International Journal of Smart Home
Vol. 10, No. 2, (2016), pp. 97-108
http://dx.doi.org/10.14257/ijsh.2016.10.2.10

A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings

Fazli Wahid[1] and DoHyeun Kim*[1]

[1]*Department of Computer Engineering Jeju National University, Jeju-Si, Republic of Korea*
E-mail : Wahid_uomian@hotmail.com and kimdh@jejunu.ac.kr
** Corresponding author: DoHyeun Kim, kimdh@jejunu.ac.kr*

Abstract

In order to manage efficiently the energy production, storage and management system, it is very important to analyze accurately the energy requirements for residential sector because the residential sector consumes a considerable amount of total energy produced. The main aim of the paper is the assurance of energy production according to the consumer demands in an efficient manner. The energy market is an important tool for setting prices between the energy producers, suppliers and the consumers. An excellent precision in the prediction of next day consumption is required for the suppliers to get good prices in the energy traded. The main aim of this paper is to facilitate the energy suppliers to make decisions for the provision of energy to different apartments according to their demand. In this paper, we have utilized K-Nearest Neighbors classifier for daily energy consumption prediction based on classification. The process consists of five stages namely data collection, data processing, prediction, and validation and performance evaluation. The historical data containing hourly consumption of 520 apartments of Seoul, Republic of Korea has been used in the experimentation. The data has been divided into different training and testing ratios and different qualitative and quantitative measures have been applied to find the performance and efficiency of the predictor. The highest accuracy has been observed for 60-40% training and testing ratio giving 95.9615% accurate results. The effectiveness of the model has been validated using 10-Fold and 5-Fold cross validation.

Keywords*: Energy consumption, daily prediction, K nearest neighbor, residential buildings*

1. Introduction

The management of modern power system in an efficient manner is a challenging problem. In order to handle different challenges in the power management system, the role of smart grid is very important. The smart grid provides assistance in the automation process of the energy management system by integrating different components of the energy management system including sensors, actuators, controls and computational ability [1]. For bringing improvement in the production, planning, maintenance and operations, the smart grid makes use of modern technologies [2]. Many governments all over the world sustain modern communication systems with keeping different environmental and energy saving factors. The smart grid performs different functionalities to cope with all these issues. The US department of energy gives definitions of functions performed by smart grid in [3]. The residential sector represents a high percentage of energy consumption; therefore, the energy suppliers mainly focus on the residential power

*Corresponding Author: DoHyeun Kim, *kimdh@jejunu.ac.kr*

ISSN: 1975-4094 IJSH

consumption. In 2007, the residential sector consumed a large amount of energy in the European Union countries. Figure 1 shows the energy consumption contribution of different sectors towards the total energy consumption in European countries [4]. Similar to energy consumption in residential buildings in the European Union countries, the residential sector in South Korea is also utilizing a considerable amount of energy.

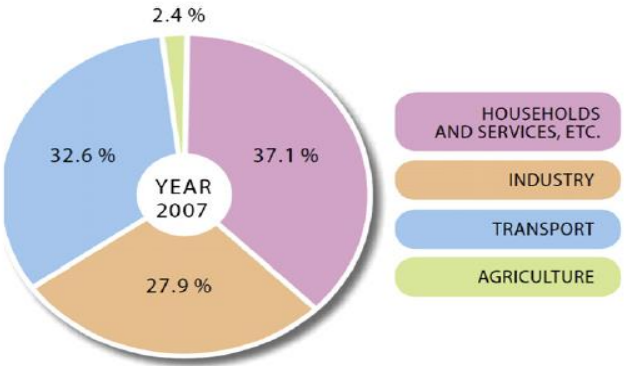


Figure 1. Distribution of Energy Consumption by Sector in European Countries in 2007

For obtaining good prices for traded energy, the next day energy consumption prediction with higher precision is very important for the power suppliers involved. Figure 2 shows the relationships among different components involved in the power supply, energy market, power retailer and power consumption and the energy prediction is shown in the following figure to show the importance of prediction of the energy in the energy management system.

