A Two-Tier Forecasting Power Generation Model in Smart Grids

Yen-Wen Wang, San-Jung Peng, and Meng-Hui Chen

***Abstract*—**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. The main idea in this study is to use cluster approach to cluster the villages and towns as different zones and cluster the data of each zone again to enhance the performance of electricity demand forecasting. If we could predict the electricity demand effective, we would decrease the power consumption. So we use the K-means to cluster the electricity substation, and collect the demand of electricity by each villages and towns to develop the forecasting model. According to the experimental result, the prediction of electricity demand is effective.

***Keywords*—**Power forecasting, K-means, Self-organizing map, Support vector regression.

I. Introduction

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 this paper, we develop a model for electricity demand forecasting of villages and towns in Taiwan. The main idea is to using K-means [7] to cluster all cities in Taiwan and the electricity substation as the central of each cluster. By this way, we can find the cities with high similarity habits of electricity demand.

Yen-Wen Wang is with the Department of Marketing and Distribution Management, Chien Hsin University of Science and Technology. Tao-Yuan City, Taiwan (corresponding author’s phone:+886922486019 ; e-mail: ywwang@uch.edu.tw).

San-Jung Peng is with the Department of Marketing and Distribution Management, Chien Hsin University of Science and Technology. Tao-Yuan City, Taiwan(e-mail: sjpeng@uch.edu.tw).

Meng-Hui Chen, is a PHD student with the Dept. of Information Management, Yuan-ze university. Tao-Yuan City, Taiwan (e-mail: bbb422@hotmail.com).

II. Literature review

A. Smart grid

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. 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 contain 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 [14] [17]. 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 [5]. Lo [8] has pointed out that a good predictive model needs to have the following five characteristics:

1) In different load problems, choose the best predictive models.
2) In different models, choose the best combination of parameters.

http://dx.doi.org/10.15242/IIE.E0315065

3) The predictive model needs to include external factors that can affect the electrical load (such as temperature, cold and other external factors).
4) Have a complete data analysis and interpretation capabilities.
5) 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 [12] 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 [15]). Multiple linear regression (MLR) is most frequently used in the short and medium load forecasting models (Papalexopoulos [1], Krogh [2], Heinemann [4]). 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 is used constantly improving. Hong [16] and Baran [11] have created different forecasting models based on the use of MLR.

B. Self-organizing map

Self-organizing map (SOM) is a two-layer neural network that maps multidimensional data on to a two dimensional topological grid. The method of grouping is based on the similarities between the data, where the measure of similarities is Euclidean distance. Vesanto and Alhoniemi [6] proposed that SOM is an effective tool in data mining, and if the units are large, the quantitative analysis of the map and the data is similar to need to be grouped. The authors [3] used SOM to identify patent trends that they analyze patent knowledge to identify research trends. They tested on patents from the United States Patent and Trademark Office (USPTO) and result both an overview of the directions of the trends and a drill-down perspective of current trends. Another algorithm using SOM that is evolving self-organizing map (ESOM), which features an evolving network structure and fast on-line learning. Their result shows that ESOM achieved better or comparable performance with a much shorter learning process [13]. Collan et al. [10] presents how transition economies can be positioned, and their development tracked, with the SOM. The result shows that SOM can be used to instead the traditional methods, especially SOM are easy to understand to represent the multidimensional data by graphical.

III. Methodology

The effects of electricity demand are various, and the geographical location is one of that. Therefore, the electricity substation is used to be as the central of clusters in this study and adopts the K-means [7] to cluster all the villages and towns. Next step is to select a cluster and use SOM to cluster the data of the cluster again, and using PCA(principal component analysis)[9] to do the feature selection to enhance the accuracy of proposed model. Finally, we adopt the multi factor line regression method to forecast the electricity demand. As following, the proposed model is shown in Fig. 1.

