ISSN 2250-3153

Prediction of Daily Peak Load for Next Month by a New Hybrid Algorithm

Vahid Behravesh[1], S.M.R. Farshchi[2]

[*] Department of Electrical Engineering, Islamic Azad University, Bardaskan Branch, Bardaskan, Iran
[**] Department of Electrical Engineering, Islamic Azad University, Bardaskan Branch, Bardaskan, Iran

Abstract- Prediction of daily peak load for next month is an important type of Medium Term Load Forecast (MTLF) for electrical power systems. In this study, a new MTLF strategy is proposed, which is composed of a new two stage feature selection technique and hybrid forecast engine. new nonlinear feature selection algorithm composed of Modified Relief (MR) and Mutual Information (MI) techniques is proposed for the MTLF and proposed forecast engine has MLP structure and a new hybrid training mechanism composed of LM learning algorithm and Differential Evolution (DE). Two test cases are considered to evaluate the efficiency of the proposed MTLF strategy: EUNITE competition data and Iran's power system.

Index Terms- Medium Term Load Forecast (MTLF), Modified Relief (MR), Mutual Information (MI), Differential Evolution (DE), Long Term Load Forecasting (LTLF)

I. Introduction

Electrical power systems Introduction need to operate at high efficiency in deregulated electricity markets wherein the participating companies such as electricity generators and retailers have to compete in order to maximize their profits. A forecast that exceeds the actual load may lead to extra power being generated and therefore may result in excessive investment in electric plant that is not fully utilized. On the other hand, a forecast that is too low may lead to some revenue loss from sales to neighboring utilities. Hence, for economically efficient operation and planning of power systems, load forecasting should be accomplished over a broad spectrum of time intervals. In general, the load forecast of electric power systems can be categorized into short term, medium term and long term forecasts. Short Term Load Forecasting (STLF) represents a great saving potential for economic and secure operation of power systems [1]. Medium Term Load Forecasting (MTLF) provides important information for power system planning and operations . Long Term Load Forecasting (LTLF) is useful for planning operations and investment planning.

As STLF is important for the daily operation of generation and distribution facilities, it has been widely studied in the last decades [2]. On the other hand, only a few works can be found in the literature about mid-term demand forecasting. A brief review of the previous MTLF research works can be found in our previous study. This study focuses on the daily peak load prediction of power systems for up to one month ahead. This kind of MTLF is important for electric power utilities as it is used in e.g., maintenance scheduling, management of limited energy units, hydrothermal coordination, adequacy evaluation, security assessment, development of cost efficient fuel purchasing strategies and advantageously negotiates contracts with other companies and reduces financial risks. Moreover, the economic impact of this load forecast is significant and more pronounced in a deregulated energy market. In competitive markets like California, where energy is traded, the accurate MTLF can provide an advantage in negotiations and assist in the development of bilateral contracts. Prediction of daily peak loads can be also useful for congestion management of electricity markets. On the other hand, daily peak load is a complex signal due to e.g., its nonlinear, non-stationary and volatile behavior including high frequency components. Besides, lack of sufficient data usually further complicates this kind of MTLF. So, there is an essential need for more accurate MTLF methods.

It is noted that some previous study considered daily peak load forecasting as a part of STLF. However, their forecast horizon is usually limited to one step ahead, i.e., the next day (Amjady, 2001), or at most one week ahead (Yalcinoz and Eminoglu, 2005). We consider daily peak load prediction for one month ahead, a MTLF problem also mentioned in (Chen *et al*., 2001; Doveh *et al*., 1999). Moreover, 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 (Amjady, 2001) or a limited number of lagged daily peak loads (e.g., 5 daysago) (Yalcinoz and Eminoglu, 2005). 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.

In this study, a well-organized feature selection technique is proposed, which can significantly enhance the accuracy and stability of the MTLF. In spite of previous linear feature selection techniques, such as principle component analysis and correlation analysis, it can consider nonlinearities of the daily peak load signal. The proposed feature selection technique can evaluate both relevancy and redundancy of candidate inputs and select the most informative subset among them. The selected features by the proposed technique are fed to the proposed forecast engine owning MLP structure and a new hybrid training mechanism composed of LM learning algorithm and Differential Evolution (DE). The proposed hybrid forecast engine can benefit from both good convergence behavior and high exploration capability.

The remaining parts of the study are organized as follows. In the 2 section, complexities of daily peak load signal and its relevant inputs are described. Then our proposed data model is presented. In the 3 section, at first a candidate set of inputs is constructed based on the data model. By removing irrelevant and redundant candidate inputs among the set by a new two stage feature selection algorithm, a subset of the most effective inputs is obtained and considered for the forecast engine. In the 14 section, the proposed forecast engine with a new hybrid training mechanism is introduced. Obtained numerical results from the proposed MTLF strategy, including the two stage feature selection algorithm and hybrid forecast engine, for EUNITE and Iran's power system are presented and discussed in section 5. Section 6 concludes the study.

The problem description: MTLF strategies and methods may be classified into two broad categories, namely, conditional models and autonomous models. In the first category, economic analysis plus management, planning and forecasting of energy demand and energy policies are the main focus of researchers. These forecasts are used to develop medium term management strategies and minor infrastructure adjustments. Conditional models attempt to relate the electricity demand growth to the other variables, namely economic indicators. The second category (autonomous models) attempt to relate future growth of electricity demand of a system based on its past growth. The methods in this category use past loads and weather information to forecast future electricity demand. 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 (Tsekouras *et al*., 2006). However, in the daily peak load forecasting for the next month, these indicators are not effective, since the forecast horizon is too short to observe the effect of the indicators. Besides, as described in) prediction of economic indicators is itself a complex task.