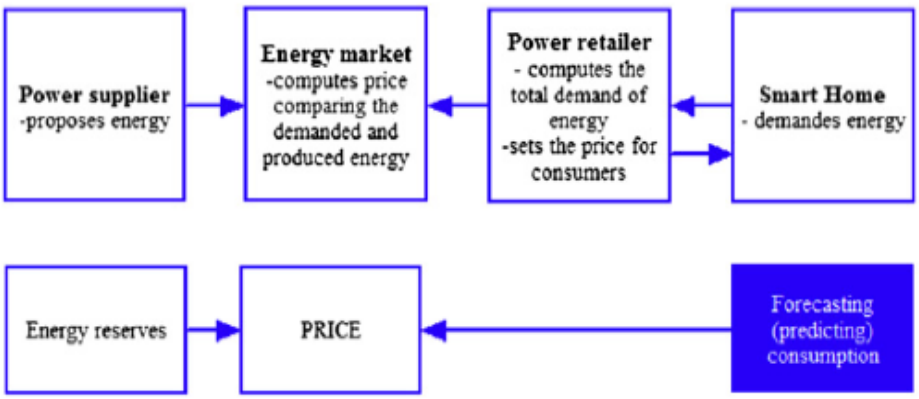


Figure 2. Interaction between Smart Homes and Smart Grid

A considerable amount of the total energy produced is consumed by the residential sector and therefore the energy consumption prediction in residential sector gains more importance. In literature, many approaches have been applied for energy consumption prediction in residential sector. An important approach of them is bottom up approach in which the energy consumption for smart appliances is predicted. Based on the consumption of all the appliances, the total hourly consumption is predicted for the whole smart home and then the daily consumption is predicted on these consumptions. Although it is important to predict the monthly, weekly, daily and even hourly consumption of every home appliance but in order to make better management of power supply to the

smart homes by the suppliers, the total monthly, weekly, daily and hourly prediction irrespective of the individual appliance is of great importance.

The purpose of this paper is to divide the power consuming apartments into two categories according to their consumed power. This prediction will be helpful for the suppliers of the power to make proper decisions regarding the demand of the power by their consumers. Also, this prediction can provide assistance in the management of smart home automation system [5]. The authors in [6] have presented a three layer energy management system for smart environment. In this energy management system, the uppermost layer is the anticipative layer, the lowermost layer is the local layer and between these two layers is the reactive layer. The anticipative layer is responsible for power prediction, price prediction and weather prediction. It is also the responsibility of the anticipative layer to provide this information to the reactive layer to assist it in its functioning. The responsibility of reactive layer is to provide the information received from the anticipative layer and different comfort criteria to the local layer to control the status of the home appliances to achieve the user comfort. The local layer is responsible for the controlling of status of home appliances. The consumed power used in the local layer by different appliances goes directly to the anticipative layer where it is used for prediction. A forth layer called external layer has been introduced in which the suppliers reside to supply power according to the demand. The suppliers obtain energy from the market provided with energy form different energy sources including thermal power plant, hydro power plant, nuclear power plant and renewable resources[7].

In the literature, there are many different approaches for power consumption prediction. The authors in [8] and [9] have used artificial neural network (ANN) for short-term energy prediction. For load forecasting, [10] used regression models and times series. For different types of forecasting, general methodologies have been discussed by the authors in [11]. Kalman filter has been used by [12] for short-term load forecasting. The authors in [13] used Kalman filter for prediction whereas GA was used by [14] for prediction and management of energy in residential buildings. Fuzzy logic has been used by [15] and [16] whereas fuzzy neural networks have been used by [17] and [18] for short-term load forecasting. Many authors have used hybrid models in which they have combined artificial neural network with statistical methods. Examples of these models are hybrid models with Bayesian inference [19] and [20], wavelet transform [21] and [22], SOM (Self Organization Map) [23] and PSO (Particle Swarm Optimization) [24]. The authors in[25], [26] used Bayesian networks for prediction and classification. Decision trees have been used by[27], [28] for prediction and classification. The authors in [29] used decision tables for classification. In our work, we have used K-Nearest Neighbors for prediction based on classification.

