Expert Systems with Applications 37 (2010) 2531–2539

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Power load forecasting using support vector machine and ant colony optimization

Dongxiao Niu [a], Yongli Wang [a], Desheng Dash Wu [b,c,*]

[a] School of Business Administration, North China Electric Power University, 102206 Beijing, China
[b] RiskLab, University of Toronto, Toronto, Ontario, Canada M5S 3G8
[c] School of Science and Engineering, Reykjavik University, Kringlunni 1, IS-103 Reykjavík, Iceland

ARTICLE INFO

Keywords:
Ant colony optimization
Feature selection
Support vector machine
Power load forecasting

ABSTRACT

This paper creates a system for power load forecasting using support vector machine and ant colony optimization. The method of colony optimization is employed to process large amount of data and eliminate redundant information. The system mines the historical daily loading which has the same meteorological category as the forecasting day in order to compose data sequence with highly similar meteorological features. With this method, we reduced SVM training data and overcame the disadvantage of very large data and slow processing speed when constructing SVM model. This paper proposes a new feature selection mechanism based on ant colony optimization in an attempt to combat the aforemention difficulties. The method is then applied to find optimal feature subsets in the fuzzy-rough data reduction process. The present work is applied to complex systems monitoring, the ant colony optimization can mine the data more overall and accurate than the original fuzzy-rough method, an entropy-based feature selector, and a transformation-based reduction method, PCA. Comparing with single SVM and BP neural network in short-term load forecasting, this new method can achieve greater forecasting accuracy. It denotes that the SVM-learning system has advantage when the information preprocessing is based on data mining technology.

1. Introduction

Power load prediction has attracted a great deal of attention from both the practice and the academia. The short-term power load forecasting is very significant for the electric network's reliability and economic development. As short-term power load prediction is of crucial importance to the reliability and economic utilization of electric networks, it is drawing more and more attention from both the practice and the academia. The aim of load forecasting is to make the best use of electric energy and relieve the conflict between supply and demand.

Load forecasting has become a crucial issue for the operational planners and researchers of electric power systems. For electricity load reliance, electricity providers face increasing competition in the demand market and must pay more attention to electricity quality, including unit commitment, hydrothermal coordination, short-term maintenance, interchange and transaction evaluation, network power flow dispatched optimization and security strategies.

Inaccurate forecast of power load will lead to a great deal of loss for power companies. Bunn and Farmer pointed out that a 1% increase in forecasting error implied a 10 million increase in operating costs (Pai & Hong, 2005). The short-term forecasts refer to hourly prediction of electricity load demand for a lead time ranging from 1 h to several days ahead. In certain instances, the prediction of the daily peak load is the objective of short-term load forecasting, since it is the most important load during any given day. The quality of short-term hourly load forecasts has a significant impact on the economic operation of the electric utility since many decisions are based on these forecasts. These decisions include economic scheduling of generating capacity, scheduling of fuel purchases, system security assessment and planning for energy transactions. The importance of accurate load forecasts will increase in the future because of the dramatic changes occurring in the structure of the utility industry due to deregulation and competition. This environment compels the utilities to operate at the highest possible efficiency which as indicated earlier requires accurate load forecasts. The load has complex and nonlinear relationships with several factors such as weather factors, climatic conditions, social activities, and seasonal factors, past usage patterns, the day of the week and the time of the day.

Some researchers considered related factors such as seasonal temperature and day type in load forecasting models. Christianse (1971) and Park, Park and Lee (1991) designed exponential smoothing models by Fourier series transformation for electricity load forecasting. Mbamalu and El-Hawary (1993) proposed multiplicative autoregressive (AR) models that considered seasonal factors in load forecasting. The analytical results showed that the

* Corresponding author. Address: School of Business Administration, North China Electric Power University, 102206 Beijing, China. Tel.: +1 416 8805219.
E-mail address: DWu@Rotman.Utoronto.Ca (D.D. Wu).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.08.019

forecasting accuracy of the proposed models outperformed the unvaried model. To achieve accurate load forecasting, Brown (1983) and Gelb (1974) built the load forecasting model employed state space and Kalman filtering technology, which were developed to reduce the difference between the actual and the predicted loads. The state space and Kalman filtering technology introduced a periodic component of the load as a random process and requires historical data covering a period exceeding 3–10 years to establish the periodic load variation for estimating the dependent variables (load or temperature) of the power system. Moghram and Rahman (1989) devised a model based on the state space and Kalman filtering technology and verified that the proposed model outperformed four other forecasting methods. However, this stream of methods always fails to avoid the influence of observation noise in the forecasting. To handle this problem, Douglas, Breipohl, Lee, and Adapa (1998) considered verifying the impacts of temperature on the forecasting model. Sadownik and Barbosa (1999) proposed dynamic nonlinear models for load forecasting. The main disadvantage of these methods is that they are time consuming in computation as the number of variables increases. Very limited research has been done to provide effective approaches to consider both the time complexity and the various influential factors of the load forecasting. Feature selection is thus important when faced with large scale of computation in load forecasting task.

The main aim of feature selection (FS) is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features (Dash & Liu, 1997). Swarm intelligence (SI) is the property of a system whereby the collective behaviors of simple agents interacting locally with their environment cause coherent functional global patterns to emerge (Bonabeau, Dorigo, & Theraulez, 1999). SI provides a tool to solve collective (or distributed) problem without requiring centralized control or the provision of a global model. For example, ants are capable of finding the shortest route between a food source and their nest without the use of visual information, and hence possess no global world model, adapting to changes in the environment. SI techniques based on the behavior of real ant colonies used to solve discrete optimization problems are ant colony optimization (ACO) techniques (Bonabeau et al., 1999). These have been successfully applied to a large number of difficult combinatorial problems such as the quadratic assignment (Maniezzo & Colomi, 1999) and the traveling salesman (Dorigo, Maniezzo, & Colomi, 1996). This method is particularly attractive for feature selection because there are very limited heuristic that can guide search to the optimal minimal subset every time. Additionally, it can be the case that ants discover the best feature combinations as they proceed throughout the search space. This paper investigates how ACO is applied to yield optimal feature subsets.

The development of meta-heuristic optimization theory has been flourishing. Many meta-heuristic paradigms such as genetic algorithm (Goldberg, 1997), simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983), and tabu search (Glover, 1989) have shown their efficacy in solving computationally intensive problems. Among them, Abo-Sinna (Osman, Abo-Sinna, & Mousa, 2005) proposed a genetic algorithm for multi-objective resource allocation problem. The chromosome is resource allocated job, and the objective is to maximize the efficiency with the minimum cost. Lin and Gen (2007) presented a hybrid genetic algorithm for the same problem. It is seen that compared to mathematical programming approaches, only few multi-objective resource allocation problem researches based on meta-heuristic algorithms have been conducted.

