users are able to shift their use from times with high prices to times with low prices for saving their power expenses. Users change the power price in such pricing schemes in addition to change the power demand trend. Another demand response program is based on reliability, in which the consumers who are enrolled asked for decreasing their power use while supply shortfalls. Just data of economic-based programming is used here as the input data to design our models. The PI (Period Index) is introduced for each hour (t) in the following form:

$$PI_t = \begin{cases} 1.\ \text{if t is in a off-peak period in weekdays} \\ 2.\ \text{if t is in to a shoulder-peak period in weekdays} \\ 3.\ \text{if t is in to a peak period in weekdays} \\ 4.\ \text{if t is in to a off-peak period in weekends} \\ 5.\ \text{if t is in to a shoulder-peak period in weekends} \end{cases} \tag{38}$$

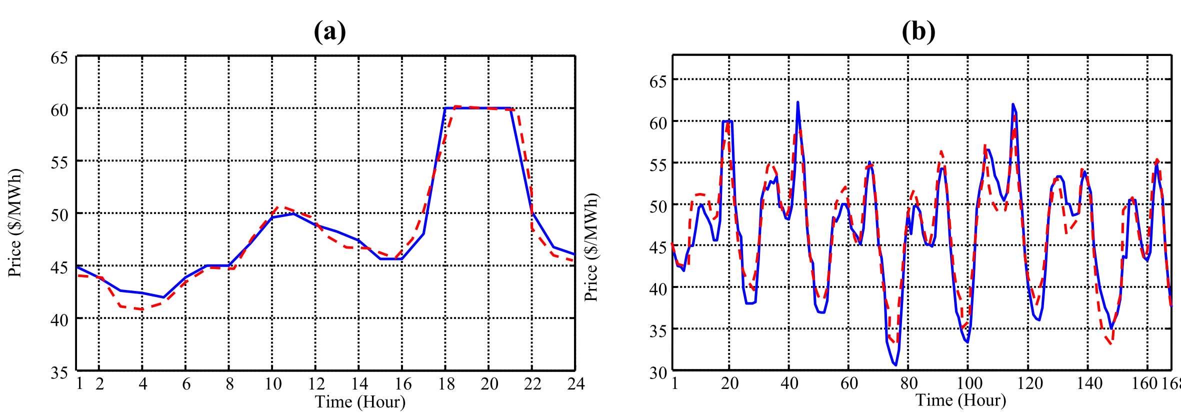

Fig. 5. Obtained forecasting outcomes of first week of Mar for the NYISO power market; the blue lines show the real cost signal and red lines show the predicted cost signal, subplots of (a) and (b) show the daily and weekly cost prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

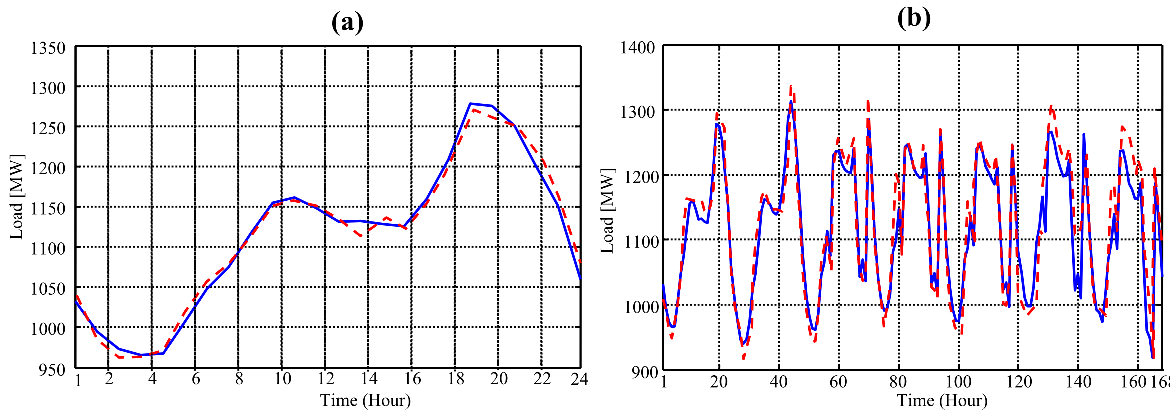

Fig. 6. Obtained forecasting outcomes for the first week of Mar for the NYISO power market; the blue lines show the real load signal and the red lines show the predicted load signals, subplots (a) and (b) show daily and weekly load prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