2. Prediction Approach

A simple model of prediction is one in which the prediction is based on the method of classification based on some historical data. The objective of this work is to predict the energy consumption based on classification of apartments according to their energy usage. The system can be used for the determination of energy plan for the assignment of energy to different apartments, which can be helpful for the power retailer in the computation of energy demand and setting the prices. The proposed methodology divides the apartments into either low power consumption apartments or high power consumption apartment according to their daily energy consumption. This prediction can be helpful for the energy retailer in their future plan of energy demands for the residential apartments. The proposed method is shown in the figure 3.

The proposed method consists of following stages

- Data collection
- Data processing

- Prediction
- Validation
- Performance evaluation

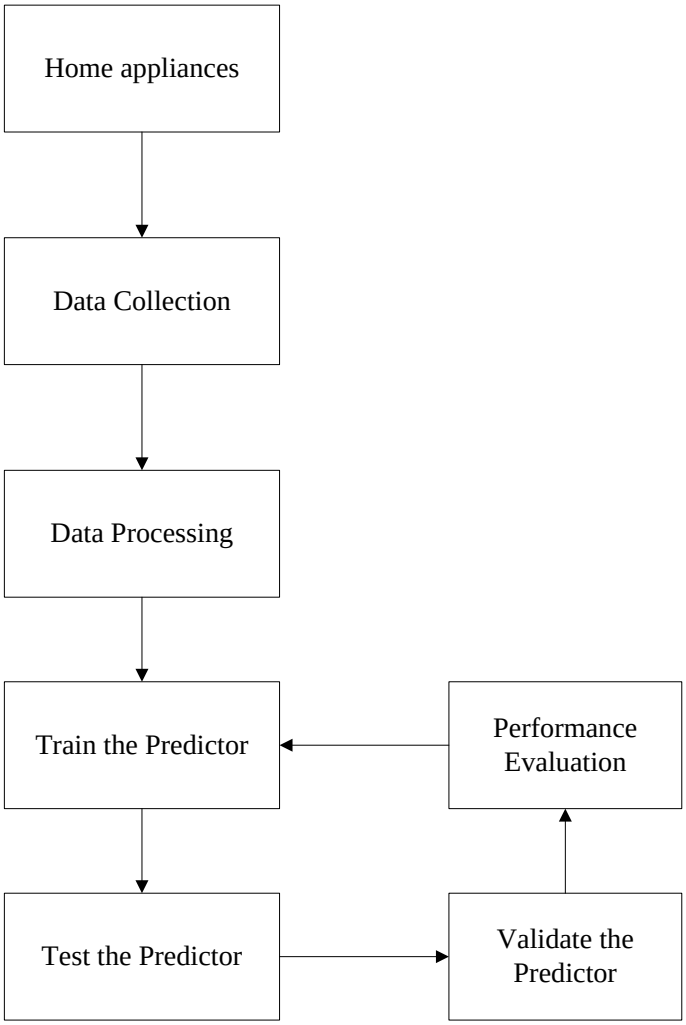


Figure 3. Proposed Model

2.1. Data Collection

In data collection stage, the energy consumed on hourly basis is collected from all the home appliances used in the residential apartments. The data containing hourly consumption of all the home appliances of 520 apartments has been used in the experimentation.

2.2. Data Processing

In the data processing stage, the data containing hourly consumption of home appliances is retrieved from the collected data and the mean, variance, skewness and kurtosis of the daily consumed power based on hourly consumption are computed. The mean, variance, skewness and Kurtosis are computed using equations 1, 2, 3 and 4.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{2}$$

$$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3 \quad (3)$$

$$K = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4 \quad (4)$$

Where x_i represents the power consumption over ith hour of the day where i = 0,1,2,…,23. N represents the total number of hours i.e. 24. μ represents mean, V represents variance, S represents skewness and K represents Kurtosis.

2.2.1. Mean

Mean represents the average of all the hourly consumed power over the whole day.

2.2.2. Variance

Variance represents the variations in the hourly consumed power over the whole day.

2.2.3. Skewness

Skewness represents the asymmetry in the hourly consumed power over the whole day.