Support vector machines (SVM) is based on the statistic theory raised by Vapnik. It is a machine learning algorithm which began to be used in the middle of ninetieth century. Statistic theory employs the criterion that minimizes the structure risk; meanwhile, it can lower the global error of model. It raises the generalization capability of the model, which is more prominent in the small-sample learning.

Through above analysis, a new idea with which the accuracy of load forecasting can be improved is presented. A new method of using Support Vector Machines based on ant colony optimization is created for power load forecasting. Computation in our case study indicates forecasting accuracy is higher than BP neural network and single SVM.

The rest of the paper is organized as follows. In Section 2, we discuss features of power load data. Sections 3 and 4 discuss ant colony optimization (ACO) for feature selection and SVM regression. A case study is demonstrated through computation and analysis in Section 5. Finally, the conclusions are presented in Section 6.

2. Feature of power load data

We discuss various influencing factors in short load forecasting in this section. To build a load forecasting model, choosing appropriate input variables is one of the most important tasks. Very few general rules have been reported to conduct this task in existing literature. Rule of thumb such as engineering judgment and experience based on repeated trial is usually useful. Other quantitative techniques such as load curve analysis and statistical analysis can be helpful for choosing key variables to build such a load forecast-

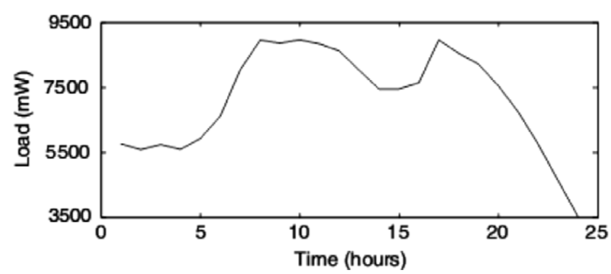

Fig. 1. Hour load during the day.

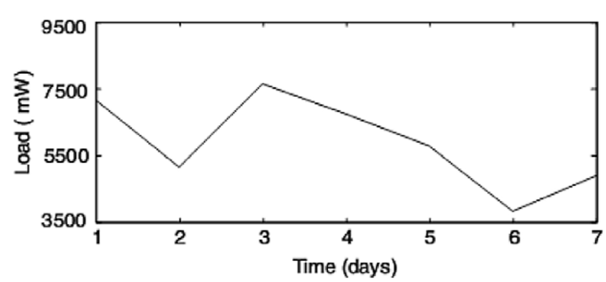

Fig. 2. Daily load during the week.

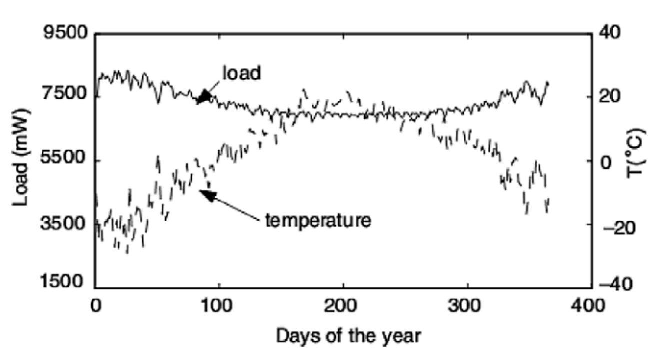

Fig. 3. The relation between temperature and load.

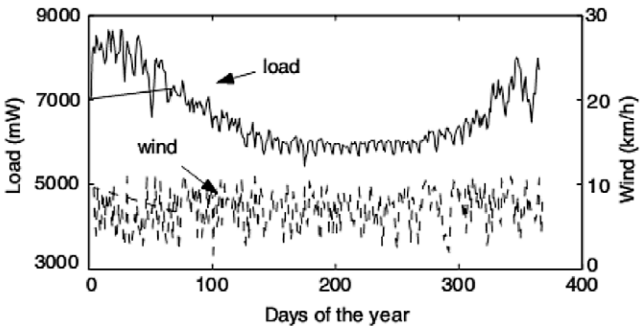

Fig. 4. The relation between wind speed and load.

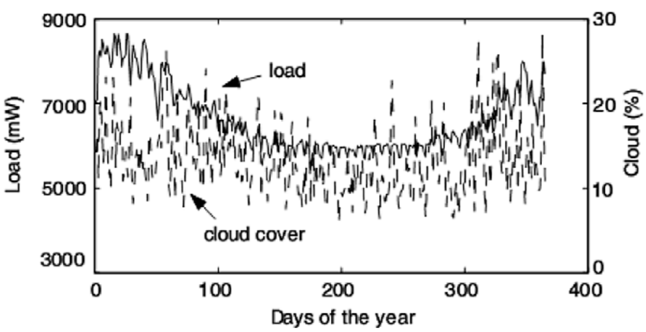

Fig. 5. The relation between cloud cover and load.

ing model. It is known that short-term load is influenced by many factors such as the weather, the development of economy and so on. In order to obtain the intuitive relation between influential factors and load swing, we observed patterns from curves depicted from raw time series data. We base our data analysis on a very typical power market from SHAANXI province in China. We note that time factors influencing load forecasting include the time of the year, the day of the week, and the hour of the day. There are big differences in load between weekdays and weekends. Weather conditions also influence the load. In fact, forecasted weather parameters are the most important factors in short-term load forecasts. Temperature is the most commonly used load predictor. We show the 2007 power load curve of SHAANXI province in China from Figs. 1–5.

2.1. Hour-forecasting feature

Daily load changes from one hour to another. Fig. 1 shows the load data for a typical day in the Chinese power market—SHAANXI province power market on 15/07/2007. Load peak occurs twice a day, one is about 9:00 AM, and the other is about 5:30 PM, and there is a significant difference in load magnitude. Therefore, an hour indicator (where the time changes from 1 to 24) is very helpful in short-term load forecasting.

2.2. Week-forecasting feature

Load also changes from one day to another during the week. Fig. 2 shows an average daily load for a typical week which is the twenty-ninth week from 23/07/2007 to 29/07/2007 in SHAANXI province. It is clear that power consumption in the weekends is much less than in the weekdays for all seasons, and it also can be seen that the power consumption of Monday and Tuesday is much more than the other weekdays. As a result, a day indicator (where n changes from 1 to 7) may be of assistance in load forecasting.

2.3. Weather-forecasting feature

For investigating the effect of weather variables on the province system load, Figs. 3–5 demonstrate the relation between the load forecasting and temperature, and wind and cloudy cover. Weather quantities are measured at more than 5 different locations in the province. It is clear that temperature is the most important weather variable, and Fig. 3 confirms this strong correlation between temperature and load. The wind speed, as shown in Fig. 4, has less effect on the load than the temperature. It can be seen that the relation between the cloud cover and the load from Fig. 5.

