Neural Comput & Applic (2012) 21:1961–1970
DOI 10.1007/s00521-011-0599-1

ORIGINAL ARTICLE

Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection

Mansour Sheikhan · Najmeh Mohammadi

Received: 29 January 2011 / Accepted: 15 April 2011 / Published online: 1 May 2011

Abstract Due to deregulation of electricity industry, accurate load forecasting and predicting the future electricity demand play an important role in the regional and national power system strategy management. Electricity load forecasting is a challenging task because electric load has complex and nonlinear relationships with several factors. In this paper, two hybrid models are developed for short-term load forecasting (STLF). These models use "ant colony optimization (ACO)" and "combination of genetic algorithm (GA) and ACO (GA-ACO)" for feature selection and multi-layer perceptron (MLP) for hourly load prediction. Weather and climatic conditions, month, season, day of the week, and time of the day are considered as load-influencing factors in this study. Using load time-series of a regional power system, the performance of ACO + MLP and GA-ACO + MLP hybrid models is compared with principal component analysis (PCA) + MLP hybrid model and also with the case of no-feature selection (NFS) when using MLP and radial basis function (RBF) neural models. Experimental results and the performance comparison with similar recent researches in this field show that the proposed GA-ACO + MLP hybrid model performs better in load prediction of 24-h ahead in terms of mean absolute percentage error (MAPE).

Keywords Short-term load forecasting · Feature selection · Ant colony optimization · Genetic algorithm · Neural network

1 Introduction

Load forecasting has always been a very important issue in reliable power systems planning and operation [1, 2]. Specifically, short-term forecasting of daily electricity demand is crucial in unit commitment and maintenance, power interchange and task scheduling of both power generation and distribution facilities. Economically, the accuracy in load forecasting can allow utilities to operate at the least cost which may contribute to significant savings in electric power companies [3]. These forecasts influence many decisions, e.g., scheduling of generating capacity and fuel purchases, and also planning for energy transactions. Because of the dramatic changes in structure of the utility industry due to deregulation, these forecasts have become more important in the recent decade.

The electric load has complex and nonlinear relationships with several factors. These factors are weather and climatic conditions, social behavior, the season, day of the week, and time of the day.

In terms of lead time, load forecasting studies can be classified into four categories [4–8]:

- Long-term forecasting with the lead time of more than 1 year
- Mid-term forecasting with the lead time of 1 week to 1 year

M. Sheikhan (✉) · N. Mohammadi
EE Department, Faculty of Engineering,
Islamic Azad University, South Tehran Branch,
P.O. Box: 11365-4435, Tehran, Iran
e-mail: msheikhn@azad.ac.ir

N. Mohammadi
Research Center of West Tehran Province Power Distribution
Company, Tehran, Iran
e-mail: njm78ir@yahoo.com

- Short-term load forecasting with the lead time of 1–168 h
- Very short-term load forecasting with the lead time shorter than 1 day

Many techniques have been proposed in the recent decades for short-term load forecasting (STLF) such as statistical models [9–11], fuzzy methods [12–14], and machine learning algorithms [15–20].

As examples researches based on statistical models, Christianse [21] and Park et al. [22] have designed exponential smoothing models by Fourier series transformation for electricity load forecasting. To achieve more accurate results in load forecasting, Gelb [23] and Brown [24] have built the load forecasting model that used state space and Kalman filtering technology. Moghram and Rahman [25] have devised a model based on this technology and verified that the proposed model outperformed other forecasting methods. However, these methods always fail to avoid the influence of observation noise in the forecasting. To handle this problem, Douglas et al. [26] have considered the impacts of temperature on the forecasting model. Sadownik and Barbosa [2] have proposed dynamic nonlinear models for load forecasting. The main disadvantage of these methods is that they are time consuming in computation when the number of variables is increased.

As example researches based on using artificial neural networks (ANNs) for STLF, Carpinteiro et al. [27] have proposed a neural model which consists of two self-organizing maps (SOMs) and Cai et al. [28] have used distributed adaptive resonance theory (ART) and hyper-spherical ARTMAP (HS-ARTMAP) neural networks. It should be noted that HS-ARTMAP is a hybrid of a radial basis function (RBF)-network-like module, which uses hyper-sphere basis function, and an ART-like module.

Also, different types of neural networks and other theories have been combined with for taking advantages of each type [15, 16, 29–31]. For example, in [15] fuzzy theory as an alternative approach has been integrated with neural network. Liao and Tsao [29] have proposed an integrated evolving fuzzy neural network and simulated annealing (SA) for load forecasting. For this purpose, they have used fuzzy hyper-rectangular composite neural networks (FHRCNNs) for the initial load forecasting. Then, evolutionary programming (EP) and SA have been used to find the optimal solution of the parameters of FHRCNNs. Furthermore, neural networks have been combined with wavelet transform and used for STLF with little influence from noise data [30, 31].

Also, support vector machines (SVMs), that successfully employed to solve nonlinear regression and time-series problems, have been combined with other technologies to their common advantages for STLF [32–34]. For example, Hong [34] has proposed a support vector regression (SVR) model with a hybrid evolutionary algorithm, called chaotic genetic algorithm (CGA), to forecast the electric loads.