2.2.4. Kurtosis

It represents peakedness or frequency of extreme hourly power consumption over the whole day.

2.3. Prediction

In the prediction stage, the apartments are predicted according to their energy consumption. Two types of apartments namely low power consumption apartments and high power consumption apartments have been identified. For the prediction, K-Nearest Neighbors has been used. The prediction stage consists of two stages namely training the predictor and testing the predictor. In the training stage, the predictor is given with known instances of the data to let the predictor learn to identify the new unknown instances. In the testing stage, the predictor is given unknown instances to identify them. The whole data has been divided into different training and testing ratios to show the prediction capability of the predictor for different training and testing ratios.

2.3.1. K-Nearest Neighbor for Prediction

KNN is considered to be one of the simplest classification techniques. In K-Nearest neighbor classification, K-closest vector based on a suitable distance vector is determined to classify the input feature vector X. The assignment of vector X to a specific class is then done by the majority of K-nearest neighbors to which they belong. K nearest neighbor algorithm is based on distance and voting function in K-nearest neighbors, the metric employed is Euclidean distance. KNN is a supervised classification technique that yields good performance results for optimal value of K. Like other supervised learning algorithms, KNN consists of two phases namely training phase and testing phase. In the training phase, data values are given to the classifier to train it. The training data has labels associated with them that represent their class. During testing phase, the KNN classifier is given unlabeled data points and the algorithm generates a list of K nearest data values. The class of unlabeled data points is determined by these values. The KNN algorithm consists of following steps.

Step1: Determine some suitable distance metric (In this case Euclidian distance)

Step2**:** In this phase, the predictor is trained for data sets whose classes are known to us. Store all the training data set X in pairs according to the selected features according to following computations X = {(xi,yi), i=1,…,n}, where xi is a training pattern in the training data set, yi is the class assigned to it and n represents amount of training pattern.

Step3: In this phase, the predictor is fed with data sets whose class is not known to us. For the identification of the class of these data sets, the distance between the newly fed feature vector and already stored features is computed.

Step4**:** K nearest neighbors are chosen and asked to vote for the class of new data set.

2.4. Predictor Validation

In order to validate the effectiveness of the predictor, 10-fold and 5-fold cross validation have been applied. In 10-fold cross-validation, the original data is randomly divided into 10 equal size subsamples. Of the 10 subsamples, one sample is used for testing and the remaining 9 samples are used for training. This process is repeated 10 times with each time one sample used for testing and remaining 9 samples for training. The 10 results obtained are then averaged to find the final result of the predictor. In this way, the predictor is validated using the whole data for testing as well as training. For 5-fold cross validation, the whole data is divided into 5 samples and the whole process is repeated as above mentioned for 10-fold cross validation.

2.5. Performance Evaluation

The performance of classifier is measured by computing Accuracy (AC), Kappa Statistics (KS), Sensitivity (SE), Specificity (SP) and ROC for the classifier.

2.5.1. Accuracy

The accuracy of the classifier has been computed using following formula

Accuracy = (Correctly Predicted Instances/Total Instances) x100 (5)

2.5.2. Kappa Statistics

The Kappa Statistics takes values between 1 and -1. As this value approaches 1, the performance of the predictor improves and as this value approaches -1, the performance of the predictor degrades. Kappa Statistics has been calculated using following formula.

$$KS = \frac{P0-PC}{1-PC} \quad (6)$$

Where *P0* represents total agreement probability, *PC* represents hypothetical probability of chance agreement.

2.5.3. Sensitivity

Sensitivity represents the proportion of positives that have been correctly identified and is given by the following formula.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (7)$$

Where TP stands for true positive and FN stands or false negative

2.5.4. Specificity

Specificity represents the proportion of negatives that are correctly identified and is given by the following formula

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (8)$$

Where TN represents True negative and FP represents false positive

3. Experimental Results and Discussion

The experiments were carried out on Intel (R) Core (TM)2 Quad CPU with 3.25 GB of RAM. For the results' computation, both MATLAB R2010a and Weka 3.7 were used. In the experiments, the input attributes are mean, variance, skewness and Kurtosis of 24 hours power consumption. The class attribute takes either 1 or 0 values corresponding to 'high power consumption apartment' or 'low power consumption apartment', respectively. The total data set was divided into different training and testing ratios. For making the algorithm more generalized, 10-Fold and 5-Fold cross validation was applied. The experimental results for different training and testing ratios and different performance evaluation for them are shown in table 1.