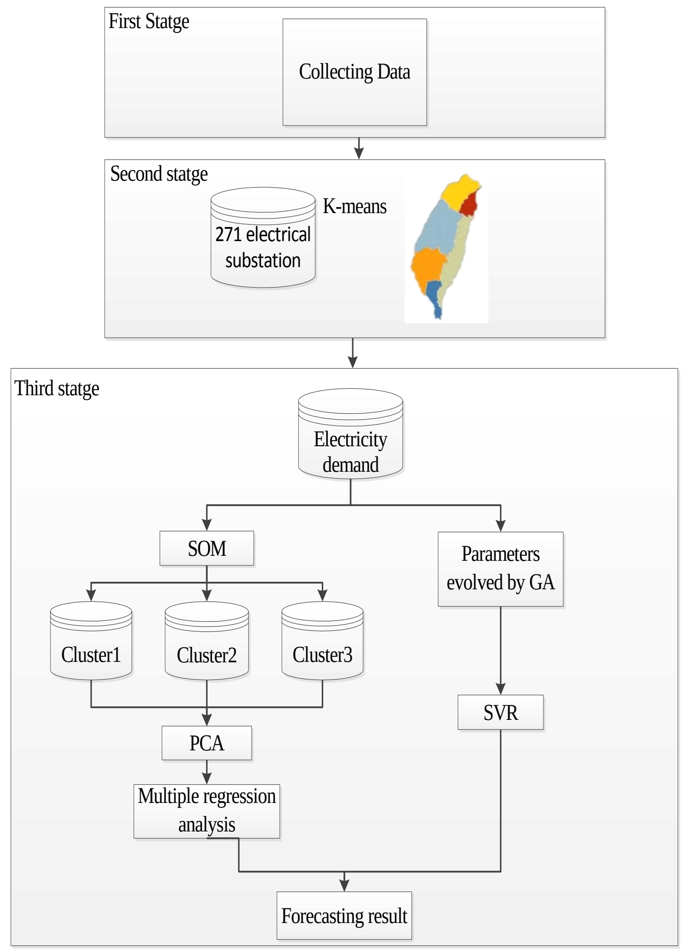

Fig. 1 The illustration of proposed model

A. Data collection

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

B. Cluster the data

After getting the data, the study is hybrid the two cluster approaches to cluster the data. The first stage is to set the electricity substation as the cluster central and using K-means to cluster the villages and towns. The second stage is to select one zone and calculate the total electricity demand of each month in the zone. Then we adopt SOM to cluster the electricity demand of each month, where SOM analysis process continues until all input vectors are processed. Convergence criterion utilized here is in terms of epochs, which defines how many times all input vectors should be fed to the SOM for analysis.

C. Forecasting of electricity demand

Multiple regression analysis is applied to learn about the relationship between the independent or predictor variables. For example, the census data has many factors affecting the electricity demand. Therefore, we collect various variables about the electricity demand in this study. Thus, multiple regression analysis is adopted to predict electricity demand. Furthermore, the data include many parameters relative to electricity demand, but there may be several noises to reduce the performance of prediction. So we adopt PCA to do feature selection to enhance the accuracy.

In addition, we also hybrid simple genetic algorithm (GA) and support vector regression (SVR) to construct a simple forecasting model to compare with the proposed forecasting model. In this model, GA is adopted to evolve effective parameters to enhance the performance.

IV. Experimental Results

The data are the electrical load of each electrical substation within 4 years and there are 271 substations in Taiwan area. From the data, there are many 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. As following, we represent the results of the substation clusters and electricity demand.

A. The substation clusters

According to the regional electricity supply and demand data, we used 15 clusters in this study, and the distribution figure is shown in Fig. 2.

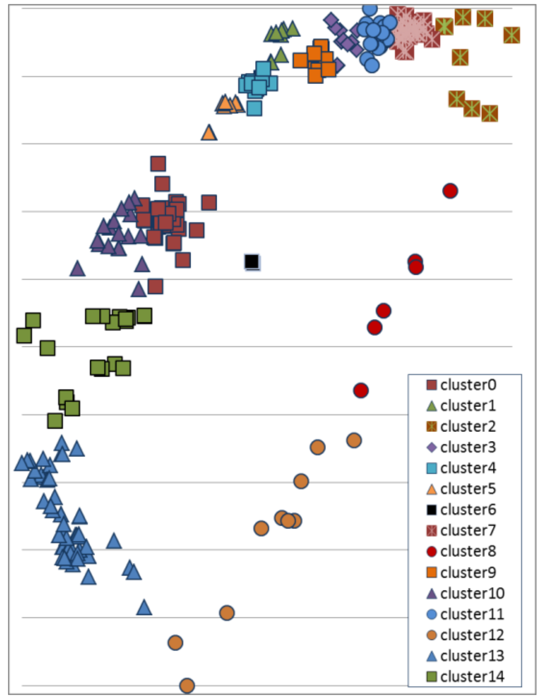

Fig. 2 A distribution of 15 clusters of K-means

B. Forecasting of electricity demand

In this study, we select one zone from the 15 cluster, and set the cluster number 3, 4 and 5 of SOM to cluster the data of the zone again. In addition, we also compare the origin data and the data adopted PCA to do the feature selection.

