International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

March 2015 | Issue 22 | Volume 7 | Number 1 | Pages 38-45

A NEW HYBRID ALGORITHM FOR SHORT TERM LOAD FORECASTING

H. Shayeghi A. Ghasemi

Technical Engineering Department, University of Mohaghegh Ardabili, Ardabil, Iran
hshayeghi@gmail.com, ghasemi.agm@gmail.com

Abstract- In restructuring the electric power industry, the load had an important role for market managers and participants when they develop strategies or make decisions to maximize their profit. Therefore, accurate short term load forecasting (STLF) becomes more and more vital for all market participants such as customer or producer in competitive electricity markets. In this paper, a new hybrid algorithm is proposed to forecast day-ahead load signals. This load-forecasting algorithm works based on two stage feature selection (TSFS) method, discrete wavelet transform (DWT), least squares support vector machine (LSSVM) optimized by a modified artificial bee colony (ABC) using chaotic local search (CLS) method namely chaotic artificial bee colony (CABC). The numerical simulation results show that the proposed hybrid algorithm improves the accuracy of electricity load forecasting in different ceases in comparison to previously-known classical and intelligent methods.

Keywords: STLF, LSSVM, DWT, Modified Artificial Bee Colony, Feature Selection.

I. INTRODUCTION

Recently, the electric power trade all around the world has undergone noteworthy deregulation processes. These new challenges have contributed to the development of a more reliable tool for energy-market participants to maximize their profits that depend on future load signals [1]. Therefore, STLF is a necessary for scheduling and planning in market. In the case of electric power corporations, electricity load forecasting is an effectual planning tool applied for the purpose of optimal bidding, planning and profit maximization, especially for power producers in deregulated or competitive markets [2].

Load forecasting is a key issue in electricity markets, and many related studies have been reported in recent years [3-4]. The available forecasting methodologies can be generally classified into two groups: hard and soft computing methods. Hard computing methods include some well-known methods such as autoregressive (AR), autoregressive moving average (ARMA), semi-parametric and similar-day models [4-7]; The traditional statistical methods generally use linear models with limited or even no capability to characterize the non-linearity of the load patterns. In addition, the stationary process considered for most of these studies cannot capture non-stationary features of the load time-series.

Soft computing methods use data-driven models, where input–output mapping is learned from historical examples. Many methods belonging of this category can be found in the literature such as support vector regression [8], wavelet transform (WT)-artificial neural network (ANN)- adaptive neural fuzzy inference system (ANFIS) [9], WESN (wavelet echo state network) with a novel ESN-based reconstruction [10].

In [11] presented a solution for STLF in micro grids, based on a three-stage architecture which starts with pattern recognition by a self-organizing map (SOM), a clustering of the previous partition via k-means algorithm, and finally demand forecasting for each cluster with a multilayer perceptron. In [12] proposed a new hybrid economic indices based short-term load forecasting (HEI-STLF) system. In which business indicators, such as the leading index or the coincide index, each combined with stock index as hybrid economic indices influencing factors for support vector regression (SVR) model, to respond to the economic dynamics and reduce its impact on forecasting accuracy.

In [13] proposed semi-parametric models to estimate the relationships between demand and the driver variables. Specifically, the inputs for these models are calendar variables, lagged actual demand observations, and historical and forecast temperature traces for one or more sites in the target power system. In addition to point forecasts, prediction intervals are also estimated using a modified bootstrap method suitable for the complex seasonality seen in electricity demand data. The proposed methodology has been used to forecast the half-hourly electricity demand for up to seven days ahead for power systems in the Australian national electricity market.

In [14] proposed a multiple linear regression model that involves treating all the seasonal cycles as the input attributes. The result helps the managers to interpret the series structure with multiple seasonal cycles. To improve the forecasting accuracy, a support vector regression model based on optimal training subset (OTS) and adaptive particle swarm optimization (APSO) algorithm is established to forecast the residual series.

Thus, a novel hybrid model combining the proposed linear regression model and support vector regression model is built to achieve the above bi-objective short-term load forecasting. This paper presents a new hybrid algorithm for day-ahead load forecasting in electricity markets. The main contributions of this paper are summarized below:

i) The proposed forecasting consists of two stage feature selection method, which can filter out both irrelevant and redundant candidate inputs. The proposed two stage feature selection method can analytically evaluate the information value of the candidate inputs and selects a minimum set of the most informative features to feed the LSSVM based forecast engine.

ii) This paper presents the LSSVM as a forecast engine to capture the nonlinear pattern in load signal.

iii) To build a LSSVM model efficiently, model parameters must be accurately determined. Therefore, proposed CABC algorithm. In other words, to cope with the disadvantage of traditional methods to obtain global optimum solution, and to ensure the solutions are not trapped in local optima when system has a large number of variables to be optimized, we propose a new modified ABC based on CLS, named CABC to be able to optimize the parameters of the LSSVM.

iv) A DWT decomposes the load components into proper levels of resolution determined by an entropy-based criterion.

II. THE PROPOSED METHODOLOGY

A. DWT

DWT is a powerful tool for noise reduction without destroying the dynamics of the original load series. The DWT can be formulated as:

$$DWT(m,n)=\frac{1}{\sqrt{2}}\int_{-\infty}^{+\infty}x(t)\phi^{*}\left(\frac{t-2^{i}k}{2^{i}}\right)dt \qquad (1)$$

where, $m=2^i$ and $n=2^ik$, $x(t)$ is the original load series, ϕ the discrete wavelet, and ϕ^* the complex conjugate of discrete wavelet (ϕ). A fast method to execute DWT has been used in this paper, which contains two stages: decomposition and reconstruction. In the decomposition phase, the low-pass filter removes the higher frequency components of the signal and high-pass filter picks up the remaining parts. Then, the filtered signals are down-sampled by two and the results are called approximate coefficients and detail coefficients [15]. The reconstruction is just a reversed process of the decomposition and for perfect reconstruction filter banks. Figure 1 depicts corresponding wavelet decomposition.