Also, some researches have been conducted to provide effective approaches that consider both the time complexity and the various influencing factors of load forecasting. So, feature selection (FS) methods have become important when we are faced with high computational load in STLF. The aim of FS is to determine a minimal feature subset while retaining high accuracy in representing the original features [35]. In this way, Wang and Cao [36] have used mutual information (MI)-based technique to choose the proper input variables of an ANN-based load forecaster. Xiao et al. [19] have employed rough set technique for attribute reduction in a back-propagation (BP)-based neural network. Also, Siwek and Osowski [37] have proposed a combination of neural predictors for 24-h ahead load pattern, in which the values of forecasted load by each predictor have been combined together using principal component analysis (PCA) method, to extract the most important information and reduce the size of vector used in the final stage of prediction. Bayesian technique has also been used for input feature selection in an ANN-based short-term load forecaster [38].

In the recent years, swarm intelligence (SI) provides a tool to solve collective (or distributed) problems without requiring centralized control. As example researches in the field of STLF using SI-based techniques [39–42], Niu et al. [39] have introduced a hybrid optimizing algorithm, called AFSA-TSGM. In this algorithm, Tabu search and genetic mutation (TSGM) operator are combined with artificial fish swarm algorithm (AFSA). This algorithm has been used in the learning of a wavelet neural network (WNN). Also, an improved particle swarm optimization (PSO) has been employed in [40] to optimize the parameters of a least squares SVM (LS-SVM). The hybrid of ant colony optimization (ACO) and SA have been presented in [41] for monthly load prediction. A parameter-wise optimization training process has also been implemented in [42] to achieve an optimal configuration of focused time lagged recurrent neural network (FTLRNN) for STLF.

In ACO algorithm, ants are capable of finding the shortest route between a food source and their nest without the use of visual information. SI techniques, based on the behavior of real ant colonies, have been used to solve optimization problems. The ACO techniques have been successfully applied to a large number of difficult combinatorial problems [43]. This method is particularly attractive for feature selection because there is very limited heuristic to guide the search to optimal minimal subset.

On the other hand, genetic algorithm (GA) is also an optimization technique which is based on natural selection

[44, 45]. GA has been also used in STLF as a component in hybrid forecast method [46]. In this paper, we propose a hybrid model of GA-ACO for feature selection with the hope of taking advantages of two techniques. It is noted that ACO has the advantage of local search and GA has a global perspective. Also, ANN which has the ability to learn complex and nonlinear relationships between load and its influencing factors is used for prediction of short-term load.

The rest of this paper is organized as follow. Section 2 explains the features of power load data. The feature reduction methods which are used in this paper are introduced in Sect. 3. These methods are ACO, hybrid of GA and ACO, and principal component analysis (PCA) which is a traditional feature transform technique. Section 4 presents ANN-based prediction and experimental results are reported in Sect. 5. The paper is concluded in Sect. 6.

2 Features of power load data

The relationship between the electricity load and its influencing factors is complex and nonlinear, making it quite difficult to represent using linear models, or even parametric nonlinear ones. For the short-term horizon, electricity loads are affected by the hour of the day, day of the week, season of the year, weather and climatic conditions, holidays, special events (e.g., strikes), energy price, and electricity network configuration changes.

Various factors that influence the system load behavior can be classified into the following major categories: weather, time, economy, and random disturbances [1, 3].

3 Feature reduction techniques

Feature reduction plays an important role in data mining, pattern recognition, and forecasting problems, especially for large scale data. This topic refers to the study of methods for reducing the number of dimensions describing the data. Its general purpose is to employ fewer features to represent data and reduce the computational cost, without deteriorating discriminative capability. Since feature reduction can bring lots of advantages to learning algorithms, such as avoiding over-fitting, resisting noise, and strengthening prediction performance, it has attracted great attention and many feature reduction algorithms have been developed during the recent years.

Generally, these algorithms can be classified into two broad categories: feature transform (or feature extraction) and feature selection. Feature transform constructs new features by projecting the original feature space to a lower dimensional one. Principal component analysis (PCA) and independent component analysis (ICA) are two widely used feature transform methods [47]. Although feature transform can obtain the least possible dimension, its major drawbacks lie in that its computational overhead is high and the output is hard to be interpreted for users [48].

On the other hand, feature selection (FS) is the process of choosing a subset of the original feature space according to discrimination capability to improve the quality of data. Unlike feature transform, the fewer dimensions obtained by feature selection facilitate exploratory of results in data analysis. Due to this predominance, FS has now been widely applied to many domains, such as load forecasting. There are three kinds of feature selection methods: filter, wrapper, and embedded methods.

Filter approaches are independent of learning algorithm. These approaches mostly include selecting features based on inter-class separability criterion. If the evaluation procedure is tied to the task (e.g., classification) of learning algorithm, the FS algorithm will be of wrapper method. This method searches through the feature subset space using the estimated accuracy from an induction algorithm as a measure of subset suitability. If the FS and learning algorithm are interleaved, then the FS algorithm will be of embedded method.

3.1 Ant colony optimization (ACO) for feature selection

ACO algorithm provides an alternative feature selection tool inspired by the behavior of ants in finding paths from the colony to food [49]. Real ants exhibit strong ability to find the shortest routes from the colony to food using a way of depositing pheromone as they travel. ACO mimics this ant seeking food phenomenon to yield the shortest path (which means the "system" of interests has converged to a single solution).

