

A study of hybrid data selection method for a wavelet SVR mid-term load forecasting model

Hamid Reza Alirezaei¹ · Abolfazl Salami² · Mohammad Mohammadinodoushan³

Received: 9 September 2015 / Accepted: 14 August 2017 / Published online: 2 September 2017
© The Natural Computing Applications Forum 2017

Abstract Mid-term load forecasting (MTLF) is used to predict the loads for the durations from a week up to a year. Many methods have been used for selecting the best input data which is a critical issue in load forecasting. Recently, two separate approaches based on fuzzy logic system and support vector machine have shown better results compared to statistical techniques. The main purpose of this paper is to employ a novel hybrid approach based on wavelet support vector machines (WSVM) and chaos theory for MTLF. First, kernel-based fuzzy clustering technique and two-step correlation analysis are separately used for selecting training samples. Moreover, chaos theory is used to find the optimum time delay constant and embedding dimension of the load time series. Furthermore, genetic algorithm is employed to optimize the parameters of the WSVM model. EUNITE competition data and Iran power system data are selected to test the proposed method. The results show the efficiency of the suggested method compared with the other methods.

Keywords Mid-term load forecasting · Wavelet support vector machine · Kernel-based fuzzy clustering · Nonlinear time series · Chaos theory

1 Introduction

Due to the energy crisis, power industries consider load forecasting as one of the most important ways of managing the electric energy consumption. Financial aspects of load forecasting in deregulated power systems are important for all participants in electricity energy markets such as GENCOs, DISCOs and consumers [1, 2]. For instance, accurate load forecasting reduces the operation costs by enhancing economic load dispatch, unit commitment and ancillary services [3, 4]. There are four types of load forecasting including very short-term, short-term, mid-term and long-term load forecasting. Utilities use mid-term load forecasting more than the other types of load forecasting in dealing with fuel energy markets. In short, accurate load forecasting leads to an economic and reliable power system planning and operation [5, 6].

Load pattern depends on many factors such as seasonal effects, holidays, local temperature, humidity percentage and wind speed, which make load forecasting a very complicated procedure [7]; however, all these factors are not of the same significance. In fact, some of these irrelevant or redundant features should be discarded while maintaining the others [8]. Therefore, the first important step in load forecasting is feature selection, which is extracting the best input data from the variables. Another key step in forecasting is extracting a function that describes the relationship between the input data and the future load behavior precisely [7]. The other problem is forecasting the loads for the countries with two different

✉ Abolfazl Salami
salami@arakut.ac.ir

Hamid Reza Alirezaei
hamid.alirezaei@gmail.com

Mohammad Mohammadinodoushan
mm3845@nau.edu

¹ Department of Electrical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

² Department of Electrical Engineering, Arak University of Technology, Daneshgah St., Arak, Markazi 3818141167, Iran

³ School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

calendars [9]. In short, the mentioned problems make the load forecasting even more complex.

Generally, there are two basic approaches in the literature for dealing with forecasting problems. The first approach consists of statistical techniques such as time series [10], smooth transition periodic autoregressive [11] and nonlinear regression [12]. The second approach consists of Artificial Intelligence algorithms such as artificial neural network-based method [13–19], fuzzy logic systems [2, 6] and gradient boosting machines and Gaussian process [20], and support vector machine (SVM) [21]. Due to its ability to deal with four kinds of problems including classification, clustering, regression and function estimation, SVM has been one of the most popular methods recently. In addition, owing to its structure, SVM can deal with a large number of variables as the input vector. Moreover, the SVM has overcome trapping in local minima, which is the problem of the traditional algorithm, and the solution is always unique and globally optimal. SVM shows good results in dealing with under-fitting and over-fitting problems which are common problems in using other artificial neural networks [22]. The ability of SVM as a regression technique has been called support vector regression (SVR). This method has been one of the most interesting regression methods in recent years dealing with forecasting problems. Moreover, this method was applied by the winner of the European network on intelligent technologies competition [23]. The other SVM-based method for the next day load forecasting is least squares support vector machine (LS-SVM) [24].

Despite SVM ability and efficacy in forecasting problems, the SVM suffers from two drawbacks, which are determining the parameters of SVM, and selecting the proper input data. Abbas and Arif [25] used GA for determining parameters of SVM. The main drawback with this paper is the huge amount of irrelevant input data. In proposed method by Hong [26], the chaotic artificial bee colony has been used to find out parameters of SVM, and a recurrent neural network has been implemented to improve the performance of mapping nonlinear load pattern. Another hybrid approach which implements SVR in order to forecast short-term load has been introduced by Duan et al. [27]. The main advantage of this reference is reducing the number of training samples by using the fuzzy clustering approach. In this reference, a fuzzy C-means (FCM) has been used to find the optimal training samples from clustered historical load data. Then, the particle swarm optimization (PSO) has been used for training these data and optimizing parameters of SVR. Yang et al. [28] improved SVM accuracy by using rough set (RS) data preprocessing method which reforms input data. In [29], Wu employed Wavelet function as the SVR kernel function. In addition, PSO algorithm with Gaussian and

adaptive mutation has been implemented to optimize the parameters of proposed Wavelet SVR model.

All aforementioned SVR-based methods have been used an evolutionary algorithm to optimize the parameters of SVR model. Among these algorithms, the GA seems more efficient than the other algorithms [30]. However, only a few references have used a powerful method to deal with nonlinearity of input data in preparing and clustering step.

Recently, kernel-based methods which can detect dependencies kind have been used in the literature. These methods can by that dominate the load properties by formulating the feature space in terms of kernels. Formulation of feature space by kernels is the advantage of kernel machines as opposed to the rest load forecasting methods mentioned earlier; it allows the modeler to control the forecasting process by selecting the kernel form and promotes model flexibility by ordering a high variety of kernels [31, 32]. Furthermore, kernel fuzzy C-mean (KFCM) clustering method based on fuzzy SVM provides good performance in classification problems with outliers or noises [33]. Moreover, two-step correlation analysis (CA–CA) method boosts the forecast accuracy by pre-forecasting when the input data are limited [34]. In order to reduce the complexity between the historical data and other related variables, KFCM is used in the present paper for data preprocessing and finding the most effective candidate as the SVR model inputs. Considering the limitation of input data in MTLF, the CA–CA method is used as another approach to data preprocessing.

The rest of the paper is organized as follows. In Sect. 2, the basic theory of KFCM, CA–CA, time-series reconstruction, SVR and WSVR methods are discussed. Section 3 explains the proposed method. Experimental results are discussed in Sects. 4, and 5 contains the conclusion.

2 Methodology

In this paper, two methods are used for data preprocessing. The first method, which has not been used before, is the kernel clustering approach, which conducts clustering in the high-dimensional feature space. The second method, which is used for data preprocessing, is two-step correlation analysis. In the next step, the time-series reconstruction technique is applied to select input data. Finally, prepared data are used as training samples of the wavelet support vector regression (WSVR) model.

