Energy Conversion and Management 49 (2008) 2678–2687

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Mid-term load forecasting of power systems by a new prediction method

Nima Amjady *, Farshid Keynia

Department of Electrical Engineering, Semnan University, Molavi, Semnan 35195-363, Iran

ARTICLE INFO

Article history:
Received 28 August 2007
Accepted 13 April 2008
Available online 5 June 2008

Keywords:
Mid-term load forecast
Daily peak load
Hybrid forecast method
Neural Network

ABSTRACT

Mid-term load forecasting (MTLF) becomes an essential tool for today power systems, mainly in those countries whose power systems operate in a deregulated environment. Among different kinds of MTLF, this paper focuses on the prediction of daily peak load for one month ahead. This kind of load forecast has many applications like maintenance scheduling, mid-term hydro thermal coordination, adequacy assessment, management of limited energy units, negotiation of forward contracts, and development of cost efficient fuel purchasing strategies. However, daily peak load is a nonlinear, volatile, and non-stationary signal. Besides, lack of sufficient data usually further complicates this problem. The paper proposes a new methodology to solve it, composed of an efficient data model, preforecast mechanism and combination of neural network and evolutionary algorithm as the hybrid forecast technique. The proposed methodology is examined on the EUropean Network on Intelligent TEchnologies (EUNITE) test data and Iran's power system. We will also compare our strategy with the other MTLF methods revealing its capability to solve this load forecast problem.

1. Introduction

Load forecast is an essential tool for power system operation and planning. Different kinds of this tool including short term load forecast (STLF), MTLF, and long term load forecast (LTLF) have their own applications. While STLF is essential for operation of power systems (e.g. unit commitment, economic dispatch, security assessment), LTLF is usually used for power system expansion and planning (e.g. construction of new power plants). Between these two extreme cases there is another kind of load forecast named MTLF, which is usually used in the operational planning of the power system such as maintenance scheduling and hydro thermal coordination [1,2].

Forecast step and forecast horizon are two key parameters for different prediction processes of the power systems. For STLF, forecast step is often one hour or a fraction of hour and forecast horizon is limited to one week ahead although its usual horizon is the next day [3,4]. In the LTLF, forecast horizon includes several years ahead (e.g. 10 years ahead) and its usual forecast step is one year (e.g. prediction of annual peal load) [5,6]. However, different forecast steps and horizons have been proposed for MTLF by the previous researchers. In [7], the forecast step of MTLF is considered one year and its horizon is a few years ahead, which looks more like LTLF. In this reference, the Authors proposed an adaptive neural network (NN) as their forecast tool for prediction of annual peak load. In [8,9], prediction of monthly energy consumption with dynamic NN and multiple regression model has been proposed, respectively, while their forecast horizon is up to 12 months ahead. In [1], the same quantity has been predicted, however their forecast horizon is one step ahead. In this reference, the original series has been decomposed to the trend and fluctuation components and different techniques have been evaluated for this purpose. Then each component has been predicted by the multi-layer perceptron (MLP) and radial basis function (RBF) neural networks. In [10], support vector machine has been proposed for prediction of daily peak load for 31 days ahead. In [11], prediction of daily peak load and daily total consumption for the next month has been considered. The authors applied various NN techniques and evaluated effect of different input features like temperature variables and calendar indicators. In [12], MLP neural network was used to predict daily peak load, daily total load and monthly electricity consumption. The forecast horizon of the daily peak and total load was one week ahead, while for the monthly electricity consumption was the next 12 months.

This paper focuses on the prediction of daily peak load for up to 31 days ahead (next month). Generally less research work has been performed on the MTLF than STLF and LTLF and most of this research is related to recent years. On the other hand, the MTLF accuracy is an important issue for both planning and operational planning of power systems like maintenance scheduling, management of limited energy units, mid-term hydro thermal coordination, adequacy evaluation, security assessment, and development of cost efficient fuel purchasing strategies. The economic impact of this load forecast is significant and more pronounced in a deregulated energy market. In competitive markets like California,

* Corresponding author. Tel.: +98 021 88797174; fax: +98 021 88880098.
E-mail address: amjady@tavanir.org.ir (N. Amjady).

0196-8904/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2008.04.008

where energy is traded, the accurate MTLF can provide an advantage in negotiations, and assist in the development of medium term generation, transmission and distribution contracts [8]. Prediction of daily peak loads can be also useful for congestion management. So, there is still an essential need for more accurate MTLF methods. This paper proposes a new forecast strategy for this purpose. Contribution of the paper can be summarized as follows:

(a) A new hybrid forecast method composed of NNs and evolutionary algorithm (EA) is proposed for the MTLF. The hybrid forecast method can find better solutions in the solution space of the training phase than a single NN.
(b) An efficient data model composed of the auto-regression part and a major calendar indicator as the exogenous variable is proposed for the MTLF problem. This data model is adaptively refined for each forecast horizon by a feature selection technique, which can consider both correlation and linear independency of the candidate inputs.
(c) The idea of preforecast is proposed to enhance the forecast accuracy, especially when the training set is limited (which is the case in the MTLF problem). Performance of different mechanisms for this purpose is evaluated and an efficient preforecast technique is introduced.
(d) The adjustable parameters of both the hybrid forecast method and feature selection technique are fine tuned by a kind of cross-validation. The cross-validation technique considers the most similar period to the forecast horizon as the validation period to simulate the real conditions of the forecast process as much as possible.

The remaining parts of the paper are organized as follows. In the second section, a more detailed evaluation of the MTLF problem and its effective inputs are presented. Then our data model for this problem is introduced. In the third section, details of the proposed hybrid forecast strategy are described. Obtained numerical results are presented and discussed in Section 4. A brief review of the paper and the future research are in Section 5.

2. The proposed data model

In [13], it has been shown that hourly load is a nonstationary stochastic time series. Daily peak load has similar characteristic, however it contains less smoothness with more volatility and sudden changes due to its longer period (one day against one hour). For instance, daily peak load of Iran's power system from January 1, 1997 up to December 31, 1999 is shown in Fig. 1. Hourly loads of this system from November 17, 1999 up to December 31, 1999 (time series with approximately the same length) are shown in Fig. 2. Horizontal axis in Figs. 1 and 2 indicates day and hour number, respectively. Vertical axis of these figures is in terms of MW. As seen, hourly loads have regular patterns and periodic behavior while daily peak loads involve with more volatility, nonstationarity and outliers (unusual loads). Besides, time series of daily peak load shows a variable rising trend. In addition to more complexity, less historical data is available for daily peak load prediction than hourly load forecasting. For instance, in the EUNITE network, which is one of our test cases, only two years data are available. In the Iran's power system more historical data is accessible. However, Iran is a fast developing country and its load pattern and behavior rapidly vary with time. So, far historical data has poor correlation with the forecast horizon and may be misleading for the prediction method. For instance, large iron foundries were commissioned in the period shown in Fig. 1, which could change the load pattern. Thus, the useful data may be again limited to two years ago. This time period usually contains sufficient historical data for the STLF [3,4,13], but for the MTLF its information content is limited. So, it is seen that a qualified forecast method is required for this MTLF problem. In the following, our solution strategy for the problem is presented.

