A Novel Hybrid Method for Short Term Load Forecasting using Fuzzy Logic and Particle Swarm Optimization

Amit Jain, Member, IEEE, M. Babita Jain, Member, IEEE and E. Srinivas

Abstract-- Load forecasting has become a very crucial technique for the efficient functioning of the power system. This paper presents a methodology for the short term load forecasting problem using the similar day concept combined with fuzzy logic approach and particle swarm optimization. To obtain the next-day load forecast, fuzzy logic is used to modify the load curves of the selected similar days of the forecast previous day by generating the correction factors for them. These correction factors are then applied to the similar days of the forecast day. The optimization of the fuzzy parameters is done using the particle swarm optimization technique on the training data set of the considered data set. A new Euclidean norm with weight factors is proposed for the selection of similar days. The proposed methodology is illustrated through the simulation results on a typical data set.

Index Terms-- Euclidean norm, Fuzzy logic approach, Particle swarm optimization, Short term load forecasting, similar day method.

I. INTRODUCTION

Load forecasting has been an integral part in the efficient planning, operation and maintenance of a non interrupting regulated power system. Short term load forecasting is necessary for the control and scheduling operations of a power system and also acts as inputs to the power analysis functions such as load flow and contingency analysis [1]. Power utilities in India still depend on the averaging of set of similar days for short term load forecasting, the results of which are very far from being accurate. Owing to this importance, various methods have been reported, that includes linear regression, exponential smoothing, stochastic process, ARMA models, and data mining models [2]-[7]. Of late, artificial neural networks have been widely employed for load forecasting. However, there exist large forecast errors using ANN when there are rapid fluctuations in load and temperatures [8]-[9]. In such cases, forecasting methods using fuzzy logic approach have been employed. In this paper, we propose an approach for short term load forecasting problem, using a hybrid technique combining the particle swarm optimization (PSO) and fuzzy logic approach. The PSO technique which roots from the concept of Computational Intelligence capable of dealing with complex, dynamic and poorly defined problems that AI has problem with, has an advantage of dealing with the nonlinear parts of the forecasted load curves, and also has the ability to deal with the abrupt change in the weather variables such as temperature, humidity and also including the impact of the day type. In this method, we select similar days from the previous days to the forecast day using Euclidean norm (based on weights, which are calculated such that they hold the essence of the earlier data) with weather variables [10]. There may be a substantial discrepancy between the load on the forecast day and that on similar days, even though the selected days are very similar to the forecast day with regard to weather and day type. Therefore, the selected similar days cannot be averaged to obtain the load forecast. To avoid this problem, the evaluation of similarity between the load on the forecast day and that on similar days is done using fuzzy logic combined with particle swarm optimization. Many methods have been reported for the load forecast using fuzzy logic [12]-[14]. This approach evaluates the similarity using the information about the previous forecast day and previous similar days. The similarity resulted with a high MAPE (mean average percentage error) value. The fuzzy inference system is used for the reduction of this MAPE and generation of suitable correction factors. The fuzzy system thus generates a correction factor for the load curve on a similar day to the shape of that on the forecast day. The fuzzy parameters of the fuzzy inference system are optimized using the PSO technique implemented in the MATLAB program of the short term load forecasting. The PSO technique is thus used to evolve the initial best parameters for the fuzzy inference system. The optimization of the fuzzy parameters using PSO is done using the training data set of the available data set and the fuzzy inference system with the optimized parameters is then used on the testing data set. For the testing data set, first the correction factors of load curves on similar days are calculated using the optimized fuzzy inference system for the similar days of the forecast previous day, then the forecast load is obtained by averaging the corrected load curves on similar days using the previous day’s correction factors. The correction factors are obtained using the training data set and then they are applied to the testing data set. The approach suitability is verified by applying it to a typical real data set. The overall MAPE was found to be very less (<0.03) for the considered data set. This paper contributes to the short term load forecasting, as it shows how the forecast can include the effect of weather variables, temperature as well as humidity and day type using a new Euclidean norm for the selection of the similar days and PSO.
optimized fuzzy logic approach for obtaining the correction factors to the similar days. The paper is organized as follows: section II deals with the data analysis; section III gives the overview of the proposed forecasting method, discussing the selection of similar days using a new Euclidean norm and the PSO optimization of the fuzzy inference system is presented; section IV presents the optimization of fuzzy parameters using particle swarm optimization; section V simulation results of the proposed forecasting methodology followed by conclusions in section VI.