When a source of food is found, ants lay some pheromone to mark the path. The quantity of the laid pheromone depends upon the distance, quantity, and quality of the food source. While an isolated ant that moves at random detects a laid pheromone, it is very likely that it will decide to follow its path. This ant will itself lay a certain amount of pheromone and hence enforces the pheromone trail of that specific path. Accordingly, the path that has been used by more ants will be more attractive to follow. In other words, the probability with which an ant chooses a path is increased with the number of ants that previously have chosen that path. This process is hence characterized by a positive feedback loop.

For a given classification task, the problem of feature selection can be stated as follows: given the original set, F, of n features, find subset S, which consists of m features ($m < n$), such that the classification accuracy is

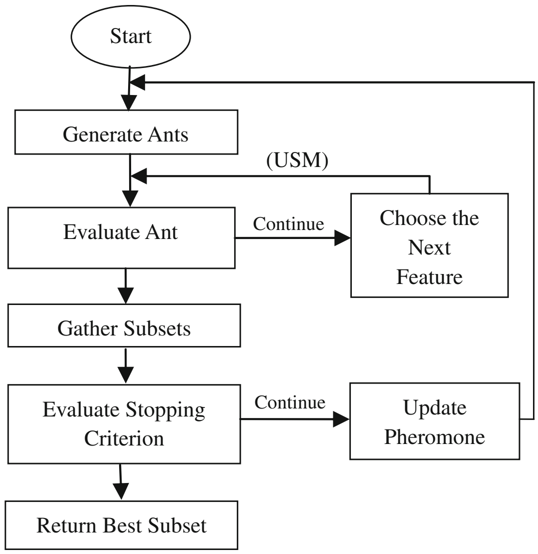

Fig. 1 ACO-based feature selection algorithm

maximized. The feature selection representation exploited by artificial ants includes the following:

- n features that constitute the original set, $F = \{f_1,\ldots,f_n\}$,
- A number of artificial ants, n_a, to search through the feature space,
- τ_i, the intensity of pheromone trail associated with feature f_i, which reflects the previous knowledge about the importance of f_i,
- For each ant j, a list that contains the selected feature subset, $S_j = \{s_1,\ldots,s_m\}$.

In the first iteration, each ant randomly chooses a feature subset of m features. Only the best k subsets, $k < n_a$, are used to update the pheromone trial and influence the feature subsets of the next iteration. In the following iterations, each ant will start with $m - p$ features that are randomly chosen from the previously selected k-best subsets, where p is an integer that ranges between 1 and $m - 1$. So, the features that constitute the best k subsets will have more chance to be present in the subsets of next iteration. However, it will still be possible for each ant to consider other features. For a given ant j, those features are the ones that achieve the best compromise between pheromone trails and local importance with respect to S_j, where S_j is the subset that consists of the features that have already been selected by ant j. The updated selection measure (USM) is used for this purpose which is defined as follows:

$$\mathrm{USM}_i^{S_j}(t) = \begin{cases} \frac{(\tau_i)^\alpha\left(\eta_i^{s_j}\right)^\beta}{\sum_{g\notin S_j}(\tau_i)^\alpha\left(\eta_i^{s_j}\right)^\beta} & \text{if } i \notin S_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\eta_i^{s_j}$ is the local importance of feature f_i given the subset S_j. The parameters α and β control the effect of pheromone trail intensity and local feature importance, respectively. The main steps of the algorithm are as follow [50]:

Step 1: Initialization

- Set the number of ants (n_a).
- Set $\tau_i = \mathrm{cc}$ and $\Delta\tau_i = 0$, $(i = 1,\ldots,n)$, where cc is a constant and $\Delta\tau_i$ is the amount of change of pheromone trial quantity for feature f_i.
- Define the maximum number of iterations.
- Define k, where the k-best subsets will influence the subsets of next iteration.
- Define p, where $m - p$ is the number of features that each ant will start with in the following iterations.

Step 2: If in the first iteration,

- For $j = 1$ to n_a,
  - Randomly assign a subset of m features to S_j.
- Go to step 4.

Step 3: Select the remaining p features for each ant

- For $mm = m - p + 1$ to m,
  - For $j = 1$ to n_a,
    - Given subset S_j, choose feature f_i that maximizes $\mathrm{USM}_i^{S_j}$.
    - $S_j = S_j \cup \{f_i\}$.
- Replace the duplicated subsets, if any, with randomly chosen subsets.

Step 4: Evaluate the selected subset of each ant using a chosen classification algorithm.

- For $j = 1$ to n_a,
  - Estimate the mean squared error (MSE) of the classification results obtained by classifying the features of S_j that is called MSE_j.
- Sort the subsets according to their MSE. Update the minimum MSE (if achieved by any ant in this iteration), and store the corresponding subset of features.

Step 5: Using the feature subsets of the best k ants, update the pheromone trail intensity and initialize the subsets for next iteration.

