An Artificial Chromosomes Embedded Genetic Algorithms for Smart Grid Power Demand Forecast

Yen-Wen Wang
Department of Marketing and Distribution Management, Chien-Hsin University, Taiwan
Email:ywwang@uch.edu.tw

Abstract—This research develops a hybrid model by integrating K-mean, Neural Network, and Artificial Chromosomes embedded Genetic Algorithms to forecast the electrical load. This hybrid model encompasses two novel concepts: 1. Under the expectations of load balancing, we clustering 272 electrical substations into different clusters, and develop a forecast model of each cluster. 2. Instead of the generation mechanisms of Genetic Algorithms, the artificial chromosome is injected to seek a higher accurate. Numerical data of various affecting factors and actual electrical load of 48 months of 272 electrical substations are collected and fed into the hybrid model for future monthly amount. Experimental results show a higher accurate of our model compare with other two traditional forecasting model.

Index Terms—power demand forecasting, genetic algorithms, smart grid

I. INTRODUCTION

With the energy shortage problem, rocketing prices, and environmental awareness, the power supply, reliability, extension of transmission lines, greenhouse gases, electricity company liberalization, and the growth of electricity technology, the world power supply system have gradually convert from centralized power to distributed power. The small distributed power supply close to the client and introducing new energy became the alternatives and substitutes for the traditional huge centralized power supply system. Because of this, Smart Grid related topics began to emerge. Smart Grid means the integration of power, electricity transmission, power transmission and its usage in the client side, combine with automation and information technology superiority. It can achieve self-monitoring, diagnosis, repair, and other functions. The power company can use the remote monitoring system to find out the electricity consumption in order to adjust the amount of power needs to be deployed. The clients can also grasp their own power consumption to adjust power usage to achieve energy saving and money saving policy.

In order to have an ideal Smart Grid deployment, the first is to understand the national and regional power trend, with the development of the basic infrastructure to integrate a complete project. In the recent years, many big countries have begun to conduct major research and construct basis in this area. Since 2001, United States have funded about 20 million in Intelligrid project. The emphasis is on the development of the software architecture and distributed control system interface maintenance and improvement of Smart Grid. From 2002 to 2006, 50 million has been funded for the research. Japan focus is to integrate power supply for a district with the new power system. The new energy system provides technological management to import power for large number of distributed power to conduct research and data analysis. The project is demonstrated in the regional power grid. The total amount of investment is more than 14.7 billion yen. In Europe, European Union has been the major region for renewable energy in the world. It is expected to continually invest 500 billion euros in the transmission and distribution upgrade and update. Therefore, the promotion of Smart Grid requires a long-term planning and investment. Under these conditions, the demand on the control of the resources required and the energy needed to be accurately forecasted has become an important task.

In general, electricity demand forecasting can be divide into long, medium, and short term power load forecasting. Since electricity cannot be stored, in order to stabilize the power supply and avoid power shortage causing significant loss to the industry. Long term electric power development plan must meet the industry’s future development and meet the people’s needs. It also needs to maintain a suitable backup power source for the changes in the load demand. Therefore, long term load forecasting provides the company with a long term planning needs for the company’s future development. Medium load forecasting is integrated below long term power development plan. It considers the need for power growth and applied to regular maintenance. It takes into account the ratio of the power structure, power supply stability, safety, environmental issues, regional factors such as supply and demand balance for electricity distribution to fully uses the power capacity. Short term load forecasting provides machine coordination, power generation reservations, economic dispatch, load management and other important basis. For example, during the summer...
peak demand season, the power company has to meet the demand of the highest electricity consumption during peak hours. Therefore, the power plant with the highest specification and power output has to be built on standby. However, after the peak hour, these power plants and equipment with the highest specification cannot be used anywhere. If there is an accurate short-term load forecasting, the power company can immediately control and suppress the electrical load during peak hour. The power plants with high specifications and equipment do not need to be built to achieve power saving and carbon reduction.