2.1 Kernel fuzzy C-mean (KFCM) clustering

Clustering approaches allow us to classify a large amount of data into two (or more) classes based on the similarity

between each pair of data. However, in most cases, it is very difficult to find a hard boundary between classes. In using fuzzy clustering approach, each data point is assigned to all clusters with a fuzzy membership degree between 0 and 1. This type of clustering approaches which do clustering into a high-dimensional feature space via a specified kernel function is known as kernel clustering approaches. The most powerful fuzzy clustering method is the FCM algorithm [31]. Recently, FCM has been generalized using Mercer’s kernel theory like any other clustering algorithm. Recent studies on classification problems in noisy datasets show the superiority of kernel fuzzy clustering approach over conventional fuzzy clustering [31, 33]. The KFCM algorithm is proposed as follows [33]:

1. Choose cluster number C and stop criterion $\varepsilon \in (0, 1)$;
2. Select kernel function $K(x_i, x_j)$ and its parameters that satisfy Mercer’s condition;
3. Initialize random centroids $V_j, j = 1, 2, \dots, C$;
4. Compute membership degree of i th vector in the j th cluster $u_{ij}, i = 1, \dots, N$ and $j = 1, \dots, C$

$$u_{ij} = \frac{(1/d^2(x_i, v_j))^{1/m-1}}{\sum_{p=1}^C (1/d^2(x_i, v_p))^{1/m-1}} \tag{1}$$

where

$$d^2(x_i, v_p) = K(x_i, x_i) - 2K(x_i, v_p) + K(v_p, v_p) = 2 - 2K(x_i, v_p) \tag{2}$$

5. Compute the new kernel matrices $K(x_i, \hat{v}_p^{\text{new}})$ and $K(\hat{v}_p^{\text{new}}, \hat{v}_p^{\text{new}})$

$$K(x_i, v_p^{\text{new}}) = \varphi(x_i) \bullet \varphi(v_p^{\text{new}}) = \frac{\sum_{k=1}^l (u_{kp})^m K(x_k, x_i)}{\sum_{k=1}^l (u_{kp})^m} \tag{3}$$

$$K(v_p^{\text{new}}, v_p^{\text{new}}) = \varphi(v_p^{\text{new}}) \bullet \varphi(v_p^{\text{new}}) = \frac{\sum_{k=1}^l \sum_{n=1}^l (u_{kp})^m (u_{np})^m K(x_k, x_n)}{(\sum_{k=1}^l (u_{kp})^m)^2} \tag{4}$$

where

$$\varphi(v_p^{\text{new}}) = \frac{\sum_{k=1}^l (u_{kp})^m \varphi(x_k)}{\sum_{k=1}^l (u_{kp})^m} \tag{5}$$

6. Update membership degree u_{ij} to u_{ij}^{new} according to Eq. (1).
7. If $\max_{i,j} |u_{ij} - u_{ij}^{\text{new}}| < \varepsilon$ stop algorithm, otherwise go to step (5).

2.2 Two-step correlation analysis (CA–CA)

In the first step of CA–CA method, the correlation between the input candidate and the output is computed. When the correlation index is higher, the candidate is more effective. If the computed correlation index is lower than a predetermined value Core1, then this candidate is deleted; otherwise, it is kept for the next step, in which the cross-correlation index between each pair of data, kept in step 1, is computed. If the computed index is less than predetermined value Core2, the selected data are saved; otherwise, the data with the lowest correlation with regard to output are deleted, while the others are saved [17, 34].

2.3 Time-series reconstruction

The phase space reconstruction technique provided by Takens, Aeyels and Sauer is a powerful application to analyze a univariate time series [35]. Suppose that each sub-partition of fuzzy clustering approach and candidate data driven from two-step correlation analysis can be assumed as a multidimensional stochastic time series including N data point. Regarding Takens embedding theory for any feature vector denoted by $\{x_i\}, i = 1, 2, \dots, N$. The n th delay vector can be reconstructed as $X_n = \{x_n, x_{n-\tau}, \dots, x_{n-(d-1)\tau}\}$ where d demonstrates the embedding dimension and τ is the time constant [36]. In this paper, AMI method [36] is used to choose optimum τ . In addition, Cao’s method [37] is applied to determine an acceptable embedding dimension d . Thus, the reconstructed phase space matrix of corresponding time series can be denote as Eq. (6)

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_l & x_{l+\tau} & \cdots & x_{l+(d-1)\tau} \end{bmatrix}_{N \times d} \tag{6}$$

where $N = N_{\text{subset}} - (d - 1) \times \tau$ and N_{subset} is the data points total number in the n th subset. The next state of Eq. (6) is the corresponding target vector as Eq. (7)

$$Y = [x_{1+d\tau} \quad x_{2+d\tau} \quad \cdots \quad x_{l+d\tau}]_{l \times 1}^T \tag{7}$$

2.4 SVR

Through mapping the input data to high-dimensional feature space, SVM uses linear regression which is less complicated compared to using nonlinear regression in low-dimensional feature space R^N [38]. For a dataset $\{(x_i, y_i)\}, i = 1, \dots, N$ where $x_i \in R^N$ denotes the i th input

vector; our objective is to find a linear estimation function in this feature space as follows [21]:

$$\hat{f}(x, \beta) = \langle \varphi(x), w \rangle + b, \text{ with } \varphi: R^N \rightarrow R^{N'}, N' \geq N \tag{8}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, $\varphi(x)$ is the mapped vectors, w denotes the weight vector and b is the balance. Since φ is fixed, these parameters are computed from the data by minimizing the following risk function. Since φ is fixed, these parameters are computed from the data by minimizing the following risk function [21]:

$$R = \frac{1}{2} \|w\|^2 + CR_{\text{emp}} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N L(\hat{f}(x_i) - y_i) \tag{9}$$

and

$$L(\hat{f}(x_i) - y_i) = \begin{cases} 0, & |\hat{f}(x_i) - y_i| \leq \varepsilon \\ |\hat{f}(x_i) - y_i| - \varepsilon, & \text{otherwise} \end{cases} \tag{10}$$

where L is Vapnik’s ε -insensitive loss function which is shown in Fig. 1, $\frac{1}{2} \|w\|^2$ is the regularization term and C is the penalty factor which is the trade-off between regularization term and the loss function.