- For $j = 1$ to k, update the pheromone trails by using the following equations:

$$\Delta\tau_i = \begin{cases} \frac{\max_{g=1:k}\left(\mathrm{MSE}_g\right) - \mathrm{MSE}_j}{\max_{h=1:k}\left(\max_{g=1:k}\left(\mathrm{MSE}_g\right) - \mathrm{MSE}_h\right)} & \text{if } f_i \in S_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\tau_i = \rho \cdot \tau_i + \Delta\tau_i \quad (3)$$

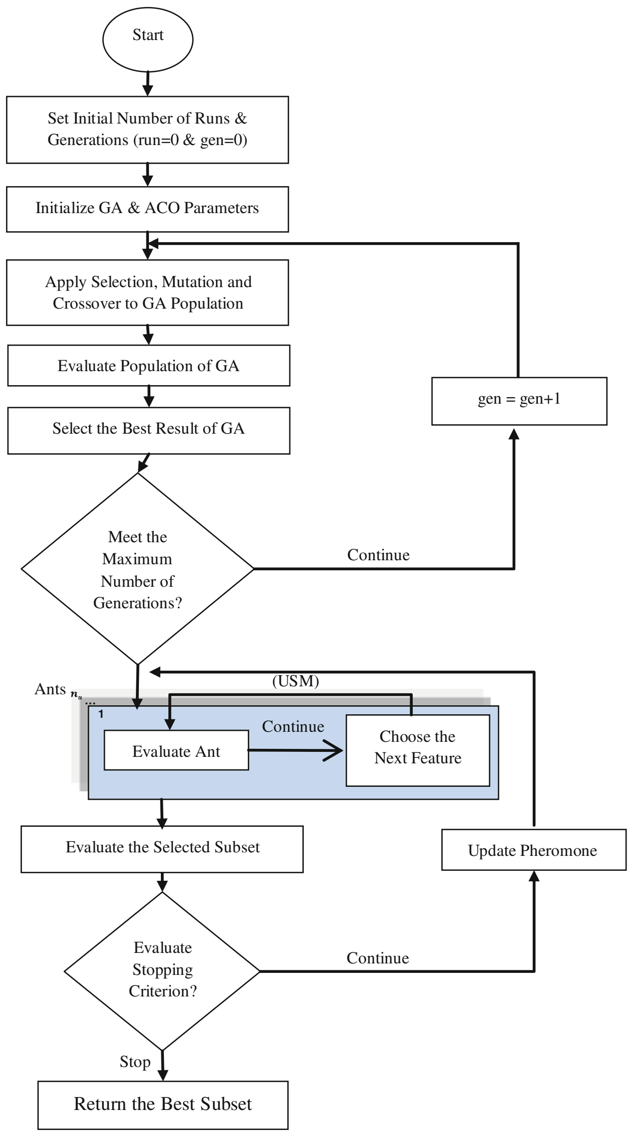

Fig. 2 Proposed GA-ACO feature selection algorithm

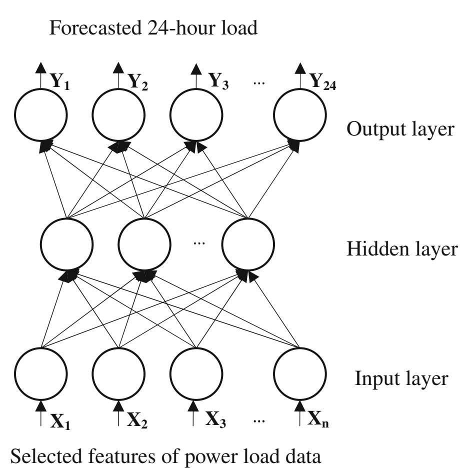

Fig. 3 Layout of MLP network for STLF

Table 1 MLP specifications

Specification	Value/type
Number of hidden-layer neurons	20
Number of output-layer neurons	24
Activation function of hidden-layer nodes	tan-sigmoid
Activation function of output-layer nodes	Linear
Number of training epochs	10,000

where ρ is a constant such that $(1 - \rho)$ represents the evaporation of pheromone trails.

- For $j = 1$ to n_a,
  - From the features of the best k ants, randomly produce $m - p$ feature subset for ant j, to be used in the next iteration, and store it in S_j.

Step 6: If the number of iterations is less than the maximum number of iterations, or the desired MSE has not been achieved, go to step 3.

The overall process of ACO feature selection is shown in Fig. 1.

3.2 Proposed GA-ACO algorithm for feature selection

In this study, the hybrid of GA and ACO is used in such a manner that they complement each other for feature selection in load forecasting. GA and ACO are used to explore the space of all subsets of given feature set. The performance of selected feature subsets is measured by invoking an evaluation function with the corresponding reduced feature space and measuring the specified classification result. The best feature subset found is then reported as the recommended set of features to be used in the actual design of classification system. The main steps of algorithm are shown in Fig. 2.

The process begins by generating a population in GA and a number of ants. GA generates feature subsets, and the resulting subsets are gathered and then evaluated at the end of iterations. The best subset is selected according to evaluation measures. If an optimal subset has been found or the algorithm has executed a certain number of runs, then the process halts and the best feature subset is reported. If none of these conditions hold, then all ants can update the pheromone according to (3), the best ant deposits additional pheromone on nodes of the best solution. After updating pheromone, the process iterates once more. It should be noted that another hybrid format of GA and ACO is proposed in [51] in which ACO and GA generate feature subsets in parallel.

3.3 Principal component analysis (PCA) for feature extraction

PCA is a well-known method for feature extraction. By calculating the eigenvectors of sample covariance matrix, PCA linearly transforms the original inputs into uncorrelated new features (called principal components), which are the orthogonal transformation of original inputs based on the eigenvectors. The obtained principal components in PCA have second-order correlations with the original inputs [3].