The experience of many utilities indicates that the weather elements, such as temperature, humidity and wind, in a decreasing order of importance, influence electricity demand. However, use of these variables in the MTLF depends on availability of their data. So, we will consider weather parameters in the candidate input set of the MTLF as much as the data of these parameters are available.

Finally, some MTLF works considered a few calendar indicators as the inputs of their forecast methods. However, for more input features, NNs usually require more training samples to extract input/output functional relationships (Amjady, 2001), while we have limitation on the training set for the MTLF problem. So, a minimum set of calendar indicators containing a daily and a seasonal indicator are included in the candidate input set of the MTLF in this research work. The Daily Calendar Indicator (DCI) is defined as follows:

$$\mathrm{DCI} = \begin{cases} -1 \text{forweekendsandpublicholidays} \\ 0 \text{forworkingdaybeforeaweekendorpublicholiday} \\ 1 \text{forworkdingdays} \end{cases} \quad (1)$$

Three types of days are discriminated by DCI. Weekends and public holidays usually have lower load levels than working days. Load level of working day before a weekend or public holiday is between these two extreme cases. It is noted that DCI = 1 indicates working days other than those which are before a weekend or public holiday (DCI = 0). The Seasonal Calendar Indicator (SCI) is as follows:

$$\mathrm{SCI} = \begin{cases} -1 \text{Wint er Season} \\ 0 \text{Fall and Spring Seasons} \\ 1 \text{Summer Season} \end{cases} \quad (2)$$

SCI discriminates cold (winter), middle (fall and spring) and warm (summer) seasons.

By aggregating the above explanations, we reach a multi-variable model for the MTLF problem including the historical information of daily peak load, weatherconditions and calendar indicators. This multi-variable model is considered for the next section.

The proposed feature selection: A. Construction of the Candidate Set of Inputs As described in the previous section, our multivariable model for the MTLF problem includes lagged values of daily peak load signal (as the auto-regression part), lagged and forecast values of weather parameters such as temperature (provided that their data is available) and calendar indicators including DCI and SCI. The weather parameters and calendar indicators constitute the cross-regression part or exogenous variables for the MTLF process. Considering such a large set of candidate inputs enables us to employ the maximum information content of the available data such that no likely informative feature is lost. For instance, suppose that only data of average daily temperature among weather parameters is available. Average daily temperature is usually recorded by weather bureaus. So, if only two months historical data are considered, we reach the following set of candidate inputs for the prediction of daily peak load for each forecast day:

{L (d-1), L (d-2), …, L(d-60),
T (d), T (d-1), T(d-2), …, (3)
T (d-60), DCI, SCI}

where, L(d-1), L(d-2), …, L(d-60) indicates lagged loads up to two months ago; T(d) represents predicted average daily temperature for the next day (forecast day); T(d-1), T(d-2), …, T(d-60) are lagged temperatures up to two months ago; DCI, SCI are two calendar indicators defined in (1) and (2). So, we reach a large set including 60+1+60+2 = 123 candidate inputs. However, this large set of inputs is not directly applicable to any forecast engine. So, the set of candidate inputs should be refined by the feature selection technique such that the most informative features are selected and the other unimportant candidates are filtered out.

Feature selection is usually a key issue for the success of a forecast process. For MTLF, where many candidate inputs are involved, the importance of an efficient feature selection technique is pronounced. However, selecting the best set of input features for MTLF in the previous works has been carried out in a discretionary way, mainly based on trial-and-error procedures or at most correlation analysis. Correlation analysis is a linear statistical analysis method. The other feature selection techniques used for load forecast (including both STLF and

MTLF) such as Principal Component Analysis (PCA) and sensitivity analysis are also linear analysis methods (Shahidehpour *et al*., 2002; Saini, 2008). Moreover, these feature selection techniques can only evaluate relevancy between target variable (here, daily peak load of the next day) and candidate inputs. In this way, redundant inputs may be selected which can complicate the training process and degrades the discrimination capability of the forecast engine.

On the other hand, considering the fact that daily peak load is generally a nonlinear mapping function of its input variables, the previous linear feature selection techniques may not correctly rank the candidate set of input features and select the best subset among them. For instance, variation of daily peak load signal with temperature for Iran's power system in year 1999 is shown in Fig. 1. This figure clearly indicates nonlinear variation of daily peak load with temperature. So, an efficient nonlinear feature selection technique is required for the MTLF. Furthermore, the feature selection technique should be able to evaluate both relevancy and redundancy of the candidate inputs.

In this study, a new nonlinear feature selection algorithm composed of Modified Relief (MR) and Mutual Information (MI) techniques is proposed for the MTLF. The proposed algorithm has two stages. In the first stage, the most relevant features among the candidate set of inputs are selected by the MR technique. Then, in the second stage, the redundant candidates among the selected subset of the first stage are filtered out by the MI technique. Hence, the outcome of the proposed feature selection technique is the most relevant features with minimum redundancy. In the following subsections B and C, the MR and MI techniques (irrelevancy and redundancy filters) are introduced, respectively. Then these two techniques are combined in the form of the proposed two stage feature selection algorithm and applied to the candidate set of inputs of the MTLF in the subsection D.