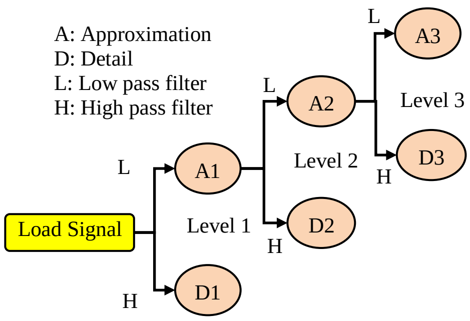

Figure 1. Discrete wavelet decomposition tree

B. LSSVM

Given a training set of N data points $\{y_k, x_k\}_{k=1}^{N}$, where $x_k \in R^n$ is the kth input pattern and $y_k \in R$ is the kth output pattern, the classifier can be constructed using the support vector method in the form:

$$y(x)=\operatorname{sign}\left[\sum_{k=1}^{N}\alpha_k y_k K(x,x_k)+b\right] \qquad (2)$$

where α_k are called support values and b is a constant. The $K(\cdot,\cdot)$ is the kernel, which can be either $K(x,x_k)=x_k^T x$ (linear SVM); $K(x,x_k)=(x_k^T x+1)^d$ (polynomial SVM of degree d); $K(x,x_k)=\tanh[\kappa x_k^T x+\theta]$ (multilayer perceptron SVM), or $K(x,x_k)=\exp\{-\|x-x_k\|_2^2/\sigma^2\}$ (RBF SVM), where $\kappa\theta$, and σ are constants. For instance, the problem of classifying two classes is defined as:

$$\begin{cases} w^T\phi(x_k)+b\geq +1 & \text{if } y_k=+1 \\ w^T\phi(x_k)+b\leq -1 & \text{if } y_k=-1 \end{cases} \qquad (3)$$

This can also be written as:

$$y_k[w^T\phi(x_k)+b]\geq 1, \quad k=1,\dots,N \qquad (4)$$

where $\phi(\cdot)$ is a nonlinear function mapping of the input space to a higher dimensional space. LS-SVM classifier:

$$\min_{w,b,e} J_{LS}(w,b,e)=\tfrac{1}{2}w^T w+\gamma\tfrac{1}{2}\sum_{k=1}^{N}e_k^2 \qquad (5)$$

subjects to the equality constraints:

$$y_k[w^T\phi(x_k)+b]=1-e_k, \quad k=1,\dots,N \qquad (6)$$

The Lagrangian is defined as:

$$L(w,b,e;\alpha)=J_{LS}-\sum_{k=1}^{N}\alpha_k\left\{y_k[w^T\phi(x_k)+b]-1+e_k\right\} \qquad (7)$$

with Lagrange multipliers $\alpha_k \in R$ (called support values). The conditions for optimality are given by:

$$\begin{cases} \dfrac{\partial L}{\partial w}=0 & \to & w=\sum_{k=1}^{N}\alpha_k y_k\phi(x_k) \\ \dfrac{\partial L}{\partial b}=0 & \to & \sum_{k=1}^{N}\alpha_k y_k=0 \\ \dfrac{\partial L}{\partial e_k}=0 & \to & \alpha_k=\gamma\, e_k \\ \dfrac{\partial L}{\partial \alpha_k}=0 & \to & y_k[w^T\phi(x_k)+b]-1+e_k=0 \end{cases} \qquad (8)$$

for $k=1,\dots,N$. After elimination of w and e one obtains the solution:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ 1_v \end{bmatrix} \qquad (9)$$

$$Z=[\phi(x_1)^T y_1;\dots;\phi(x_N)^T y_N],\ Y=[y_1;\dots;y_N],$$
$$1_v=[1;\dots;1],\ e=[e_1;\dots;e_N] \qquad (10)$$

and $\alpha=[\alpha_1;\dots;\alpha_N]$. Mercer's condition is applied to the matrix $\Omega=ZZ^T$ with:

$$\Omega_{kl}=y_k y_l\phi(x_k)^T\phi(x_l)=y_k y_l K(x_k,x_l) \qquad (11)$$

The kernel parameters, i.e. σ for RBF kernel, can be optimally chosen by optimizing an upper bound on the VC dimension [15]. The support values α_k are proportional to the errors at the data points in the LS-SVM case, while in the standard SVM case many support values are typically equal to zero.

C. TSFS

The two stage feature selection method can be summarized as follows:

- **First Stage:** Mutual information (MI) between input $CI(t) \in \{X(t),Y_1(t),Y_2(t),\ldots,Y_m(t)\}$ and target feature $x(t)$, that is $MI[CI(t),x(t)]$, is calculated [15]. Higher value of $MI[CI(t),x(t)]$ means $CI(t)$ is a more relevant input for perdiction $x(t)$. So, the feature inputs of $\{X(t),Y_1(t),Y_2(t),\ldots,Y_m(t)\}$ are sorted based on their MI with the output such that a higher $MI[CI(t),x(t)]$ value results in a higher rank. Then, the inputs with $MI[CI(t),x(t)]$ value greater than a relevancy threshold TH_1 are retained as the relevant features of the forecast process and the other candidate inputs are filtered out. The first stage of the proposed two stage feature selection technique can be considered as an irrelevancy filter.