As illustrated in Fig. 1, by introducing ξ_i and ξ_i^* two positive slack variables that represent the error between actual values and margin support vectors, Eq. (9) can be written as [21]:

$$\Phi(w, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i + \xi_i^* \tag{11}$$

$$\text{St: } \begin{cases} y_i - (\langle \varphi(x), w \rangle + b) \leq \varepsilon + \xi_i^* \\ (\langle \varphi(x), w \rangle + b) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{12}$$

Using Lagrangian multipliers and applying the Karush–Kuhn–Tucker conditions, the following dual optimization problem is defined [21]:

$$\begin{aligned} \text{Min } & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (\varphi(x_i) \bullet \varphi(x_j)) \\ & + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \end{aligned} \tag{13}$$

Subject to constraints:

$$\begin{aligned} \sum_{i=1}^l (\alpha_i^* - \alpha_i) &= 0, \\ 0 \leq \alpha_i, \alpha_i^* &\leq C, \quad i = 1, \dots, N \end{aligned} \tag{14}$$

Solving the optimization problem in Eq. (13) leads to the values α_i, α_i^* , which are the parameters of SVM regression function [21]:

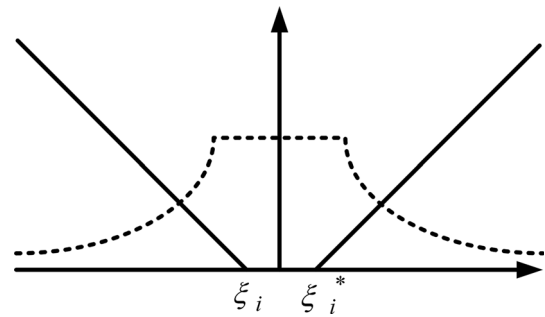


Fig. 1 Vapnik ε -insensitive loss function

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \langle \varphi(x_i), \varphi(x_j) \rangle + b \tag{15}$$

Using a different function known as kernel instead of the dot product of the point $\varphi(x_i)$ and $\varphi(x_j)$ makes the SVR algorithm nonlinear. If a kernel function $K(x_i, x_j)$ satisfies the Mercer’s conditions, then it will be called an admissible SV kernel. Choosing a good kernel function is challenging and depends on the problem stiffness and input vectors. In this paper, the wavelet kernel function introduced by Li et al. [39] is used as a kernel function in Eq. (15).

2.5 WSVR

As mentioned in Sect. 2.4, an inner dot-product kernel is the formation of an SV’s kernel, which must satisfy Mercer’s condition. However, it is difficult to decompose a translation invariant kernel (i.e. $K(x_i, x_j) = K(x_i - x_j)$) into the product of two functions and proof them as admissible SV kernels. According to [21], a translation invariant kernel is an admissible SV kernel if and only if the Fourier transform in Eq. (16) is nonnegative.

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} e^{-j(\omega x)} K(x) dx \tag{16}$$

More conditions for kernel functions are presented in [21]. According to wavelet theory, a function can be approximated by a family function produced by mother wavelet ($h(x)$), which must satisfy the following condition [30, 39]:

$$W_h = \int_0^\infty \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty \tag{17}$$

where $H(\omega)$ is the Fourier transform of $h(x)$. Any function that was proved in the condition (17) can be a dilation and translation function as:

$$h_{a,c}(x) = |a|^{-1/2} h\left(\frac{x-c}{a}\right). \tag{18}$$

According to Tensor theory, an N -dimensional wavelet function can be written as [40]:

$$h(X) = \prod_{i=1}^N h(x_i) \tag{19}$$

where every $h_i(x)$, $i = 1, \dots, N$ must separately satisfy the condition in Eq. (16). For any admissible mother wavelet function, the wavelet kernel that satisfies the translation invariant theorem can be presented as:

$$K(x, x') = \prod_{i=1}^N h\left(\frac{x_i - x'_i}{a}\right). \tag{20}$$

The Morlet’s wavelet function is used as the mother wavelet in this paper as showed in Eq. (21) [30],

$$h(x) = \cos(1.75x) \exp(-x^2/2). \tag{21}$$

Considering Eq. (21), it has been proofed that the Morlet’s wavelet kernel function in Eq. (20) can be defined as (22) which is an acceptable SV kernel function [30]

$$K(x, x') = \prod_i \left(\cos\left(1.75 \times \frac{(x_i - x'_i)}{a}\right) \exp\left(-\frac{\|x_i - x'_i\|^2}{2a^2}\right) \right). \tag{22}$$

In order to measure the forecast error, mean absolute percentage error (MAPE) is used as an index for performance evaluation. This index is given by the following equation.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|\widehat{L}_F(i) - L_{Actual}(i)|}{L_{Actual}(i)} \tag{23}$$

where i is the number of days, $\widehat{L}_F(i)$ is the forecasted load of the i th day and $L_{Actual}(i)$ is the i th day actual load.

3 The proposed method

In the first step of the proposed method, four attributes are considered as feature vectors of each day including daily peak load, daily temperature, calendar attributes and holiday index. In the first step, the best input data are selected in the clustering stage by the KFCM and the CA–CA methods. Then, for each subset driven from the second step, the reconstructed matrices of corresponding time series in phase space are represented in Eqs. (6) and (7). Having considered the time-series modeling, the training dataset for a particular day is formatted as (24):

$$\left[\{x_i, x_{i+1}, \dots, x_{i+(d-1)\tau}\}, \text{temperature, type of day,} \right. \\ \left. \times \text{holiday index} \right]. \tag{24}$$

According to Eq. (24), each training sample consists of at most $d + 9$ feature vectors, which include the following data: time delay vector corresponding to the i th normalized

peak load in the length of d , normalized average daily temperature; seven binary digits denoting the day of week and one binary digit, 0 or 1, which represent the holiday index. Consequently, the training data are arranged in a matrix of size $N_{\text{subset}} \times (d + 9)$. According to Eq. (7), the training target vector Y is the normalized daily peak load that shifted τ days ahead by considering the last vector of the reconstructed load matrix. Solving the SVR optimization problem by GA in training stage and finding the optimal α_i 's and α_i^* 's, the regression (predictor) function will be formed as:

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) k(x_i, x_j) + b \tag{25}$$

where k is the wavelet kernel defined in Eq. (22). D number of historical data are used as the input data of Eq. (25) for forecasting the next day peak load. The flowchart corresponding to the proposed methods is shown in Fig. 2.

4 Experimental results

The first dataset is tested for obtaining the minimum possible error within EUNITE 2001 competition, and the second one is Iran’s power system during 2002–2005. The main objective of this section is to analyze the data and find a relationship between consumption loads and other available information.

4.1 Data analysis

In 2001, EUNITE held a competition for load forecasting of January 1999. The input data were consumption load data (Jan. 1997–Dec. 1998), average daily temperature (Jan. 1995–Dec. 1998), calendar information and holiday indices (Jan. 1997–Dec. 1998) [41]. These data are discussed widely by Bo-Juen et al. [23].