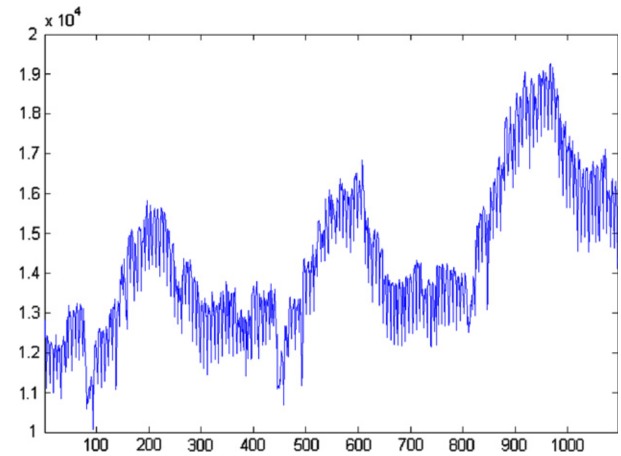

Fig. 1. Daily peak loads of Iran's power system from January 1, 1997 up to December 31, 1999.

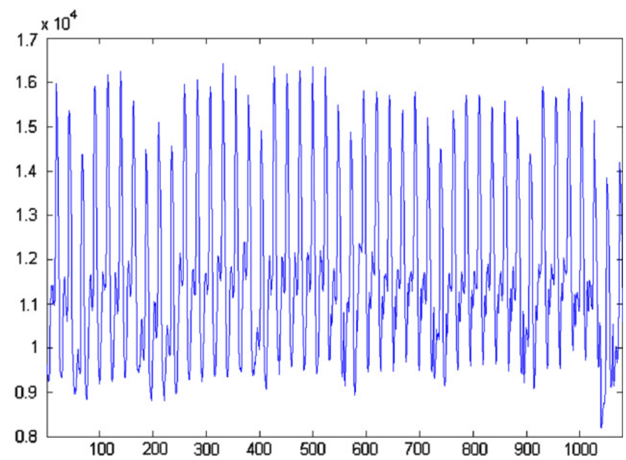

Fig. 2. Hourly loads of Iran's power system from November 17, 1999 up to December 31, 1999.

Literature on MTLF categorizes the methodologies for modeling MTLF into two general groups. The first group focuses on economic analysis, management and long term planning and forecasting of energy load and energy policies (referred to as the conditional modeling approach). Researchers in this group often use economic indicators such as gross national product (GNP), consumer price index (CPI), average salary earning (ASE), currency exchange rate (CER), and statistical indices of different industries, and/or electrical infrastructure measures such as number of connections and appliance saturation measure in addition to information on historical load data and weather related variables to forecast future energy demands [7,8,14]. The second approach (referred to as the autonomous modeling approach), primarily uses past loads and weather information to forecast future electricity demand [1,8–11]. Our methodology is in the second group, i.e. the economic indicators are not considered in our model. Economic indicators and electrical infrastructure measures are usually useful for LTLF and MTLF with long forecast horizon, e.g. prediction of annual peak load for at least one year ahead [7]. However, in the daily peak load forecasting for the next month, these indicators are not effective, since the forecast step and horizon are too short to observe their effect. A similar situation is seen in STLF (this kind of MTLF is closer to STLF than LTLF). Besides, as described in [1], prediction of economic indicators is itself a complex task.

We do not use weather data (such as temperature and relative humidity) in our model as well. The problems of using weather variables for STLF have been described in [3]. These problems for MTLF become more serious. STLF models that rely on weather information require prediction of weather parameters for the next few hours, or at most next few days. However, for our MTLF model, the weather parameters should be predicted for one month ahead, which is a significantly more challenging problem. For instance, in Iran, the weather bureaus only provide prediction of weather variables for up to the next week. Besides, the influence of weather is embedded in the periodic behavior of the load series, so that its influence may be captured if enough data are used to carry out the prediction [1]. In [8] it has been shown that, for MTLF, role of weather parameters can be implicitly considered by correct modeling of the load behavior. In [10], the authors evaluated different models for daily peak load prediction and concluded that time series modeling scheme (considering past loads) without temperature performs better than all other models with temperature.

Finally, some MTLF works considered a few input variables as calendar indicators. For instance, in [10], seven binary variables were proposed to encode weekdays, weekends, and holidays. In [9], 21 binary variables were used to indicate days and months. In [11], seven binary variables and one integer were used for this purpose. Considering calendar indicators (as proposed in the previous research works on MTLF) imposes many input variables with different nature (past loads are real values), which complicates training phase of the NN. Besides, for more input features, NNs usually require more training samples to extract input/output functional relationships [3], while we have limitation on the training set for the MTLF problem. So, in our initial modeling of the MTLF problem, we did not consider the calendar indicators and tried to implicitly capture their effect by the load behavior. For instance, daily peak load has weekly and annual periodicities, which can be also seen from Fig. 1 (this matter will be shown by the correlation analysis in the next section). So, by considering these periodicities, periodic behavior of daily peak load time series may be implicitly captured. On the other hand, information content of the calendar indicators can have a discriminatory effect for the NN. So, a compromise solution (between not considering these indicators and adding the detailed indicators) may be better for the MTLF data modeling. We found that considering major calendar indicator as a binary input variable to distinguish weekdays from weekends and public holidays can somewhat enhance the daily peak load forecast accuracy. We will show this matter in more details in the next sections.

By aggregating the above explanations, we reach to a data model including past daily peak loads plus the major calendar indicator for the MTLF problem. This indicator has value ‘1’ for weekdays and ‘0’ for weekend, which is Sunday and Friday for Christian and Islamic countries, respectively. Besides, for daily peak load forecasting, we treated public holidays like weekends [4,15] with ‘0’ value for the indicator. Hence, a minimum set of comprehensive input variables are considered for the MTLF problem, which simplify data acquisition, reduce input data noise (such as prediction error of economic indicators and weather variables) and avoid from redundancy in input data and less correlated inputs. Besides, the minimum set of inputs enhance training efficiency of the forecast method. Based on this data modeling, the proposed hybrid forecast method is described in the next section.

Some previous STLF works also considered daily peak load forecasting [4,12]. However, their forecast horizon is usually limited to one step ahead, i.e. the next day [4], or at most one week ahead [12]. We consider daily peak load prediction for one month ahead, which is a MTLF problem also mentioned in [10,11]. Major differences are seen between these two forecasts. For instance, peak load forecast of the next day or at most next week can be performed by means of predicted hourly loads around it [4] or a limited number of lagged daily peak loads (e.g. 5 days ago) [12]. However, for one month ahead, prediction of hourly loads is not practicable due to propagation of forecast error. Besides, we will show that much more lagged peak loads should be considered for this kind of forecast.

3. The proposed forecast method

As described in the previous section, our data model (excluding the major calendar indicator) is based on the past daily peak loads as the candidate input variables. However, it is still a large set and the best candidates should be selected among them. The best input features for our forecast process are those which have the highest correlation with the output variable (i.e. peak load of the next day) and the highest degree of linear independency, from linear algebra viewpoint. In this way, the most effective candidate inputs with minimum redundancy are selected. For this purpose, a two step correlation analysis has been used in this paper. At the first step, the correlation between each candidate input and the output variable is computed, where more correlation indicates a more effective candidate. If correlation index between a candidate variable and the output feature is greater than a prespecified value $cor1$, then this candidate is retained for the next step; else, it is not considered any further. In the second step, for the retained candidates, a cross-correlation analysis is performed. If the correlation index between any two candidate variables is smaller than a prespecified value $cor2$, then both variables are retained; else, only the variable with the largest correlation with respect to the output is retained, while the other is not considered any further. Less correlation between candidate variables results in more linear independency and so more information content of these variables [3,16]. This correlation analysis is applied to the candidate inputs, i.e. the past daily peak loads. Retained candidates after the two steps of the correlation analysis are selected as the input features of the forecast method [7]. We also examined other feature selection techniques such as regression analysis [17], Akaike’s information criterion (AIC) [18] and autocorrelation functions [19]. However, for the MTLF problem, the proposed two step correlation analysis resulted in the best feature selection and forecast accuracies, due to consideration of linear independency of candidate inputs in addition to their correlation.