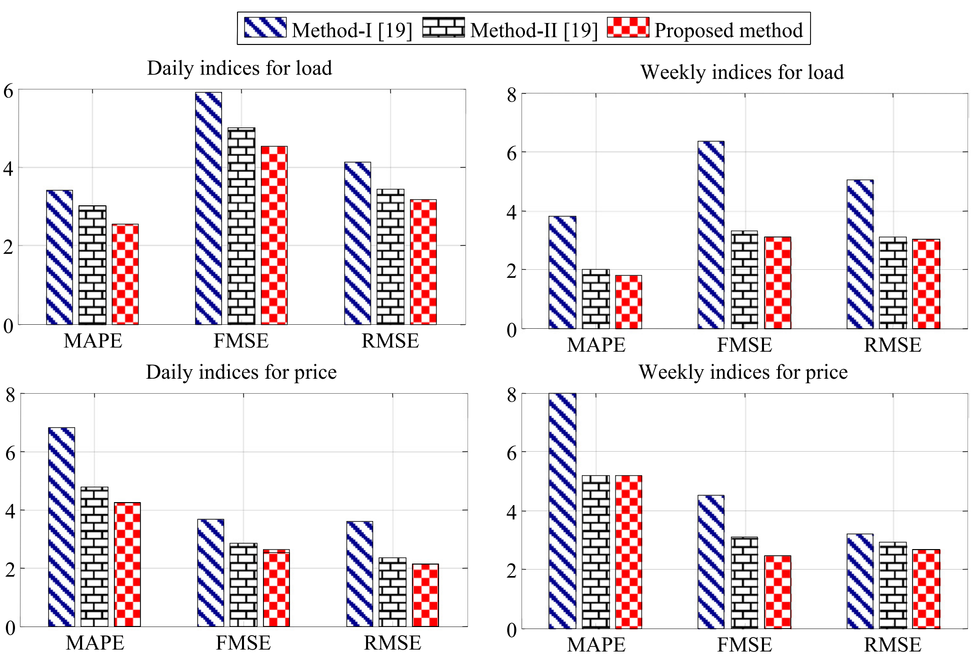

Fig. 7. Comparison of three suggested method with the assessed indexes of daily and weekly power cost and load prediction, Method-I: MIMO-LSSVM type I and Method-II: MIMO-LSSVM type II.

The PI is a metric that determines the related time interval for each hour regarding the TOU pricing rates. It also presents a MIMO engine with additional data respect to the existing demand response program in NSW zone.

The suggested framework is implemented for 4 test weeks in the considered market. Some rules are identified for each one of test weeks, and obtained rules changed from a week to another one. As an instance, 5 rules are obtained for first week in March that have high importance regarding the pre-determined thresholds of rule support, lift and confidence. One of the obtained rules is presented as:

$$(D^{f,i}_{t,d} \to H \wedge P^{f,i}_{t,d} \to H \Rightarrow DDF_{t,d} \to HD \wedge DPF_{t,d} \to HD) \tag{39}$$

According to this rule, once the predicted demand and price have high values in hour t, real demand and price of hour t significantly will drop in comparison with the predicted value. The allocated confidence of this rule is 100 percent and it is happened 7 times in February. Anther obtained rule is presented as:

$$(D^{f,i}_{t,d} \to VL \wedge P^{f,i}_{t,d} \to VL \Rightarrow DDF_{t,d} \to LR \wedge DPF_{t,d} \to LR) \tag{40}$$

According to this rule, once the predicted demand and price have high values in hour t, a low growth will be occurred in the real demand and price in comparison with the predicted value. Table 1 provided the predicted results for 4 test weeks.

Regarding Table 1, the modified forecast can result in lower demand and price prediction errors in comparison with other ones. Totally, about 2.10% and 7.57% enhancement are achieved for demand and price prediction errors, respectively. Complication of price and demand interactions can influence this enhancement. Regarding the limited interactions in existing markets, this enhancement will support the performance of the suggested approach.