- **Second Stage:** Suppose that $S_1 \subset \{X(t),Y_1(t),Y_2(t),\ldots, Y_m(t)\}$ is a subset of candidate inputs selected by the irrelevancy filter (first stage) by filtering out the irrelevant features. Higher value of MI between two selected candidates $CI_k(t) \in S_1$ and $CI_m(t) \in S_1$, i.e. $MI[CI_k(t),CI_m(t)]$, means more common information between the candidate inputs $CI_k(t)$ and $CI_m(t)$ and so these candidates have a higher level of redundancy. The following redundancy criterion $RC[.]$ measures the redundancy of each selected candidate $CI_k(t) \in S_1$ with the other candidate inputs of S_1:

$$RC\left[CI_k(t)\right] = \max_{CI_m(t)\in S_1}\left(MI\left[CI_k(t),CI_m(t)\right]\right) \tag{12}$$

We can rank the candidate inputs of S_1 according to the redundancy measure of (5) such that a higher value of $RC[CI_k(t)]$ means $CI_k(t)$ is a more redundant feature or equivalently a less informative candidate input. If $RC[CI_k(t)]$ becomes greater than a redundancy threshold $TH2$, $CI_k(t)$ is considered as a redundant candidate input and so between this candidate and its partner, one feature should be filtered out. For instance, suppose that

$$\arg\max_{CI_m(t)\in S_1}\left(MI\left[CI_k(t),CI_m(t)\right]\right) = CI_r(t)\ ,\ CI_r(t)\in S_1 \tag{13}$$

In other words, $CI_r(t)$ has the highest mutual information with $CI_k(t)$ among the features of S_1. Between $CI_k(t)$ and its partner $CI_r(t)$, one variable should be eliminated. For this purpose, the relevancy factors of these features, i.e. $MI[CI_k(t),x(t)]$ and $MI[CI_r(t),x(t)]$, are considered and the feature with less relevancy factor (less relevant feature or less effective feature for the forecast process) is filtered out. It is possible that more than two features be redundant such that only one of them is eligible to be retained. So, the redundancy filtering process is repeated for all features of S_1 until no redundancy measure of (12) becomes greater than $TH2$.

The redundancy filter, as described above, constitutes the second stage of the proposed two stage feature selection method. The subset of features $S_2 \subset S_1$ that pass the redundancy filter are finally selected candidate inputs by the proposed technique. The candidate features of S_2 are considered as the inputs of the forecast engine.

D. CABC

In this section, the standard ABC is briefly reviewed. Interested readers are referred to [16] for more details. The process of the ABC algorithm is presented as follows:

- Step 1. Initialization: Generate random population and calculate their fitness values. This population and fitness values called employed bees and nectar amounts, respectively.

- Step 2. Move the Onlookers: An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount by "Equation (14)", this method, known as roulette wheel selection method. The movement of the onlookers follows:

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_i} \tag{14}$$

where, p_i and SN are probability of selecting the ith employed bee and number of employed bees, and fit_i is the fitness value of the solution.

$$x_{ij}(t+1) = \theta_{ij} + \phi(\theta_{ij}(t) - \theta_{kj}(t)) \tag{15}$$

where, $k \in \{1,2,...,BN\}$ and $j \in \{1,2,...,D\}$ are randomly chosen indexes and x_i, t, θ_k and $\phi()$ are the position of the ith onlooker bee, the iteration number, the randomly chosen employed bee and random variable between (-1,1), respectively, and D is the number of dimension of optimization problem. BN is number of onlooker bee.

- Step 3. Move the Scouts: When selected a food source, all the employed bees associated with it abandon the food source, and become scout. The scouts are moved by:

$$\theta_{ij} = \theta_{ij_{\min}} + r(\theta_{ij_{\max}} - \theta_{ij_{\min}}) \tag{16}$$

where, r denotes a random factor and $\in$ [0, 1], $\theta_{ij\max}$ and $\theta_{ij\min}$ are lower and upper boundary of x_i, respectively.

- Step 4. Update the Best Food Source Found So Far: Memorize the best food source found so far.

- Step 5. Termination Checking: checking termination criteria satisfied, if it is satisfied then stop algorithm otherwise go to step 2.

To improve the exploration and exploitation abilities of the standard ABC, a novel operator called 'CLS', originating from astrophysics, is proposed in this paper. The disruption operator is inspired by nature and, with the least computation, can improve the ability of ABC to further explore and exploit the search space. In other words, we introduce a CABC approach that combines two ABC with CLS mechanisms. Also, we see that the instability of the specified motions is exactly what should make them useful CLS. Therefore, we define a new CLS based on pendulum law as follows:

$$c_{i+1}^{j}=\begin{cases}2c_i^j\times(1+\dfrac{g_{best}^{k-1}}{g_{best}^{k}})\times\cos(2\pi\dfrac{g_{best}^{k-1}}{g_{best}^{k}}), & 0.5<c_i^j\le 1\\ 0.1c_i^j\times(1-\cos((1+\dfrac{g_{best}^{k-1}}{g_{best}^{k}}))), & 0<c_i^j\le 0.5\end{cases} \quad (17)$$

where, g^{k}_{best} is best optimal value for k^{it} iteration and $g_{best}^{k-1}/g_{best}^{k}$ denotes the fine tuning necessary to achieve the desired sequence of gyrations. The CLS-based ABC algorithm can be summarized as follows:

i) Generate an initial chaos population randomly for CLS mechanism.