The data is collected from the villages and towns during 2009/1 to 2012/12 in Taiwan. Therefore, we select the data during 2009/1 to 2011/12 as the training data, and the data during 2012/1 to 2012/12 as the testing data.

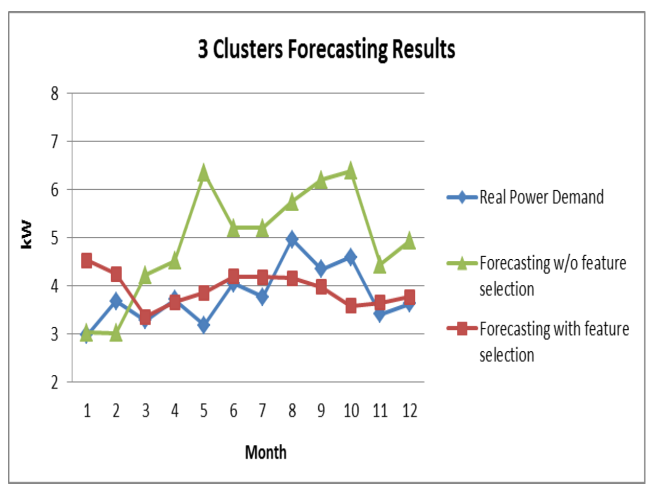

Fig. 3 The forecasting result of cluster-3

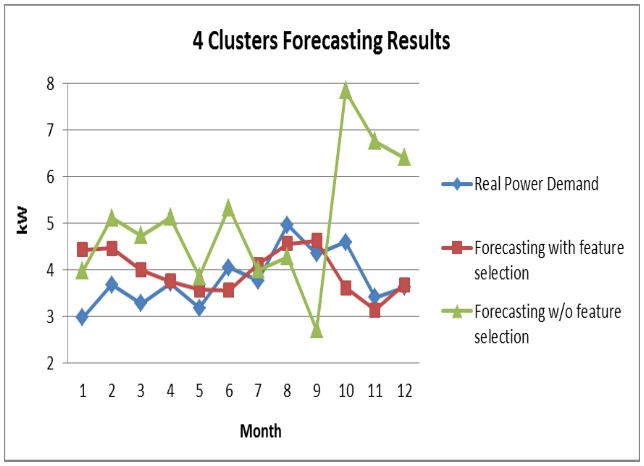

Fig. 4 The forecasting result of cluster-4

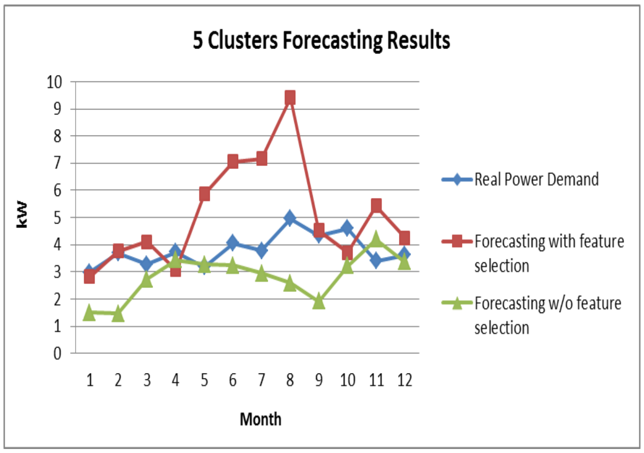

Fig. 5 The forecasting result of cluster-5

The result as shown in Table I is represented different cluster number in SOM compared. The measure is adopted mean absolute percentage error (MAPE) given as following:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{ED_i' - ED_i}{ED_i} * 100\% \quad (1)$$

Where n is the total numbers of testing months.

TABLE I

THE FORECASTING RESULT COMPARED WITH DIFFERENT FORECASTING MODEL

Forecasting model		SOM+MRA			GA+SVR
		# clusters of SOM			
		3	4	5	
MAPE	w/o feature selection	21.31 %	24.55 %	31.58 %	17.77%
	with feature selection	13.75 %	14.29 %	28.60 %	

As mentioned, according to the results shown in Fig. 3, Fig. 4, Fig. 5 and Table I, it is represented that when using the cluster number-3 by SOM to cluster the data and adopting PCA to do feature selection can get the best result.

V. CONCLUSION AND FURTHER WORKS

In this study, we use two-stage cluster mechanism and multiple regression analysis to develop the electricity demand forecasting model. According to different purpose, we adopt different cluster approaches. Finally, in order to find the key parameters of electricity demand, PCA is adopted for feature selection mechanism. Accrding to results, the best MAPE which is 13.75% represents the performance is effective.