4.6. Sensitivity evaluates of the proposed model

The impacts of proposed strategy rules, variables considering and finally the prediction machine are studied on the performance of the proposed prediction strategy. Two previous models contain; MIMO prediction model of Amjady (2007), NARX model of Espinoza et al. (2007) have been replace by the proposed model. Also, in order to price/load prediction the modified ref Espinoza et al. (2007) model

Table 1
Obtained value for MAPE (%) of NEW power market.

Month	Demand			Price		
	ANN-MIMO (Amir et al., 2012)	LSSVM-MIMO (Shayeghi et al., 2015)	Proposed method	ANN (Amir et al., 2012)	LSSVM-MIMO (Shayeghi et al., 2015)	Proposed method
Mar.	2.54	2.14	2.01	9.85	7.03	5.93
Jun.	2.27	2.20	2.12	15.70	12.03	10.51
Sep.	2.61	2.54	2.53	7.75	6.59	5.69
Dec.	1.81	1.75	1.75	6.76	6.34	5.65
Mean	2.31	2.15	2.10	10.01	8.00	6.94

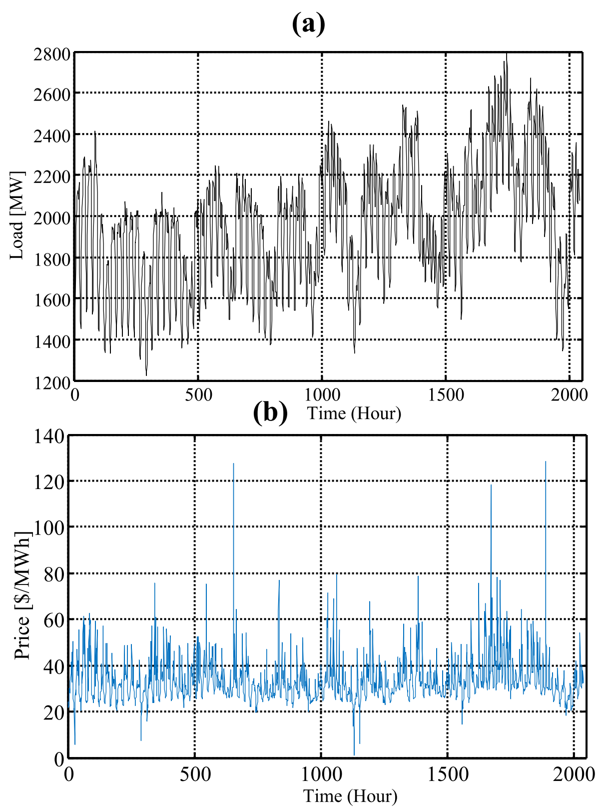

Fig. 8. (a) Standardized load, (b) standardized power cost, from the NYISO power marketplace.

is applied to those models. In Espinoza et al. (2007), it is reported that NARX model is a robustness method against exogenous inputs in prediction of variables, therefore we analysis effects of load and price together for analysis of our robustness against NARX method. The proposed method and NARX method is applied to NSW zone real data, the results of simulation shows the 14.12% improvement for demand and 15.21% improve for load prediction using the proposed strategy.

Table 2
MAPE (%) of the section analysis prediction error of PJM electricity market.

Methods	April,	July	October	January	Mean
WT + MI + MIMO + GSA	4.644	5.627	4.93	4.617	4.954
WT + MMI + MIMO + GSA	4.084	3.92	4.414	4.021	4.109
DWT + MMI + MIMO + GSA	3.318	3.819	3.849	3.516	3.625
DWT + MMI + MIMO + MGSA	2.815	3.243	3.011	3.105	3.043

4.7. Analyze of different section effects

In this section, the effects and performance of each layer of proposed model have been studied. These sections are pre-prediction section, learning center, and optimization method, which the results of this analysis are presented in Table 2.

As seen from of results, obtained error of proposed model in price prediction 16% is better than other studied models. Also, the results of WT, MI, MIMO, GSA is not acceptable. The main factor for this shortage, can be result from the small historical data size and variation of them. From of results, when any part of the proposed strategy is ignored, the results accuracy is reduced. Also, the any section of the proposed strategy effects on results are different with other parts and by replacing of them, average MAPE changes from 4.962% to 3.051%. The experimental results show that the proposed method can predict the price and load with high accuracy. In sum up, the advantage of the proposed strategy could be descripted in 3 features; 1- the DWT is transfer the time series of price and load to better structure series in different domain, 2- the proposed MIMO model without considering linear/nonlinear behavior of signals can predict the associated load/price signals. 3- the optimal variables values of LSSVM model can be determined by applying MGSA method.