Iran’s power system data are as follows. The hourly power consumption data from 2002 to 2005 [42], the calendar information and the holiday indices [43]. In this paper, one of system specification, i.e., Iran meteorological data are not considered due to Iran geographical breadth and the presence of all three cold, tropical and temperate climates; as the south region of Iran has tropical climate, while the northwest region is rather cold. Therefore, the long-term meteorological forecasting is a highly complicated task.

As Fig. 3 shows, the power consumption increases significantly within a 3-year period, indicating that the country is in development state. Amjady and Keynia [34] recommendation is using a 2-year historical data for short-term prediction of Iran’s power system, considering its

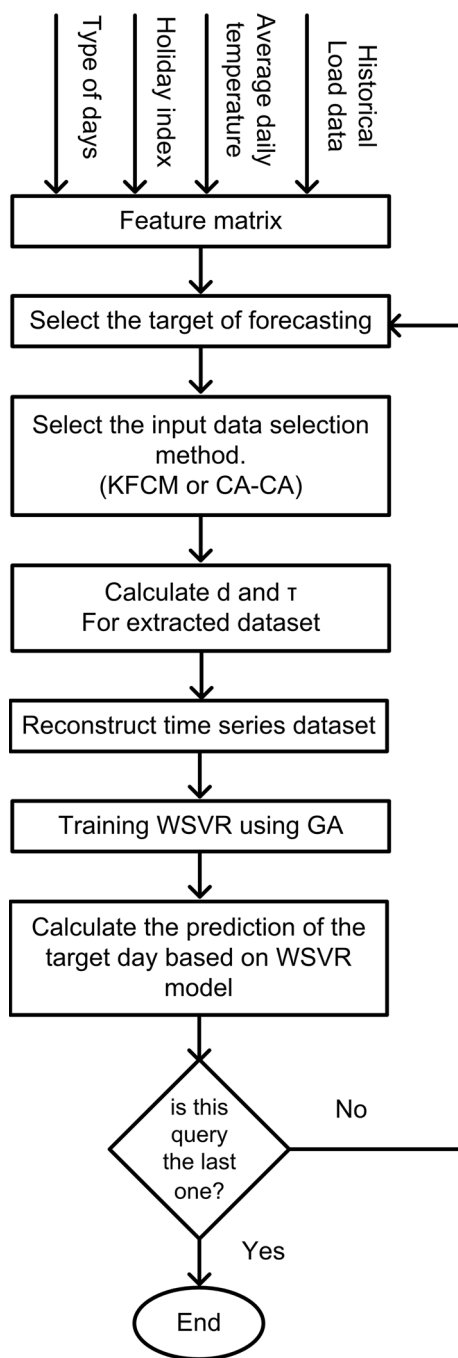


Fig. 2 Flowchart of proposed MTLF

rapidly changing consumption pattern. Based on Fig. 4, it can be noticed that although average power consumption rises significantly, it follows a similar pattern for working days during the period March 21, 2002–March 20, 2005. However, this pattern is different for weekends inasmuch as the maximum power consumption decreases during Thursdays, Fridays and holidays.

In Iran, most of the religious holidays are based on the lunar calendar. However, the official calendar in Iran is the

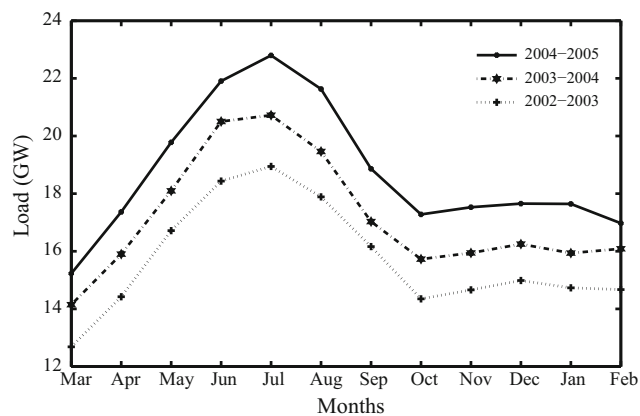


Fig. 3 Monthly averages of the load—Iran's power system (March 21, 2002–March 20, 2005)

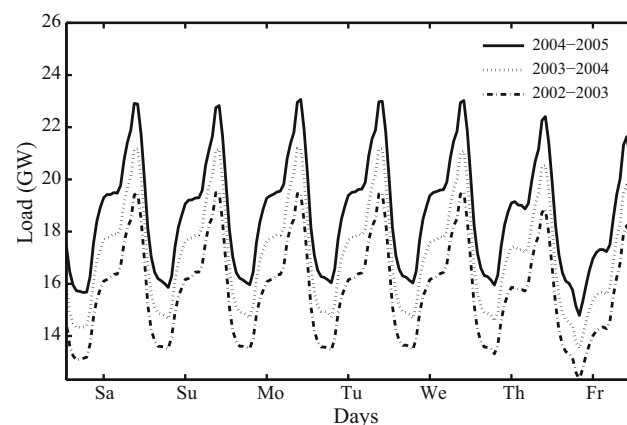


Fig. 4 Hourly load averages for each day of the week—Iran's power system (March 21, 2002–March 20, 2005)

solar calendar. Given the fact that a lunar year is 11 days shorter than a solar year, holidays are dissimilar each year in Iran. In fact, the calendar information and holiday indices are the most important factors in predicting the electrical power of Iran's power system [34]. Therefore, it is required to adjust the solar calendar for most holidays except Fridays, which is Iran's weekend. Such complications make modeling Iran power consumption a difficult task.

4.2 Numerical results

The proposed MTLF methods are programmed in MATLAB and are tested on two datasets. The test results are discussed in this section.

4.2.1 EUNITE competition data

In this section, the kernel-based fuzzy clustering approach, introduced in Sect. 2.1, is employed to divide the EUNITE

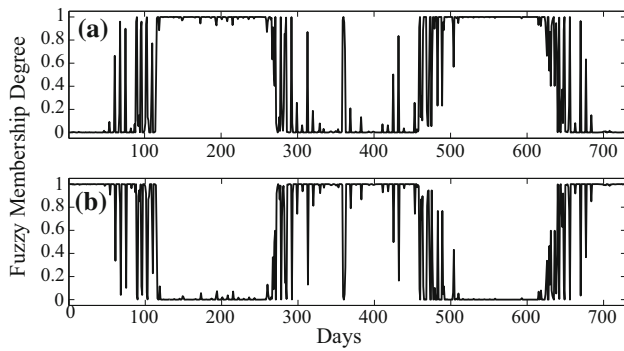


Fig. 5 Fuzzy membership degree for **a** subset 1, **b** subset 2—EUNITE data (1997–1998)