In this study, PCA algorithm is performed on the data set using two functions of MATLAB Software: "prepca" and "trapca." Then the data is normalized to [−1,1] range by using the "prestd" function in MATLAB Software. The function "trapca" processes the network input training set by applying the principal component transformation that has been previously computed by "prepca."

4 ANN-based prediction

ANN-based methods are good candidates for STLF problem, as they are characterized by not requiring explicit models to represent the complex relationship between the load and factors that influence its behavior.

The advantage of neural networks lies in their ability to represent complex nonlinear relationships and learning these relationships directly from the data being modeled.

The data set used in this study is a real electricity load time-series, which includes the electricity load of West-Tehran Province Power Distribution Company, recorded every hour and also weather conditions. The size of this data set is 300, which are the data in the May 4 2009 to March 1 2010 time interval.

We used 48 features for forecasting the load of 24 h ahead. These features are as follow:

1. $f_1,\ldots,f_8$: $d - i$ day's maximum temperature ($i = 0,\ldots,7$)
2. $f_9,\ldots,f_{16}$: $d - i$ day's minimum temperature ($i = 0,\ldots,7$)
3. f_{17}: d day's rainfall
4. f_{18}, f_{19}: minimum and maximum of d day's humidity percentage
5. f_{20}: d day's month
6. f_{21}: d day's season
7. f_{22}: d day's week
8. f_{23}: indicates whether d day is weekend or not, 1 if it is weekend, 0 otherwise
9. f_{24}: indicates whether d day is holiday or not, 1 if it is holiday, 0 otherwise
10. $f_{25},\ldots,f_{48}$: $h - i$ hour's history load ($i = 1,\ldots,24$)

The selected features of the above list form the input variables to ANN-based load forecaster. The layout of three-layer multilayer perceptron (MLP) network, as one of the simulated ANNs in this study, is depicted in Fig. 3.

Table 2 RBF specifications

Specification	Value
Number of output-layer neurons	24
Number of hidden-layer neurons	260
Mean squared error goal	0.001
Spread of radial basis functions	3.4

Table 3 GA parameters

Parameter	Value
Population size	200
Number of generations	20
Probability of crossover	0.9
Probability of mutation	0.04

Table 4 ACO parameters

Parameter	Value
Number of ants	150
Number of selected features	20
$\alpha = \beta$	1
K	50

The input variables are the selected features of load data by using mentioned feature selection algorithms discussed in Sect. 3. The outputs are the predicted load of 24 h ahead.

5 Experimental results

In this paper, the performance of five different models is investigated for 24-h ahead load forecasting. These models are as follow:

1. MLP-based forecasting, with no-feature selection (NFS), abbreviated as NFS + MLP
2. RBF-based forecasting, with no-feature selection, abbreviated as NFS + RBF
3. PCA-based feature extraction + MLP-based forecasting, abbreviated as PCA + MLP
4. ACO-based feature selection + MLP-based forecasting, abbreviated as ACO + MLP
5. Hybrid of GA and ACO for feature selection + MLP-based forecasting, abbreviated as GA-ACO + MLP

The MLP and RBF specifications in our simulations, which have been run in MATLAB Software environment, are reported in Tables 1 and 2, respectively.

The parameters of GA and ACO algorithms are also reported in Tables 3 and 4, respectively.

Table 5 Forecasted electricity load of 24 h using five simulated models (MW)

Time point	Actual load	NFS + MLP		NFS + RBF		PCA + MLP		ACO + MLP		GA-ACO + MLP	
		Forecasted load	Error (%)	Forecasted load	Error (%)	Forecasted load	Error (%)	Forecasted load	Error (%)	Forecasted load	Error (%)
1	789.79	811.86	2.79	795.98	0.78	795.79	0.76	799.05	1.17	776.82	−1.64
2	724.72	739.67	2.06	732.64	1.09	724.52	−0.03	718.17	−0.90	707.57	−2.37
3	687.15	701.16	2.04	702.82	2.28	697.52	1.51	687.69	0.08	683.12	−0.59
4	673.18	690.36	2.55	684.65	1.70	689.21	2.38	671.30	−0.28	671.63	−0.23
5	669.37	690.49	3.16	687.28	2.68	684.86	2.31	678.03	1.29	659.01	−1.55
6	683.85	697.02	1.93	685.71	0.27	693.24	1.37	684.70	0.12	676.40	−1.09
7	703.83	713.67	1.40	699.44	−0.62	728.33	3.48	702.82	−0.14	696.49	−1.04
8	777.69	765.94	−1.51	742.12	−4.57	798.45	2.67	768.48	−1.18	768.13	−1.23
9	887.94	853.98	−3.83	868.98	−2.14	900.05	1.36	875.46	−1.41	867.98	−2.25
10	954.67	929.28	−2.66	940.48	−1.49	972.69	1.89	941.95	−1.33	938.20	−1.73
11	1,004.39	981.55	−2.27	1,009.45	0.50	1,021.01	1.65	988.80	−1.55	979.35	−2.49
12	1,020.41	1,011.59	−0.86	1,058.50	3.73	1,038.33	1.76	1,018.86	−0.15	1,013.71	−0.66
13	993.30	1,004.15	1.09	1,068.96	7.62	1,019.95	2.68	1,013.29	2.01	1,003.79	1.06
14	967.62	968.52	0.09	1,005.13	3.88	995.78	2.91	972.68	0.52	965.77	−0.19
15	980.48	965.17	−1.56	956.68	−2.43	1,003.96	2.39	981.18	0.07	975.18	−0.54
16	979.09	954.31	−2.53	939.35	−4.06	976.70	−0.24	961.42	−1.80	957.86	−2.17
17	963.98	938.32	−2.66	896.36	−7.01	986.47	2.33	941.24	−2.36	952.73	−1.17
18	987.46	965.78	−2.20	960.69	−2.71	1,063.19	7.67	950.98	−3.69	980.14	−0.74
19	1,081.82	1,058.38	−2.17	1,069.50	−1.14	1,138.83	5.27	1,078.58	−0.30	1,088.93	0.66
20	1,109.31	1,083.32	−2.34	1,091.40	−1.61	1,118.63	0.84	1,076.48	−2.96	1,100.53	−0.79
21	1,101.40	1,066.08	−3.21	1,101.01	−0.04	1,079.46	−1.99	1,052.64	−4.43	1,072.85	−2.59
22	1,062.22	1,018.48	−4.12	1,075.29	1.23	1,035.89	−2.48	1,021.22	−3.86	1,030.37	−3.00
23	995.91	956.68	−3.94	1,003.11	0.72	977.85	−1.81	960.56	−3.55	964.30	−3.17
24	896.07	874.49	−2.41	903.52	0.83	901.80	0.64	882.38	−1.53	878.31	−1.98
MAPE			2.31		2.30		2.18		1.53		1.46