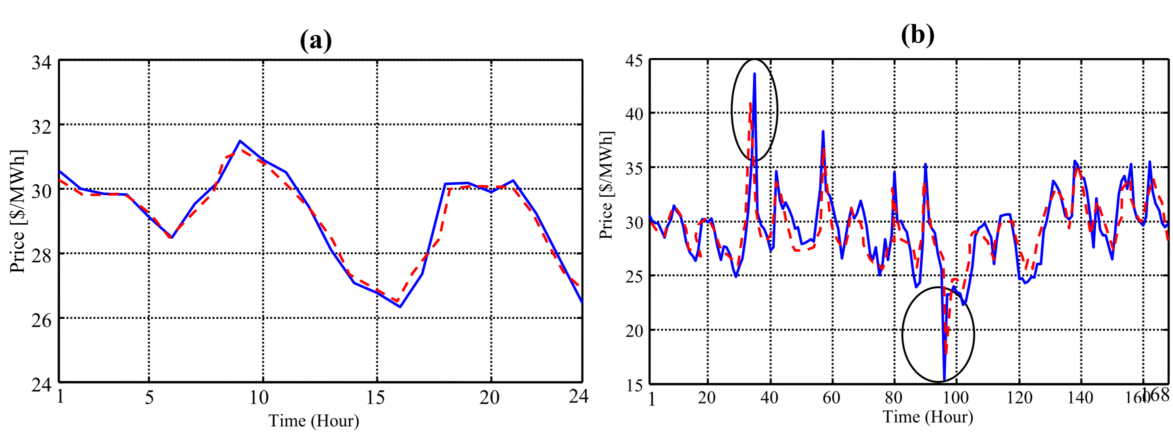

Fig. 9. The obtained outcomes of forecasting from a day an week of the PJM power marketplace; the blue lines show the real cost signal and the red lines show the predicted cost signal, subplots (a) and (b) shows daily and weekly cost prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

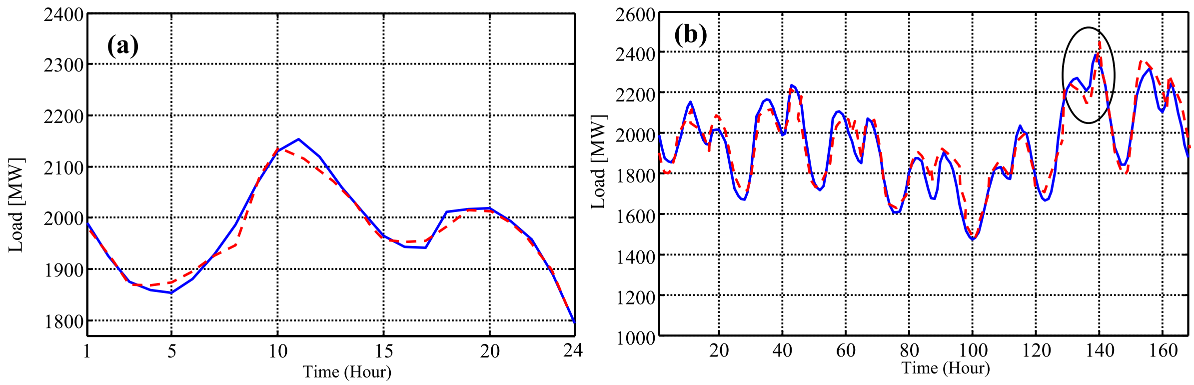

Fig. 10. The obtained outcomes of forecasting from a day an week of the PJM power marketplace; the blue lines show the real load signal and the red lines show the predicted load signal, subplots (a) and (b) shows daily and weekly load prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