Figs. 1–5 also demonstrate the variable tendency of the load curve. From Figs. 1 and 2, the regularity of the load curve is very apparent, these can be an important feature for load forecasting; Fig. 3 indicates that the load curve changes evidently with the temperature in a year, this factor will be an important factor to influence the load forecasting. However, from Figs. 4 and 5, the relation between the load curve and the wind, cloud cover, rainfall and so on cannot be seen. It needs some methods to select these features and calculate the relation extent with those influential factors. In this paper, we use ant colony optimization model to select the influential factors and analyze the relationship of these factors.

3. ACO for feature selection

The ant colony optimization algorithm (ACO) provides an alternative feature selection tool inspired by the behavior of ants in finding paths from the colony to food. 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 mimic this ant seeking food phenomenon to yield the shortest path (which means the "system" of interests has converged to a single solution). Different equally shortest paths can exist. An ACO algorithm can be generally applied to any combinatorial problem as far as it is possible to define:

Firstly, the problem can be described in a set of nodes and edges between nodes to form a graph. So the problem can be seen easily to find the main problem and overcome it.

Secondly, heuristic desirability of paths: it is a suitable heuristic measure which can find better paths from one node to every other connected node, and it can be described in a graph details.

Thirdly, construction of solutions: a feasible and complete solution of the formulated inter-cell layout problem is considered as a permutation of manufacturing cells. Each part of this solution is termed state. In the optimum process, each ant initially assigns a cell to location 1 then assigns another cell to location 2 and so on till a complete solution is obtained.

Fourthly, pheromone updating rule: firstly, it is the area pheromone updating rule, the effect is to make the desirability of edges change dynamically in order to shuffle the tour. The nodes in one ant's tour will be chosen with a lower probability in building other ants' tours. As a consequence, ants will favor the exploration of edges not yet visited and prevent converging to a common path. Next, go to global updating rule, this process is performed after all ants have completed their tours. Therefore, only the globally best ant that found the best solution up to the current iteration of the algorithm is permitted to deposit pheromone.

Fifthly, probabilistic transition rule: the rule determines the probability of an ant traversing from one node in the graph to the next. The heuristic desirability of traversal and edge pheromone levels is combined to form the so-called probabilistic transition rule. It can be denoted in Fig. 6.

The feature selection task may be reformulated into an ACO-suitable problem. ACO requires a problem to be represented as

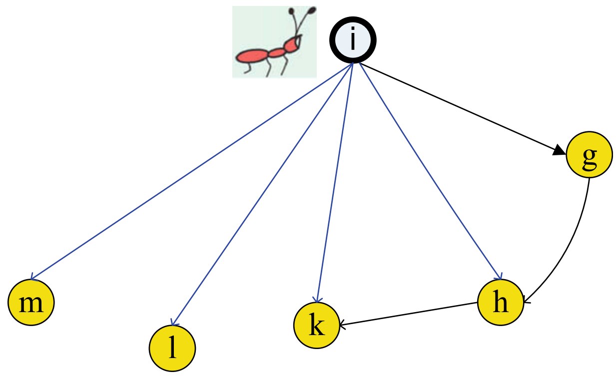

Fig. 6. ACO problem representation for FS.

the graph-here nodes represent features, with the edges between them denoting the choice of the next feature. The search for the optimal feature subset is then an ant traversal through the graph where a minimum number of nodes are visited that satisfies the traversal stopping criterion. Fig. 6 illustrates this setup – the ant is currently at node i and has a choice of which feature to add next to its path (dotted lines). It chooses feature g next based on the transition rule, then h and then k. Upon arrival at k, the current subset $\{i, g, h, k\}$ is determined to satisfy the traversal stopping criterion. The ant terminates its traversal and outputs this feature subset as a candidate for data reduction (Lin & Gen, 2007).

Probability of an ant at feature T_m choosing to travel to feature T_n at time r:

$$o^p_{mn}(r) = \frac{[\kappa_{mn}(r)]^\alpha \cdot [\mu_{mn}]^\beta}{\sum_{j\in L^p_m}[\kappa_{mj}(r)]^\alpha \cdot [\eta_{mj}]^\beta} \tag{1}$$

where p is the number of ants, L^k_m is the set of ant p's unvisited features, μ_{mn} is the heuristic desirability of choosing feature n when at feature i and $\kappa_{mn}(r)$ is the amount of virtual pheromone on edge (m, n). The choice of parameters α and β is determined experimentally. Several parameter values are chosen in the range [0, 1] and evaluated by experimentation.

The time complexity of the ant-based approach to feature selection is $U(IBv)$, where I is the number of iterations, B the number of original features and v the number of ants. This can be seen from Fig. 7. In the worst case, each ant selects all the features. As the heuristic is evaluated after each feature is added to the reduct candidate, this will result in B evaluations per ant. After one iteration in this scenario, Bv evaluations will be performed. After I iterations, the heuristic will be evaluated IBv times.

Dependency function may also be chosen as the heuristic desirability measure, but this is not necessary. In fact, it may be of more use to employ a nonrough set related heuristic for this purpose to avoid the pitfalls of a QUICKREDCT style search. By using an alternative measure such as an entropy-based heuristic (Goldberg, 1997), the method may avoid feature combinations that may mislead the rough set-based heuristic. Again, the time complexity of this fuzzy-rough ant-based method will be the same as that mentioned earlier, $U(ABv)$ is the time complexity of the ant-based approach to feature selection, in which A is the number of iterations, B the number of original features and v the number of ants.

The pheromone on each edge is updated according to the following formula:

$$\kappa_{mn}(r+1) = (1-\rho)\cdot\kappa_{mn}(r) + \Delta\kappa_{mn} \tag{2}$$

where

$$\Delta\kappa_{mn}(r) = \sum_{p=1}^{i}(e'(D^v)/|D^v|) \tag{3}$$

This is the case if the edge (m, n) has been traversed. In Eqs. (2) and (3), $\Delta\kappa_{mn}(r)$ is 0 otherwise. $\kappa_{mn}(r)$ is defined in Eq. (1). $\kappa_{mn}(r+1)$ is the amount of virtual pheromone on edge (m, n) at time $r+1$. The value ρ is the decay constant used to simulate the evaporation of the pheromone, D^v is the feature subset found by ant v. The