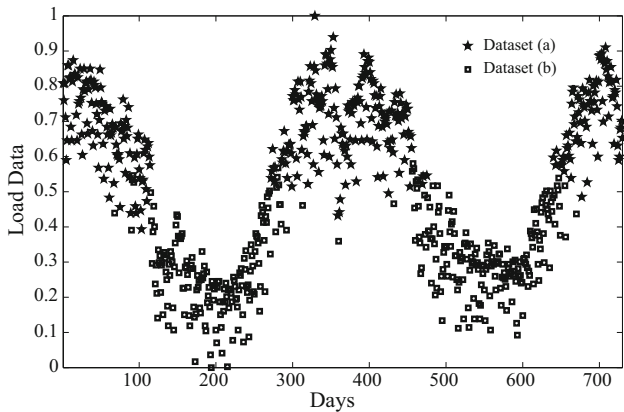


Fig. 6 Kernel-based fuzzy C-mean clustering results—EUNITE data (1997–1998)

dataset into two subsets. The obtained results for each day fuzzy membership degree are shown in Fig. 5. Owing to the negative correlation between electrical power data and average daily temperature, according to Fig. 6, the load dataset is divided into two subsets: data of cold months of the year (October–March) and data for spring and summer (April–September).

By using the AMI method, the time delay value is separately computed for the maximum daily power load time series and average daily temperature time series. Subsequently, the embedding dimension values are computed using Cao’s method. The computed results are

Table 1 Embedding dimension and the time delay constant for various time series

Case study	Datasets (time series)	Time delay (t)	Embedding dimension (d)
EUNITE	Maximum daily load subset (a)	$\tau = 1$	$d = 7$
	Average daily temperature subset (a)	$\tau = 1$	$d = 5$
	Maximum daily load subset (b)	$\tau = 1$	$d = 7$
	Average daily temperature subset (b)	$\tau = 1$	$d = 5$
Iran’ power system	Maximum daily load subset (a)	$\tau = 1$	$d = 6$
	Maximum daily load subset (b)	$\tau = 1$	$d = 6$

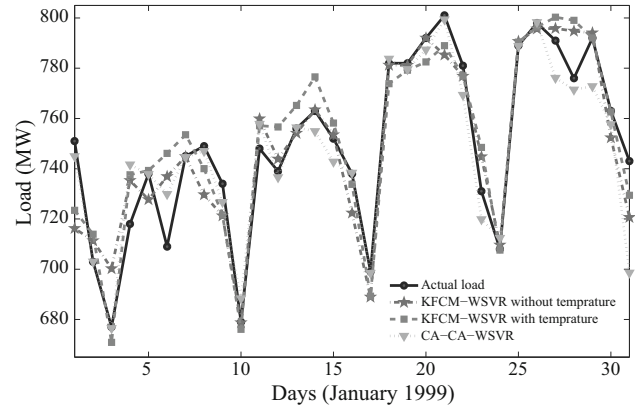


Fig. 7 Forecasts and actual load—EUNITE data (January 1999)

presented in Table 1. The results of the proposed methods are explained below:

1. Forecasting using KFCM clustering and WSVR (KFCM-WSVR)

In this section, clustering and retrieving the desired time series is performed by using the time-series modeling with and without considering the average daily temperature data as a feature vector. Due to the lack of access to temperature values of January 1999, the average daily temperature data of two past Januaries extracted from historical data were used.

Using the proposed method, which is described in Sect. 3, the MAPE value is 1.34 and 1.31% for the dataset with and without temperature, respectively. As illustrated in Fig. 7, once the temperature inputs are ignored, the obtained results are relatively more accurate. The modeling in Ref. [23], which was the winner of EUNITE competition with the error value of 1.95%, also proves that even using real temperature values does not improve the prediction results.

2. Forecasting by using the two-step correlation analysis and WSVR (CA-CA-WSVR)

In this method, the power data of January 1998 were selected as the target vector. Next, two analytical

Table 2 Comparison of the proposed models and other models—EUNITE network data

Method	MAPE (%)
SVM	1.95
SVM + GA	1.93
CA–CA + LM + EM	1.60
LSVM	1.51
KFCM–WSVR with temperature	1.34
LWSVM	1.34
KFCM–WSVR without temperature	1.31
LS-SVM + chaos theory	1.10
CA–CA-WSVR	1.04

correlative steps were applied in order to find the data of maximum linear correlation with target vector.

A comparison between the optimum results obtained from two proposed methods and January 1999 actual load is illustrated in Fig. 7. As shown in Table 2, the results obtained using CA–CA-WSVR indicate lower MAPE as compared to the results of KFCM–WSVR; however, the maximum error value obtained from this method reveals a rise as compared to previous method. This difference in error values can be attributed to several reasons. First, appropriate input vectors are not chosen properly due to the nonlinear nature of the load. Secondly, applying the past year data as a sample of target vector which generates error in the next steps of modeling process. Thirdly, choosing the insufficient embedding dimension, which might increase error value owing to the computations uncertainty through

value of 1.60%. El-Attar et al. [44] reached an error value of 1.52% by using a local SVM model (LSVM). They presented a weighted model for their previous method (LWSVM), and improved error value to 1.34% [35], which is almost equal to this research MAPE for KFCM–WSVR with temperature data. Having used least square SVM (LS-SVM) and chaos theory, Haishan and Xiaoling [45] reached to MAPE value of 1.1% which is in the range of the error values generated using KFCM–WSVR and CA–CA-WSVR in this paper.

4.2.2 Iran's power system

The historical data of electrical power network within March 21, 2002 to March 20, 2005 are used to predict maximum daily power consumption of March 21, 2004 to April 20, 2004.

In the KFCM–WSVR method, the optimum embedding dimension is extracted for the developed time series. The results of implementing of Cao's method are shown in Fig. 8. We calculate the $E1(d)$ and the $E2(d)$ using 369 data driven from KFCM method where we let the time delay τ equal to 1. Due to the random nature of load data, the future load values are independent of the past load values, and thus $E2(d)$ will be equal to 1 for any given d in this case; where d is embedding dimension. As shown in Fig. 8, $E1(d)$ stops changing when d is greater than d_0 . Therefore, $d_0 + 1$ is the minimum embedding dimension we look for. Using the embedding dimension values which are described in Fig. 8, Eqs. 6 and 7, the training matrix is calculated as below.

$$\text{Training Matrix} = \begin{bmatrix} L_1 & L_2 & \cdots & L_6 & L_7 & T_8 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ L_2 & L_3 & \cdots & L_7 & L_8 & T_9 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ L_{728} & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ L_{722} & L_{723} & \cdots & L_{727} & L_{728} & T_{729} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ L_{723} & L_{724} & \cdots & L_{728} & L_{729} & T_{730} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}_{369 \times 16} \quad (26)$$

Cao's method. The simulation results are compared with some published methods based on SVM in Table 2. Bo-Juen et al. [23] reduced the MAPE to 1.95% by using their proposed SVM method and won the EUNITE competition in 2001. Abbas and Arif [25] applied genetic algorithm in optimizing and obtaining the parameters of SVM and improved the obtained results. Amjady and Keynia [34] applied two correlative analytical steps in order to select target features of the Evolutionary Algorithm and Levenberg–Marquardt Algorithm-based predictions. The proposed algorithm known as CA–CA–EA–LM has an error

$$\text{Target Matrix} = [L_8 \ L_9 \ \dots \ L_{729} \ L_{730}]_{1 \times 369}^T \quad (27)$$

Results of Table 1 are used to retrieve dynamic state of the studied system. Due to embedding dimension uncertainty, the values $5 < d < 8$ are tested. As show in Table 3, the MAPE value is equal to 2.00% for $d = 6$.