Table 1. Different Types of Accuracies' Measures for Different Training and Testing Ratio of KNN

Ratios (Training, Testing)	Prediction Accuracy	Kappa Statistics	Sensitivity	Specificity	ROC Value
30-70%	93.75	0.8615	0.938	0.938	0.974
40-60%	95.1923	0.8921	0.952	0.952	0.976
50-50%	95	0.8887	0.95	0.95	0.979
60-40%	95.9615	0.9069	0.96	0.96	0.982
70-30%	95.7692	0.9027	0.958	0.958	0.985
75-25%	95.3846	0.8935	0.954	0.954	0.977
80-20%	95.1923	0.8862	0.953	0.952	0.975
10-Fold Cross Validation	94.6154	0.8745	0.946	0.946	0.978
5-Fold Cross Validation	92.6923	0.8313	0.927	0.927	0.916

4. Critical Analysis of KNN Performance

In this section, we show a graphical representation of comparisons for all training and testing ratios and cross validation and accuracy, sensitivity, specificity and Kappa Statistics for both the percentage split and cross validations have been graphically compared. Figure 4, 5, 6 and 7 shows different performance evaluation measurements for different training-testing ratios as well as cross validation. The figures show that the KNN classifier gives best performance for 60-40% training and testing ratios for all performance evaluation measurement parameters.