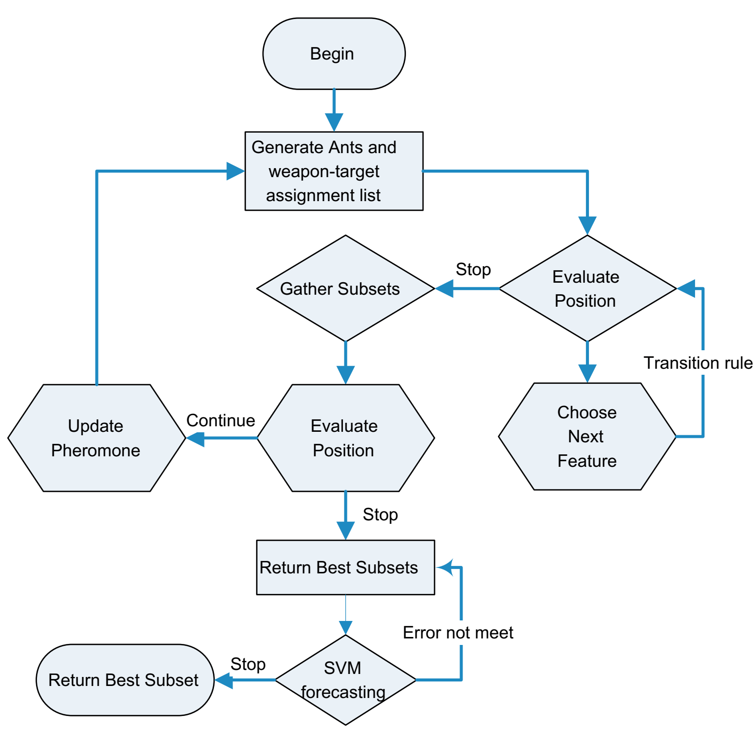

Fig. 7. ACO–SVM forecasting process.

pheromone is updated according to both the fuzzy-rough measure of the "goodness" of the ant's feature subset (e') and the size of the subset itself. By this definition, all ants update the pheromone. Alternative strategies may be used for this, such as allowing only the ants with the best feature subsets to proportionally increase the pheromone. These are, however, beyond the scope of this paper.

4. SVM regression

This section presents the fundamental knowledge of SVM regression. Suppose a set of data (x_i, y_i), $i = 1, 2 \cdots n$, $x_i \cdots \boldsymbol{R}^n$ are given as inputs, $y_i \cdots \boldsymbol{R}$ are the corresponding outputs. SVM regression theory is to find a nonlinear map from input space to output space and map the data to a higher dimensional feature space through the map, then the following estimate function is used to make linear regression (Jensen & Shen, 2005).

$$f(x) = [\omega \cdot \phi(x)] + b \quad \phi : \boldsymbol{R}^m \to \boldsymbol{F}, \quad \omega \in \boldsymbol{F} \tag{4}$$

$f(x)$ is the regression estimate function which constructed through learning of the sample set. ω is weight vector, b is the threshold value, $\phi(x)$ is the nonlinear mapping from input space to high-dimensional feature space which is the only hidden space. The problem of the function approximate is equivalent with the minimizing the following problem.

$$R = \frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}|y_i - f(x_i)|_\varepsilon \tag{5}$$

$||\omega||^2$ is the weights vector norm, which is used to constrain the model structure capacity in order to obtain better generalization performance. C is the regularized constant determining the trade-off between the empirical error and the regularization term. In addition, in (5), we adopted Vapnik's linear loss function with ε – intensive zone as a measure for empirical error which is shown in (6).

$$|y - f(x)|_\varepsilon = \begin{cases} 0 & \text{if } |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon & \text{otherwise} \end{cases} \tag{6}$$

After we introduce positive slack variables ε_i and ε_i^*, minimizing the risk function R in (2) is equivalent to minimizing the objective function shown as follows:

$$R = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) \tag{7}$$

Subject to

$$\begin{cases} (\omega \cdot \phi(x)) + b - y_i \le \varepsilon + \xi_i^* & i = 1, 2, \ldots, l \\ y_i - (\omega \cdot \phi(x)) - b \le \varepsilon + \xi_i & i = 1, 2, \ldots, l \\ \xi_i, \xi_i^* \ge 0 & i = 1, 2, \ldots, l \end{cases} \tag{8}$$

where ξ_i is the upper training error (ε_i^* is the lower) subjected to the ε-intensive tube y_i-$(\omega \cdot \varphi(x)) + b \leqslant \varepsilon$. With the Lagrange multipliers introduced, the decision function given in (8) can be expressed as the following explicit form:

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)k(X_i, X) + b, \tag{9}$$

where a_i and a_i^* are Lagrange multipliers with $a_i \times a_i^* = 0$ and $a_i, a_i^* \ge 0$ for any $i = 1, 2, \ldots, l$. Using Mercer's theorem, the regression is obtained by solving a finite dimensional QP problem in the dual space avoiding explicit knowledge of the high-dimensional mapping and using only the related kernel function. In (9), we introduced a kernel function $K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$, which is the inner product of two vectors in feature space $\phi(x_i)$ and $\phi(x_j)$. It can be shown that any symmetric kernel function K satisfying Mercer's condition corresponds to a dot product in some feature space. A common kernel is RBF kernel adopted in the paper which is shown as follows.

$$K(x, y) = \exp(-\|x - y\|^2/(2\sigma^2)) \tag{10}$$

Thus, the Lagrange multipliers can be obtained by maximizing the following form:

$$\begin{aligned} R(a_i, a_i^*) = &-\frac{1}{2}\sum_{i,j=1}^{n}(a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) - \varepsilon\sum_{i=1}^{l}(a_i + a_i^*) \\ &+ \sum_{i=1}^{l}y_i(a_i - a_i^*)] \end{aligned} \tag{11}$$

Subject to

$$\begin{aligned} &\sum_{i=1}^{n}a_i = \sum_{i=1}^{n}a_i^* \\ &0 \le a_i^* \le \boldsymbol{C}, \quad i = 1, 2, \ldots, l, \\ &0 \le a_i \le \boldsymbol{C}, \quad i = 1, 2, \ldots, l \end{aligned} \tag{12}$$

Through adjusting the two parameters C and ε, the generalized performance can be controlled in high-dimension space. According to Karush–Kuhn–Tucker (KKT) conditions, only some of coefficients in $(a_i - a_i^*)$ differ from zero, and the corresponding training data are referred to as support vector, which can be regarded as the number of neurons in hidden layer of the network structure.

5. Computation and analysis

5.1. SVM based on ACO

Our process of using ACO feature selection and SVM for load forecasting is demonstrated in Fig. 7. To tailor this mechanism to find fuzzy-rough set reducts, it is necessary to use the dependency measure as the stopping criterion. This means that an ant will stop building its feature subset when the dependency of the subset reaches the maximum for the dataset. If the data of the first attribute subset cannot meet the required value, then it returns the best subsets to choose the data of another attribute subset, it does not stop until finding the minimum error. The optimum process of the SVM–ACO forecasting process can be seen in Fig. 7.