II. LITERATURE REVIEW

In the development of smart grid related fields, the power load forecasting has a very important role. The predicted results can serve as a future basis for national power development plan, electricity demand-side management (DSM, Demand Side Management), and energy procurement. This type of issue can be divided into long, medium and short-term forecasts (Willis 2002). Long-term forecast generally can be applied to the ratio of the power structure, stable fuel supply, energy security, environmental protection, regional supply and demand balance for future long-term power development plan assessment and planning (Hong 2009) (Hong 2008). Medium-term and short-term forecasts can be the discussion topic for energy procurement, DSM, and other topics.

If we distinguish the power load forecasting based on the size of the region, it can be divided into large regional and small regional forecasts. Small regional forecasting problems contains considerable uncertainties especially in short-term load forecasting. Therefore, the majority of small regional forecasting studies have mainly focused on the long-term forecasting. The previous literature indicates that there are many problems when performing load forecasting. The regional development and land use situation has a significant influence on the region's future growth in electricity consumption. Therefore the use of geographic information systems (GIS, Geographic Information System) for analysis of land use in a small regional load forecasting related research is fairly often [1][2]. Most of the studies only focus on the region overall electricity load demand for forecasting. It cannot focus on individual demand and provide forecasting based on each unit[3]. Lo [4] has pointed out that a good predictive model needs to have the following five characteristics:

- In different load problems, choose the best predictive models.
- In different models, choose the best combination of parameters.
- The predictive model needs to include external factors that can affect the electrical load (such as temperature, cold and other external factors).
- Have a complete data analysis and interpretation capabilities.
- Models need to have the ability to previously forecast.

In general prediction model can be divided into two types. The first one belongs to the statistical model, the other is the predictive model developed based on artificial intelligence tools. Linear statistical model assumes the predicted subject can be applied with linear analysis. The conventional methods include Similar day, time series, regression, and others. Similar day is one of the earliest prediction model. It is easy to use, so many predictions related research still use it to make predictions in order to get a good prediction results. Weron [5] have discussed about this method applied to power forecasting in his book. Time Series is a model used in the short-term forecasting. It is often criticized that the analysis interpreted by such a method does not have a very intuitive relationship with the data itself, but it makes good predictions and many studies still use it for prediction (Huang [6]). Multiple linear regression (MLR) are most frequently used in the short and medium load forecasting models (Papalexopoulos [7], Krogh [8], Heinemann [9]). From the literature review, we learn that although MLR provides good prediction, but in the growing demand for high accuracy in the case, many of the traditional forecasting methods have gradually been retired. In recent years, the linear model used are constantly improving. Hong [10] and Baran [11] have created different forecasting models based on the use of MLR.

III. RESEARCH METHODS

This research will focus on the current domestic demand for electricity and a variety of power sources while considering the situation of load balancing.

![Study flow chart](image)

Figure 1. Study flow chart

There are three main stages in this research (as shown in Fig. 1) and the first stage is the variables selection stage. This stage is to select many possible variables, which may influence power supply. The second stage is the data preprocessing stage. K-mean cluster technique would be adopted. The parameter of cluster number $k$ need to be determined first, after experimental design, the best cluster...
number will be adopted to measure the testing data. The last stage is the forecasting stage, the Artificial Chromosome in an Evolving Neural Network (referred to in this study AC-ENN) will be used to forecast regional electricity load. After being compared with other three forecasting models, the superior model will be recommended to the decision makers. The details of each stage will be described as follows:

This study will be divided into the following three stage:

A. First Stage: Data Collection

This study will first consider load balancing in Taiwan national power plant to subdivide into regions. First of all we target at all levels of Taiwan's current power plants (such as: nuclear, fire, water, wind ... etc.) and investigate its ability to provide power supply. The data for regional demand for electricity across the country also needs to be collected.

B. Second Stage: Cluster the Regional Electricity Supply and Demand Data

In the past, the traditional power division in Taiwan is based on the administrative and geographical position, so the region can be divided into northern, central, southern, and eastern regions. However, Smart Grid architecture concept is that the regional electricity supply can support each other to meet peak load. The power loss from the transmission between regions became a great cost. Therefore, this study examines the district electricity demand and supply information to apply K-mean models to cluster the data. We expect to reach load balancing expectations in the districts and minimize the district electricity power transmission losses achieve optimal partition group.