The calendar indicator is excluded from the two step correlation analysis, since this indicator is a binary variable and has a low correlation with the output feature, even in the normalized form. However, the indicator can be still useful to separate different kinds of days. So, the input features of the forecast method includes past daily peak loads selected by the two step correlation analysis plus the calendar indicator. Now, training samples can be constructed for the forecast method. In order to avoid saturation phenomenon and provide better conditions for NN training, the input and output variables of the training samples are linearly normalized to be within the range [1].

Some researchers tried seasonal modeling for MTLF. For instance, in [8,10], year has been divided into hot and cold months with the aim of classifying similar data in one group. Then each group is dedicated to one NN (or any other used forecast method). We used this technique in our previous work on STLF [4]. However, for the MTLF problem, as previously described, we have limitation on the number of training samples and so we decided not to decrease the training set any further.

For construction of the forecast method, we simply begin with a MLP neural network due to its flexibility as a nonlinear predictor and ease of implementation [3]. According to Kolmogorov's theorem, the MLP can solve a problem by using one hidden layer, provided it has the proper number of neurons [7]. So, one hidden layer has been considered in the MLP structure of the NN. Different learning algorithms like Levenberg–Marquardt (LM), Broyden, Fletcher, Goldfarb, Shanno (BFGS), Bayesian regularization (BR), gradient descent back propagation, gradient descent with momentum backpropagation and conjugate gradient back propagation have been examined for the MTLF problem [20]. Among them the LM was selected due to its best results for daily peak load prediction. This algorithm is an approximation of Newton's method and it computes the approximate Hessian matrix. Mathematical details of the LM algorithm can be found in [20,21].

The results of the training phase, i.e. the NN weights, are given to an evolutionary algorithm (EA). From mathematical viewpoint, training of a NN in general, is a search process in a vector space where its dimension is equal to the number of degrees of freedom (such as weights) of the NN. Search path and criterion are dependent on the training mechanism of the NN. For the MTLF, we found that near optimum solutions are usually close to each other in the vector space. In other words, the optimum solution is surrounded by a set of semi-optimum ones. At the end of the training phase, the learning algorithm may find one of these near optimum solutions while the better ones might be in its vicinity and unseen for the NN, since the learning algorithms usually search the solution space in a special direction (like the steepest descent). So, this makes the motivation to search around the final solution of the learning algorithm in various directions as much as possible to find a better solution. EA can be a suitable candidate for what is required. Since a local search around the solution of the NN is required, we use from EA with momentum, which has smoother search paths avoiding from sudden changes. In the rare cases that the EA cannot find a better solution, the NN's one will be restored. The evolution of the proposed EA is based on the following relations:

$$\Delta W_{(n+1)} = m \cdot \Delta W_{(n)} + (1-m) \cdot g \cdot W_{(n)} \tag{1}$$

$$W_{(n+1)} = W_{(n)} + \Delta W_{(n+1)} \tag{2}$$

where W is an adjustable parameter (weight or bias value) of the NN and ΔW indicates change of it. The subscripts n and $n+1$ represent two successive generations (parent and child, respectively) of the EA. g is a small random number separately generated for each adjustable parameter. m is momentum constant. Use of the momentum can smooth the search path decreasing sudden changes. In our examinations $m = 0.5$ and g is selected in the range of (0,0.1) for all generations of the EA. In other words, a uniform search is selected for the proposed EA, without using localizing techniques, such as hill climbing operator [22,23]. Non-uniform searches are suitable when there is one optimum point in the solution space and it is desired that the stochastic search technique, like EA, converges to it (e.g. the unit commitment problem [22]). However, in our problem there are several optimums and each one may be better than the other. At the beginning of the EA, $W_{(0)}$ is the obtained value from the learning algorithm for the adjustable parameter W and $\Delta W_{(0)} = 0$. In each cycle, the EA repeats (1) and (2) until the next generation of all adjustable parameters is obtained. Then, the error function of the NN (validation error) is evaluated for the new generation. If the child has less error than its parent, the parent is replaced by the child, otherwise the parent is restored and the next cycle of the EA is executed. So, at the end of the EA, the best examined solution among all generations will be selected. We considered 100 generations for the EA in our solution strategy for the MTLF problem. Although efficiency of the EA depends on the position of the near optimum solutions in the vector space, however as a rule of thumb for the MTLF problem, in about 80% of our experiments, EA greatly enhanced the obtained solution of the NN, in about 10% little improvement was observed and in about 10% no improvement was obtained at all and so the NN's solution was restored. Detailed numerical results of the proposed MTLF strategy will be presented in the next Section.

Daily peak load has sudden changes, volatile behavior and multiple periodicities, which can be seen from Fig. 1. To correctly learn all characteristics of such a complex signal, NNs usually require a large number of hidden nodes and weights; otherwise the learning capability of the NN is saturated and even the learning algorithm may diverge. On the other hand, number of training samples of the MTLF is limited. In these conditions it is very likely that the NN traps in the overfitting problem, where the NN learns well the training samples, however its performance for the unseen test samples is poor. In other words, the NN memorizes training samples instead of learning them and so it loses its generalization capability. To solve the overfitting problem, a kind of cross-validation technique has been incorporated in our solution strategy in which the progress of the training phase is controlled by the validation error instead of training error. Validation set is a part of training samples which is removed from the training set and so becomes unseen for the NN. Minimum of the validation set error is selected as the optimal point of the training phase, where it is expected that the generalization capability of the NN be maximized. For instance, training and validation error of a training phase of the NN for Iran's power system is shown in Fig. 3. Here, the forecast horizon is August 1999 and training set is its two years ago. In Fig. 3, the training and validation errors are in terms of MSE (mean square error). As seen, the training error decreases along with the number of iterations while the validation error decreases at first (approximately coincided to the training error), bounces around, and then starts to increase and continues in this way. The validation error in training iteration 300 is about 10 times larger than its minimum (in training iteration 11), revealing occurrence of serious overfitting. By extending the training phase from iteration 300, the validation error further increases. Similar curves for training and validation errors have been obtained in the other training phases.