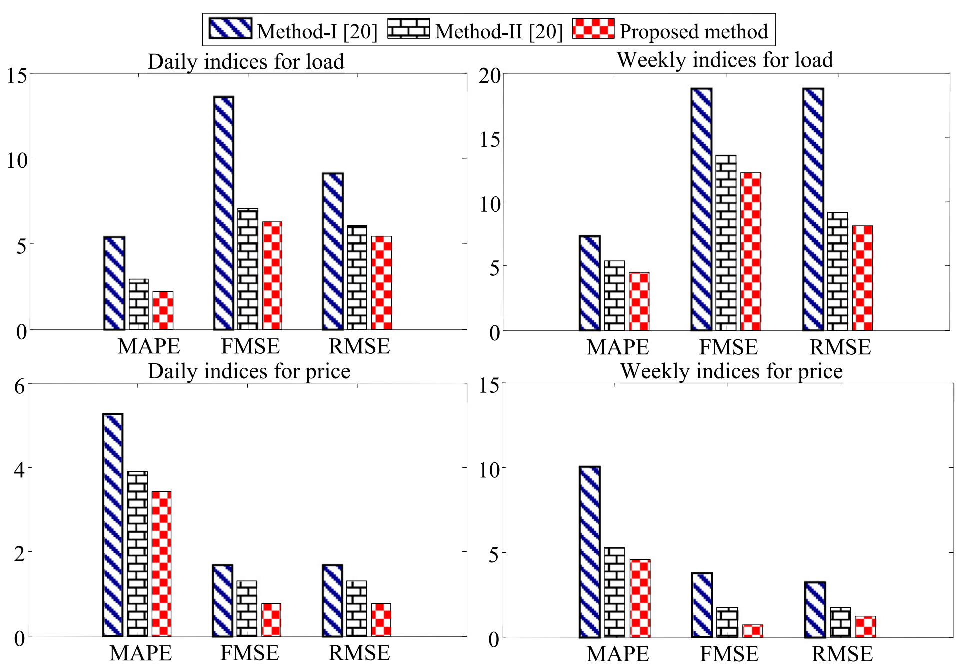

Fig. 11. Comparison of three models with assessment indexes in daily and weekly power cost and load prediction Method-I: MIMO-LSSVM type I and Method-II: MIMO-LSSVM type II.

5. Conclusion

In this work, a new composed prediction model for both price and load signals is presented. The suggested approach is working by bidirectional effects of both signals on each other. In this model, the signal is decomposed by proposed DWT and then filtered through the MMI feature selection approach. After that, the signal is predicted by proposed forecast engine which is coupled by MGSA algorithm, for MIMO approach. The suggested model is compared with various well-known forecasting models which proof the validity of proposed model in terms of high accuracy, better statistically behavior and high performance. Furthermore, the sensitivity analysis has been applied to this model which shows better accuracy of proposed model in comparison with other approaches. As a future work, some additional items can be added to the proposed model i.e., weather condition, available ancillary services, power exchanges, and maintenance outages of generators and transmission lines but, the over-fitting problem can be occurred as well.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 51867019), Science and technology project of Jiangxi Provincial Department of Education, China (Grant No. GJJ180961), China Postdoctoral Fund Project (Grant No. 2019M652278), Jiangxi Postdoctoral Science and Technology project, China (Grant No. 2018KY13).

References

Aggarwal, Sanjeev Kumar, Saini, Lalit Mohan, Kumar, Ashwani, 2009. Electricity price forecasting in deregulated markets: A review and evaluation. Int. J. Electr. Power Energy Syst. 31 (1), 13–22.