Table 6 Performance comparison of five load forecasting methods-MAPE for 1 week

Method	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Week average
NFS + MLP	3.80	2.69	3.88	2.64	1.98	1.96	3.80	2.96
NFS + RBF	1.24	2.05	1.84	2.3	2.74	2.6	2.5	2.18
PCA + MLP	1.97	3.14	3.69	1.40	2.15	3.26	1.81	2.49
ACO + MLP	1.58	1.12	1.66	1.58	1.65	1.63	1.77	1.57
GA-ACO + MLP	1.80	1.43	1.48	1.16	1.66	1.45	1.56	1.51

The results of load forecasting for 24 h of 1 day using five mentioned models (i.e., NFS + MLP, NFS + RBF, PCA + MLP, ACO + MLP, and GA-ACO + MLP) are reported in Table 5. The relative error is reported in Table 5 which is defined as follows:

$$\text{Error} = \left[\frac{P_{\text{predicted}}(i) - P_{\text{actual}}(i)}{P_{\text{actual}}(i)}\right] \times 100\% \tag{4}$$

in which $P_{\text{actual}}(i)$ is the actual load, and $P_{\text{predicted}}(i)$ is the forecasting load at time i. Also, the mean absolute percentage error (MAPE) is reported at the last row of Table 5 which is defined as follows:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_{\text{predicted}}(i) - P_{\text{actual}}(i)}{P_{\text{actual}}(i)}\right| \times 100\% \tag{5}$$

in which N is the time interval of averaging, e.g., $N = 24$ in Table 5.

As can be seen in Table 5, the maximum and minimum differences between the forecasted load and actual load by using five mentioned methods are as follow. In

Table 7 Performance comparison of proposed models for STLF and similar recent researches

Model/research group	Prediction model	Hybrid-party algorithm	Role of hybrid-party algorithm	Number of inputs to model	Number of model outputs	MAPE
Mishra and Patra [52]	MLP	GA	MLP training	9	1	3.19
	MLP	PSO	MLP training	9	1	4.21
	MLP	AIS	MLP training	9	1	5.28
Liao and Tsao [29]	FHRCNN	Fuzzy system	Modification of MLP structure	83	24	1.84
	MLP	EP	MLP training	83	24	1.81
	MLP	EP + Fuzzy system	MLP training	83	24	1.77
	MLP	EP + SA	MLP training	83	24	1.76
	MLP	EP + Fuzzy system + SA	MLP training	83	24	1.62
Wang et al. [53]	SVM	AQPSO	SVM training	24	24	2.43
	SVM	QPSO	SVM training	24	24	2.13
	SVM	PSO	SVM training	24	24	1.83
Hippert and Taylor [38]	MLP	Bayesian evidence	Optimizing MLP structure	17	24	1.59
Kelo and Dudul [42]	FTLRNN	Cross correlation analysis	Feature selection	3 (Summer)	4	1.75
				4 (Rainy)	5	2.10
				3 (Dry)	4	1.72
Amjady and Keynia [46]	MLP	Evolutionary algorithm	Modification of MLP weights	37	24	2.04
Xiao et al. [19]	MLP	Rough set	Feature selection	7	2	3.42 (Daily minimum load)
						0.50 (Daily maximum load)
Hong [34]	SVR	CGA	Determining parameters of SVR model	168	24	3.03
Cai et al. [28]	HS-ARTMAP	–	–	660	24	2.65
	dART & HS-ARTMAP			660	24	1.90
Proposed in this study	MLP	ACO	Feature selection	20	24	1.57
	MLP	GA-ACO	Feature selection	20	24	1.51