In CA–CA-WSVR method, the similarity of the target day specifications and the selected sample vector is the factor that must be considered. For a numerical example, in order to forecast the load of March 21, 2004, the similar day in 2003 is considered as a target vector. In the first step

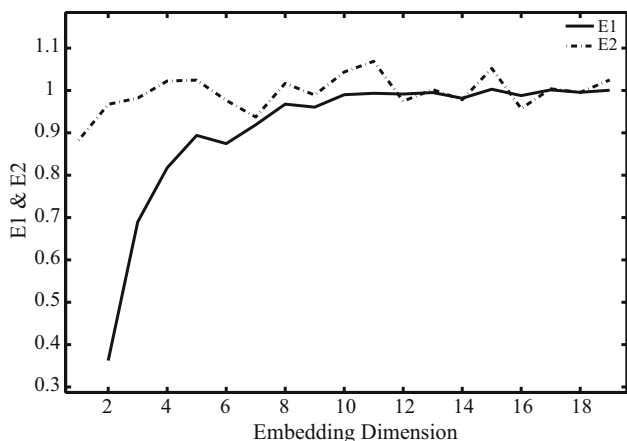


Fig. 8 Values *E1* and *E2* for the data from Iran’s power system

Table 3 Comparison of obtained result for different value of embedding dimension—Iran’s power system

Embedding dimension	MAPE (%)	MAX (GW)
<i>d</i> = 5	2.28	2.30
<i>d</i> = 6	2.00	2.45
<i>d</i> = 7	2.11	2.04
<i>d</i> = 8	2.10	1.84

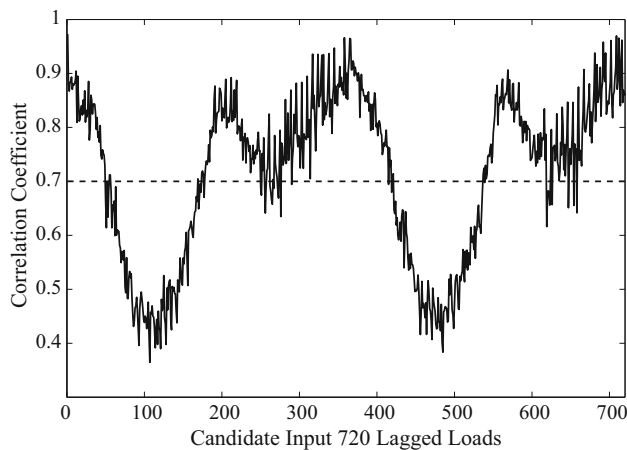


Fig. 9 Results of the first step of the correlation analysis for March 21, 2004—Iran’s power system

of CA–CA method, the last two years daily load data are used as candidate inputs. The correlation between the target day and these data is calculated which are shown in Fig. 9. Then, the data with correlation index equal to or

Table 4 Results of the correlation analysis for the Iran’s power system (March 21, 2002)

Core1	Core2	Selected inputs
0.7	0.7	1, 32, 33, 39, 40, 47, 53, 54, 60, 61, 67, 82, 305, 359, 360, 369, 383, 390, 396, 397, 404, 411, 438, 691, 705, 712, 724, 725, 726

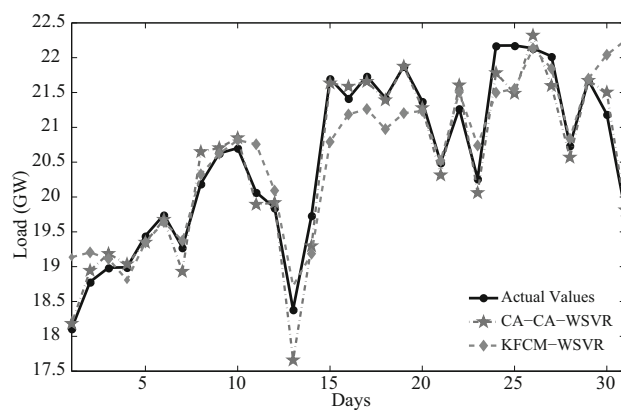


Fig. 10 Actual load (solid line), forecast with KFCM–WSVR (dash dot line with pentacle sign), forecast with CA–CA–WSVR (dash line with square sign)—Iran’s power system (March 21, 2004–April 20, 2004)

less than Core1, which is selected 0.7, will be deleted, while other data are kept for step 2. In step 2, the cross-correlation index between each pair of data, kept in step 1, is computed. Then, the data with correlation index equal to or less than Core2, which is selected 0.7, will be deleted while keeping other data. After the two-step correlation, the data shown in Table 4 are selected as input vectors.

In training phase for each SV, the same day/month in the last year must be deleted from the training set and considered as the unseen validation. Therefore, the number of validation data is equal to number of test data. In the training stage, the optimal SVM parameters are found using GA. The population size, crossover ratio, mutation percentage and mutation ratio in GA are 50, 0.8, 0.3 and 0.04, respectively. After 200 iterations, the penalty factor, error margin and scaling parameter of wavelet kernel calculated by GA are 1.4739, 0.01636 and 0.4051, respectively.

The results obtained from KFCM–WSVR and CA–CA–WSVR are presented in Fig. 10. Table 5 shows that our CA–CA–WSVR model has better forecast accuracy and stability than the proposed KFCM–WSVR data model, due to less value of average MAPE. According to Table 2 and Table 5, it is clear that the MAPE value of our proposed method showed 45% improvement comparing to all average MAPE values of Table 2. This enhanced accuracy is mainly related to the forecast block and especially its preprocessing mechanism. Ref. [45] which used chaos theory and time-series reconstruction as its preprocessing method with SVM has shown results with less accuracy comparing to the proposed model in this paper. Therefore,

Table 5 Obtained results with proposed forecast models—Iran's power system

Proposed method	MAPE (%)	MAX (GW)
KFCM-WSVR without temperature	2.00	2.423
CA-CA-WSVR	1.07	0.727

selecting efficient preprocessing is an important for forecasting. Moreover, Ref. [34] which used the same preprocessing method as our paper, but with Levenberg–Marquardt, has also shown results with larger MAPE value. This indicates that choosing an efficient preprocessing method is not enough.