Finally, we select one zone of the 15 cluster to test the performance of proposed forecasting model. Thus the futher works suggested is to collect more data to verify the robust of proposed forecasting model and try to hybrid different cluster approaches or feature selection methods to enhance the forecasting accroacy.

ACKNOWLEDGMENT

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

REFERENCE

[1] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," IEEE Transactions on Power Systems, vol. 5(4), pp. 1535-1547, 1990. http://dx.doi.org/10.1109/59.99410

[2] B. Krogh, Llinas, E. S. de and D. Lesser, "Design and Implementation of An on-Line Load Forecasting Algorithm," IEEE Transactions on Power Apparatus and Systems, vol. 101(9), pp, 3284-3289, 1982.

[3] Deng, D. & Kasabov, N., On-line pattern analysis by evolving self-organizing maps, Neurocomputing, 51, pp. 87-103, 2003. http://dx.doi.org/10.1016/S0925-2312(02)00599-4

[4] G. T. Heinemann, D. A. Nordmian, E. C. Plant, "The Relationship Between Summer Weather and Summer Loads - A Regression Analysis," IEEE Transactions on Power Apparatus and Systems , vol. 85(11), pp. 1144-1154, 1966. http://dx.doi.org/10.1109/TPAS.1966.291535

[5] H.S. Hippert, C.E. Pedreira and R. Castro, "Neural networks for short-term load forecasting: A review and evaluation," IEEE Trans. Power Systems, vol.16(1), pp. 44-55, 2001. http://dx.doi.org/10.1109/59.910780

[6] J. Vesanto, E. Alhoniemi. "Clustering of the Self-Organizing Map,"IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 3, MAY, 2000.

[7] J. Z.C. Lai, Y.C. Liaw, "Improvement of the k-means clustering filtering algorithm, Pattern Recognition, " 41, 3677-3681, 2008. http://dx.doi.org/10.1016/j.patcog.2008.06.005

[8] K.L. Lo and Y.K. Wu, "Risk assessment due to local demand forecast uncertainty in the competitive supply industry," IEE Proceedings-Generation, Transmission and Distribution, vol. 150(5), pp. 573-581, 2003. http://dx.doi.org/10.1049/ip-gtd:20030641

[9] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," Philosophical Magazine, vol. 2 (11), pp. 559–572, 1901. http://dx.doi.org/10.1080/14786440109462720

[10] M. Collan, T. Eklund, and B. Back, "Using the Self-Organizing Map to Visualize and Explore Socio-Economic Development," EBS Review, 22(1), 6-15, 2007.

[11] M. E. Baran, L.A. Freeman, F. Hanson, V. Ayers, "Load estimation for load monitoring at distribution substations," IEEE Transactions on Power Systems, vol. 20(1), pp. 164-170, 2005. http://dx.doi.org/10.1109/TPWRS.2004.840409

[12] R. Weron, "Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach," John Wiley & Sons Ltd, England, vol. 79, 2006. http://dx.doi.org/10.1002/9781118673362

[13] A. Segev, J. Kantola, "Identification of trends from patents using self-organizing maps," Expert System Application, vol. 39(18), pp. 13235-13242, 2012. http://dx.doi.org/10.1016/j.eswa.2012.05.078

[14] S. Fan, Y.K. Wu, W.J. Lee, and C.Y. Lee, "Comparative study on load forecasting technologies for different geographical distributed loads," International Conference of Power and Energy Society General Meeting, pp. 24-29 July, 2011.

[15] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," IEEE Transactions on Power Systems, vol. 18(2), pp. 673-679, 2003. http://dx.doi.org/10.1109/TPWRS.2003.811010

[16] T. Hong, M. Gui1, M.E Baran. and H. L. Willis, "Modeling and Forecasting Hourly Electric Load by Multiple Linear Regression with Interactions," International Conference of Power and Energy Society General Meeting, pp. 25-29 July, 2010.

[17] Y.K. Wu, C.L. Lin, B.R. Ke, Y.K. Huang, K.W. Hsu, "Development of an industrial forecasting tool in wind power," the 5th IEEE Conference on Industrial Electronics and Applications, pp. 614-619, 2010.

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 Intelligence 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.

San-Jung Peng is a Lecturer in the Department of Marketing and Distribution Management in Chien-Hsin University. His fields of research include customer relationship management, marketing management and retail management.

Meng-Hui Chen is a PHD student with the Dept. of Information Management, Yuan-ze University. Tao-Yuan City, Taiwan. He has published some related researches in Journal of Intelligent Manufacturing, Journal of Soft Computing, and Journal of Applied Soft Computing.