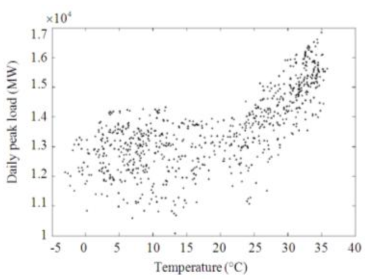

Fig. 1: Variation of daily peak load signal with temperature for Iran's power system in year 1999

The first stage: Irrelevancy filter: The first stage of the proposed feature selection technique is an irrelevancy filter, which determines the subset of relevant features by removing irrelevant ones. The irrelevancy filter is based on a modified form of the Relief algorithm. The Relief algorithm is a non-linear feature selection method that uses instance based learning to assign a relevance weight to each feature. However, the original Relief algorithm can only be used for two class problems, but daily peak load is a continuous variable. In our previous work (Amjady and Daraeepour, 0000), we presented a modified version of the Relief algorithm for price forecast of electricity market that can be used for continuous output variables. Besides, the modified Relief algorithm can perform a more thoroughly search of the feature space to determine relevant candidate inputs compared with the original Relief algorithm. Mathematical details of the modified Relief algorithm can be found in (Amjady and Daraeepour, 0000). In this study, we use from this algorithm as the irrelevancy filter or first stage of the proposed two stage feature selection technique.

The modified Relief algorithm assigns a relevance weight to each candidate input. Then the candidates are ranked based on their weight values. The candidate inputs with the weight value larger than a threshold, denoted by irrelevancy threshold or ITH, are selected by the modified Relief algorithm and the other candidates are filtered out. This threshold is determined by the cross-validation technique, which will be described in the next section.

The second stage: Redundancy filter: The selected features by the irrelevancy filter may have redundant information. A redundant set of candidate inputs not only increases the computation burden of the learning algorithm, but also complicates construction of input/output mapping function for the forecast engine leading to degradation of its accuracy. Especially, considering limited historical data for price forecast, redundant features should be filtered out. The second stage of the proposed feature selection technique includes a redundancy filter, which filters out the redundant candidate inputs among the selected subset by the first stage (the irrelevancy filter). The redundancy filter is based on MI technique, which can measure common information between random variables without making any assumption about the nature of their underlying relationships (Peng *et al*., 2005); bearing in mind, it is essential to also consider nonlinear relations between random variables in the MTLF.

The MI technique, presented in this study, is a modified version of the MI method proposed in our previous work for price forecast of electricity markets (Amjady and Daraeepour, 0000). In the following, at first a brief description of MI criterion is presented and then our previous approach and current one are introduced.

The MI between two random variables $x \in X$ and $y \in Y$ can be interpreted as the information about y that we get by studying x, or vice versa. Therefore, MI (x, y) is the information found commonly in two random variables x and y, which is defined as follows (Peng *et al*., 2005):

$$\mathrm{MI}(x,y) = \int_{x\in X}\int_{y\in Y} P(x,y)\log_2\left(\frac{p(x,y)}{P(x)P(y)}\right)dydx \qquad (4)$$

MI technique suffers from the curse of dimensionality for the MTLF problem. MI requires individual probability distribution of candidate inputs (e.g., P(x) and P(y)) and their joint probability distribution (P(x,y)). In the MTLF problem, candidate inputs (except DCI and SCI defined in (1) and (2), respectively) are continuous variables. Some of these candidates are considered relevant and selected by the irrelevancy filter (modified Relief) at the first stage. So, the MI based redundancy filter involves continuous candidate inputs. As described in the previous section, available history of the candidate inputs for

ISSN 2250-3153

MTLF is limited. Constructing individual and joint probability distributions for the continuous candidate inputs with limited data is a complex task and suffers from low accuracy (Amjady and Daraeepour, 0000). Besides, MI also requires numerical integration of these probability distributions, as shown in (4), which intensifies the computational complexity of the MI based redundancy filter. Our solution for this problem is to incorporate data discretization as a preprocessing task. So, the continuous candidate features need to be properly discretized. For discretized random variables, (4) becomes as follows:

$$MI(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2\left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right) \qquad (5)$$

where, x_i, i =1,…, m and y_j, j = 1,…,n are discrete values of random variables x and y. In our previous work for electricity price. we used a binary discretization technique based on median for each variable in MI based feature selection. Half of the values of each variable are more than its median, which are rounded to 1 and the other half are less than it which are rounded to 0. The mathematical details of this MI based feature selection method can be found in (Amjady and Daraeepour, 0000). However, the binary discretization technique may not reflect all properties of the population (feature values). So, a better discretization method is proposed here. Suppose that for each variable x selected by the irrelevancy filter, μx and σx represent its mean value and standard deviation, respectively. Values of x less than $\mu_x - \sigma_x$, between $\mu_x - \sigma_x$ and $\mu_x + \sigma_x$ and greater than $\mu_x + \sigma_x$ are rounded to -1, 0 and 1, respectively. More than three states can be also considered in the discretization of the continuous candidate inputs increasing the accuracy of the quantization process (analog to digital conversion). However, with more states, less historical data can be assigned to each state. So, the construction of the individual and joint probability distributions crequired for the computation of mutual information in (5) becomes more complex and inaccurate. Besides, the threefold discretization technique is also compatible with our discrete candidate inputs, i.e., calendar indicators, which are three state variables.