Step 1. Except the history load data, we select the daily typical (such as weekend or holiday), daily highest temperature, daily average temperature, daily lowest temperature, daily rainfall, daily wind speed and cloudy cover and daily seasonal attribute as random disturbance factors to short-term load, and it can be seen in Table 1. Although the daily wind speed and cloudy cover cannot make obvious influence on the load curve through above analysis, these influential factors have been chosen. The historical load days are clustered by ACO model. Get the weather information through weather forecast, and the most similar historical day class could be selected. Using the combined ant colony optimization algorithm which is introduced earlier to input the history data and preprocess, get the training and testing samples which have the most similar features.

Step 2. Using SVM model to forecast with the data which have been disposed and optimized by ACO method and it also adding some features of the forecasting day (in which d day represents the forecasting day), such as the d day's highest temperature, average temperature, lowest temperature, rainfall, wind speed and cloudy cover, seasonal attribute and whether it is holiday or weekend. The forecasting day's influential factors would be input beforehand. This method can consider many influential factors of the forecasting points in order to improve the load forecasting

Table 1
Initial decision table.

Attribute name	Attribute meaning
$Z_1, \cdots, Z_7$	$T_{d-i}^{\max}(i = 1, \ldots, 7)$, express $d - i$ day's max-temperature
$Z_8, \cdots, Z_{14}$	$T_{d-i}^{\min}(i = 1, \ldots, 7)$, express $d - i$ day's min-temperature
$Z_{15}, \cdots, Z_{21}$	$T_{d-i}^{avg}(i = 1, \ldots, 7)$, express $d - i$ day's average temperature
$Z_{22}, \cdots, Z_{27}$	$T_d^l(l = 1, 5, 9, 13, 17, 21)$, express d day's six temperature point on l o'clock
Z_{28}	R_d, express d day's rainfall
Z_{29}	Wd_d, express d day's wind speed
Z_{30}	Hum_d, express d day's humidity
Z_{31}	Cld_d, express d day's cloudy cover
Z_{32}	M_d, express d day's month
Z_{33}	Sea_d, express d day's season
Z_{34}	Wk_d, express d day's week
Z_{35}	H_d, express whether d day is holiday, 0 is holiday, 1 is not holiday
Z_{36}	Wnd_d, express whether d day is weekend, "−" is weekend, "+" is not weekend

accuracy. Parameters in SVM model should be initialized. α_i α_i^* and b are assigned random values.

Step 3. The objective functions from (8)–(10) are established by using training sample. Then, previous functions are converted into their dual problems like (11) and (12), thus α_i α_i^* and b will be worked out. And then they are substituted into (13), by this way the predicted value of the subsequent time points will be worked out. The preceding algorithm is shown in Fig. 8.

5.2. Sample selection

Data are chosen from the database of Inner Mongolia region. The power load data from 0:00 on 5/1/2004 to 12:00 on 3/31/2006 are selected as training sample and used to establish the single-variable time series. And the power load data from 13:00 on 3/31/2006 to 24:00 on 5/28/2006 as testing sample.

Inner Mongolia Power Grid is located across three major economic zones: the north, northeast and northwest of China. The power supply of this area covers about 100 cities except Chifeng and Tongliao Cities. Selection of the great Inner Mongolia area for the load forecasting is thus very representative in China due to the following three reasons. Firstly, the great Inner Mongolia area has distinct seasons, which is quite common all over China. Secondly, this area belongs to the land-locked region in the north and is seldom affected by extreme climate conditions. This is important to study the changing pattern of the load forecasting. Thirdly, the Inner Mongolia Power Grid has double tasks: first is satisfying regional power load demand and second is supplying the demand of Jing-Jin-Tang power grid, which is of the main power grid in North China.

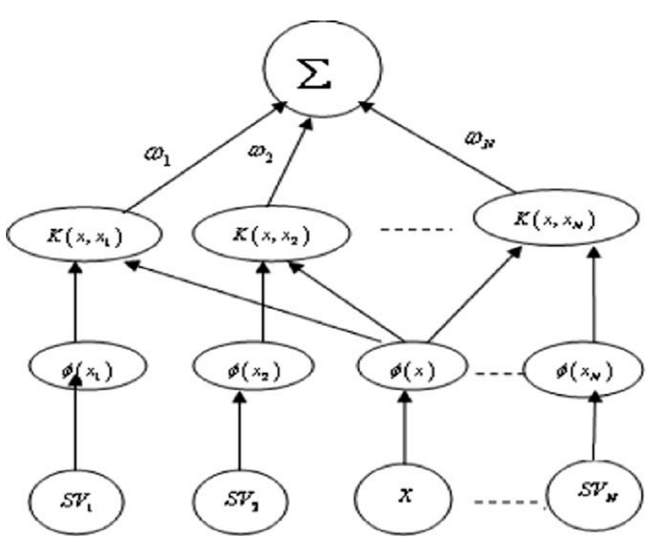

Fig. 8. Structure of support vector machines based ant colony algorithm.

In 2004, the generating capacity of Inner Mongolia region is 64 billions kWh, in which 30 billions kWh that is 45% is supplied for north and northeast of China, and one fifth demand of Beijing city is supplied by Inner Mongolia Power Grid. In 2005, it transmitted 45 billions kWh which took up a half of the whole generating capacity of the region to these areas.

Inner Mongolia Power Grid transmitted 64.29 billions to north and northeast of China in 2007. The Inner Mongolia region has become an important energy terminal of the economic development of north and northeast of China. As the Inner Mongolia Power Grid has to supply power for regional and several outside markets, Inner Mongolia Power Grid has paid more and more attention to the load forecasting to improve the security and stability of its electric network.

In order to investigate the performance of the forecasting system thoroughly, the Inner Mongolia region power system with different typical load characteristics and weather conditions are considered in this research. Inner Mongolia region is a big power supply system in China, supplying the electrical needs of the North of China and the Northeast of China. The electrical demand of the Inner Mongolia region is mainly for industry and resident in a large area. Fig. 9 illustrates the quarter-by-quarter electricity demands of the system for more than two years of Inner Mongolia region.

From Fig. 9, the following common characteristics of the loads can be observed: the load series have multiple seasonal patterns, corresponding to a daily, weekly and monthly periodicity, respectively; the loads are also influenced by the calendar effect, i.e., weekend and holidays; sometimes, the demand presents high volatility and no constant means. From the above observations, it can be concluded that there exist different regimes or dynamics in the load series during different periods.

