Analysis and predicting electricity energy consumption using data mining techniques-A case study I.R. Iran - Mazandaran province

Conference Paper · April 2015

DOI: 10.1109/PRIA.2015.7161634

Noorollah Karimtabar
Technical and Vocational University of Iran

Siavash Alipour
Kharazmi University

2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA 2015) March 11-12, 2015

Analysis and predicting electricity energy consumption using data mining techniques- A case study I.R. Iran - Mazandaran province

Noorollah Karimtabar*
Ministry of Education – Mazandaran
Babol – I. R. Iran
Saeidkarim.kk66@gmail.com

Sadegh Pasban
Department of Computer Engineering
Kharazmi University
Tehran, I. R. Iran
sadegh.info@gmail.com

Siavash Alipour
Department of Electronic Engineering
Malek Ashtar University
Tehran, I. R. Iran
siavash.alipur@gmail.com

Abstract--- The electricity consumption forecast is especially important with regard to policy making in developing countries. In this paper, the electricity consumption rate is predicted using the data mining techniques. The datasets that were collected for predicting the electricity consumption are related to Islamic Republic of Iran - Mazandaran province pertaining to the years 1991 to 2013. The research objective is analyzing the electricity consumption rate in recent years and predicting future consumption. According to a study the electricity consumption growth rate between the years 2006 to 2013 and the years 1999 to 2006 equaled 28.41 and 73.53, respectively. The results of the research conducted using the regression model indicate a 2.48 relative error. The output of this prediction shows that the total electricity consumption rate increases about 3.2% annually on average and will reach 7076796 megawatts by the year 2020 that shows a 22.28% growth comparing to the year 2013.

Keywords: Data-mining; electricity consumption rate; dataset; prediction; relative error.

1. INTRODUCTION

Energy is the foundation of economic development, and electricity is one of the major energy sources. A nation's energy policy is thus of crucial importance, as it will not only guide the development of the country, but will also affect the operating environment of various industries. Due to the huge amount of capital investment and the lengthy construction time required in electricity capacity expansion projects, poor decisions can lead to very negative outcomes. A good demand forecast is a prerequisite for the effective development of energy policies, as it can reduce the possibility of errors during electricity system planning. Therefore, producing accurate electricity consumption forecasts is very important [1].

Also, worldwide energy consumption is rising fast because of the increase in human population, continuous pressures for better living standards, emphasis on large-scale industrialization in developing countries and the need to sustain positive economic growth rates. Given this fact, a sound forecasting technique is essential for accurate investment planning of energy production generation and distribution.

Energy consumption in Islamic Republic of Iran has risen remarkably in the past few decades due to the city's increasing population and economic development. In particular, there is a substantial increase in the total annual electricity energy consumption in the domestic sector.

In this study, used three methods supervised numerically predict. Traditionally, regression analysis has been the most popular modeling technique in predicting energy consumption [2]. Artificial neural networks have also been used for the prediction of energy consumption, such as in the study by Kalogirou and Bojic [3]. Simulated data were used to train an artificial neural network in order to generate a mapping between the input and output, and the model was subsequently used to predict energy consumption. In fact, neural networks have established their role as data analysis tools in different areas. A SVM[1] performs classification by constructing an *N*-dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. [4]. Comparison of these different data analysis and modeling techniques has been considered in various applications, but rarely in energy consumption prediction [5].

[1] Support Vector Machine

978-1-4799-8445-9/15/$31.00 ©2015 IEEE

2. Background to the study

A review of the literature shows that there are numerous studies on the relationship among energy consumption and prediction of electricity usage rate. The researches that were conducted on predicting the electricity consumption can be divided into two categories, namely the short-term prediction, and the long-term prediction. The conducted researches indicate that the type of climate in each region is effective in selecting the variables. Further, we study some of these researches.

In 2007, Kelvin and Geoffry[5] presented a paper on predicting the power consumption in Hong Kong using the three models of regression, neural networks and the decision tree. To collect the research data, the monthly power consumption rate was studied in summer and winter seasons in the years 1999 and 2000. The variables employed in this research include the temperature, moisture, the number of home appliances, type of housing, number of family members and the domestic properties. In summer, the average temperature was 26.7 % at the lowest and 31.3 % at the highest and the average moisture was 80.7 %. In winter, the average temperature was 13.6 % at the lowest and 17.4 % at the highest and the moisture was 80.7 %. In order to collect the number of home appliances, the questionnaire method was used. During the research, houses were classified into four types including the rental houses, private houses, corporate houses, and rural houses. 38% of the houses were rental, 15% corporate, 41% private and 6% rural. They came to the conclusion that the proposed model could be improved by adding factors such as temperature and the wind speed. In this research, the SAS Enterprise Miner software was used to build the predictive models. The results indicate that the type of housing and the domestic properties are the two most effective factors in power consumption. Finally, the results obtained from a comparison of the models outputs showed that the decision tree have more accurate prediction.