Considering the threefold discretization for the continuous candidate inputs, an auxiliary variable uij corresponding to each pair of random variables xi, I = 1,2,3 and y_j, j = 1,2,3 is defined as follows:

$$u_{ij} - 3x_i + y_j \qquad (6)$$

Since each of x_i and y_j can have three states {-1,0,+1}, the auxiliary variable u_{ij} can have 9 values, ranging from 4-4, as shown in Table 1. In this table, u_{ij},-4 counts number of data points (out of all samples) in which u_{ij}= -4. u_{ij},-3 up to u_{ij},4 are similarly defined. L indicates number of all historical data points. For instance, for the daily peak load prediction problem with two years historical data, L = 2×365 = 730. So, the nine states of the joint probability P (x_i,y_j) can be easily computed as shown in Table 1. Besides, the three state individual probabilities P (x_i) and P (y_j) can be also determined as follows:

$$P(x_i = -1) = (u_{ij}, -4 + u_{ij}, -3 + u_{ij}, -2)/L \qquad (7)$$

$$P(x_i = 0) = (u_{ij}, -1 + u_{ij}, -0 + u_{ij}, 1)/L \qquad (8)$$

$$P(x_i = 1) = (u_{ij}, -2 + u_{ij}, 3 + u_{ij}, 4)/L \qquad (9)$$

$$P(y_j = -1) = (u_{ij}, -4 + u_{ij} - 1 + u_{ij}, 2)/L \qquad (10)$$

$$P(y_j = 0) = (u_{ij}, -3 + u_{ij}0 + u_{ij}, 3)/L \qquad (11)$$

$$P(y_j = 1) = (u_{ij}, -2 + u_{ij}1 + u_{ij}, 4)/L \qquad (12)$$

Based on the above individual and joint probabilities, (5) can be written as follows:

$$\begin{aligned}
MI(x,y) &= \sum_{i=1}^{3}\sum_{j=1}^{3} P(x_i, y_j)\log_2\left[\frac{P(x_i, y_j)}{P(x_i)\times P\left(y_j\right)}\right] \\
&= P(x_i = -1, y_j = 1)\log_2\left[\frac{P(x_i = -1, y_j = -1)}{P(x_i = -1)\times P(y_j = -1)}\right] \\
&+ P(x_i = -1, y_j = 0)\log_2\left[\frac{P(x_i = -1, y_j = 0)}{P(x_i = -1)\times P(y_j = 0)}\right] \\
&+ P(x_i = -1, y_j = 1)\log_2\left[\frac{P(x_i = -1, y_j = 1)}{P(x_i = -1)\times P(y_j = 1)}\right] \\
&+ P(x_i = 0, y_j = -1)\log_2\left[\frac{P(x_i = 0, y_j = -1)}{P(x_i = 0)\times P(y_j = -1)}\right] \\
&+ P(x_i = 0, y_j = 0)\log_2\left[\frac{P(x_i = 0, y_j = 0)}{P(x_i = 0)\times P(y_j = 0)}\right] \\
&+ P(x_i = 0 y_j = 1)\log_2\left[\frac{P(x_i = 0, y_j = 1)}{P(x_i = 0)\times P(y_j = 1)}\right] \\
&+ P(x_i = 1, y_j = -1)\log_2\left[\frac{P(x_i = 1, y_j = -1)}{P(x_i = 1)\times P(y_j = -1)}\right] \\
&+ P(x_i = 1, y_j = 0)\log_2\left[\frac{P(x_i = 1, y_j = 0)}{P(x_i = 1)\times P(y_j = 0)}\right] \\
&+ P(x_i = 1, y_j = 1)\log_2\left[\frac{P(x_i = 1, y_j = 1)}{P(x_i = 1)\times P(y_j = 1)}\right] \qquad (13)
\end{aligned}$$

The individual probabilities of (13) are computed according to (7)-(12) and its joint probabilities are calculated based on Table 1. This formulation of MI technique can be implemented with low computation burden. By means of (13), we can compute MI between pairs of candidate inputs, which have been selected at the first stage by the modified Relief algorithm. Higher MI value of two variables means more common information content between them. If MI between two candidate inputs is larger than a threshold, denoted by redundancy threshold or RTH, these two candidates are considered redundant features.

Table 1: Representation of the auxiliary variable uij in terms of discretized candidate inputs xi and yj

y_j, x_i	$x_i = -1$	$x_i = 0$	$x_i = 1$
$y_j = 1$	$P(x_i = -1, y_j = -1) = u_{ij},-4 /L$	$P(x_i = 0, y_j = -1) = u_{ij},-1 /L$	$P(x_i = 1, y_j = -1) = u_{ij},2 /L$
$y_j = 0$	$P(x_i = -1, y_j = 0) = u_{ij},-3 /L$	$P(x_i = 0, y_j = 0) = u_{ij},0 /L$	$P(x_i = 1, y_j = 0) = u_{ij},3 /L$
$y_j = 1$	$P(x_i = -1, y_j = 1) = u_{ij},-2 /L$	$P(x_i = 0, y_j = 1) = u_{ij},1 /L$	$P(x_i = 1, y_j = 1) = u_{ij},4 /L$

Between two redundant candidates, the less relevant one with lower relevance weight value, obtained from the first stage (modified Relief algorithm), is eliminated and the more relevant one with higher weight value is retained. This process is repeated until no redundant candidate is found. In this way, the MI based redundancy filter can filter out redundant features among the selected set by the modified Relief algorithm. The threshold of the redundancy filter RTH is determined by the crossvalidation technique like the threshold of theirrelevancy filter ITH.