The data used in this study were provided by the Inner Mongolia Power Grid, which is the transmission company in Inner Mongolia region. The data consists of 108 weeks of quarter-by-quarter observations of electricity demand in Inner Mongolia region, from Saturday 1 May 2004 to Sunday 28 May 2006. The series consists of 72768 observations. The power load data from 0:00 on 5/1/2004 to 12:00 on 3/31/2006 are selected as training sample and used to establish the single-variable time series. And the power load data is shown in Fig. 9. We used the first 100 weeks of data to estimate the various parameters and the remaining 8 weeks to evaluate the postsample forecast accuracy. This gave 67,200 quarter-by-quarter observations for estimation, and 5568 for evaluation.

5.3. Error analysis

Relative error and root-mean-square relative error are used as the final evaluating indicators:

$$e = \frac{A(i) - F(i)}{A(i)} \times 100\% \tag{13}$$

$$RMSRE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A(i) - F(i)}{A(i)}\right)^2}. \tag{14}$$

In the Eqs. (13) and (14), e is relative error between the actual load and the forecasting load at the time i, $RMSRE$ is root-mean-square

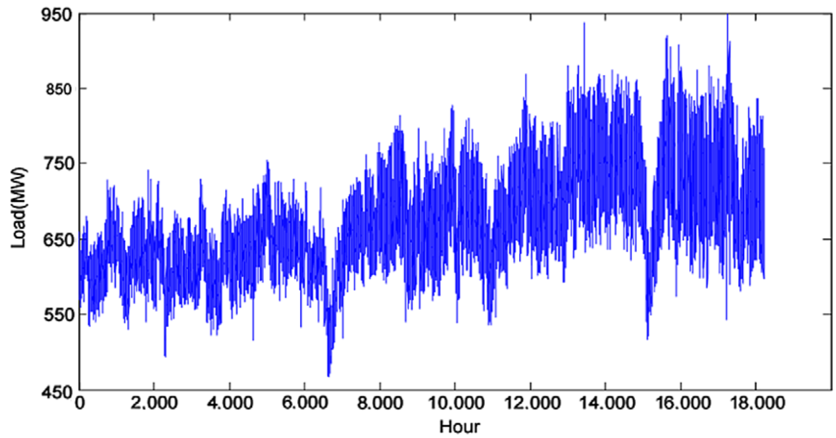

Fig. 9. The distribution of load about Inner Mongolia region 1May 2004–28 May 2006.

relative error of the forecasting load error. $A(i)$ is the actual load at time i. $F(i)$ is the forecasting load at time i.

5.4. Ant colony optimization for attribute reduction

In this section, we assess the performance of the improved algorithm. The platform used for conducting following experiments is a PC with Pentium 4 2.0 GHz CPU, 1 Gmemory, Windows XP as operating system. The experimental parameters are confirmed by large number of experiments. The parameters are set as follows: No_of_ants = 1000, Min_cases_per rule = 5, uncovered_cases = 10, No rules_converg = 10 and circles time = 50.

Fig. 10 is the reduced attribute evolution curve for the sample data. In Fig. 10, the Attribute Number curve describes the number of attribute that acquired by the new ACO model at different evolution iterations; the Reduced No is the number of attribute in reduced process at different evolution iterations with the new method. Fig. 10 contains 6 conditional attributes between 18th iteration and 22nd iteration. The reduced number of attributes increases gradually when the evolution goes on. Until it evolute the 22nd iteration, the algorithm searches the best result and achieves the minimum reduction in the sample data. Then, we remove redundant samples and save the optimum solution into the optimum solution set. In the 29th evolution iteration, the reduced number is stable, and we can see that 21 attributes are dropped from the original 38 attributes.

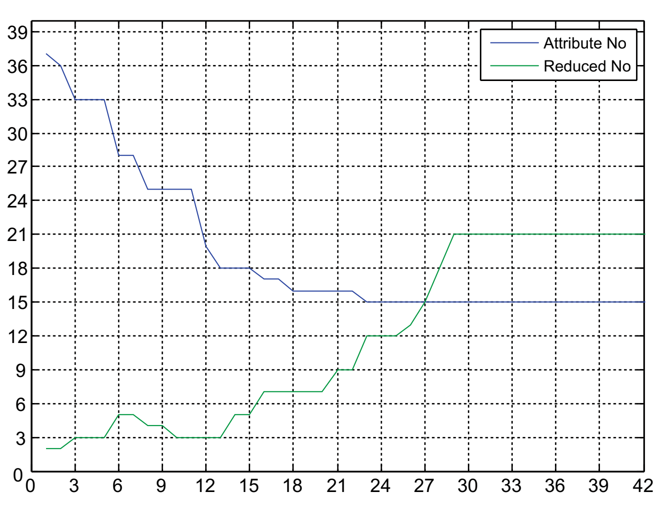

Fig. 10. The curve of algorithm convergence and global optimum.

5.5. SVM forecasting

SVM is used to make prediction after the samples are normalized. Matlab toolbox is used to compute the results, and radial basis function is chosen as the kernel function. The parameters are chosen as the following: $C = 79.83$, $\varepsilon = 0.001$ and $\sigma^2 = 5.3$. In a single SVM without using ant colony, the parameters are chosen as $C = 26.30$, $\varepsilon = 0.006$ and $\sigma^2 = 1.60$.

BP algorithm is used to make prediction with sigmoid function. Empirical trial suggests the following parameter values are appropriate: the node number of input layer, interlayer and output layer are 11, 8 and 1 respectively. The system error is 0.001, and the maximal interactive time is 5000.

With the ACO–SVM model, SVM model and ANN model, the training time of the data are 14s, 32s, 23s, respectively. It can be seen that the training time of each model on disposing the training data is very different. With the SVM model where the data has not been preprocessed, the training time is 32 s. This is 9 s longer than ANN computation. However, in ACO–SVM model, the training time is only 14 s.

Table 2 and Fig. 11 give the result of load forecasting with ANN, SVM and ACO–SVM models. Several patterns are observed.

Firstly, the maximum and minimum deviations are captured between the forecasting value and the actual value with three models. With ANN method, the maximum deviation is −30.67 MW which is on 11:00, and the minimum deviation is 4.54 MW on 1:00. In SVM method, the maximum and minimum deviations are 23.93 MW and −3.57 MW which are on 10:00 and 1:00, respectively. With the new ACO–SVM model that was built in this article, the maximum forecasting deviation is only −23.08 MW on 11:00 in 24 time points, and the minimum deviation is 2.8MW on 4:00. Compared with other two models, the advantage of ACO–SVM model is obvious. It makes more accurate load forecasting with 7 MW than ANN and less accurate load forecasting than SVM in maximum deviation. As to the minimum deviation of ACO–SVM model, it makes advantage forecasting than ANN and SVM models with 1.74 MW and 0.77 MW.