PSO particle swarm optimization, *AIS* artificial immune system, *FHRCNN* fuzzy hyper-rectangular composite neural network, *EP* evolutionary programming, *SA* simulated annealing, *AQPSO* adaptive quantum-behaved PSO, *FTLRNN* focused time lagged recurrent neural network, *SVR* support vector regression, *CGA* chaotic genetic algorithm, *HS-ARTMAP* hyper-spherical ARTMAP neural network, *dART* distributed adaptive resonance theory neural network

NFS + MLP model, the maximum difference is 43.74 MW which is occurred at 22:00 and the minimum difference is 0.9 MW at 14:00. In NFS + RBF model, the maximum difference is 75.66 MW which is occurred at 13:00 and the minimum difference is 0.39 MW at 21:00. In PCA + MLP model, the maximum and minimum differences are 75.73 and 0.2 MW which are occurred at 18:00 and 2:00, respectively. In ACO + MLP model, the maximum and minimum differences are 48.76 MW at 21:00 and 0.54 MW at 3:00, respectively. Finally, in GA-ACO + MLP model, the maximum and minimum differences are 31.61 MW and 1.85 MW which are occurred at 23:00 and 14:00, respectively.

As can be seen, the performance of MLP and RBF neural networks in the case of no-feature selection (NFS) is in the same way. So, in the rest of our simulations we have only used MLP as the neural load forecaster to investigate the performance of different feature selection methods. The experimental results show that GA-ACO + MLP model performs better in terms of prediction error as compared to three other hybrid models reported in Table 5. Also, the GA-ACO + MLP model offers the minimum MAPE as compared to other mentioned models.

In addition, if we consider the [−15 MW, 15 MW] interval for the difference between forecasted load and actual load then each of the NFS + MLP and PCA + MLP models offer 8 points of forecasted loads in this interval, and NFS + RBF model offers 12 points, but each of the ACO + MLP and GA-ACO + MLP models offer 15 points of forecasted loads in this interval.

The performance of the mentioned methods in terms of MAPE for 1 week from March 4 2010 to March 11 2010 is also compared and reported in Table 6. As shown in Table 6, the GA-ACO + MLP model performs better in terms of averaged over week MAPE.

The performance of the proposed ACO-based feature selection models, i.e., ACO + MLP and GA-ACO + MLP, in STLF application is compared with some recent researches in this field (Table 7).

6 Conclusion

In this paper, ACO and hybrid model of GA and ACO feature selection algorithms have been used for feature selection with the MLP neural network as the classifier in STLF application. Experimental results have shown that the feature selection based on GA-ACO performs better when compared with other simulated methods in this study (i.e., MLP and RBF without feature selection, PCA + MLP, and ACO + MLP). The proposed GA-ACO hybrid feature selection algorithm besides an MLP-based load forecaster performs better in terms of the accuracy of STLF and offers MAPE = 1.51 that is an outstanding performance when compared with some recent similar researches in this field (as shown in Table 7).

References

1. Gross G, Galiana F (1987) Short term load forecasting. Proc IEEE 75:1558–1573
2. Sadownik R, Barbosa E (1999) Short-term forecasting of industrial electricity consumption in Brazil. Int J Forecast 18:215–224
3. Afshin M, Sadeghian A (2007) PCA-based least squares support vector machines in week-ahead load forecasting. In: Proceedings of IEEE industrial and commercial power systems technical conference, pp 1–6
4. Kothari DP, Nagrath IJ (2003) Modern power system analysis, 3rd edn. Tata McGraw Hill, New Delhi
5. Feinberg EA, Genethliou D (2005) Load forecasting. In: Chow JH, Wu FF, Momoh JJ (eds) Applied mathematics for restructured electric power systems: optimization, control and computational intelligence, power electronics and power systems. Springer, US, pp 269–285
6. Gonzalez-Romera E, Jaramillo-Moran MA, Carmona-Fernandez D (2006) Monthly electric energy demand forecasting based on trend extraction. IEEE Trans Power Syst 21:1946–1953
7. Kyriakides E, Polycarpou M (2007) Short term electric load forecasting: a tutorial. In: Chen K, Wang L (eds) Trends in neural computation, studies in computational intelligence. Springer, Berlin, pp 391–418
8. Hahn H, Meyer-Nieberg S, Pickel S (2009) Electric load forecasting methods: tools for decision making. Eur J Oper Res 199:902–907
9. Amjady N (2001) Short-term hourly load forecasting using time-series modeling with peak load estimation capability. IEEE Trans Power Syst 16:498–505
10. Huang S-J, Shih K-R (2003) Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. IEEE Trans Power Syst 18:673–679
11. Al-Hamadi HM, Soliman SA (2004) Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model. Electr Power Syst Res 68:47–59
12. Gao J, Sun H-B, Tang G-Q (2001) Application of optimal combined forecasting based on fuzzy synthetic evaluation on power load forecast. J Syst Eng 16:106–110
13. Kodogiannis VS, Anagnostakis EM (2002) Soft computing based techniques for short-term load forecasting. Fuzzy Sets Syst 128:413–426
14. Al-Kandari AM, Soliman SA, El-Hawary ME (2004) Fuzzy short-term electric load forecasting. Int J Electr Power Energy Syst 26:111–122
15. Kim KH, Park JK, Hwang KJ, Kim SH (1995) Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert-systems. IEEE Trans Power Syst 10:1534–1539
16. Satish B, Swarup KS, Srinivas S, Rao AH (2004) Effect of temperature on short-term load forecasting using an integrated ANN. Electr Power Syst Res 72:95–101
17. He W (2008) Forecasting electricity load with optimized local learning models. Int J Electr Power Energy Syst 30:603–608
18. Alves da Silva A, Ferreira VH, Velasquez MG (2008) Input space to neural network based load forecasters. Int J Forecast 24:616–629