Combination of the irrelevancy and redundancy filters: The proposed two stage feature selection technique, composed of the irrelevancy and redundancy filters, can be summarized as the following step by step algorithm named preprocessing algorithm:

Step 1: Construct the candidate set of inputs as described in the subsection A, such as the set of candidate inputs shown in (3). All candidate features (except the calendar indicators) are linearly normalized to eliminate the effect of different ranges of the candidates such as different ranges of load and temperature variables

Step 2: The irrelevancy filter assigns a relevance weight value to each candidate input. The candidate inputs with the relevance weight larger than a threshold ITH are selected by the modified Relief algorithm and the other candidates are filtered out

Step 3: The selected subset of candidate inputs by the irrelevancy filter is given to the second stage filter. In the redundancy filter, mutual information between each pair of candidate inputs among the subset is calculated based on (13). If MI between two candidate inputs is larger than a threshold RTH, these two candidates are considered redundant features. Between two redundant candidates, the less relevant one with lower relevance weight is filtered out and the more relevant one with higher weight is retained. This process is repeated until no redundant candidate is found

II. MATERIALS AND METHODS

RESULTS

Two test cases are considered to evaluate the efficiency of the proposed MTLF strategy: EUNITE competition data and Iran's power system. In 2001, EUNITE network organized a competition aiming at mid-term load forecasting, i.e., predicting daily maximum load of the next 31 days. 56 competitors participated in the competition where the authors of proposed the winning entry. Their MTLF method was based on the Support Vector Machine (SVM) and can reach to minimum MAPE (Mean Absolute Percentage Error) among all competitors. In the EUNITE test case, forecast horizon was January 1999. Two years load data (from 1997-1998), four years average daily temperature data (from 1995-1998) and dates of holidays (from 1997-1999) have been provided by the organizer of the competition (Chen *et al*., 2001). EUNITE data can be obtained from their website (EUNITE data Website: http://www.neuron.tuke.sk/competition/).

Our second test case is power system of Iran, a fast developing country. We considered load data of 1997 and 1998 as the training set and those of 1999 as the testing set, like the EUNITE test case. 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 (Iranian Grid Management Company Website: http://www.igmc.ir/dis).

Although the data of 1999 may not seem up to date data, for the sake of a fair comparison, we should use the same data in order to compare our method with the other MTLF techniques. Moreover, we obtained similar MTLF results for more recent years of Iran's power system test case. Indeed, if the proposed MTLF method can correctly learn the input/output mapping function of daily peak load signal and predict its next values, then it can repeat this process for the other time periods with slight variations in the forecast accuracy.

For the first test case, i.e., EUNITE competition data, considering the availability of only load and temperature data, the candidate set of inputs is as shown in (3) including 123 candidates. This candidate set is refined by the proposed two stage feature selection technique based on the preprocessing algorithm. The obtained results from the first and second stages are shown in Table 2 and 3, respectively. In Table 2, the selected features plus their rank and normalized relevance weights are shown. The features of Table 2 are selected by the irrelevancy filter (the modified Relief algorithm) eliminating the irrelevant features. The relevance weights, obtained from the modified Relief algorithm, are normalized with respect to their maximum to yield the normalized relevance weights. The selected features by the second stage (the MI based redundancy filter) plus their rank are shown in Table 3. These features are selected after filtering out the redundant candidate inputs among the set of candidates of Table 2. As seen, the filtering ratio of the first and second stages is 123/42 = 2.93 and 42/24 = 1.75, respectively, resulting in2.93×1.75 = 123/24 = 5.13 filtering ratio of the proposed two stage feature selection technique. So, the proposed technique can effectively eliminate the irrelevant and redundant candidate inputs and select a minimum set of the most informative features.

The selected features in Table 3 are applied as inputs to the hybrid forecast engine.

Table 2: Obtained results from the first stage of the two stage featureselection technique for the EUNITE test case

Selected features	Rank	Normalized relevance weight
SCI	1	1.0000
DCI	2	0.6943
L(d-1)	3	0.5565

L(d-7)	4	0.5140
L(d-2)	5	0.5050
L(d-14)	6	0.4826
T(d-2)	7	0.4690
T(d-3)	8	0.4566
T(d-1)	9	0.4483
L(d-3)	10	0.4470
T(d)	11	0.4376
L(d-5)	12	0.4322
L(d-21)	13	0.4307
T(d-4)	14	0.4258
L(d-4)	15	0.4250
L(d-6)	16	0.4187
L(d-8)	17	0.4127
L(d-61)	18	0.3945
L(d-13)	19	0.3922
T(d-5)	20	0.3910
L(d-15)	21	0.3879
L(d-28)	22	0.3857
L(d-10)	23	0.3856
L(d-60)	24	0.3825
L(d-56)	25	0.3817
L(d-59)	26	0.3816
L(d-11)	27	0.3779
L(d-58)	28	0.3765
L(d-9)	29	0.3748
T(d-6)	30	0.3746
L(d-57)	31	0.3725
L(d-49)	32	0.3688
L(d-12)	33	0.3682
L(d-54)	34	0.3660
T(d-7)	35	0.3634
L(d-55)	36	0.3633
T(d-9)	37	0.3560
L(d-35)	38	0.3557
L(d-52)	39	0.3542
T(d-8)	40	0.3527
L(d-16)	41	0.3518
L(d-48)	42	0.3502

This table shows the calendar indicators, including SCI and DCI, have the highest ranks among the candidate inputs, indicating information content of these features.