II. VARIABLES IMPACTING THE LOAD PATTERN

The analysis on the monthly load and weather data helps in understanding the variables, which may affect load forecasting. The data analysis is carried out on data containing hourly values of load, temperature, and humidity of 7 months. In the analysis phase, the load curves are drawn and the relationship between the load and weather variables is established [11]. Also, the week and the day of the week impact on the load is obtained.

A. Load Curves

The load curve for the month of May is shown in Fig. 1. The observations from the load curves are as follows:
1. There exists weekly seasonality but the value of load scales up and down.
2. The load curves on week days are similar
3. The load curves on the weekends are similar.
4. Days are classified based on the following categories:
   a. Normal week days (Tuesday - Friday)
   b. Monday
   c. Sunday
   d. Saturday

Monday is accounted to be different to weekdays so as to take care for the difference in the load because of the previous day to be weekend.

B. Variation of Load with Temperature

Temperature is the most important weather variable that affects the load. The deviation of the temperature variable from a normal value results in a significant variation in the load. Fig. 2 and Fig. 3 show the relationship between the temperature and load. Fig. 2 shows a plot between the average temperatures versus maximum demand and the average demand versus average temperature graphs as shown. Fig 4 shows the plot between the average humidity versus maximum demand. Fig 5 shows the plot between the average humidity versus average demand. From the graphs it can be seen that there exists a positive correlation between load and humidity for the forecast month of July i.e. demand increases as the humidity increases.

C. Variation of Load with Humidity

Another weather variable that affects the load level is humidity. To study the effect of this particular weather variable on load we plot the maximum demand versus average humidity and the average demand versus average humidity graphs as shown. Fig 4 shows the plot between the average humidity versus maximum demand. Fig 5 shows the plot between the average humidity versus average demand. From the graphs it can be seen that there exists a positive correlation between load and humidity for the forecast month of July i.e. demand increases as the humidity increases.

---

**Fig. 1 Load Curve for the month of May**

**Fig. 2 Plot between Maximum Load versus Average Temperature**

**Fig. 3 Plot between Average Load versus Average Temperature**

**Fig. 4 Plot between Maximum Load versus Average Humidity**

**Fig. 5 Plot between Average Load versus Average Humidity**
D. Autocorrelation of Load

It is known that the load at a given hour is dependent not only on the load on the previous hour but also on the load at the same hour of the previous day. Hence, it is assumed that the load curve is more or less similar to the load curve on the previous day.

III. LOAD FORECASTING USING FUZZY LOGIC

A. Similar Day selection

In this paper, Euclidean norm with weight factors is used to evaluate the similarity between the forecast day and the searched previous days. Euclidean norm makes us understand the similarity by using the expression based on the concept of norm. Decrease in the Euclidean norm results in the better evaluation of the similar days i.e., smaller the Euclidean norm the more similar are the days to the forecast day. In general, the Euclidean norm using maximum and minimum temperatures along with the day type variable is used for the evaluation of the similar days. But, the norm using maximum and minimum temperatures is not efficient for the selection of the similar days because humidity is also an important weather variable as also shown in section II C.

In the present work, we have proposed a new Euclidean norm to account for the humidity. The new Euclidean norm uses average temperature, average humidity and day type with weight factors to evaluate the similarity of the searched previous days. The expression for the new Euclidean norm is as follows:

$$EN = \sqrt{w_1 T_{\text{max}}^2 + w_2 H_{\text{avg}}^2 + w_3 D^2}$$

(1)

$$\Delta T_{\text{max}} = T_{\text{max}} - T_p$$

(2)

$$\Delta H_{\text{avg}} = H_{\text{avg}} - H_p$$

(3)

$$\Delta D = D - D_p$$

(4)

Where, Tavg and Havg are the forecast day average temperature and average humidity respectively. Also, Tpavg and Hpavg are the average temperature and average humidity of the searched previous days and w1, w2, w3 are the weight factors determined by least squares method based on the regression model constructed using historical data [2]. The similar days are selected from the previous 30 days of the forecast day. The data selection is limited to account for the seasonality of the data. The day types considered for the methodology are 4(Tuesday-Friday), 3 (Monday), 2(Saturday), 1(Sunday).