$$X_{cls}^{0}=[X_{cls,0}^{1},X_{cls,0}^{2},\ldots,X_{cls,0}^{Ng}]_{1\times N_g}$$
$$cx_0=[cx_0^1,cx_0^2,\ldots,cx_0^{Ng}] \quad (18)$$
$$cx_0^j=\frac{X_{cls,0}^{j}-X_{j,\min}}{X_{j,\max}-X_{j,\min}},\ j=1,2,\ldots,N_{\dim}$$

where, the chaos variable can be generated as follows:

$$x_{cls}^{i}=[x_{cls,i}^{1},x_{cls,i}^{2},\ldots,x_{cls,i}^{Ng}]_{1\times N_g},\ i=1,2,\ldots,N_{chaos}$$
$$x_{cls,i}^{j}=cx_{i-1}^{j}\times(X_{j,\max}-X_{j,\min})+P_{j,\min} \quad (19)$$
$$j=1,2,\ldots,N_{\dim}$$

ii) Determine the chaotic variables as follows:

$$cx_i=[cx_i^1,cx_i^2,\ldots,cx_i^{Ng}],\quad i=0,1,2,\ldots,N_{choas}$$
$$cx_{i+1}^{j}=\text{base}\ \ CLS\quad j=1,2,\ldots,N_{\dim} \quad (20)$$
$$cx_0^j=\text{rand}(0)$$

where, in Equations (18)-(20), N_{chaos} is the number of individuals for CLS, N_{dim} is the optimization parameters, cx_i^{Ng} the ith chaotic variable, rand(.) generate a random number between (0, 1), and X_{cls}^{i} is current position of bee based on chaos theory.

iii) Map the decision variables.

iv) Convert the chaotic variables to the decision variables.

v) Evaluate the new solution with decision variables.

III. THE PROPOSED FORECASTING

Structure of the proposed load forecasting, including a preprocessor and suggested hybrid forecast engine, is shown in Figure 2. Focus of this paper is on the new module of hybrid forecast engine. However, before of this module, the applied preprocessor should be first introduced to describe the performance of the proposed short term load forecasting strategy. The preprocessor receives the input data of the proposed strategy, divides the input data as detail and approximate section then normalizes the data to bring all inputs to the same range, refines the inputs by the feature selection process and feeds the hybrid forecast engine by its selected inputs. The input data of the proposed strategy, shown by $S(t)$ in Figure 2, is as follows:

$$S(t)=\ \{L(t-1),\ldots,L(t-N_L),EX_1(t),EX_1(t-1),\ldots,$$
$$,EX_1(t-N_1),\ldots,EX_P(t),EX_P(t-1),\ldots,EX_P(t-N_P)\} \quad (21)$$

where $L(t-1),\ldots,L(t-N_L)$ are the historical values of load, since electrical load is dependent on its past values.

The output of the proposed strategy is load forecast of the next time interval, denoted by $L(t)$ in Figure 2. The time interval depends on the load forecast step; for instance, for hourly load forecast, t is measured in terms of hour. Electrical load is also dependent on exogenous variables (such as temperature and humidity) in addition to its past values. These exogenous variables are shown by EX_1 to EX_P in (21). Since, the inputs of (21) have different ranges (such as load and temperature); we linearly normalize all inputs and output to be within the range [0,1] to avoid the masking effect. Linear normalization is a simple and well-known mathematical transformation. Suppose that an input x (such as load, temperature, humidity, etc.) is in the range of [$x_{\min}$, $x_{\max}$]. Linear normalization of x to be within the range of [0,1] is as follows:

$$x_n=(x-x_{\min})/(x_{\max}-x_{\min}) \quad (22)$$

where x_n shows normalized value of x in the range of [0,1]. Thus, each input is separately normalized based on its own minimum and maximum values. Also, the normalized variable x_n can easily be returned to the original range [$x_{\min}$, $x_{\max}$] by means of inverse transform:

$$x=x_{\min}+x_n.(x_{\max}-x_{\min}) \quad (23)$$

The output of the proposed hybrid forecast engine is in the normalized form, which is returned to the actual range by the inverse transform of (23).

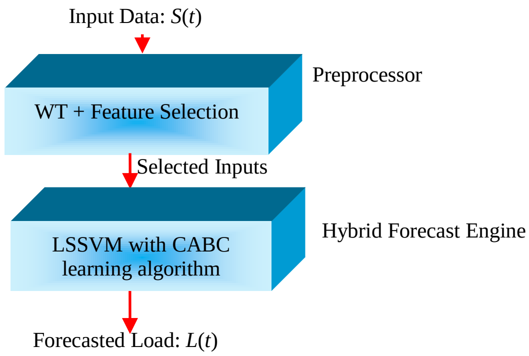

Figure 2. Structure of the proposed load forecasting strategy including the preprocessor and hybrid forecast engine

For each exogenous variable i, both its forecast value $EX_i(t)$ and past values $EX_i(t-1),\ldots,EX_i(t-N_i)$ (such as temperature forecast and past temperatures) are considered as the input data in (21). Choosing these exogenous variables is dependent on the engineering judgment and availability of data. For instance, while residential customers usually have high sensitivity to weather conditions (such as temperature), industrial loads are not so sensitive to weather parameters. In (21), N_L and N_1 to N_P indicate order of back shift for load L and P exogenous variables EX_1 to EX_P, respectively. From a data mining view point, these orders should be considered high enough so that no useful information is missed. After, the selected inputs by the preprocessor are given to the proposed hybrid forecast engine (Figure 2). The proposed forecast engine is a LSSVM trained by a new stochastic search technique, i.e. ABC algorithm. Finally, the performance of the proposed load forecasting can be summarized as the following step by step procedure:

- Step 1) Preprocessor normalizes input data and selects the most informative inputs for the load forecasting. Frequency domain features are obtained from DWT of electricity load time series. These features include $A3$ and $D1$, i.e. the lowest and highest frequency components of the DWT with three decomposition levels, respectively. Time domain candidate features contain electricity load load (L). Forecast and 200 lagged values of $A3$, $D1$ and L are considered in the data model as shown in Figure 2.