Secondly, the deviation between forecasting value and actual value with these models can be defined the scope of the deviation as [−10 MW, 10 MW]. In ANN method, there are 8 points of forecasting values in this scope, and 9 points with SVM model. But in ACO–SVM method, 14 points are in the scope [−10 MW, 10 MW], and 4 points in other 10 forecasting values are in the scope [10 MW, 11 MW]. It can also be seen that use of ACO–SVM model

Table 2
ACO–SVM forecast results and comparison MW.

Time point	Actual load	ANN		SVM		ACO–SVM	
		Forecast load	Error	Forecast load	Error	Forecast load	Error
T 00_00	551.87	542.38	−1.72%	559.98	1.47%	546.96	−0.89%
T 01_00	540.42	544.96	0.84%	536.85	−0.66%	534.64	−1.07%
T 02_00	523.73	534.26	2.01%	531.22	1.43%	531.27	1.44%
T 03_00	518.40	511.97	−1.24%	527.52	1.76%	523.17	0.92%
T 04_00	518.49	529.53	2.13%	510.45	−1.55%	521.29	0.54%
T 05_00	558.00	549.57	−1.51%	563.47	0.98%	545.28	−2.28%
T 06_00	635.72	657.72	3.46%	621.03	−2.31%	646.46	1.69%
T 07_00	648.51	656.88	1.29%	660.12	1.79%	640.53	−1.23%
T 08_00	682.36	689.05	0.98%	696.76	2.11%	692.80	1.53%
T 09_00	762.83	780.60	2.33%	748.26	−1.91%	771.37	1.12%
T 10_00	737.48	752.08	1.98%	760.49	3.12%	721.92	−2.11%
T 11_00	759.07	728.40	−4.04%	778.12	2.51%	735.99	−3.04%
T 12_00	686.83	700.64	2.01%	664.58	−3.24%	696.86	1.46%
T 13_00	628.71	620.16	−1.36%	638.52	1.56%	634.31	0.89%
T 14_00	646.22	662.57	2.53%	653.26	1.09%	660.50	2.21%
T 15_00	704.81	723.63	2.67%	685.99	−2.67%	697.20	−1.08%
T 16_00	748.12	734.13	−1.87%	762.48	1.92%	755.45	0.98%
T 17_00	764.69	786.33	2.83%	788.62	3.13%	773.71	1.18%
T 18_00	763.03	775.54	1.64%	772.49	1.24%	752.50	−1.38%
T 19_00	772.51	763.86	−1.12%	758.22	−1.85%	779.23	0.87%
T 20_00	845.07	873.89	3.41%	825.13	−2.36%	863.24	2.15%
T 21_00	882.00	869.12	−1.46%	894.97	1.47%	869.39	−1.43%
T 22_00	773.68	799.44	3.33%	794.80	2.73%	782.89	1.19%
T 23_00	634.36	660.50	4.12%	613.93	−3.22%	655.04	3.26%
T 24_00	575.67	591.16	2.69%	588.85	2.29%	567.21	−1.47%
RMSRE			2.18%		2.01%		1.50%

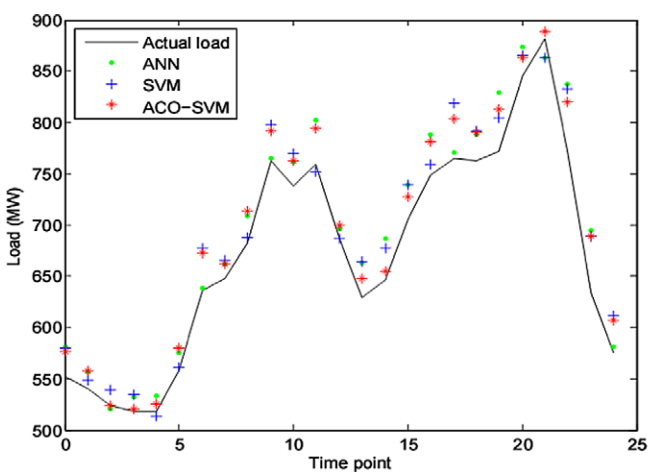

Fig. 11. The forecasting results using the proposed method.

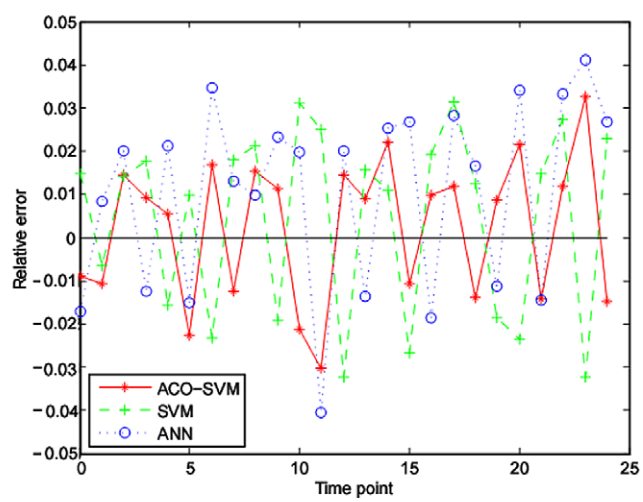

Fig. 12. Error analysis.

to forecast has better stability and accuracy than the other methods. Because it makes pretreatment on the data with the ant colony optimization, the data can be classed into many groups according to different features with this method, which select different features and remove many influential factors that make no or little effect on forecasting. This improves the computer speed and the forecasting accuracy.

Thirdly, we analyze the comparison between forecasting values and actual values with the equation $E = \sum_{i=1}^{n}|A(i) - F(i)|/n$ in these methods. With ANN model, E is 15.58 MW and 14.45 MW with SVM method, but it is 10.63 MW with ACO–SVM model which is 31.77% lower and 28.72% lower than ANN and SVM, respectively. Apparently, the new model can narrow the scope of the deviation and improve the stabilization of load forecasting.

The 24-hour-ahead load prediction and the actual load at 6/3/2006 are shown in Table 2. The 24-hour-ahead load prediction curves and the actual load curve are shown in Fig. 12. It also shows the error analysis of short-term load forecast of power systems using the propose method. In Table 2, the Relative Error comparison is made between the general ANN, the SVM and the ACO–SVM.

We can see from the above computation that comparing with the ANN and SVM method, the ACO–SVM model has higher forecast precision. Evidently, the forecasting errors of the proposed model are lower than the other models.