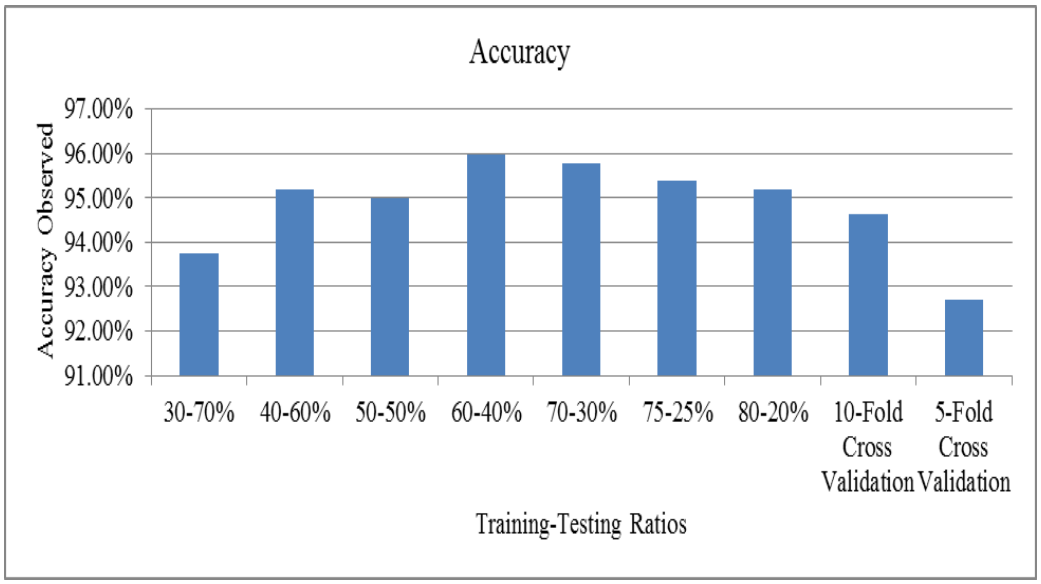


Figure 4. Accuracy Comparison for Different Training and Testing Ratios

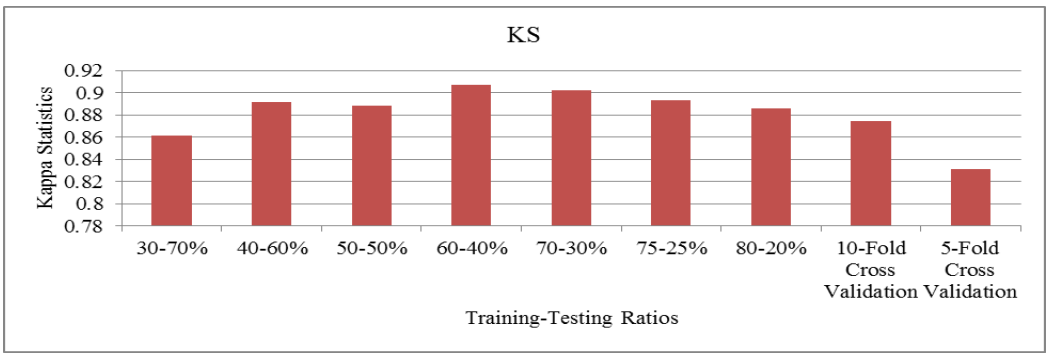


Figure 5. Kappa Statistics Comparison for Different training and Testing Ratios

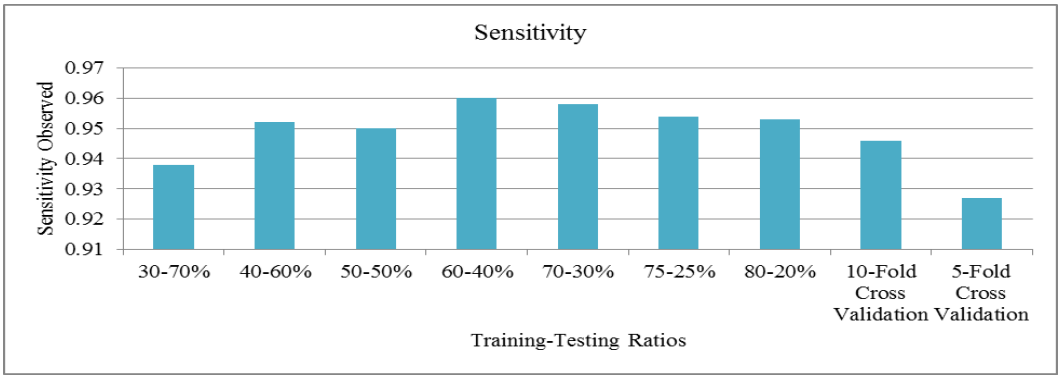


Figure 6. Sensitivity Comparison for Different Training and Testing Ratios

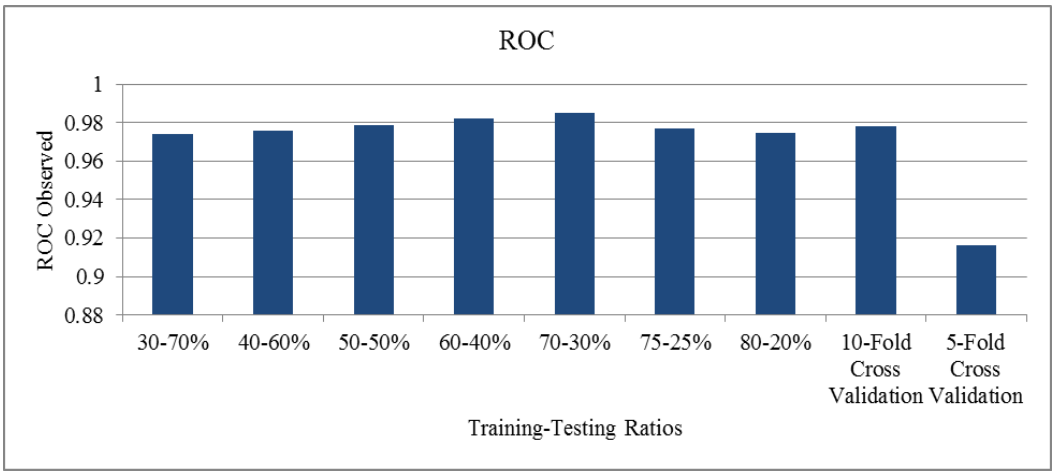


Figure 7. ROC Comparison for Different Training and Testing Ratios

5. Conclusion

In this paper, a simple approach for daily energy consumption of residential apartments based on classification has been carried out. The proposed approach divides the apartments into either high power consumption apartments or low power consumption apartments according to their daily energy usage based on hourly consumption. The proposed architecture consists of five stages namely data collection, data processing, prediction, and validation and performance evaluation. The predictor used in the model is K-Nearest neighbor. This prediction will be helpful for the suppliers of the power to make proper decisions regarding the demand of the power by their consumers. Also, this prediction can provide assistance in the management of smart home automation system. The historical data containing hourly consumption of 520 apartments of Seoul, Republic of Korea has been