In addition to the determination of the optimal point of the training phase, tunable parameters of the whole method including $cor1$ and $cor2$ of the correlation analysis and number of hidden nodes N_H of the NN are also adjusted by the cross-validation technique. For this purpose, the training phase with different values for

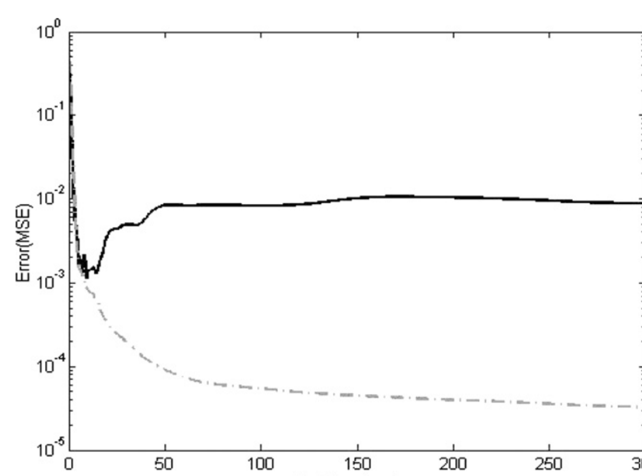

Fig. 3. Training error (dash dot line) and validation error (solid line).

the tunable parameters is executed and the execution resulted in the least validation error is selected. To reduce number of combinations that should be examined, the two step procedure of [7] has been also used. For MTLF, we found that efficiency of the cross-validation technique is highly dependent on the selection of the validation set. The signal behavior (here, daily peak load) in the validation set should be as similar as possible to the signal behavior in the forecast horizon so that the validation error be a true representation of the forecast error. To do this for the MTLF problem, we evaluated correlation between daily peak load of different months and found that the signal behavior in each month usually has the highest correlation with the same month in the last year. So, this time period is selected as the validation set. For instance, in Fig. 3, the validation set is August 1998, removed from the two years training set. The validation error is used in both learning phase of the NN and evolution of the EA (for assessment of different generations), since the generalization capability of the whole forecast method should be evaluated.

Due to importance of the validation set data, after adjustment of the tunable parameters and determination of the termination point of the learning phase, the validation month is returned to the training set. Then, the final training phase of the NN is executed by the whole training data and the obtained parameters of the cross-validation. Obtained weights of this learning phase are fine tuned by the EA and then loaded to the MLP structure. Now, the NN is ready for the test phase, where prediction of each unseen test sample can be obtained by a forward propagation of the MLP.

Another technique, which is used in our solution strategy to enhance the MTLF accuracy is preforecast. Tracking a nonstationary signal with rapid variations, e.g. the daily peak load, for NNs is usually very hard and large deviations may be seen in the estimation of the NN [3]. On the other hand, if the MLP has an initial forecast following trend of the target signal, then it can much easier learn behavior of the signal. In this case, the MLP should learn difference between the two trajectories instead of the global values of the target trajectory. This makes the motivation of preforecast. However, the initial trajectory or preforecast must have a close correlation with the target one, otherwise it has no use and even be misleading for the MLP. Especially, for the MTLF problem, our training data is limited and so the learning process should be as simple as possible. In other words, the preforecast should be as close as possible to the target trajectory. Thus, to enhance the accuracy of the preforecast, we cascaded a few NNs, shown in Fig. 4. Architecture and data flow of the proposed hybrid intelligent system can be seen from this figure. All preforecast and forecast NNs have the same set of lagged daily peak loads (selected by the two step correlation analysis) plus the calendar indicator. Besides, each NN sends its peak load prediction (PLP) for the next one. We examined different architectures for the preforecast block and sample results for the Iran's power system are shown in Table 1. Learning algorithm and number of cascaded MLP neural networks (in the parentheses) have been indicated. Up to 6 MLP neural networks with LM learning algorithm are shown in the columns 1 to 6. For comparison, up to 3 MLPs with BR and BFGS learning algorithms are also shown in the next six columns. In the last three columns, obtained results for LM → BR, LM → BFGS, and LM → BR → BFGS are shown. Preforecast errors shown in the second row of Table 1 are in terms of mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{d=1}^{N}\frac{|L_{(d)} - FL_{(d)}|}{L_{(d)}} \times 100 \tag{3}$$

where $L_{(d)}$ and $FL_{(d)}$ are actual and forecast peak load of day d, respectively and N is the number of days of the forecast horizon (e.g. one month). In the third row of Table 1, total learning time of the cascaded NNs in terms of second (s) is shown. For each NN, when validation error starts to increase, its training phase is terminated and weights of the NN in the training iteration with minimum validation error are returned as the results of the training phase. This termination mechanism is used as the convergence criterion for the training phase of NNs in all examinations of this paper. As seen from Table 1, the learning times of the cascaded NNs do not increase linearly, since training phase of NNs may be terminated in different iterations (e.g. training iteration 11 in Fig. 3). The MAPE values and learning times shown in Table 1 are average values for 12 months of 1999, where for each month its two years ago has been considered as the training set.

As seen from Table 1, the best preforecast accuracies are related to three and four cascaded MLPs with LM. For more cascaded MLPs, preforecast accuracy decreases. Also, other learning algorithms, such as BR and BFGS have higher preforecast errors. By further increasing the number of cascaded MLPs, e.g. more than LM(6), preforecast errors further increase. We examined other data sets, but always LM(3) or LM(4) were the best choices. So, we consid-

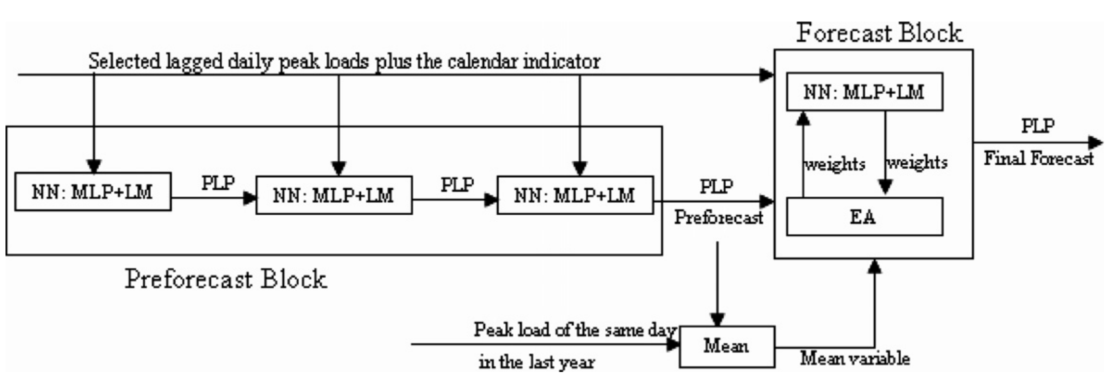

Fig. 4. Architecture and data flow of the proposed hybrid intelligent system including preforecast and forecast parts (PLP stands for peak load prediction).

Table 1
MAPE values (%) and learning times (s) for different preforecast mechanisms (average values for 12 months of 1999 for Iran's power system)

LM (1)	LM (2)	LM (3)	LM (4)	LM (5)	LM (6)	BR (1)	BR (2)	BR (3)	BFGS (1)	BFGS (2)	BFGS (3)	LM–BR	LM–BFGS	LM–BR–BFGS
3.20	3.07	2.21	2.41	3.48	3.99	4.54	4.68	4.40	4.19	4.53	3.55	4.35	4.36	5.20
10	27	35	55	75	81	18	33	52	15	27	49	34	47	64

ered LM(3) as the preforecast block in the proposed hybrid intelligent system, shown in Fig. 4. For the MTLF problem, number of samples and so information content of the training data set is limited. Thus, by increasing the number of cascaded NNs, preforecast accuracy saturates and then begins to decrease. We also examined time series methods such as ARIMA (auto-regressive integrated moving average) for preforecast, however their prediction accuracy were usually less than the NN techniques. This is due to the fact that time series techniques are linear forecast methods, while the daily peak load often has highly nonlinear behavior like Fig. 1.