B. Fuzzy Inference System

The load forecasting at any given hour not only depends on the load at the previous hour but also on the load at the given hour on the previous day. Also, the Euclidean norm alone is not sufficient for the load forecast as the selected similar days for the forecast day have considerably large mean absolute percentage error (MAPE). Assuming same trends of relationships between the previous forecast day and previous similar days as that of the forecast day and its similar days, the similar days can thus be evaluated by analyzing the previous forecast day and its previous similar days.

The fuzzy inference system is used to evaluate the similarity between the previous forecast days and previous similar days resulting in correction factors, used to correct the similar days of the forecast day to obtain the load forecast. To evaluate this degree of similarity, three fuzzy input variables for the fuzzy inference system are defined [10].

$$E_L^k = L_p - L_{ps}$$

(5)

$$E_T^k = T_p - T_{ps}$$

(6)

$$E_H^k = H_p - H_{ps}$$

(7)

Where, Lp and Lps are the average load of the previous forecast day and the previous kth similar day, Tp, Tps, Hp, Hps show the value corresponding to temperature and humidity respectively. E_L, E_T, E_H take three fuzzy set values; Low (L), Medium (M), High (H). The membership functions of the input variables and output variable are as shown in Figs 6-7. The fuzzy rules for the inference system for the given fuzzy variables are based on the generalized knowledge of the effect of each variable on the load curve [11]. If the membership of E_L is U_E_L, that of E_T is U_E_T and that of E_H is U_E_H, the firing strength, u, of the premise is calculated based on the min operator. The firing strength of each rule is calculated as follows:

$$\mu_i = \min(\mu_{E_L}, \mu_{E_T}, \mu_{E_H})$$

(8)
The membership function of an inferred fuzzy output variable is calculated using a fuzzy centroid defuzzification scheme to translate fuzzy output statements into a crisp output value, \( W_k \).

\[
W_k = \sum_{i=1}^{27} \alpha_i \mu_i^k / \sum_{i=1}^{27} \mu_i^k
\]

(9)

The output value is expressed by \( W_k \) which is the correction factor for the load curve on the kth similar day to the shape on the forecast day. \( W_k \) is applied to each similar day and corrects the load curve on similar days. The forecast next day load curve \( L(t) \) is then given by averaging the corrected loads on similar days.

\[
L(t) = \frac{1}{N} \sum_{k=1}^{N} (1 + W_k) L_k(t)
\]

(10)

Where \( L_k(t) \) is the power load at \( t \) o'clock on the kth corrected similar day, \( N \) is the number of similar days and \( t \) is the time from 1 to 24.

### IV. OPTIMIZATION OF FUZZY PARAMETERS USING PARTICLE SWARM OPTIMIZATION

PSO is a population-based optimization method first proposed by Eberhart and Colleagues [15, 16]. Some of the attractive features of PSO include the ease of implementation and the fact that no gradient information is required. It can be used to solve a wide array of different optimization problems. Like evolutionary algorithms, PSO technique conducts search using a population of particles, corresponding to individuals. Each particle represents a candidate solution to the problem at hand. In a PSO system, particles change their positions by flying around in a multidimensional search space until computational limitations are exceeded.

The PSO technique is an evolutionary computation technique, but it differs from other well-known evolutionary computation algorithms such as the genetic algorithms. Although a population is used for searching the search space, there are no operators inspired by the human DNA procedures applied on the population. Instead, in PSO, the population dynamics simulates a 'bird flock's' behavior, where social sharing of information takes place and individuals can profit from the discoveries and previous experience of all the other companions during the search for food. Thus, each companion, called particle, in the population, which is called swarm, is assumed to 'fly' over the search space in order to find promising regions of the landscape. For example, in the minimization case, such regions possess lower function values than other, visited previously. In this context, each particle is treated as a point in a d-dimensional space, which adjusts its own 'flying' according to its flying experience as well as the flying experience of other particles (companions). In PSO, a particle is defined as a moving point in hyperspace. For each particle, at the current time step, a record is kept of the position, velocity, and the best position found in the search space so far.