used in the experimentation. The data has been divided into different training and testing ratios and different qualitative and quantitative measures have been applied to find the performance and efficiency of the predictor. The highest accuracy has been observed for 60-40% training and testing ratio giving 95.9615% accurate results. The effectiveness of the model has been validated using 10-Fold and 5-Fold cross validation.

Acknowledgement

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.10043907, Development of high performance IoT device and Open Platform with Intelligent Software). And this research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2015-H8501-15-1017) supervised by the IITP(Institute for Information & communications Technology Promotion). Corresponding author; DoHyeun Kim (e-mail: kimdh@jejunu.ac.kr).

References

[1] C. W. Gellings, "The smart grid: enabling energy efficiency and demand response", The Fairmont Press, Inc, **(2009).**

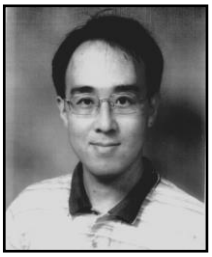
[2] A. Ipakchi, and F. Albuyeh, "Grid of the future", Power and Energy Magazine, IEEE., vol. 7, no. 2, **(2009)**, pp. 52-62.

[3] M. C. Williams, J. P. Strakey, and W.A. Surdoval, "The US department of energy, office of fossil energy stationary fuel cell program", Journal of Power Sources., vol. 143, no. 1, **(2005)**, pp. 191-196.