Comparing with the performance of other two models, it can be seen that the relative errors of the proposed model in this paper are around zero. Out of 24 19 points that means 80% of the forecasting points are in the scope of [−0.02, 0.02]. In general, the preferred interval is [−0.03, 0.03], and the root-mean-square relative error is only 1.50%, which indicates the accuracy of load forecasting is satisfied. But in ANN model, the relative errors are more fluctuant, some of them are very small (e.g., close to 0), but others are very big (more than 4%). We repeat the calculation of the same

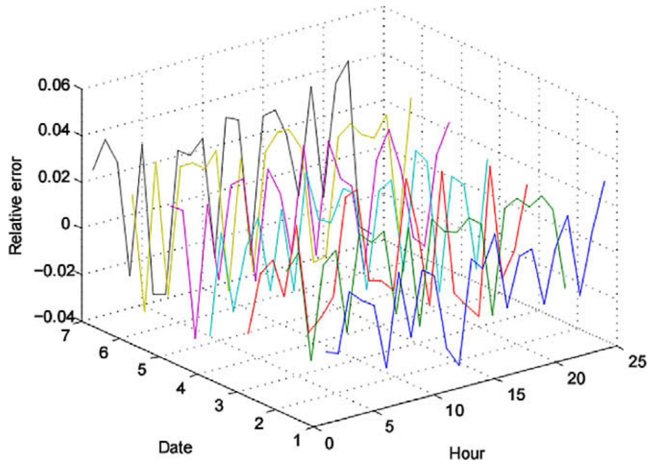

Fig. 13. The surface figure of the performance of model ACO–SVM.

group of data and find that the relative errors have great changes. Sometimes, the relative error is close to 2%, while sometimes it is close to 10%, which shows the instability and local convergent characteristics of ANN model. As can be seen in Table 2, 12 out of 24 points are in the scope [−0.02, 0.02], and 5 of them exceed 3% which do not meet the preferred interval. Similarly, the SVM model also can hardly yield satisfactory calculation. For example, the RMSRE is 2.01%. Out of 24 points 15 are in the scope [−0.02, 0.02], which is not better than ANN. Obviously, the excellence of ACO–SVM model is apparent for the global convergence and preciseness in logic and mathematics. This is validated from Fig. 12. The forecasting load of seven days has been shown in Fig. 12. It can be seen that the load forecasting curve makes small fluctuation from first to fourth day, where the scope of the relative error for the most time points is [−0.025, 0.025]. Only in the last two days, the error of some time point would be more than 3%. The ACO–SVM model makes not only accurate load forecasting, but also the load forecasting curve to be stable in a relatively long time.

Fig. 13 is the surface curve of the load forecasting deviation performance of ACO–SVM model in the next seven days. It depicts the forecasting deviation of every day's 24 time points of the following seven days which ranges from 29/05/2006 to 04/06/2006. From Fig. 13, it can be seen that the forecasting accuracy is very high, and the variable curve of deviation is very stable. With ACO–SVM model, the load forecasting relative error of the second forecasting day is 1.50% which keeps on a high forecasting accuracy, and the following days' relative errors are 2.01%, 1.98%, 2.00%, 2.18%, 2.63% and 2.81%. From the relative error of the seven day computation, the highest error rate is 2.81%, which is in the scope [−3%, −3%] and indicates a well-satisfied load forecasting.

6. Conclusions

We have developed SVM based on ant colony optimization models tp forecast short-term power load. To use this proposed approach, we employed ant colony optimization algorithm to preprocess the data which is subjected to influence of many uncertain factors in short-term power load forecasting. Data are divided into many groups according to the different features. Using Ant Colony Optimization method, we removed some factors with no or little influence on the forecasting. For example, the load of power systems is influenced by factors such as day sort, week sort, day weather sort, day temperature. In this paper, these factors are included when processing the history data with ACO algorithm. The real load data prediction shows that the model is effective in short-term power load forecasting. The computation and comparing with the single ANN and SVM validate the effectiveness for short-term power load forecasting. Further research includes comparison with various traditional optimization models (Wu, 2009a, 2009b).

Acknowledgements

Natural Science Foundation of China (70671039), The Ministry of Education to support the new century talents plan (NCET-07-0281).

References

Bonabeau, E., Dorigo, M., & Theraulez, G. (1999). *Swarm intelligence: From natural to artificial systems*. NY, USA: Oxford University Press Inc.
Brown, R. G. (1983). *Introduction to random signal analysis and Kalman filtering*. New York: John Wiley.
Christianse, W. R. (1971). Short term load forecasting using general exponentials smoothing. *IEEE Transactions of Power Apparatus Systems, PAS-90*, 900–911.
Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*(3), 131–156.
Dorigo, M., Maniezzo, V., & Colomi, A. (1996). The ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems Man Cybernetics Part B, 26*(1), 29–41.
Douglas, A. P., Breipohl, A. M., Lee, F. N., & Adapa, R. (1998). The impact of temperature forecast uncertainty on Bayesian load forecasting. *IEEE Transaction on Power Systems, 13*, 1507–1513.
Gelb, A. (1974). *Applied optimal estimation*. Massachusetts: The MIT Press.
Glover, F. (1989). Tabu search – Part I. *ORSA Journal of Computing, 1*, 190–206.
Goldberg, D. E. (1997). *Genetic algorithms in search optimization and machine learning*. Reading, Massachusetts: Addison Wesley.
Jensen, R., & Shen, Q. (2005). Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets and Systems, 149*, 5–20.
Kirkpatrick, S., Gelatt, C., Jr., & Vecchi, M. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.
Lin, C.-M., & Gen, M. (2007). Multi-objective resource allocation problem by multistage decision-based hybrid genetic algorithm. *Applied Mathematics and Computation, 187*, 574–583.
Maniezzo, V., & Colomi, A. (1999). The ant system applied to the quadratic assignment problem. *Knowledge Data Engineering, 11*(5), 769–778.
Mbamalu, G. A. N., & El-Hawary, M. E. (1993). Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. *IEEE Transaction on Power Systems, 8*, 343–348.
Moghram, I., & Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transaction on Power Systems, 4*, 1484–1491.
Osman, M. S., Abo-Sinna, M. A., & Mousa, A. A. (2005). An effective genetic algorithm approach to multi-objective resource allocation problems. *Applied Mathematics and Computation, 163*, 755–768.
Pai, P.-F., & Hong, W.-C. (2005). Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management, 46*, 2669–2688.
Park, J. H., Park, Y. M., & Lee, K. Y. (1991). Composite modeling for adaptive short-term load forecasting. *IEEE Transaction on Power Systems, 6*, 450–457.
Sadownik, R., & Barbosa, E. P. (1999). Short-term forecasting of industrial electricity consumption in Brazil. *Journal of Forecasting, 18*, 215–224.
Wu, D. (2009a). Performance evaluation: An integrated method using data envelopment analysis and fuzzy preference relations. *European Journal of Operational Research, 194*(1), 227–235.
Wu, D. (2009b). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications, 36*(5), 9105–9112.