Moreover, the effect of short-run trend (e.g., selection of L (d-1)) and weekly periodicity (such as selection of L (d-7) and L(d-14)) are seen. Finally, some temperature features are also among the selected candidates of Table 3 indicating the information value of these features for the MTLF problem.

To better illustrate the effectiveness of the proposed feature selection technique for the MTLF, obtained prediction errors for the first test case, i.e., EUNITE competition data, are shown in Table 4. The results of Table 4 are in terms of MAPE, since it was the error metric of the competition. This error metric is defined as follows:

$$\mathrm{MAPE}(\%) = \frac{1}{N}\sum_{i=1}^{N}\frac{|\,L_{iACT} - L_{iFOR}\,|}{L_{iACT}} \times 100 \qquad (17)$$

where, LiACT and LiFOR represent actual and forecasted daily peak load for ith day; N indicates number of days of the forecast horizon. For the EUNITE competition data, forecast horizon was January 1999 and so N = 31. In Table 4, obtained prediction errors, in terms of MAPE, with different feature selection techniques are shown.

Principle Component Analysis (PCA) and Correlation Analysis (CA) are well-known linear feature selection techniques used for feature selection of electricity load forecast (Espinoza *et al.*, 2005; Reis *et al.*, 2005), prediction of power system imbalance volume (Garcia and Kirschen, 2006) and electricity price forecast (Li *et al.*, 2007; Pindoriya *et al.*, 2008; Zareipour *et al.*, 2006). PCA and CA are single stage feature selection methods. CA-CA, shown in the fourth column of Table 4, is a two stage feature selection method including correlation analysis for both irrelevancy and redundancy filters. We proposed this method in our previous works for STLF (Our Energy Paper, 0000) and MTLF (Amjady and Keynia, 0000). The obtained MAPE values from nonlinear single stage feature selection techniques (including only the irrelevancy filter) based on the Mutual Information (MI) and Modified Relief (MR) algorithms are shown in the fifth and sixth columns of Table 4, respectively. In the last column of this table, obtained MAPE value from the proposed two stage feature selection technique, combining the MR and MI algorithms, is represented (MR-MI). Characteristics of different feature selection techniques are briefly shown in the second row of Table 4.

For the sake of a fair comparison, all feature selection techniques of Table 4 have the same hybrid forecast engine, training data and test data. As seen, the proposed MR-MI method leads to the least prediction error for the MTLF, which reveals superiority of this method with respect to the other feature selection techniques of Table 4.

Also, to illustrate the effectiveness of the proposed hybrid forecast engine, obtained results in three cases are shown in Table 5. In the first two cases, shown in the second and third columns of this table, a single DE and a single LM are used as the training mechanism of the NN forecast engine without the iterative training algorithm. In the last Column of Table 5, the obtained results from the proposed hybrid forecast engine with the iterative training algorithm are shown.

Table 3: Obtained results from the second stage of the two stage feature selection technique for the EUNITE test case

Selected features	Rank
SCI	1
DCI	2
L(d-1)	3
L(d-7)	4
L(d-14)	5
T(d-2)	6
L(d-3)	7
T(d) 8	8
L(d-5)	9
L(d-21)	10
T(d-4)	11
L(d-61)	12
L(d-28)	13

L(d-10)	14
L(d-56)	15
L(d-59)	16
T(d-6)	17
L(d-49)	18
L(d-12)	19
L(d-54)	20
T(d-9)	21
L(d-35)	22
L(d-52)	23
L(d-16)	24

Table 4: Obtained prediction errors for the EUNITE competition data with different feature selection techniques

Feature selection technique	Characteristics	MAPE (%)
PCA	Linear single stage	2.09
CA	Linear single stage	1.98
CA-CA	Linear two stage	1.61
MI	Linear single stage	1.57
MR	Linear single stage	1.54
MR-MI	Linear two stage	1.40

Table 5: Obtained results for the EUNITE competition data with three different forecast engines

Forecast engine	MAPE (%)	PAPE (%)
Single DE	2.38	8.98
Single LM	1.90	4.15
Iterative training algorithm	1.40	3.52

Table 6: Obtained prediction errors for the EUNITE competition data with different MTLF methods

MTLF method	MAPE (%)
The winning entry of the EUNITE competition (Chen *et al*., 2001)	1.95
The best method in our previous study (Amjady, 2001)	1.60
Proposed method in this paper	1.40

For the sake of a fair comparison, all three forecast engines have the same MR-MI feature selection technique, training data and test data. In addition to MAP values, Peak Absolute Percentage Error (PAPE) values are also shown in Table 5, which give an insight about the stability of the predictions:

$$\mathrm{PAPE}(\%) = \underset{1 \leq i \leq N}{\mathrm{Max}} \left(\frac{L_{iACT} - L_{iFOR}}{L_{iACT}} \right) \tag{18}$$

As seen, the hybrid forecast engine with the iterative training algorithm has both lower MAPE and PAPE values than the two other forecast engines of Table 5.

In Table 6, MTLF error of the winning entry of the EUNITE competition (Chen *et al*., 2001), our previous work (Amjady and Keynia, 0000) and the proposed method in this study (MR-MI + hybrid forecast engine) are presented. In our previous work, we presented different MTLF models that their MAPE values for the EUNITE competition data vary in the range of 2.61-1.60%.

Among these models, the MAPE value of the best one (1.60%) is shown in the third column of Table 6. As seen, the proposed method of this study has better MTLF accuracy than both the winning entry of the EUNITE competition and our previous work. PAPE is not used in Table 6, since its value for the winning entry of the EUNITE competition is not available. The three adjustable parameters of the proposed MTLF method, including ITH, RTH and NH, are fine-tuned for the EUNITE competition data by the cross-validation technique as 0.35, 0.55 and 19, respectively.