In order to evaluate information value of the predicted peak load as an input feature, we analyze it along with the other candidate variables, i.e. the lagged daily peak loads, by the two step correlation analysis. Even the predicted value of the first NN (which usually has the highest prediction error among the cascaded NNs) was always selected for all examined data sets. In other words, not only the predicted values have high correlation with the output variable (more than *cor*1), but also in competition with the other competitor candidates, they always win and are selected by the second step of the correlation analysis. This evaluation confirms our preforecast idea. Besides, for the forecast NN, i.e. the last one, another input feature named mean variable (Fig. 4) is also used, which is average of the preforecast value and peak load of the same day in the last year. Due to annual periodicity, output variable has a high correlation with peak load of the same day in the last year (we will show it numerically in the next section). These results make the motivation to construct another candidate variable from the two variables with the highest correlations by the averaging operator (Fig. 4). Although from the mathematical view point, the obtained mean variable is linearly dependent on its constructing variables, however when we examine it by the two step correlation analysis, is usually selected due to its high correlation with the output variable (its high correlation overcomes the linear dependency). As seen from Fig. 4, the EA further modifies the weights of the last NN to find a better solution in the vector space.

The training samples are shuffled a few hundred times before training the NN to enhance the randomness of the data and to eliminate the sequential bias effect. Now, preparation of the proposed method can be summarized as the following step by step algorithm:

(1) Select initial values for the adjustable parameters including *cor*1 and *cor*2 of the correlation analysis and number of hidden nodes of the NNs (both preforecast and forecast blocks).
(2) Consider a set of lagged daily peak loads for a sufficiently long period so that the weekly and annual periodicities can be considered (e.g. one year ago). Among this set, select the input features owning the highest correlation with the output variable (more than *cor*1) and the lowest degree of linear dependency (less than *cor*2) by means of the proposed feature selection technique. These input features plus the calendar indicator are used for the training samples of the first NN of the preforecast block, while for its next two NNs, peak load prediction of the previous NN is also considered as the input feature (Fig. 4). For the forecast NN, both the preforecast and mean variable are added to the inputs. After determination of the input features of each NN, training samples of the NNs can be constructed by the available historical data (e.g. two years ago).
(3) For each NN, remove the samples related to the cross-validation period, i.e. the same month in the last year, from the training set and consider them as the unseen validation samples.
(4) Train the NNs (with LM learning algorithm) and EA. Optimal point of the training phase of the NNs is selected according to the cross-validation technique.
(5) If all combinations of the adjustable parameters have been examined, select the best one according to the validation error and go to the next step; otherwise change the parameters based on the two step procedure of [7] and go back to step 2.
(6) Return the validation set to the training data and perform the final training phase by means of the obtained values for the adjustable parameters. Now prediction of daily peak loads of the next month can be obtained by the forward propagation of the NNs.

4. Test results

The proposed MTLF method has been examined on two test cases: EUNITE competition data and Iran's power system. In 2001, EUNITE network organized a competition aiming at mid-term load forecasting (predicting daily maximum load of the next 31 days). Fifty-six competitors participated in the competition where the authors of [10] proposed the winning entry. Their MTLF method was based on the support vector machine (SVM) and can reach to minimum MAPE among all competitors. In the EUNITE test case, forecast horizon was January 1999 and so n becomes 31. Two years load data (from 1997 to 1998), four years average daily temperature data (from 1995 to 1998) and dates of holidays (from 1997 to 1999) have been provided by the organizer of the competition [10]. EUNITE data can be easily obtained from their website [24].

Our second test case is power system of Iran, a fast developing country. As seen from Fig. 1, load demand of Iran's power system rapidly increases. Similarly, we considered load data of 1997 and 1998 as the training set and those of 1999 as the testing set. Load pattern of Iran's power system rapidly varies and so farther daily peak loads (before 1997) have poor correlation with those of 1999. The load data of Iran's power system can be obtained from [25].

Obtained results from the first and second steps of the correlation analysis for the EUNITE network data are shown in Fig. 5 and Table 2, respectively. This correlation analysis is related to the forecast NN (the last NN of Fig. 4) and so its candidate inputs comprise preforecast, mean variable and lagged daily peak loads (up to one year ago):

$$\{\text{Preforecast},\ \text{Mean Variable},\ L_{(d-1)}, L_{(d-2)}, \ldots, L_{(d-365)}\} \tag{4}$$

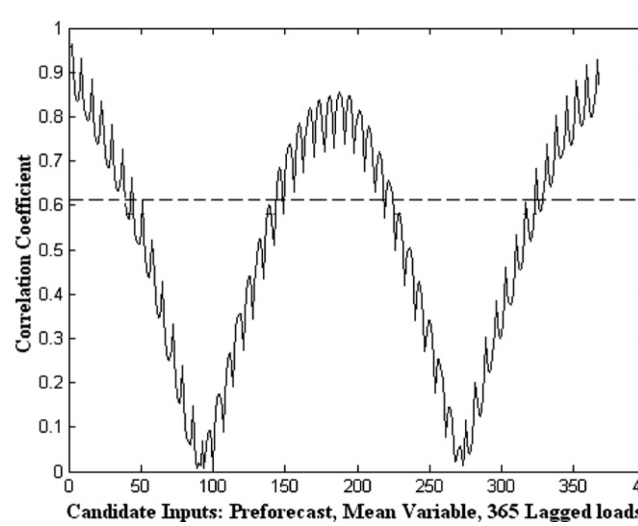

Fig. 5. Results of the first step of the correlation analysis for EUNITE network data.

Table 2
Results of the correlation Analysis for the EUNITE network data

*Cor*1	*Cor*2	Number of selected inputs	The best case
0.2	0.9	152(2)	
0.2	0.8	36(1)	
0.2	0.7	14(1)	
0.2	0.6	9(1)	
0.4	0.9	112(2)	
0.4	0.8	28(1)	
0.4	0.7	11(1)	
0.4	0.6	7(1)	
0.6	0.9	74(2)	
0.6	0.8	20(1)	
0.6	0.7	9(1)	
0.6	0.6	6(1)	
0.8	0.9	28(2)	
0.8	0.8	9(1)	
0.8	0.7	3(1)	
0.8	0.6	2(1)	
0.61	0.83	37(2)	(*)

(1) The preforecast is among the selected inputs.
(2) Both the preforecast and mean variable are among the selected inputs.

Table 3
Obtained results for the EUNITE network data

Method	MAPE (%)
Winning entry of the EUNITE competition	1.95
Proposed technique without calendar indicator	1.65
Proposed technique with 1 calendar indicator (*)	1.60
Proposed technique with 6 calendar indicators	1.79
Proposed technique with 6 + 11 calendar indicators	2.61

where $L_{(d-k)}$ represent daily peak load of k days ago. The candidate inputs of (4) are shown in the same order on the horizontal axis of Fig. 5. Hence, there are totally 1 + 1 + 365 = 367 variables in the initial set of candidate inputs. The vertical axis of Fig. 5 indicates correlation coefficient with the output variable. For the EUNITE data, the best values of *cor*1 = 0.61, *cor*2 = 0.83, and N_H = 25 have been obtained by the cross-validation technique like the procedure described for the Iran's power system. As described in the previous Section, the calendar indicator is excluded from the two step correlation analysis. However, the cross-validation technique has been performed on the final data model including the calendar indicator, since this technique should simulate the real conditions of the forecast process as much as possible. The value of *cor*1 is indicated on Fig. 5 by a horizontal dashed line. Candidate inputs with correlation coefficient higher than *cor*1 are selected for the second step. As seen, preforecast, i.e. the first candidate, has the highest correlation with the output variable. Besides, the second one, i.e. the mean variable, also has a high correlation. Correlation coefficient of candidate inputs from 3 to 367, show weekly (small oscillations), half year (middle peak) and annual (last peak) periodicities of the daily peak load.