The K-mean clustering method starts with k initial seeds of clustering, one for each cluster. All the n objects are then compared with each seed by means of the Euclidean distance and assigned to the closest cluster seed. The accuracy of the K-mean procedure is very dependent upon the choice of the initial seeds. To obtain better performance the initial seeds should be very different among themselves. One efficient strategy to improve the K-mean performance is to use, for example, the Ward’s procedure first to divide the n objects into k groups and then use the average vector of each of the k groups as the initial seeds to start the K-mean. As all the agglomerative clustering procedures, this method is available in a majority of statistical software.

C. Third Stage: Forecast the Electrical Load for Each Region

In the second stage, we cluster the regional electricity supply and demand data, then we forecast the electrical load for each region. This study uses AC-ENN forecasting model.

In an optimization problem, GA is one of the most commonly used algorithms. GA utilizes the concept of natural selection with evolution process and through mating, mutation mechanisms to optimize the results. The algorithm has been developed for many years, but its often criticized for its slow convergence and cannot escape local optimum. Therefore, in recent years, our team specialize in Mining Gene Structures (MGS) to attain a better solution and improve its quality [12]-[14]. The following will introduce the development process of MGS by our team.

The research team proposed a gene structure mining method in 2005 (Chang [12]). We improve the algorithm from two factors of the genes, duplication and fabrication. The duplication process speeds up the convergence and avoid falling into local optimum. We gather the elite chromosomes and the information contained within to generate a new set of artificial solution set, which is the artificial chromosome collection. In the collection of information, a dominance matrix is used to create votes for elite solution to record each visit to various cities in correct sequence. This dominance matrix with the recorded information is then used to construct an artificial chromosome. This study used this algorithm in Traveling Salesman Problem, TSP. The study confirmed by applying the gene structure mining method on the genetic algorithm, it can accelerate the speed for convergence and obtain a high quality.

```
Main Procedure ()
Generations: Number of generations
Interval: The interval produced AC
Initiate Population
Construct Initial Population(Population)
Removed Identical Solution()
counter = 0
while counter < generations do
    Evaluate Objectives and Fitness()
    Find Elite Solutions(i)
    Selection with Elitism Strategy()
    If counter < Interval or counter % Interval != 0
        do
            Crossover()
            Mutation()
            Total Replacement()
        Else
            Calculate Average Fitness()
            Collect Gene Information()
            Generate Artificial Chromosomes()
            Replacement()
        End if
    counter = counter + 1
End while
```

Figure 2. GA-AC’s pseudo code

In 2007, MGISPNGA is proposed to solve combinatorial problems [15]. It is focused on improving some operations for SPGA. Lastly, it is combined with genetic structure mining mechanism to solve multi-objective flowshop scheduling problem, multi-objective parallel machine scheduling problem, and multi-objective knapsack problem. Its gene structure mining method can be divided into simple genetic structure mining (SMGS), weighted gene structure mining (WMGS), and threshold-type gene structure mining (TWMGs). It is compared with three evolutionary algorithms, SPGA, NSGA-II, and SPEA2. Their performance of the solution is discussed with three
different metrics, D1r metric, R metric and C metric. The experimental results show that MGISPGA can obtain a solution that contains both good convergence and scalability Plato solution. Within the three different gene structure mining approach, TWMGS obtains the best solution.

This research will use forecasting model and apply AC as a parameter for evolution mechanism. It provides power load forecasting based on the clustering result and make adjustments to improve the accuracy of the forecasting model. The detail introduce of AC mechanism can refer to the reference [14].

The GA-AC pseudo code is shown in Fig. 2, and the flow chart is shown in Fig. 3:

The data in this research are the electrical load of each electrical substation within 4 years. There are 271 substations in Taiwan area and monthly electrical load is considered as an objective of the forecasting model. Other variables might influence the electrical load include: Metal/Machinery Industry production Index, Industrial production index, Chemical Industry production index, Traditional industries production index, Electricity and gas supply production index, Water supply production index, Buildings construction production index, Composite leading index, Industrial production index, Electric power consumption, Index of Producer's Shipment for Manufacturing and the last year power supply of 271 electrical substation.