[4] D. TREN, "EU energy and transport in figuresâ€"Statistical pocketbook 2009, Directorate-General for Energy and Transport", Office for Official Publications of the European Communities, Luxemburg, **(2009).**
[5] D. L. Ha, S. Ploix, E. Zamai, and M. Jacomino, "Real times dynamic optimization for demand-side load management", International Journal of Management Science and Engineering Management., vol. 3, no. 4, **(2008)**, pp. 243-252.
[6] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage", Energy and Buildings, vol. 67, **(2013)**, pp. 668-679.
[7] N. Arghira, L. Hawarah, S. Ploix, and M. Jacomino, "Prediction of appliances energy use in smart homes", Energy, **(2012)**, vol. 48, no. 1, pp. 128-34.
[8] H.S. Hippert, C.E. Pedreira, and R.C. Souza, "Neural networks for short-term load forecasting: A review and evaluation. Power Systems", IEEE Transactions, vol. 16, no. 1, **(2001)**, p. 44-55.
[9] O. A. Carpinteiro, A. J. Reis and A.P. da Silva, "A hierarchical neural model in short-term load forecasting", Applied Soft Computing, vol. 4, no. 4, **(2004)**, pp. 405-412.
[10] G. Gross, and F. D. Galiana, "Short-term load forecasting", Proceedings of the IEEE, **(1987)**, vol. 75, no. 12, pp. 1558-1573.
[11] WASU, M.R.V, "An Approach to Electricity Load Forecasting and Techniques: A Review ".
[12] G. Irisarri, S. Widergren, and P. Yehsakul, "On-line load forecasting for energy control center application", IEEE Power Engineering Review, **(1982)**, vol. 1, no PER-2, pp. 23.
[13] Ali, S. and D.-H. Kim, "Effective and comfortable power control model using Kalman filter for building energy management", Wireless personal communications, **(2013)**, vol. 73, no. 4, pp. 1439-1453.
[14] Ali, S. and D.-H. Kim, "Optimized Power Control Methodology Using Genetic Algorithm." Wireless personal communications, **(2015)**, pp. 1-13.
[15] S. J. Kiartzis, A. G. Bakirtzis, J. B. Theocharis and G. Tsagas, "A fuzzy expert system for peak load forecasting. Application to the Greek power system", Electrotechnical Conference, MELECON 2000. 10th Mediterranean, vol. 3, **(2000)**, pp. 1097-1100.
[16] V. Miranda, and C. U. Monteiro, "Fuzzy inference in spatial load forecasting.", in Power Engineering Society Winter Meeting, IEEE, (**2000).**
[17] A. G. Bakirtzis, J. B. Theocharis, S. J. Kiartzis, and K. J. Satsios, "Short term load forecasting using fuzzy neural networks", Power Systems, IEEE Transactions, vol. 10, no. 3, **(1995)**, pp. 1518-1524.
[18] D. Srinivasan, C. Chang and A. Liew, "Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting", Power Systems, IEEE Transactions, vol. 10, no. 4, **(1995)**, pp. 1897-1903.
[19] L. M. Saini, "Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks.", Electric Power Systems Research, vol. 78, no. 7, **(2008)**, pp. 1302-1310.
[20] P. Lauret, E. Fock, R. N. Randrianarivony and J. F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting", Energy conversion and management, **(2008)**, vol. 49, no. 5, pp. 1156-1166.
[21] S. J. Yao, Y. H. Song, L. Z. Zhang, and X. Y. Cheng, "Wavelet transform and neural networks for short-term electrical load forecasting", Energy Conversion and Management, vol. 41, no. 18, **(2000),** pp. 1975-1988.
[22] T. Nengling, J. R. Stenzel and W. Hongxiao, "Techniques of applying wavelet transform into combined model for short-term load forecasting", Electric Power Systems Research, vol. 76, no. 6, **(2006)**, pp. 525-533.
[23] M. Amin-Naseri, and A. Soroush, "Combined use of unsupervised and supervised learning for daily peak load forecasting", Energy conversion and management, **(2008**), vol. 49, no. 6, pp. 1302-1308.
[24] M. El-Telbany and F. El-Karmi, "Short-term forecasting of Jordanian electricity demand using particle swarm optimization", Electric Power Systems Research, vol. 78, no. 3, **(2008)**, pp. 425-433.
[25] L. Hawarah, M. Jacomino and S. P. Ploix, "Smart Home-From User's Behavior to Prediction of Energy Consumption", ICINCO no. 1.
[26] I. H. Witten and E. Frank, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, **(2005)**.
[27] J. R. Quinlan, "Induction of decision trees", Machine learning, vol. 1, no. 1, **(1986)**,pp. 81-106.
[28] L. B. J. F. R. Olshen and J. Charles, Stone. "Classification and regression trees", Wadsworth International Group, vol. 93, no. 99, **(1984)**, pp. 101.
[29] R. Kohavi, "The power of decision tables, in Machine Learning", ECML-95. Springer. **(1995)**, pp. 174-189.

Authors

Fazli Wahid, He received his B.S (4 Years Computer Science) from University of Malakand, Pakistan in 2006, M.S (Computer Science) from SZABIST; Pakistan in 2015. He is currently a PhD student at Jeju National University, South Korea. His area of interest is energy consumption prediction, optimization and management using multi-layer perceptron, Artificial Bee Colony, Ant Colony and other machine learning techniques.



Do-Hyeun Kim, He received the B.S. degree in electronics engineering from the Kyungpook National University, Korea, in 1988, and the M.S. and Ph.D. degrees in information telecommunication the Kyungpook National University, Korea, in 1990 and 2000, respectively. He joined the Agency of Defense Development (ADD), from Match 1990 to April 1995. Since 2004, he has been with the Jeju National University, Korea, where he is currently a Professor of Department of Computer Engineering. From 2008 to 2009, he has been at the Queensland University of Technology, Australia, as a visiting researcher. His research interests include sensor networks, M2M/IOT, energy optimization and prediction, intelligent service, and mobile computing.