- Step 2) Using the selected inputs of the preprocessor, the hybrid forecast engine is trained by the proposed ABC. The decision variables $x_1,\ldots,x_{ND}$ of the final solution of the CABC (the best answer of the last iteration) are considered as the weights of the LSSVM based forecast engine.

- Step 3) After training the LSSVM and determining its weights, it is ready to forecast the future hourly loads. The LSSVM has kernel function in its output layer for predicting load of the next time interval, i.e. $L(t)$ (Figure 2). Multi-period STLF (e.g., prediction of load for the next 24 hours) is reached via recursion, i.e. by feeding input variables with the forecaster's outputs. For instance, predicted load for the first hour is used as $L(t-1)$ for the load forecast of the second hour provided that $L(t-1)$ is among the selected inputs for the forecast engine.

IV. SIMULATION AND DISCUSSION

A. New England

As in any research area, in the load forecast, it is important to allow the reproduction of one's results. The only way of doing that is using public domain data sets. The utility-specific segmentation of the electricity grid resulted in a system where fuel, resources and, most importantly, power were difficult to share among utilities. The systemic vulnerability of this design was made fully apparent during the Great Northeast Blackout of 1965. Since then, utilities have formed "power pools" to ensure regional system reliability. In the Northeast, the New England Power Pool (NEPOOL) was established in 1971 to facilitate collaboration among the utilities in Massachusetts, New Hampshire, Vermont, Rhode Island, Connecticut, and Maine.

After three decades of development and operation, NEPOOL finally produced an electricity grid with its own system operator and sufficient generations to ensure that the New England region of the United States never again experiences a full system failure. Since the NEPOOL market is well recognized in the U.S. and beyond, the proposed forecast strategy is tested using the data from the day-ahead energy market [17]. In modeling calendar terms, we make the assumption that day of the week, month of the year, and nonworking days have the effect of shifting electricity usage in a parallel fashion. Each parameter associated with the day-of-week and month dummy variables simply shifts the intercept of the model. Figure 3 plots daily load curves by each weekday. This graphic was created using hourly observations of regional demand from January 1, 2012 to Dec. 31, 2012.

Hour 1 on the horizontal axis of each graph is equivalent to 1:00 a.m., hour 3 to 2:00 a.m., and so no. Each plot in Figure 3 retains the same general shape, yet each is of a slightly different size. The sample results of the feature selection technique (correlation analysis) are presented. The candidate inputs for each subseries $A3$, $D3$, $D2$, and $D1$ include lagged values of that subseries load (L) up to 200 hours age, respectively (totally 800 candidates). For instance, the 800 candidate inputs for $A3$ subseries are $\{L_{(h-1)}, L_{(h-2)}, \ldots, L_{(h-200)}\}$.

The candidate inputs with correlation coefficient more are selected for the feature selection technique. $A3$ is the low frequency component of the original load signal and follows trend of the signal. Usually the most useful information of the original signal can be found in the $A3$ subseries. On the other hand, detail components contain high frequency components of the original signal such that $D1$ is the highest frequency component.

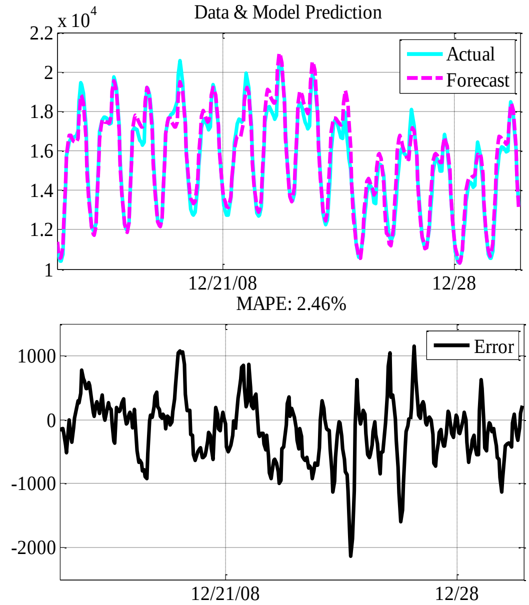

Figure 3. Hourly load forecasting in New England electricity market

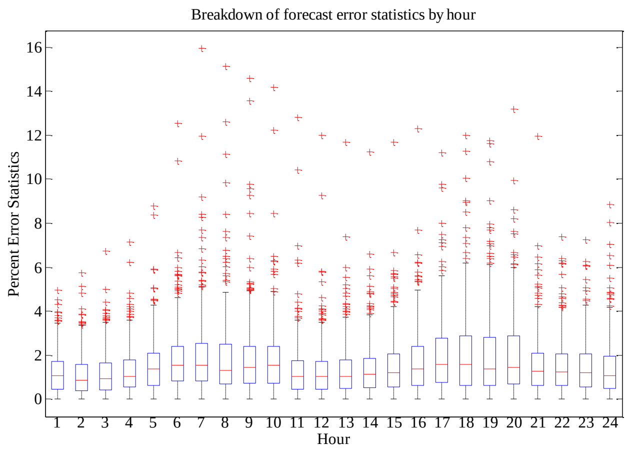

Figure 4. Boxplots of Hourly Error Percent for the New England electricity market