In Table 2, sample results for different values of *cor*1 and *cor*2 including the best ones are shown. As seen, by increasing *cor*1, i.e. focusing on more correlation between candidate inputs and output, and by decreasing *cor*2, i.e. focusing on more linear independent candidates, number of selected inputs decreases. However, for all values of *cor*1 and *cor*2, the preforecast is among the selected candidates. In some cases and especially the best case, the mean variable is also chosen. The 37 selected candidate inputs (the last row of Table 2) plus the calendar indicator comprise input features of the forecast NN (totally, 38 inputs), which are as follows:

$$\{\text{Calendar Indicator},\ \text{Preforecast},\ \text{Mean Variable},\ L_{(d-1)}, L_{(d-7)}, L_{(d-9)}, L_{(d-11)}, L_{(d-13)}, L_{(d-14)}, L_{(d-17)}, L_{(d-19)}, L_{(d-21)}, L_{(d-23)}, L_{(d-25)}, L_{(d-28)}, L_{(d-31)}, L_{(d-33)}, L_{(d-36)}, L_{(d-37)}, L_{(d-39)}, L_{(d-41)}, L_{(d-43)}, L_{(d-50)}, L_{(-315)}, L_{(d-320)}, L_{(d-323)}, L_{(d-325)}, L_{(d-328)}, L_{(d-331)}, L_{(d-333)}, L_{(d-336)}, L_{(d-339)}, L_{(d-341)}, L_{(d-344)}, L_{(d-347)}, L_{(d-349)}, L_{(d-356)}, L_{(d-364)}\} \quad (5)$$

The other candidate inputs not shown in (5) (367 − 37 = 330 candidates), have been eliminated by the two step correlation analysis. So, the filtering ratio of the proposed feature selection technique for the EUNITE data is $(1-(37/367)) \times 100$ = 89.9% or 89.9% of inefficient candidate inputs has been removed, which can greatly enhance the training phase of the proposed method. In Table 3, MTLF error of the proposed method (obtained from the forecast NN) and the winning entry of the EUNITE competition [10] are represented. To better illustrate the effect of calendar indicators, as exogenous variables, MTLF error of the proposed technique in four cases are shown in Table 3. The 6 calendar indicators (shown in rows 5 and 6 of Table 3) are binary variables, which encode weekdays. The six binaries stand for Monday to Saturday, respectively and Sunday is represented as all six variables are set to zero. The 11 calendar indicators (shown in the last row of Table 3) similarly encode months of a year. Excluding the calendar indicators, the other input variables of the four cases of the proposed technique are the same. The results of Table 3 are in terms of MAPE, defined in (3), since it was the error metric of the competition. This table shows that among the four cases, our data model with one calendar indicator produces the best results, which is represented by (*). Besides, by increasing the calendar indicators, MTLF accuracy of the proposed method decreases, since its training process becomes more complicated. The MTLF error of our initial data model (without calendar indicator) is close to the best one.

The authors of [10] examined their method, i.e. SVM, with different data models and the MAPE value shown in Table 3 was their best result, which won the competition. The other competitors of the EUNITE had generally lower forecast accuracies, e.g. the MAPE of the second place was 2.11% [24]. As seen, our proposed method with both initial data model (1.65% MAPE) and final data model (1.60% MAPE) has a considerably better forecast accuracy than the winning entry. As another comparison, combination of wavelet transform and NN has been proposed in [19] for load forecast. The Authors examined different combinations of input features and their best MAPE values for the EUNITE network data were 2.8% and 3.1%, respectively.

To better illustrate effect of the preforecast mechanism, actual load and predicted values of the forecast strategy with 3 different preforecasts (LM(3), LM(6), and BFGS(1)) for the EUNITE network data are shown in Fig. 6. As seen, predicted values with LM(3) preforecast overall follows the actual load better than the other forecasts. MAPE of these three forecasts are 1.60%, 2.08% and 4.19% for preforecast of LM(3), LM(6), and BFGS(1), respectively. So, LM(3) is a good candidate for preforecast, which is also in accordance with the results of Table 1. Performance of LM(6) is less than it and BFGS(1) is a poor candidate. Despite the mentioned comparisons in Table 1 and Fig. 6, it cannot be claimed that LM(3) is the best preforecast mechanism for the MTLF; it is only a good candidate. Indeed, finding the best preforecast mechanism is a challenging task that demands further research.

Sample results of the correlation analysis for Iran's power system (August, 1999) are shown in Fig. 7 and Table 4, like Fig. 5 and Table 2, respectively. Here, the 367 candidate inputs are as described in (4). As seen from Fig. 7, daily peak load of Iran's power system has high dependency on the preforecast, mean variable, its short-run trend (daily peak loads close to the forecast day) and annual periodicity, like the EUNITE network data. However, effect of half year periodicity for Iran's power system is less than EUNITE. Results of Table 4 follow the same trend of Table 2. Similar correlation results have been obtained for the other months of