In the cluster the regional electricity supply and demand data stage, we used 15 clusters in the following research. The number of electrical substations in each area is shown in Table I and the distribution figure is shown in Fig. 4.

<table>
<thead>
<tr>
<th>Cluster</th>
<th># of electrical substations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>53</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
</tr>
</tbody>
</table>

This research develops a clustering AC-ENN for electrical load and we compare this method with other traditional methods such as Back-propagation network (BPN) and Evolving Neural Network (ENN).

MAPE is applied as a standard performance measure for all different models in this research. The forecasting result of each cluster is shown in Table II, AC-ENN is found to be quite accurate. These results are found to be superior to other three forecasting models. By the advanced non-linear problem forecasting ability of neural network and the powerful searching capability of genetic algorithm, AC-ENN model achieves very good accuracy for electrical load problem.

**Figure 3.** A flowchar of GA-AC algorithm

**Figure 4.** A distribution of 15 clusters of K-mean

**IV. RESEARCH RESULTS**

The data in this research are the electrical load of each electrical substation within 4 years. There are 271 substations in Taiwan area and monthly electrical load is considered as an objective of the forecasting model. Other variables might influence the electrical load include: Metal/Machinery Industry production Index, Industrial production index, Chemical Industry production index, Traditional industries production index, Electricity and gas supply production index, Water supply production index, Buildings construction production index, Composite leading index, Industrial production index, Electric power consumption, Index of Producer's Shipment for Manufacturing and the last year power supply of 271 electrical substation.

This research develops a clustering AC-ENN for electrical load and we compare this method with other traditional methods such as Back-propagation network (BPN) and Evolving Neural Network (ENN).

MAPE is applied as a standard performance measure for all different models in this research. The forecasting result of each cluster is shown in Table II, AC-ENN is found to be quite accurate. These results are found to be superior to other three forecasting models. By the advanced non-linear problem forecasting ability of neural network and the powerful searching capability of genetic algorithm, AC-ENN model achieves very good accuracy for electrical load problem.
Although, the experimental in this research still can be improved, but the experimental results in section 4 demonstrated the effectiveness of the AC-ENN that is superior to other traditional approaches. The AC-ENN approach also provides another informing tool to the decision maker in electrical load.

As shown in Table II, we infer that most of bad forecasting results appear in the densely populated area, such as cluster 0 (Taichung city, the most central big city in Taiwan), cluster 7, cluster 11 (Taipei city, the most northern big city in Taiwan) and cluster 13 (Tainan city and Kaohsiung City, the most southern big cities in Taiwan). It shows a cluster approach can be improved in the densely populated area.

This research applies AC-ENN to forecast the electrical load. The intensive experimental results show the following: 1. in encompassing test, ENN and AC-ENN models are much better to BPN. It shows the optimization model will perform better than the general model. 2. AC-ENN is the most accurate model with the smallest MAPE and it shows our AC mechanism is better than the original generation in Genetic Algorithms.

ACKNOWLEDGMENTS

This research work was supported by the National Science of Council in Taiwan with the contract No. NSC 101-2221-E-231 -007 -MY3.

REFERENCES


Yen-Wen Wang was born in Taipei city in Taiwan. And graduated from the Department of Industrial Engineering and Management, Yuan-Ze University in Taoyuan city, Taiwan, in 2007. The major fields of study are artificial intelligence, production scheduling, and soft computing. He is an associate professor in the department of Marketing and Distribution Management in Chien-Hsin University, and also a director of center of teaching and learning excellent. He has published some related researches in Journal of Production Economic, Journal of Production Research, Journal of Intelligenct Manufacturing, Journal of Expert Systems with Applications etc. A representative work is published in International Journal of Production Economics, “Combining Independent Component Analysis and Growing Hierarchical Self-Organizing Maps with Support Vector Regression in Product Demand Forecasting,” vol. 128(2), 603–613 in 2010. His research interests are applied AI model to deal with forecasting or scheduling problems. Dr. Wang has earned the most cited articles 2006 to 2008 from Expert System with Applications.