While the average hourly forecasting errors may indicate some differences in central tendency between models, looking at the distribution of forecasting errors by hour reveals extremely similar results for both approaches. A convenient way of investigating the distribution of forecast errors is with a box-and-whisker diagram (or just boxplot). According to Figure 4, boxplots of absolute percent error by hour illustrate that even the pattern of outliers for each hour is consistent across models for both within-sample and out-of sample time periods. To evaluate and compare the performance of the proposed technique than other techniques, a Mean Absolute Percentage Error (MAPE) index is proposed in this study.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_{iACT} - P_{iFOR}|}{P_{AVE-ACT}} \tag{24}$$

$$P_{AVE-ACT} = \frac{1}{N}\sum_{i=1}^{N} P_{iACT} \tag{25}$$

where, P_{iACT} and P_{iFOR} are actual and predict value and $P_{AVE\text{-}ACT}$ denotes the average of actual load as given in (23). The average of the MAPE index of the proposed algorithm is 1.58% that is significantly acceptable for a forecasting method. Figure 5 shows the distribution forecasting error in load forecasting of New England electricity market.

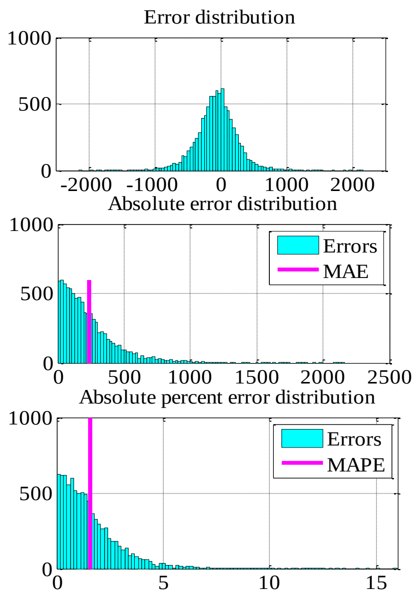

Figure 5. Distribution of error forecasting in New England electricity market

B. Iranian Electricity Market

In this case study, the proposed Hybrid forecasting algorithm shown in Figure 2 is used to forecast the day-ahead load in Iranian Electricity Market. The effectiveness of the proposed algorithm is demonstrated by forecasting load for Iran's market over a month of 2012 (in Farvardin calendar 1391). The data of Iran's electricity market can be found in [39]. The simulation mechanism is similar to New England market. Also, load data of the Iran electricity market in year 2012 is shown in Figure 6.

For the sake of a fair comparison, Multi-Layer Perceptron (MLP) neural network, WT-MI+NN and the proposed hybrid algorithm is constructed for load forecasting. In order to provide a graphical outlook on the forecast accuracy of the proposed hybrid algorithm, its results for the Iran electricity market and the fall's test week are exposed in Figures 7 and 8, respectively. Figure 9 shows error of the proposed hybrid algorithm, MLP and WT-MI+NN methods for the 24 hour and 168 hour. The circle axis denotes the plotting hours while the vertical axis refers to the error percentage. Among all the compared techniques, the performance of NN is sometimes far from satisfactory.

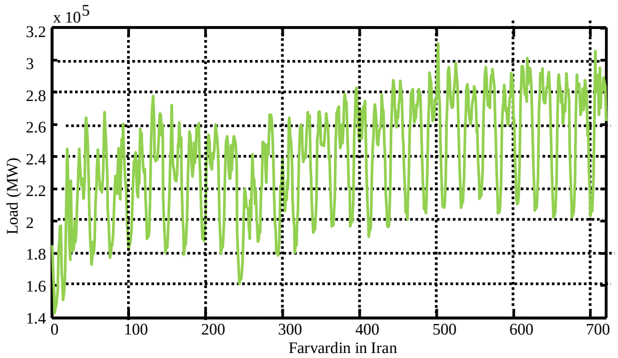

Figure 6. Load data of Iran electricity market in year 2012

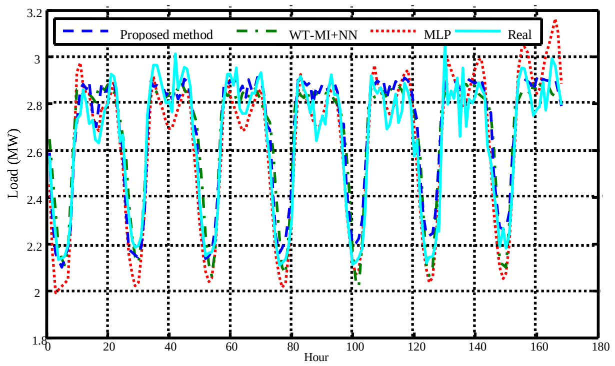

Figure 7. Spring week 2012 results for the Iran market: The Cyan (solid) show the actual load signal, while, Blue (dash), Greed-dark (dash-dot) and red (dot) lines represent the forecasted load by proposed method, WT-MI+NN and MLP, respectively