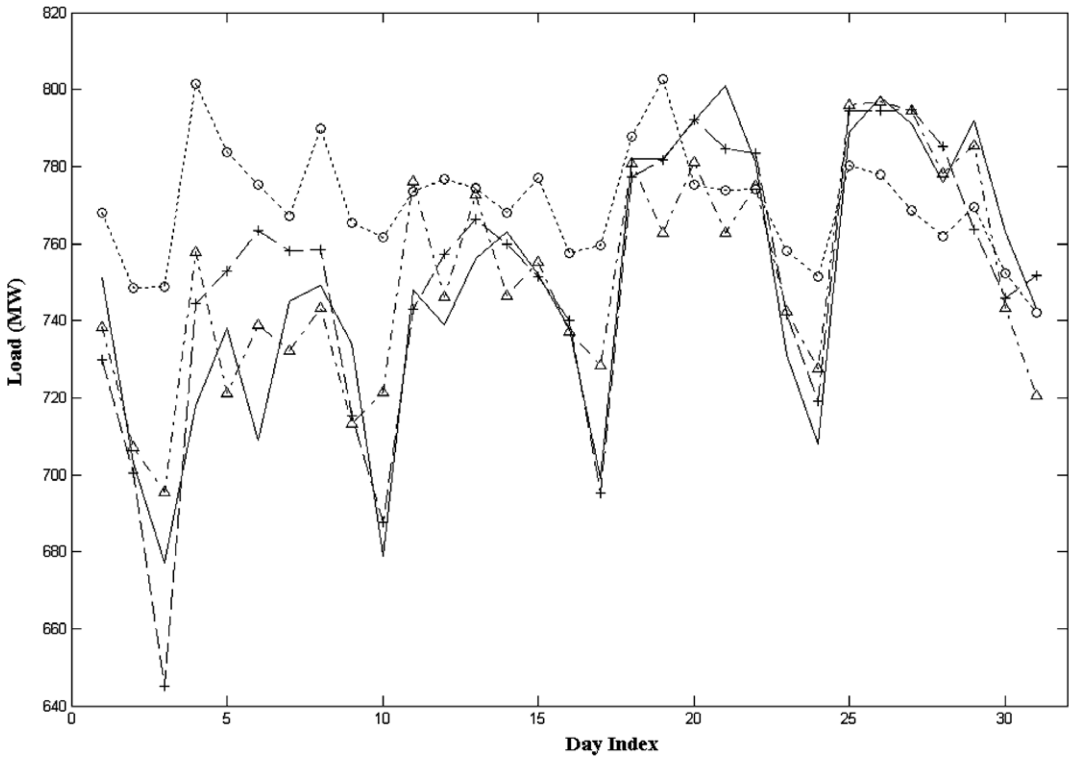

Fig. 6. Actual loads (solid line), forecast with LM(3) preforecast (dashed line with plus sign), forecast with LM(6) preforecast (dash dot line with triangle sign), and forecast with BFGS(1) preforecast (dotted line with circle sign).

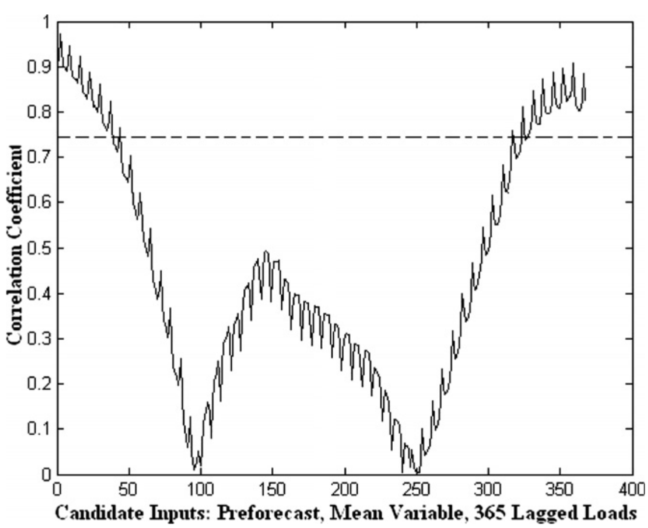

Fig. 7. Results of the first step of the correlation analysis for Iran's power system data (August, 1999).

Table 4
Results of the correlation Analysis for the Iran's power system data (August, 1999)

*Cor*1	*Cor*2	Number of selected inputs	The best case
0.2	0.9	243(2)	
0.2	0.8	64(1)	
0.2	0.7	20(1)	
0.2	0.6	11(1)	
0.4	0.9	117(2)	
0.4	0.8	28(1)	
0.4	0.7	9(1)	
0.4	0.6	5(1)	
0.6	0.9	70(2)	
0.6	0.8	11(1)	
0.6	0.7	4(1)	
0.6	0.6	3(1)	
0.8	0.9	35(2)	
0.8	0.8	6(1)	
0.8	0.7	1(1)	
0.8	0.6	1(1)	
0.74	0.88	26(2)	(*)

(1) The preforecast is among the selected inputs.
(2) Both the preforecast and mean variable are among the selected inputs.

1999 in Iran's power system. The 26 selected candidate inputs indicated in the last row of Table 4 plus the calendar indicator comprise input features of the forecast NN (totally, 27 inputs), which are as follows:

$$\{\text{Calendar Indicator},\ \text{Preforecast},\ \text{Mean Variable},\ L_{(d-1)}, L_{(d-14)}, L_{(d-35)}, L_{(d-49)}, L_{(d-55)}, L_{(d-63)}, L_{(d-308)}, L_{(d-315)}, L_{(d-316)}, L_{(d-318)}, L_{(d-322)}, L_{(d-324)}, L_{(d-326)}, L_{(d-329)}, L_{(d-335)}, L_{(d-337)}, L_{(d-339)}, L_{(d-341)}, L_{(d-347)}, L_{(d-350)}, L_{(-357)}, L_{(d-360)}, L_{(d-362)}, L_{(d-364)}\} \tag{6}$$

The other past daily peak loads not shown in (6) (367 − 26 = 341 candidates), have been eliminated by the two step correlation analysis. So, the filtering ratio of the feature selection technique for this test case is $(1-(26/367))\times 100 = 92.9\%$ indicating the effectiveness of the technique. MTLF error of the proposed method with the four data models for Iran's power system are shown in Table 5, like Table 3. However, the MAPE values shown in Table 5 are average values for 12 months of 1999. The same trend of Table 3 is also seen in Table 5. For a better comparison, detailed results of our initial and final data models (owning less MAPE values in Table 5) for all months of 1999 are shown in Table 6. To give an insight about stability of predictions, PAPE (peak absolute percentage error) values have been also mentioned in Table 6:

$$\text{PAPE} = \underset{1\leqslant d\leqslant N}{\text{Max}}\left(\frac{|L_{(d)} - FL_{(d)}|}{L_{(d)}}\right)\times 100 \tag{7}$$

Table 5
Obtained results for Iran's power system in 1999

Method	MAPE (%)
Proposed technique without calendar indicator	1.85
Proposed technique with 1 calendar indicator (*)	1.76
Proposed technique with 6 calendar indicators	2.09
Proposed technique with 6 + 11 calendar indicators	3.19

Table 6
Obtained results for Iran's power system in all months of 1999

Month	Proposed technique without the calendar indicator		Proposed technique with the calendar indicator	
	MAPE (%)	PAPE (%)	MAPE (%)	PAPE (%)
1	2.15	8.69	1.44	6.65
2	1.99	6.53	2.15	7.45
3	2.01	8.11	1.49	5.58
4	1.85	9.25	1.92	7.12
5	1.84	5.35	1.77	5.11
6	1.75	3.96	1.63	3.84
7	1.68	4.33	1.68	6.49
8	1.70	4.03	1.41	3.62
9	1.77	5.12	1.59	5.19
10	1.89	4.22	2.36	7.84
11	1.69	5.14	1.42	3.93
12	1.92	9.01	2.33	7.71
Average	1.85	6.14	1.76	5.88

Description of parameters of (7) is as explained for (3). All MAPE and PAPE values are in terms of percentage. Table 6 shows that our data model with the calendar indicator overall has better forecast accuracy and stability (less value of average MAPE and PAPE) than the initial data model. Iran is a developing country and so its daily peak load has a more volatile and time variant behavior than EUNITE load data. Thus, MAPE values of Iran's power system are a little larger than that of EUNITE.

As another comparison, it is noted that average MAPE value of Table 6, i.e. 1.76%, is considerably less than all average MAPE values of Table 1. This enhanced accuracy is mainly related to the forecast block and especially its EA part (Fig. 4), which is added to the preforecast mechanism of LM(3). *cor*1, *cor*2, N_H, and number of selected variables (NSV) for Iran's power system are represented in Table 7. NSV indicates number of all inputs including those selected by the two step correlation analysis plus the calendar indicator. For the NNs in the EUNITE and Iran's test cases, activation function of the input layer nodes is identity function and for the nodes of the hidden and output layers is hyperbolic tangent.