19. Xiao Z, Ye S-J, Zhang B, Sun C-X (2009) BP neural network with rough set for short term load forecasting. Expert Syst Appl 36:273–279
20. Niu D, Wang Y, Dash Wu D (2010) Power load forecasting using support vector machine and ant colony optimization. Expert Syst Appl 37:2531–2539
21. Christianse WR (1971) Short term load forecasting using general exponentials smoothing. IEEE Trans Power Apparatus Syst 90:900–911
22. Park JH, Park YM, Lee KY (1991) Composite modeling for adaptive short-term load forecasting. IEEE Trans Power Syst 6:450–457
23. Gelb A (1974) Applied optimal estimation. MIT Press, Massachusetts
24. Brown RG (1983) Introduction to random signal analysis and Kalman filtering. Wiley, New York
25. Moghram I, Rahman S (1989) Analysis and evaluation of five short-term load forecasting techniques. IEEE Trans Power Syst 4:1484–1491
26. Douglas AP, Breipohl AM, Lee FN, Adapa R (1998) The impact of temperature forecast uncertainty on Bayesian load forecasting. IEEE Trans Power Syst 13:1507–1513
27. Carpinteiro OAS, Reis AJR, da Silva APA (2004) A hierarchical neural model in short-term load forecasting. Appl Soft Comput 4:405–412
28. Cai Y, Wang J-Z, Tang Y, Yang Y-C (2011) An efficient approach for electric load forecasting using distributed ART (adaptive resonance theory) & HS-ARTMAP (Hyper-spherical ARTMAP network) neural network. Energy 36:1340–1350
29. Liao G-C, Tsao T-P (2004) Application of fuzzy neural networks and artificial intelligence for load forecasting. Electr Power Syst Res 70:237–244
30. Yao SJ, Song YH, Zhang LZ, Cheng XY (2000) Wavelet transform and neural networks for short-term electrical load forecasting. Energy Convers Manag 41:1975–1988
31. Zhang B-L, Dong Z-Y (2001) An adaptive neural-wavelet model for short-term load forecasting. Electr Power Syst Res 59:121–129
32. Pai P-F, Hong W-C (2005) Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. Electr Power Syst Res 74:417–425
33. Pai P-F, Hong W-C (2005) Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers Manag 46:2669–2688
34. Hong W-C (2009) Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. Electr Power Energy Syst 31:409–417
35. Dash M, Liu H (1997) Feature selection for classification. Intell Data Anal 1:131–156
36. Wang Z, Cao Y (2006) Short-term load forecasting based on mutual information and artificial neural networks. Lecture Notes Comput Sci 3972:1246–1251
37. Siwek K, Osowski S (2010) Neural network ensemble for 24-hour load pattern prediction in power systems. Stud Comput Intell 302:151–169
38. Hippert HS, Taylor JW (2010) An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting. Neural Netw 23:386–395
39. Niu D, Gu Z, Zhang Y (2009) An AFSA-TSGM based wavelet neural network for power load forecasting. Lecture Notes Comput Sci 5553:1034–1043
40. Niu D, Kou B, Zhang Y, Gu Z (2009) A short-term load forecasting model based on LS-SVM optimized by dynamic inertia weight particle swarm optimization algorithm. Lecture Notes Comput Sci 5552:242–250
41. Ghanbari A, Hadavandi E, Abbasian-Naghneh S (2010) An intelligent ACO-SA approach for short term electricity load prediction. Lecture Notes Comput Sci 6216:623–633
42. Kelo SM, Dudul SV (2011) Short-term Maharashtra state electrical power load prediction with special emphasis on seasonal changes using a novel focused time lagged recurrent neural network based on time delay neural network model. Expert Syst Appl 38:1554–1564
43. Maniezzo V, Colomi A (1999) The ant system applied to the quadratic assignment problem. Knowl Data Eng 11:769–778
44. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison Wesley, New York
45. Srinivas M, Patnik LM (1994) Genetic algorithms: a survey. IEEE Computer Society Press, Los Alamitos
46. Amjady N, Keynia F (2009) Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. Energy 34:46–57
47. Kohavi R, John JH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324
48. Fodor IK (2002) A survey of dimension reduction techniques, Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, US Department of Energy
49. Rashidy Kanan H, Faez K (2008) An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. Appl Math Comput 205:716–725
50. Al-Ani A (2005) Feature subset selection using ant colony optimization. Int J Comput Intell 2:53–58
51. Nemati S, Basiri ME, Ghasem-Aghaee N, Hosseinzadeh Aghdam M (2009) A novel ACO–GA hybrid algorithm for feature selection in protein function prediction. Expert Syst Appl 36:12086–12094
52. Mishra S, Patra SK (2008) Short term load forecasting using a neural network trained by a hybrid artificial immune system. In: Proceedings of IEEE third international conference on industrial and information systems, pp 1–5
53. Wang J, Liu Z, Lu P (2008) Electricity load forecasting based on adaptive quantum-behaved particle swarm optimization and support vector machines on global level. In: Proceedings of international symposium on computational intelligence and design, pp 233–236