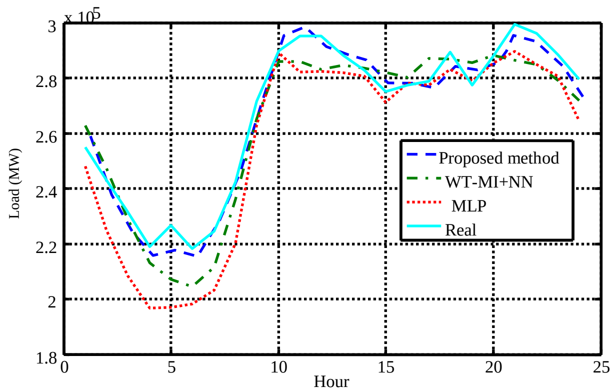

Figure 8. Spring hourly 2012 results for the Iran market: The Cyan (solid) show the actual load signal, while, Blue (dash), Greed-dark (dash-dot) and red (dot) lines represent the forecasted load by proposed method, WT-MI+NN and MLP, respectively

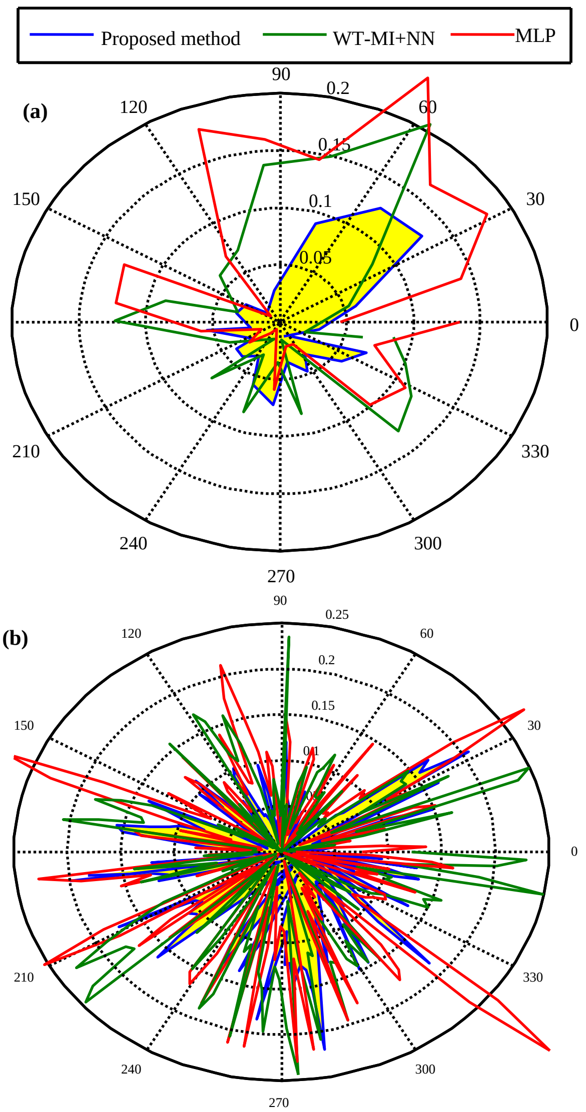

Figure 9. Error for the three selected methods, (a) daily forecast, (b) weekly forecast

Overall set-up time of the proposed technique including the DWT-TSFS, training of the forecaster (LSSVM+CABC) and fine-tuning of the adaptable parameters takes about 28 minutes on a Pentium P4 3.2 GHz PC with 1 GB RAM which is acceptable within a day-ahead decision making framework. Table 1 compares the proposed method, MLP and WT-MI+NN techniques for forecasting system load signals using the MAPE, FMSE and ESD criteria. As the proposed forecasting algorithm shows lower MAPE (FMSE and ESD), its predictions is more stable. During the valley hours, the volatility of the spot load increases, and, thus, MLP shows considerably larger prediction errors, but the load volatility during specific off-peak hours is also less accurate.

Table 1. Statistical analysis of forecasting error

Method	MAPE	FMSE (EUR/MWh)	ESD (EUR/MWh)
Proposed Method	3.35	0.1099	0.0874
NN	3.62	0.1160	0.1126
WT-MI+NN	3.53	0.1141	0.119

Moreover, the FMSE is the square root of the average of 24 (daily) squared differences between the actual load and the forecasted ones;

$$FMSE_{day} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{iACT} - P_{iFOR}\right)^2} \tag{31}$$

The ESD, one of the important performance criteria, is given by:

$$\begin{cases} ESD_{day} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_i - E_i^{Ave}\right)^2} \\ E_i = P_{iACT} - P_{iFOR} \\ E_i^{Ave} = \frac{1}{N}\sum_{i=1}^{N} E_i \end{cases} \tag{32}$$

V. CONCLUSIONS

Short-term load forecast (STLF) is an important operational function in both regulated power systems and deregulated open electricity markets. However, STLF is not easy to handle due to nonlinear and random-like behaviors of system loads, weather conditions, and variations of social and economic environments. Despite the performed research works in the area, more accurate and robust STLF methods are still demanded due to STLF importance and complexity. To improve the accuracy of load forecasting a novel algorithm is proposed in this paper for load forecasting problem. The proposed algorithm consists of three units where the first unit, the preprocessing unit, is responsible for detection of change of seasons, for tagging days, and for searching for similar days.

The second unit is a LSSVM-based hourly predictor, while the third unit optimizes the LSSVM parameters based on the CABC algorithm. The proposed forecasting algorithm is examined using data from the New England and Iranian electricity markets, which is evaluated as two successful electricity markets. The results show that the proposed algorithm has high accurate and superior forecast ability. The TSFS only considers two-way interactions. Extension of this algorithm to also include higher order interactions will be considered in the future work.