The assumption is a basic concept of PSO [16]. In the PSO algorithm, instead of using evolutionary operators such as mutation and crossover, to manipulate algorithms, for a d-variables optimization problem, a flock of particles are put into the d-dimensional search space with randomly chosen velocities and positions knowing their best values so far (Pbest) and the position.

In the d-dimensional space, the velocity of each particle, adjusted according to its own flying experience and the other particle's flying experience. For example, the \( i \)th particle is represented as \( x_i = (x_{i1}, x_{i2},..., b_{id}) \) in the dimensional space. The best previous position of the \( i \)-th particle is recorded and represented as: \( \text{Pbest}_i = (\text{Pbest}_{i1}, \text{Pbest}_{i2},... \text{Pbest}_{id}) \)

The index of best particle among all of the particles in the group is \( \text{gbest}_{id} \). The velocity for particle \( i \) is represented as \( \text{vi} = (v_{i1}, v_{i2},..., v_{id}) \). The modified velocity and position of each particle can be calculated using the current velocity and the distance from \( \text{Pbest}_{id} \) to \( \text{gbest}_{id} \) as shown in the following formulas [17]:

\[
v[i] = v[i] + c1 * \text{rand()} * (\text{pbest}[i] - \text{present}[i]) + c2 * \text{rand()} * (\text{gbest}[i]-\text{present}[i])
\]

(11)

\[
\text{present}[] = \text{present}[] + v[i]
\]

(12)

Where

\( v[i] \) is the particle velocity, \( \text{present}[] \) is the current particle (solution), \( \text{pbest}[] \) and \( \text{gbest}[] \) are defined as stated before. \( \text{rand()} \) is a random number between (0,1). \( c1, c2 \) are learning factors. Usually \( c1 = c2 = 2 \).

The evolution procedure of PSO Algorithms is shown in Fig. 8. Producing initial populations is the first step of PSO. The population is composed of the chromosomes that are real codes. The corresponding evaluation of populations called the "fitness function". It is the performance index of a population. The fitness value is bigger, and the performance is better.

After the fitness function has been calculated, the fitness value and the number of the generation determine whether or not the evolution procedure is stopped (Maximum iteration number reached?). In the following, calculate the Pbest of each particle and Gbest of population (the best movement of all particles). The update the velocity, position, gbest and pbest of particles give a new best position (best chromosome in our proposition).
For the data set considered the fuzzy inference system has been optimized for six parameters (maxima and minima of each of the input fuzzy variable EL, ET, EH), considering 49 particles. Hence each particle is a six dimensional one. The initial values of the fuzzy inference system are obtained by using the May and June data. These values are incorporated into the fuzzy inference system to obtain the forecast errors of June month.

The particle swarm optimization function accepts the training data i.e. May and June 99, and the objective is to reduce the RMS MAPE error of the forecast days(June 99) using May data. The RMS MAPE is taken as the fitness function and the particle swarm optimizer function is run for 100 iterations(by then the RMS MAPE) is more or less fixed. After every iteration the optimizer updates the latest particle position using the optimizer equations based on the PBest and Gbest of the previous iteration if the fitness function value is better than the previous one. The parameters thus obtained after the PSO optimization are the final fuzzy input parameters and these are used to forecast the load of the testing data set i.e. July month.

V. SIMULATION RESULTS

The performance of the method for the short term load forecast is tested by using the 7 months data, from January to July of a particular data set used. The method has been simulated using the fuzzy logic toolbox available in MATLAB. Load forecasting is done for the month of July.