In order to also give a graphical view about the MTLF accuracy of the proposed method, its obtained results for the EUNITE competition data are shown in Fig. 3. As seen, forecasted load curve is very close to actual load curve and MTLF error values are very low.

For the second test case, i.e., Iran's power system, the candidate set of inputs is as shown in (3), like the EUNITE competition data, since only the daily peak load data, daily average temperature data and calendar indicators are available for this test case as well. Similar feature selection results (like Table 2 and 3) are obtained for the second test case. So, for the sake of conciseness, only MTLF errors for this test case are presented in Table 7. The obtained results from the best MTLF model among those presented in our previous work (Amjady and Keynia, 0000) and the proposed method in this study for Iran's power system are shown in the second and third columns of Table 7, respectively.

In Table 7, MAPE and PAPE values for 12 months of 1999 are presented. The obtained values for the three adjustable parameters of the proposed MTLF method by the cross-validation technique, including ITH, RTH and NH, for each month are shown in the last three columns of Table 7, respectively. As seen, the proposed method in this study overall has both better MAPE and PAPE values than the best method in our previous work. For instance, average MAPE and PAPE values of the proposed MTLF method are 1.51 and 5.46%, respectively (indicated in the last row of Table 7) while those of the best method in our previous work are 1.76 and 5.88%, respectively.

Table 7: Obtained MTLF errors for Iran's power system

	The best method in our previous work (Amjady, 2001)		Proposed method in this paper				
Month	MAPE (%)	(%)	PAPE (%)	MAPE (%)	PAPE TH	RTH	NH
1	1.44	6.65	1.4	4.18	0.4	0.45	18
2	2.15	7.45	1.29	3.82	0.42	0.49	18
3	1.49	5.58	1.66	7.00	0.34	0.49	21
4	1.92	7.12	1.29	3.01	0.58	0.50	20
5	1.77	5.11	1.48	4.25	0.50	0.40	18
6	1.63	3.84	1.71	8.37	0.42	0.52	21
7	1.68	6.49	1.30	6.01	0.48	0.43	20
8	1.41	3.62	1.57	9.94	0.44	0.4	18
9	1.59	5.19	1.52	4.68	0.36	0.45	17
10	2.36	7.84	1.84	4.75	0.35	0.42	20

ISSN 2250-3153

11	1.42	3.93	1.56	4.06	0.47	0.59	18
12	2.33	7.71	1.57	5.47	0.51	0.71	18
Average	1.76	5.88	1.51	5.46	-	-	-

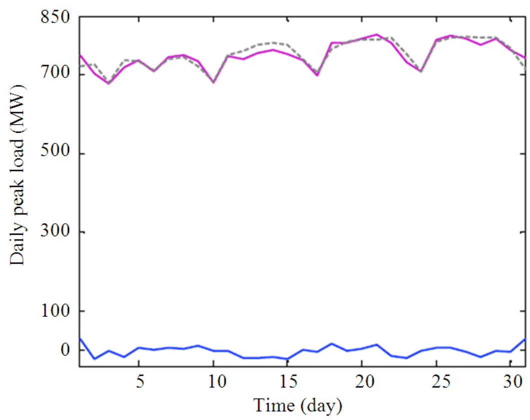

Fig. 3: Actual daily peak load (solid line-purple), forecasted daily peak load (dashed line-gray) and forecast error (blue) of the proposed method for EUNITE competition data

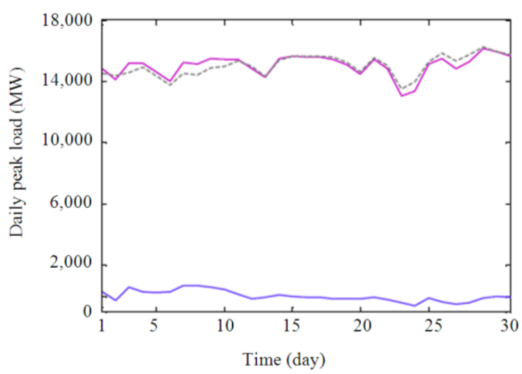

Fig. 4: Actual daily peak load (solid line-purple), forecasted daily peak load (dashed line-gray) and forecast error (blue) of the proposed method for 10th month of 1999 for Iran's power system

Results of the proposed MTLF method for the 10th month of 1999 with the highest MAPE value in Table 7 are graphically shown in Fig. 4. As seen, even for the worst test month of Table 7, the proposed method has good accuracy and the forecast curve acceptably follows the actual curve.

Total setup time of the proposed method including execution of the two stage feature selection algorithm, training phase of the hybrid forecast engine and execution of the cross-validation technique for our test cases is about 50 m on a simple hardware set including a Pentium P4 3.2 GHz personal computer with 1 GB RAM. This setup time is completely reasonable within a month-ahead decision making framework. In some forecast problems, such as prediction of electricity 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.