Due to high dependency of the daily peak load on the annual periodicity, which can be seen from Figs. 5 and 7, lagged loads up to one year ago are usually among the selected input features of the correlation analysis. So, available data for preparation of training samples reduces to one year, resulting in a maximum of 365 training samples (by consideration of validation samples, this number further decreases). This limited set of training samples intensifies the overfitting problem and necessitates use of the cross-validation technique as discussed in the previous section.

The major part of the setup time of the whole method is related to the cross-validation technique where the adjustable parameters including *cor*1, *cor*2, and N_H of NNs are determined. However, in practice, this setup time can be greatly reduced. At first, it is noted that LM, used for the training of the NNs (Fig. 4), is a fast learning algorithm (Table 1). As seen from Table 7, optimum values of *cor*1, *cor*2, and N_H for different test months are close to each other. Besides, selected values for the EUNITE network data are also close to those of Iran's power system. So approximate ranges for the adjustable parameters can be easily determined (e.g. [0,5,0.7] for *cor*1). Then, by the cross-validation technique, these parameters are fine tuned. Changing the parameters is based on the two step procedure of [7] to reduce the number of combinations. The whole setup time of the proposed method for our examined cases on a personal computer Pentium P4 3.2 GHz, with 1 GB RAM memory is about one hour, which is reasonable within a one month ahead decision making framework. In some forecast problems, such as prediction of price in a real time market, setup time may be an important factor. However, in the MTLF problems, forecast accuracy is the key issue, while enough computation time is usually available.

Table 7
Parameters of the forecast strategy for Iran's power system

Month	*cor*1	*cor*2	N_H	NSV
1	0.63	0.86	20	26
2	0.68	0.86	20	24
3	0.61	0.86	25	30
4	0.61	0.88	20	33
5	0.59	0.86	25	42
6	0.64	0.85	20	27
7	0.53	0.87	25	39
8	0.74	0.88	20	27
9	0.60	0.85	20	29
10	0.62	0.85	20	22
11	0.62	0.87	20	35
12	0.63	0.85	20	32
Average	0.61	0.85	22	29

5. Conclusion

MTLF is a newer aspect of power system load forecasting. Different kinds of MTLF such as prediction of daily peak load, daily energy consumption, monthly electricity consumption, and annual peak load have been proposed by the previous researchers. This paper focuses on daily peak load prediction for the next month, due to its importance especially for operational planning of power systems. However, daily peak load is a complex, volatile and non-stationary signal with limited information content of the training set. Here, a hybrid solution strategy has been proposed for this purpose, composed of an efficient data model, preforecast mechanism and a hybrid forecast method for prediction. Different data models for this MTLF problem have been evaluated and the best ones are introduced. Obtained results for the EUNITE network data and Iran's power system reveals the capability of the proposed strategy for daily peak load prediction.

The research work is under way in order to develop an optimization method, like dynamic programming, to find optimum values of the adjustable parameters and the best preforecast structures. Besides, consideration of the other kinds of MTLF, like prediction of monthly electricity consumption, can be another matter of the future research.

6. List of symbols

W	adjustable parameter (weight or bias value) of the NN
ΔW	change of W
N_H	number of hidden nodes of the NN
n	generation number of the EA
g	random number
m	momentum constant of the EA
$L_{(d)}$	actual peak load of day d
$FL_{(d)}$	forecast peak load of day d
N	number of days of the forecast horizon

References

[1] Gonzalez-Romera E, Jaramillo-Moran MA, Carmona-Fernandez D. Monthly electric energy demand forecasting based on trend extraction. IEEE Trans Power Syst 2006;21:1946–53.

[2] Amjady N, Farrokhzad D, Modarres M. Optimal reliable operation of hydrothermal power systems with random unit outages. IEEE Trans Power Syst 2003;18:279–87.

[3] Amjady N. Short-term bus load forecasting of power systems by a new hybrid method. IEEE Trans Power Syst 2007;22:333–41.

[4] Amjady N.Short-term hourly load forecasting using time series modeling with peak load estimation capability. IEEE Trans Power Syst 2001;16:798–805.

[5] Shakouri H, Rastad M, Nazarzadeh J. A hybrid nonlinear model for the annual maximum simultaneous electric power demand. IEEE Trans Power Syst 2006;21:1069–78.

[6] Kandil MS, El-Debeiky SM, Hasanien NE. Long-term load forecasting for fast developing utility using a knowledge-based expert system. IEEE Trans Power Syst 2002;17:491–6.

[7] Tsekouras GJ, Hatziargyriou ND, Dialynas EN. An optimized adaptive neural network for annual midterm energy forecasting. IEEE Trans Power Syst 2006;21:385–91.

[8] Ghiassi M, Zimbra DK, Saidane H. Medium term system load forecasting with a dynamic artificial neural network model. Electric Power Syst Res 2006;76:302–16.

[9] Mirasgedis S, Sarafidis Y, Georgopoulou E, Lalas DP, Moschovits M, Karagiannis F. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 2006;31:208–27.

[10] Chen BJ, Chang MW, Lin CJ. Load forecasting using support vector machines: A study on EUNITE competition 2001. IEEE Trans Power Syst 2004;19:1821–30.

[11] Doveh E, Feigin P, Greig D, Hyams L. Experience with FNN models for medium term power demand predictions. IEEE Trans Power Syst 1999;14:538–46.

[12] Yalcinoz T, Eminoglu U. Short term and medium term power distribution load forecasting by neural networks. Energy Convers Manage 2005;46:1393–405.

[13] Fan S, Chen L. Short-term load forecasting based on an adaptive hybrid method. IEEE Trans Power Syst 2006;21:392–401.

[14] Ranaweera DK, Karady GG, Farmer RG. Economic impact analysis of load forecasting. IEEE Trans Power Syst 1997;12:1388–92.

[15] Song KB, Baek YS, Hong DH, Jang G. Short-term load forecasting for the holidays using fuzzy linear regression method. IEEE Trans Power Syst 2005;20:96–101.

[16] Amjady N. Generation adequacy assessment of power systems by time series and fuzzy neural network. IEEE Trans Power Syst 2006;21:1340–9.

[17] Garcia MP, Kirschen DS. Forecasting system imbalance volumes in competitive electricity markets. IEEE Trans Power Syst 2006;21:240–8.

[18] El-Keib A, Ma X, Ma H. Advancement of statistical based modeling techniques for short-term load forecasting. Electric Power Syst Res 1995;35:51–8.

[19] Rocha Reis AJ, Alves da Silva AP. Feature extraction via multiresolution analysis for short-term load forecasting. IEEE Trans Power Syst 2005;20:189–98.

[20] Amjady N. Introduction to intelligent systems. Semnan, Iran: Semnan University Press; 2002.

[21] Hagan MT, Mehnaj MB. Training feed forward networks with Marquardt algorithm. IEEE Trans Neural Networks 1994;5:989–93.

[22] Damousis IG, Bakirtzis AG, Dokopoulos PS. A solution to the unit-commitment problem using integer-coded genetic algorithm. IEEE Trans Power Syst 2004;19:1165–72.

[23] Michalewicz Z. Genetic algorithms + data structures = evolution programs. New York/USAYork, USA: Springer-Verlag; 1996.

[24] <http://neuron.tuke.sk/competition/>.

[25] <http://www.igmc.ir/dis/>.