The fuzzy inference system is designed for the computer simulation. First, fuzzy parameters are s based on the existing training data set, as described in the above section. Hence, the data of the month June has been used for the selection of similar days. The number of similar days used for the forecasting is five.
The forecast results deviation from the actual values are represented in the form of MAPE. Mean Absolute Percentage Error (MAPE) is defined as:

\[ MAPE = 100 \times \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_A^i - P_F^i}{P_A^i} \right| \]  

(13)

\( P_A, P_F \) are the actual and forecast values of the load. \( N \) is the number of the hours of the day i.e. \( N = 1, 2, \ldots, 24 \). With the proposed method the MAPE error for the considered days, for which forecasted results are shown in Fig. 9-12, is calculated and these are as follows in Table 5:

<table>
<thead>
<tr>
<th>Day</th>
<th>MAPE Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 July (Saturday)</td>
<td>2.3574</td>
</tr>
<tr>
<td>25 July (Sunday)</td>
<td>1.9855</td>
</tr>
<tr>
<td>26 July (Monday)</td>
<td>1.7404</td>
</tr>
<tr>
<td>27 July (Tuesday)</td>
<td>1.391</td>
</tr>
</tbody>
</table>

In the present work, a novel PSO optimized fuzzy logic system for short term load forecasting is developed to forecast the load for next day. In the fuzzy inference system, the optimization of membership functions was required, as it reduces the MAPE and so the input parameters limits are adjusted in optimal values so as to give a MAPE error of low values. The computer MATLAB simulation demonstrates that the fuzzy inference system associated to the PSO algorithm approach became very strong, giving good results and possessing good robustness and the results will improve in quality with the use of historical data of more number of years.

VI. CONCLUSION

In this paper an optimal fuzzy logic is designed using PSO algorithm. The fuzzy inference system with PSO algorithm is better than the traditional fuzzy inference system without PSO algorithms as we observed during our simulation study and proposed system provides better improvement and good robustness. The system takes into account the effect of humidity as well as temperature on load and a fuzzy logic based method is used to correct the similar day load curves of the forecast day to obtain the load forecast. Also, a new Euclidean norm with weight factors is proposed, which is used for the selection of similar days. Fuzzy logic is used to evaluate the correction factor of the selected similar days to the forecast day using the information of the previous forecast day and the previous similar days.

To verify the forecasting ability of the proposed methodology, we did simulation to do load forecasting for the month of July in a data set of 7 months and results for four representative days of a week in the month of July are given. The results obtained from the simulation show that the proposed forecasting methodology, which proposes the use of weather variables i.e. temperature as well as humidity, gives load forecasting results with reasonable accuracy. The proposed methodology would certainly give improved results when it is implemented on historical data set of many years.
which we will present in our next research paper. Authors hope that the proposed methodology will be helpful in using more weather variables, which will certainly be better than using only temperature as the weather variable that affects the load, in short term load forecasting and will further propagate research for short term load forecasting using new optimization techniques like PSO.

VII. REFERENCES


VIII. BIOGRAPHIES

Amit Jain graduated from KNIT, India in Electrical Engineering. He completed his master’s and Ph.D. from Indian Institute of Technology, New Delhi, India. He was working in Alstom on the power SCADA systems. He was working in Korea in 2002 as a Post-doctoral researcher in the Brain Korea 21 project team of Chungbuk National University. He was Post Doctoral Fellow of the Japan Society for the Promotion of Science (JSPS) at Tohoku University, Sendai, Japan. He also worked as a Post Doctoral Research Associate at Tohoku University, Sendai, Japan. Currently he is Head of Power Systems Research Center at IIIT, Hyderabad, India. His fields of research interest are power system real time monitoring and control, artificial intelligence applications, power system economics and electricity markets, renewable energy, reliability analysis, GIS applications, parallel processing and nanotechnology.

M Babita Jain is PhD student at the International Institute of Information Technology, Hyderabad, India since 2006. The areas of interest include IT Applications to Power Systems, Load forecasting and Multi Agent Systems. She has actively involved in establishing IEEE student branch and Women in engineering affinity group activities.

E. Srinivas is a Master's Student in Power System research Center, International Institute of Information Technology, Hyderabad, India. He received his B. Tech degree from Sri Indu College of Engineering and Technology, Ibrahimpatnam. His areas of interest include applications of computational intelligence techniques to power systems and operations planning of power systems.