Both proposed methods, KFCM and CA-CA, provide acceptable results except for only some days such as March 21. The error in this day is mainly due to selecting inappropriate candidate as input vectors for the model. In fact, considering the optimal embedding dimension, the load data of March 13–20 are used as input vectors in KFCM model in order to forecast the load of March 21. However, there is not enough similarity between input vectors, which are the last days of a year, and the target vector, which is the first day of a year. This is mostly because the commercial load consumption is high in the last days of the year. The errors of forecasting for the next days are higher in KFCM method due to the fact that it uses the last day's data for forecasting the next day, and thus there always exists an accumulated error for forecasting the next days. For example, the error of predicted data for March 21 is used for March 22. Nevertheless, the CA-CA model does not use embedding windows, and thus the error in it is much lower.

5 Conclusion

A new method is suggested in this work for MTLF, which is an important part of today power systems studies. Two hybrid SVR-based methods are proposed for reducing the complexity between the load data and other related variables and improve MTLF accuracy. A kernel-based fuzzy clustering technique and a two-step correlation analysis are separately used to extract the optimal subset from historical data. Simulation results on EUNITE competition data show that the suggested methods are more efficient in comparison with other related methods. Moreover, the suggested methods are evaluated on Iran's power system data. The advantage of the proposed methods is using Mercer's kernel theory, which allows us to deal with nonlinear data. The performance of the proposed KFCM-WSVR method shows its less sensitivity to weather data. Furthermore, the

simulations confirm that the CA-CA-WSVR model shows better results in comparison with KFCM-WSVR mode.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alamaniotis M, Ikononopoulos A, Tsoukalas LH (2012) Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting. *IEEE Trans Power Syst* 27(3):1477–1484. doi:[10.1109/tpwrs.2012.2184308](https://doi.org/10.1109/tpwrs.2012.2184308)
- Çevik H, Çunkaş M (2015) Short-term load forecasting using fuzzy logic and ANFIS. *Neural Comput Appl* 26(6):1355–1367. doi:[10.1007/s00521-014-1809-4](https://doi.org/10.1007/s00521-014-1809-4)
- Sugianto LF, Lu X-B (2002) Demand forecasting in the deregulated market: a bibliography survey. School of Business Systems, Victoria
- Fan S, Chen L, Lee WJ (2008) Short-term load forecasting using comprehensive combination based on multi-meteorological information. In: Industrial and commercial power systems technical conference, 2008. ICPS 2008. IEEE/IAS, 4–8 May 2008, pp 1–7. doi:[10.1109/icps.2008.4606288](https://doi.org/10.1109/icps.2008.4606288)
- Gonzalez-Romera E, Jaramillo-Moran MA, Carmona-Fernandez D (2006) Monthly electric energy demand forecasting based on trend extraction. *IEEE Trans Power Syst* 21(4):1946–1953. doi:[10.1109/tpwrs.2006.883666](https://doi.org/10.1109/tpwrs.2006.883666)
- Khosravi A, Nahavandi S, Creighton D, Srinivasan D (2012) Interval type-2 fuzzy logic systems for load forecasting: a comparative study. *IEEE Trans Power Syst* 27(3):1274–1282. doi:[10.1109/tpwrs.2011.2181981](https://doi.org/10.1109/tpwrs.2011.2181981)
- Feinberg E, Genethliou D (2005) Load forecasting. In: Chow J, Wu F, Momoh J (eds) Applied mathematics for restructured electric power systems. Power electronics and power systems. Springer, New York, pp 269–285
- Zhao M, Fu C, Ji L, Tang K, Zhou M (2011) Feature selection and parameter optimization for support vector machines: a new approach based on genetic algorithm with feature chromosomes. *Expert Syst Appl* 38(5):5197–5204. doi:[10.1016/j.eswa.2010.10.041](https://doi.org/10.1016/j.eswa.2010.10.041)
- Kwang-Ho K, Hyoung-sun Y, Yong-Cheol K (2000) Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method. *IEEE Trans Power Syst* 15(2):559–565. doi:[10.1109/59.867141](https://doi.org/10.1109/59.867141)
- Ding N, Besanger Y, Fdr Wurtz (2015) Next-day MV/LV substation load forecaster using time series method. *Electr Power Syst Res* 119:345–354. doi:[10.1016/j.epsr.2014.10.003](https://doi.org/10.1016/j.epsr.2014.10.003)
- Amaral LF, Souza RC, Stevenson M (2008) A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *Int J Forecast* 24(4):603–615. doi:[10.1016/j.ijforecast.2008.08.006](https://doi.org/10.1016/j.ijforecast.2008.08.006)
- Tsekouras GJ, Dialynas EN, Hatziargyriou ND, Kavatza S (2007) A non-linear multivariable regression model for midterm energy forecasting of power systems. *Electr Power Syst Res* 77(12):1560–1568. doi:[10.1016/j.epsr.2006.11.003](https://doi.org/10.1016/j.epsr.2006.11.003)
- Ghayekhloo M, Menhaj MB, Ghofrani M (2015) A hybrid short-term load forecasting with a new data preprocessing framework. *Electr Power Syst Res* 119:138–148. doi:[10.1016/j.epsr.2014.09.002](https://doi.org/10.1016/j.epsr.2014.09.002)
- López M, Valero S, Senabre C, Aparicio J, Gabaldon A (2012) Application of SOM neural networks to short-term load