Altan, A., Karasu, Seçkin, Bekiros, Stelios, 2019. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos Solitons Fractals 126, 325–336.
Amir, M., Hamidreza, Z., William, D.R., 2012. Electricity price and demand forecasting in smart grids. IEEE Trans. Smart Grid 3 (2), 664–674.
Amjady, N., 2007. Short-term bus load forecasting of power systems by a new hybrid method. IEEE Trans. Power Syst. 22 (1), 333–341.
Amjady, N., Daraeepour, A., Keynia, F., 2010. Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. IET Gener. Transm. Distrib. 4 (3), 432–444.
Duan, Min, Darvishan, Ayda, Mohammaditab, Rasoul, Wakil, Karzan, Abedinia, Oveis, 2018. A novel hybrid prediction model for aggregated loads of buildings by considering the electric vehicles. Sustainable Cities Soc. 41, 205–219.
Elattar, Ehab Elsayed, 2013. Day-ahead price forecasting of electricity markets based on local informative vector machine. IET Gener. Transm. Distrib. 7 (10), 1063–1071.
Espinoza, M., Suykens, J., Belmans, R., Moor, B.D., 2007. Electric load forecasting. IEEE Control Syst. Mag. 27 (5), 43–57.
Giladi, C., Sintov, A., 2020. Manifold learning for efficient gravitational search algorithm. Inform. Sci. 517, 18–36.
Grossi, L., Nan, F., 2019. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. Technol. Forecast. Soc. Change 141, 305–318.
Hafeez, G., Saleem Alimgeer, K., Khan, I., 2020. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. Appl. Energy 269, 114915.
Han, X., Han, L., Gooi, H., Pan, Z., 2012. Ultra-short-term multi-node load forecasting – a composite approach. IET Gener. Transm. Distrib. 6 (5), 436–444.
Lin, Jie, et al., 2020. A novel multitype-users welfare equilibrium based real-time pricing in smart grid. Future Gener. Comput. Syst. 108, 145–160.
Muniain, P., Ziel, F., 2020. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. Int. J. Forecast. 1–18. http://dx.doi.org/10.1016/j.ijforecast.2019.11.006, (in press).
Muthukrishnan, A., Charles Rajesh kumar, J., Vinod Kumar, D., Kanagaraj, M., 2019. Internet of image things-discrete wavelet transform and gabor wavelet transform based image enhancement resolution technique for IoT satellite applications. Cogn. Syst. Res. 57, 46–53.
Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. society for industrial and applied mathematics, philadelphia, pennsylvaniaa.
Noradin, Ghadimi, Akbarimajd, Adel, Shayeghi, Hossein, Abedinia, Oveis, 2018. Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting. Energy 161, 130–142.
NSW: Australian Energy Market Operator. Available: http://www.aemo.com.au/.
NYISO: NYISO Electricity Market Data. Available: http://www.nyiso.com/.
Oveis, Abedinia, Amjady, Nima, 2016. Short-term load forecast of electrical power system by radial basis function neural network and new stochastic search algorithm. Int. Trans. Electr. Energy Syst. 26 (7), 1511–1525.
Oveis, Abedinia, Amjady, Nima, Shafie-Khah, Miadreza, Catalão, Joao. P.S., 2015. Electricity price forecast using combinatorial neural network trained by a new stochastic search method. Energy Convers. Manage. 105, 642–654.
Oveis, Abedinia, Amjady, Nima, Zareipour, Hamidreza, 2016. A new feature selection technique for load and price forecast of electrical power systems. IEEE Trans. Power Syst. 32 (1), 62–74.
Pablo, A.E., Michel, T., Claudio, A.P., Jacek, M.Z., 2009. Normalized mutual information feature selection. IEEE Trans. Neural Netw. 20 (2), 189–201.
PJM: PJM Electricity Market Data. Available: http://www.pjm.com/.
Salehpour, Mohammad Javad, Moghaddas Tafreshi, S.M., 2019. The effect of price responsive loads uncertainty on the risk-constrained optimal operation of a smart micro-grid. Int. J. Electr. Power Energy Syst. 106, 546–560.
dos Santos, G.S., Justi Luvizotto, L.G., Mariani, V.C., dos Santos Coelho, L., 2012. Least squares support vector machines with tuning based on chaotic differential evolution approach applied to the identification of a thermal process. Expert Syst. Appl. 39 (5), 4805–4812.
Sharmin, S., Shoyaib, M., Ali, A.A., Hossain Khan, M.A., Chae, O., 2019. Simultaneous feature selection and discretization based on mutual information. Pattern Recognit. 91, 162–174.
Shayeghi, H., Ghasemi, A., Moradzadeh, M., Nooshyar, M., 2015. Simultaneous day-ahead forecasting of electricity price and load in smart grids. Energy Convers. Manage. 95, 371–384.
Wu, L., Shahidehpour, M., 2014. A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid. IET Gener. Transm. Distrib. 8 (12), 1937–1950.
Yang, D., Li, G., Cheng, G., 2007. On the efficiency of chaos optimization algorithms for global optimization. Chaos Solitons Fractals 34 (4), 1366–1375.
Yang, W., Wang, J., Niu, T., Du, P., 2020. A novel system for multi-step electricity price forecasting for electricity market management. Appl. Soft Comput. 88, 106029.
Zhao, J.H., Dong, Z.Y., Li, X., Wong, K.P., 2007. A framework for electricity price spike analysis with advanced data mining methods. IEEE Trans. Power Syst. 22 (1), 376–385.