III. DISCUSSION

CONCLUSION

In this study, a new MTLF strategy is proposed, which is composed of a new two stage feature selection technique and hybrid forecast engine. At first, a multi-variable data model is constructed for the MTLF problem including, for instance, the historical information of daily peak load, weather conditions and calendar indicators. This data model is refined by the proposed two stage feature selection technique. The proposed technique has two successive filters to remove irrelevant and redundant candidate inputs among the large set of candidate features of the MTLF. So, the technique selects a minimum set of the most informative features for the forecast process. The hybrid forecast engine has a new iterative training algorithm to combine LM and DE training mechanisms. The proposed forecast engine can benefit from both high convergence rate of LM and good global search ability of DE. The obtained results from extensive experiments confirm the validity of the developed MTLF approach. The research work is under way in order to develop better feature selection techniques, forecast engines and cross-validation methods.

REFERENCES

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

[2] Energy Policy, 0000. A Fuzzy Inference Model for Short-Term Load Forecasting

[3] Our Energy Paper, 0000.

[4] Amjady, N. and F. Keynia, 0000. Mid-Term Load Forecasting of Power Systems by a New Hybrid Prediction Strategy,

[5] Iran Energy Policy, 0000. A simulated-based neural network algorithm for forecasting electrical energy consumption

[6] Willis, H.L., 1997. Power distribution planning reference book. New York: Marcel Decker Inc.,

[7] Energy Policy Paper, 0000. Forecasting of Turkey's net electricity energy consumption on sectoral bases,

[8] Yalcinoz, T. and U. Eminoglu, 2005. Short term and medium term power distribution load forecasting by neural networks. Energy Conversion Manage., 46: 1393-1405.

[9] Chen, B.J., M.W. Chang and C.J. Lin, 2001. Load forecasting using support vector machines: A studyon EUNITE competition. IEEE Trans. Power Syst., 19: 1821-1830.

[10] Doveh, E., P. Feigin, D. Greig and L. Hyams, 1999. Experience with FNN models for medium term power demand predictions. IEEE Trans. Power Syst., 14: 538-546.

[11] Tsekouras, G.J., N.D. Hatziargyriou and E.N. Dialynas, 2006. An optimized adaptive neural network for annual midterm energy forecasting. IEEE Trans. Power Syst., 21: 385-391.

[12] Ghiassi, M., D.K. Zimbra and H. Saidane, 2006. Medium term system load forecasting with a dynamic artificial neural network model. Electric Power Syst. Res., 76: 302-316.

[13] Gonzalez-Romera, E., M. A. Jaramillo-Moran and D. Carmona-Fernandez, 2006. Monthly electric energy demand forecasting based on trend extraction. IEEE Trans. Power Syst., 21: 1946-1953.

[14] Mirasgedis, S., Y. Sarafidis, E. Georgopoulou, D.P. Lalas and M. Moschovits *et al*., 2006. Models for mid-term electricity demand forecasting incorporating weather influences. Energy, 31: 208-227.

[15] Shahidehpour, M., H. Yamin and Z. Li, 2002. Market Operations in Electric Power Systems, A John Wiley Sons, Inc., Publication

[16] Saini, L.M., 2008. Peak load forecasting using bayesian regularization, Resilient and adaptive back propagation learning based artificial neural networks. Electric Power Syst. Res., 78: 1302-1310.

[17] Amjady, N. and A. Daraeepour, 0000. Design of Input Vector for Day-Ahead Price Forecasting of Electricity Markets"

[18] Peng, H., F. Long and C. Ding, 2005. Feature selection based on mutual information: Criteria of max-dependency, max-relevance and minredundancy. IEEE Trans. Pattern Anal. Machine Int., 27: 1226-1238.

[19] Hagan, M.T. and M.B. Menhaj, 1994. Training feed forward networks with marquardt algorithm. IEEE Trans. Neural Netw., 5: 989-993.

[20] Storn, R., 1999. System design by constraint adaptation and differential evolution. IEEE Trans Evol Comput 3: 22-34

[21] Engelbrecht, A.P., 2007. Computational Intelligence: An Introduction. 2rd Edn., John Wiley Sons,

[22] EUNITE Data, 0000. http://www.neuron.tuke.sk/competition/

[23] Iranian Grid Management Company, 0000. http://www.igmc.ir/dis/

[24] Espinoza, M., C. Joye, R. Belmans and D.B. Moor, 2005. Short-term load forecasting, profile identification and customer segmentation: A Methodology Based Periodic Time Series. IEEE Trans. Power Syst., 20: 1622-1630.

[25] Reis, R.A.J. and A.A.P. Silva, 2005. Feature extraction via multiresolution analysis for short-term load forecasting. IEEE Trans. Power Syst., 20: 189-198.

[26] Garcia, M.P. and D.S. Kirschen, 2006. Forecasting System imbalance volumes in competitive electricity markets. IEEE Trans. Power Syst., 21: 240-248.

[27] Li, G., C.C. Liu, C. Mattson and J. Lawarree, 2007. Day-ahead electricity price forecasting in a grid environment. IEEE Trans. Power Syst., 22: 266-274.

[28] Pindoriya, N.M., S.N. Singh and S.K. Singh, 2008. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans. Power Syst., 23: 1423-1432.

[29] Zareipour, H., C.A. Canizares, K. Bhattacharya and J. Thomson, 2006. Application of publicdomain market information to forecast Ontario's wholesale electricity prices. IEEE Trans. Power Syst., 21: 1707-1717.

AUTHORS

First Author – Vahid Behravesh, Department of Electrical Engineering, Islamic Azad University, Bardaskan Branch, Bardaskan, Iran, Email: vahidbehravesh@yahoo.com

Second Author – S.M.R. Farshchi, Department of Electrical Engineering, Islamic Azad University, Bardaskan Branch, Bardaskan, Iran, Email: shiveex@gmail.com