- forecasting: the Spanish electricity market case study. *Electr Power Syst Res* 91:18–27. doi:[10.1016/j.epsr.2012.04.009](https://doi.org/10.1016/j.epsr.2012.04.009)
15. Dong J-r, Zheng C-y, Kan G-y, Zhao M, Wen J, Yu J (2015) Applying the ensemble artificial neural network-based hybrid data-driven model to daily total load forecasting. *Neural Comput Appl* 26(3):603–611. doi:[10.1007/s00521-014-1727-5](https://doi.org/10.1007/s00521-014-1727-5)
 16. Sheikhan M, Mohammadi N (2012) Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection. *Neural Comput Appl* 21(8):1961–1970. doi:[10.1007/s00521-011-0599-1](https://doi.org/10.1007/s00521-011-0599-1)
 17. Tsekouras GJ, Hatziazgyriou ND, Dialynas EN (2006) An optimized adaptive neural network for annual midterm energy forecasting. *IEEE Trans Power Syst* 21(1):385–391. doi:[10.1109/TPWRS.2005.860926](https://doi.org/10.1109/TPWRS.2005.860926)
 18. Li Y-y, Niu D-x (2005) An optimum credibility based integrated optimum gray neural network model of monthly power load forecasting. *Power Syst Technol* 5:16–19
 19. Chang P-C, Fan C-Y, Lin J-J (2011) Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach. *Int J Electr Power Energy Syst* 33(1):17–27. doi:[10.1016/j.ijepes.2010.08.008](https://doi.org/10.1016/j.ijepes.2010.08.008)
 20. Lloyd JR (2014) GEFCom2012 hierarchical load forecasting: gradient boosting machines and Gaussian processes. *Int J Forecast* 30(2):369–374. doi:[10.1016/j.ijforecast.2013.07.002](https://doi.org/10.1016/j.ijforecast.2013.07.002)
 21. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222
 22. Hippert HS, Bunn DW, Souza RC (2005) Large neural networks for electricity load forecasting: are they overfitted? *Int J Forecast* 21(3):425–434. doi:[10.1016/j.ijforecast.2004.12.004](https://doi.org/10.1016/j.ijforecast.2004.12.004)
 23. Bo-Juen C, Ming-Wei C, Chih-Jen L (2004) Load forecasting using support vector machines: a study on EUNITE competition 2001. *IEEE Trans Power Syst* 19(4):1821–1830. doi:[10.1109/tpwrs.2004.835679](https://doi.org/10.1109/tpwrs.2004.835679)
 24. Espinoza M, Suykens JK, Moor B (2006) Fixed-size least squares support vector machines: a large scale application in electrical load forecasting. *CMS* 3(2):113–129. doi:[10.1007/s10287-005-0003-7](https://doi.org/10.1007/s10287-005-0003-7)
 25. Abbas SR, Arif M (2006) Electric load forecasting using support vector machines optimized by genetic algorithm. In: Multitopic conference, 2006. INMIC '06. IEEE, 23–24 December 2006, pp 395–399. doi:[10.1109/inmic.2006.358199](https://doi.org/10.1109/inmic.2006.358199)
 26. Hong W-C (2011) Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* 36(9):5568–5578. doi:[10.1016/j.energy.2011.07.015](https://doi.org/10.1016/j.energy.2011.07.015)
 27. Duan P, Xie K, Guo T, Huang X (2011) Short-term load forecasting for electric power systems using the PSO-SVR and FCM clustering techniques. *Energies* 4(1):173–184
 28. S-x Yang, Cao Y, Liu D, C-f Huang (2011) RS-SVM forecasting model and power supply-demand forecast. *J Cent South Univ Technol* 18(6):2074–2079. doi:[10.1007/s11771-011-0945-6](https://doi.org/10.1007/s11771-011-0945-6)
 29. Wu Q (2010) Power load forecasts based on hybrid PSO with Gaussian and adaptive mutation and Wv-SVM. *Expert Syst Appl* 37(1):194–201. doi:[10.1016/j.eswa.2009.05.011](https://doi.org/10.1016/j.eswa.2009.05.011)
 30. Wu Q (2011) Hybrid model based on wavelet support vector machine and modified genetic algorithm penalizing Gaussian noises for power load forecasts. *Expert Syst Appl* 38(1):379–385
 31. Wu Z-d, Xie W-x, Yu J-p (2003) Fuzzy C-means clustering algorithm based on kernel method. In: Proceedings of fifth international conference on computational intelligence and multimedia applications, 2003. ICCIMA 2003, 27–30 September 2003, pp 49–54. doi:[10.1109/iccima.2003.1238099](https://doi.org/10.1109/iccima.2003.1238099)
 32. Alamaniotis M, Bargiotas D, Tsoukalas LH (2016) Towards smart energy systems: application of kernel machine regression for medium term electricity load forecasting. *SpringerPlus Eng* 5:1–15. doi:[10.1186/s40064-016-1665-z](https://doi.org/10.1186/s40064-016-1665-z)
 33. Xiaowei Y, Guangquan Z, Jie L, Jun M (2011) A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. *IEEE Trans Fuzzy Syst* 19(1):105–115. doi:[10.1109/TFUZZ.2010.2087382](https://doi.org/10.1109/TFUZZ.2010.2087382)
 34. Amjady N, Keynia F (2008) Mid-term load forecasting of power systems by a new prediction method. *Energy Convers Manag* 49(10):2678–2687. doi:[10.1016/j.enconman.2008.04.008](https://doi.org/10.1016/j.enconman.2008.04.008)
 35. Elattar EE, Goulermas J, Wu QH (2010) Electric load forecasting based on locally weighted support vector regression. *IEEE Trans Syst Man Cybern Part C Appl Rev* 40(4):438–447. doi:[10.1109/TSMCC.2010.2040176](https://doi.org/10.1109/TSMCC.2010.2040176)
 36. Lau KW, Wu QH (2008) Local prediction of non-linear time series using support vector regression. *Pattern Recogn* 41(5):1539–1547. doi:[10.1016/j.patcog.2007.08.013](https://doi.org/10.1016/j.patcog.2007.08.013)
 37. Cao L (1997) Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D* 110(1–2):43–50. doi:[10.1016/S0167-2789\(97\)00118-8](https://doi.org/10.1016/S0167-2789(97)00118-8)
 38. Müller K-R, Smola AJ, Rätsch G, Schölkopf B, Kohlmorgen J, Vapnik V (1997) Predicting time series with support vector machines. In: Artificial neural networks—ICANN'97. Springer, pp 999–1004
 39. Li Z, Weida Z, Licheng J (2004) Wavelet support vector machine. *IEEE Trans Syst Man Cybern B Cybern* 34(1):34–39. doi:[10.1109/TSMBCB.2003.811113](https://doi.org/10.1109/TSMBCB.2003.811113)
 40. Smola AJ (1996) Regression estimation with support vector learning machines. Master's thesis, Technische Universit at M unchen
 41. Eunite (2001) European network of excellence on intelligent technologies for smart adaptive systems. <http://neuron.tuke.sk/competition/>. Accessed 1 Mar 2012
 42. IGMCo (2008) Iran Grid Management Co. <http://www.igmc.ir/>. Accessed 20 Apr 2008
 43. Iranian-Calendar-Center (2012) Institute of Geophysics, University of Tehran. <http://calendar.ut.ac.ir/>. Accessed 20 Mar 2012 (**in Farsi**)
 44. El-Attar EE, Goulermas JY, Wu QH (2009) Forecasting electric daily peak load based on local prediction. In: Power & energy society general meeting, 2009. PES '09. IEEE, 26–30 July 2009, pp 1–6. doi:[10.1109/pes.2009.5275587](https://doi.org/10.1109/pes.2009.5275587)
 45. Haishan W, Xiaoling C (2006) Power load forecasting with least squares support vector machines and chaos theory. In: The sixth world congress on Intelligent control and automation, 2006. WCICA 2006, 0-0 0 2006. pp 4369–4373. doi:[10.1109/wcica.2006.1713202](https://doi.org/10.1109/wcica.2006.1713202)