REFERENCES

[1] Y. Aslan, S. Yavasca, C. Yasar, "Long Term Electric Peak Load Forecasting of KUTAHYA Using Different Approaches", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 7, Vol. 3, No. 2, pp. 87-91, June 2011.
[2] M. Ghayekhloo, M.B. Menhaj, M. Ghofrani, "A Hybrid Short-Term Load Forecasting with a New Data Preprocessing Framework", Electric Power Systems Research, Vol. 119, pp. 138-148, 2015.
[3] A. Selakov, D. Cvijetinovic, L. Milovic, S. Mellon, D. Bekut, "Hybrid PSO-SVM Method for Short-Term load Forecasting during Periods with Significant Temperature Variations in City of Burbank", Applied Soft Computing, Vol. 16, pp. 80-88, 2014.

[4] D.X. Niu, H.F. Shi, D.D. Wu, "Short-Term Load Forecasting Using Bayesian Neural Networks, Learned by Hybrid Monte Carlo Algorithm", Applied Soft Computing, Vol. 12, No. 6, pp. 1822-1827, 2012.
[5] Y. Gao, M. Er, "NARMAX Time Series Model Prediction: Feedforward and Recurrent Fuzzy Neural Network Approaches", Fuzzy Sets and Systems, Vol. 150, pp. 331-350, 2005.
[6] M.L.M. Lopez, C.R. Minussi, A.D.P. Lutufo, "Electric Load Forecasting Using a fuzzy ART & ARTMAP Neural Network", Applied Soft Computing, Vol. 5, No. 2, pp. 235-244, 2005.
[7] R. Mamlook, O. Badran, E. Abdulhadi, "A Fuzzy Inference Model for Short-Term Load Forecasting", Energy Policy, Vol. 37, pp. 1239-1248, 2009.
[8] E. Ceperic, V. Ceperic, A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines", IEEE Trans. Power Syst., Vol. 28, No. 4, pp. 4356-4364, 2013.
[9] R. A. Hooshmand, H. Amooshahi, M. Parastegari, "A Hybrid Intelligent Algorithm Based Short-Term Load Forecasting Approach", Electrical Power and Energy Systems, Vol. 45, pp. 313-324, 2013.
[10] A. Deihimi, O. Rang, H. Showkati, "Short-Term Load and Temperature Forecasting Using Wavelet Echo State Networks with Neural Reconstruction", Energy, Vol. 57, pp. 382-401, 2013.
[11] L. Hernandez, C. Baladron, J.M. Aguiar, B. Carro, A. Sanchez Esguevillas, "Artificial Neural Networks for Short-Term Load Forecasting in Microgrids Environment", Energy, Vol. 75, pp. 252-264, 2014.
[12] C.T. Lin, L.D. Chou, Y.M. Chen, L.M. Tseng, "A Hybrid Economic Indices Based Short-Term Load Forecasting System", Electrical Power Energy Systems, Vol. 54, pp. 293-305, 2014.
[13] S. Fan, R.J. Hyndman, "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model", IEEE Trans. Power Syst., Vol. 27, No. 1, pp. 134-141, 2012.
[14] J.X. Che, "A Novel Hybrid Model for Bi-Objective Short-Term Electric Load Forecasting", Electrical Power Energy Systems, Vol. 61, pp. 259-266, 2014.
[15] H. Shayeghi, 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.
[16] H. Shayeghi, A. Ghasemi, "MOABC Algorithm for Economic/Environmental Load Dispatch Solution", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 13, Vol. 4, No. 4, pp. 82-88, December 2012.
[17] http://www.iso-ne.com/.
[18] "Operating Report of Electricity Market", Tehran, Iran, http://edipg.igmc.ir:8070/edipg.

BIOGRAPHIES

Hossein Shayeghi received the B.Sc. and M.Sc. degrees in Electrical and Control Engineering in 1996 and 1998, respectively. He received his Ph.D. degree in Electrical Engineering from Iran University of Science and Technology, Tehran, Iran in 2006. Currently, he is a full Professor in Technical Engineering Department of University of Mohaghegh Ardabili, Ardabil, Iran. His research interests are in the application of robust control, artificial intelligence and heuristic optimization methods to power system control design, operation and planning and power system restructuring. He has authored and co-authored of 7 books and two book chapters in Electrical Engineering area and more than 290 papers in international journals and conference proceedings. In addition, he collaborates with several international journals as reviewer boards and works as editorial committee of eight international journals. He has served on several other committees and panels in governmental, industrial, and technical conferences. He was selected as distinguished researcher of the University of Mohaghegh Ardabili several times. In 2007, 2010, 2011 and 2013 he was also elected as distinguished researcher in engineering field in Ardabil province of Iran. Also, he is a member of Iranian Association of Electrical and Electronic Engineers (IAEEE) and IEEE.

Ali Ghasemi received the B.Sc. and M.Sc. (Honors with first class) degree in Electrical Engineering from Isfahan University of Technology, Isfahan, and University of Mohaghegh Ardabili (UMA), Ardabil, Iran, in 2009 and 2011, respectively. Currently he is pursuing the Ph.D. degree in the Electrical Engineering and Computer Science of UMA. His research interests are application of forecast methods, operation adaptive and robust control of power systems, planning, power system restructuring and applications of heuristic techniques. He has authored of a book in Electrical Engineering area in Farsi, and more than 120 papers in reputable international journals and conference proceedings. Also, he collaborates as editorial committee and reviewer of 22 international journals. He is a member of the Iranian Association of Electrical and Electronic Engineers (IAEEE). In 2012 and 2013, he received the award of the 4th Electric Power Generation Conference and UMA for his M.Sc. thesis. He is the recipient honor M.Sc. and Ph.D. student award of UMA, 2014. He received the 2013 best young researcher award of the Young Researcher and Elite Club.