However, Chang[6] proposed a paper on the short-term power consumption prediction using the AGM (Amendatory Grey Model) method or the revised gray method. The gray system theory was proposed by Deng and it is only applicable when the information is inadequate and unknown. In other words, the gray model can be useful when the dataset is low. Therefore, this method compared to the SVR and BPN methods requires less data whereas the SVR and BPN methods need a vast amount of data to yield the desired results. In this research, the power consumption rates in APEC countries are utilized using the gray model in order to predict the power consumption. The results indicate that the mean absolute error of the predication model is less that 3.1 [1].

In 2008, Kusian[7] proposed a paper under the title of data mining methods to predict the windmills consumption rate. In this paper, the multiple time series models were employed for the predictions. Also, five data mining algorithms were used to build the PRR mode, namely MLP, SVR, Random Forest, Classification, and Step-by-Step Regression. The studies show that the SVR algorithm yield better results compared to other algorithms. The data employed in this research was obtained from a wind farm with more than 100 turbines. The shortcoming of building multiple models of the time series is that its calculations are costly due to the utilization of various parameters and therefore the error rate increases accordingly.

Researchers such as Bianco, Manca and Nardani[8] used the long-term data to predict the power consumption rate in Italy. The utilized model in this research was linear and its data was considered from the year 1970 to 2007. In this research, the indoor, outdoor and the total power consumption rate were studied. In the first part, the paper on the indoor and outdoor power consumption rate was investigated and in the second part, various regression models were studied based on different data. Various statistical tests were employed to investigate the credibility of the suggested models. The results indicate that the indoor and outdoor sections will experience 2% increase yearly in the power consumption rate in future.

In 2010, Ekonomou[9] presented a paper under the title 'the long-term prediction of the power consumption in Greece using the neural networks'. In this research, he employed the multilayer perceptron model method for prediction. The network input includes data that have been collected from the years 1992 to 2008. The aim was to predict the power consumption rate between the years 2009 and 2015. The data related to the years 2005 to 2008 are all tentative. The relative error of the estimated values compared to the real values equals 2%.

In 2011, a paper was published by Kavaklioglu[10] under the title 'modeling and predicting the power consumption rate in Turkey using the SVR method'. In this research, Turkey's power consumption rate is modeled based on four indices, namely population, gross domestic product, imports and exports. The existing data for building the model are related to the years 1975 to 2006 which are collected yearly. 80% of this data is intended for training and 20% for experiment. The prediction made in this research is pertinent to the years 2007 to 2026. All calculations in this research were made using the MATLAB software. Based on the predictions made in this research, the power consumption rate in the year 2026 will reach 248.05 TW/h. The relative error related to this prediction equals 1.51.

3. DATASET

The case study of this research was in Iran and Mazandaran province. Therefore, the research data was collected from this region. The dataset employed in this research was collected annually and is related to the years 1991 to 2013. Moreover, power consumption prediction results are based on long-term when are annually. The dataset variables are shown in table 1. These variables are classified in two groups.

TABLE I. The dataset variables

Independent Variables	Population, Temperatures, Moisture, Electricity price
Target variable	Electricity consumption Rate(total amount, domestic, Commercial, industrial, Agriculture)

The first group of data includes independent variables that are effective on the electricity consumption rate. The independent variables consist of population, temperature, and moisture and electricity consumption price. The second group includes the target variable, i.e. the total electricity consumption rate. The above statistics were collected from Iran's statistics center. Before proceeding to data analysis, the data is to be preprocessed. Data preprocessing means preparing the data for the main process, namely knowledge discovery, to start. In this stage, the data will be analyzed in terms of having errors such as the lost variables, noisy data, repetitive data, and false data. In case any of such errors exist, it should be resolved [9]. Data integration, data selection and data conversion are of other tasks that are to be performed in the preprocessing stage [10]. In figures 1 to 5, the dataset is depicted as a graph.

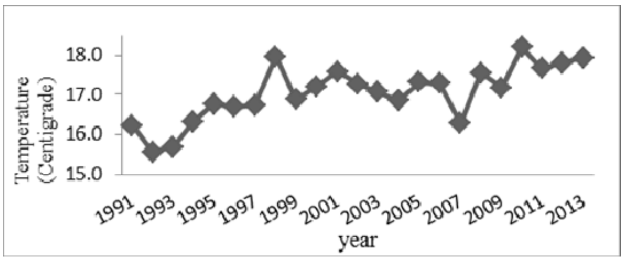

Figure 1. The average temperature from 1991 to 2013

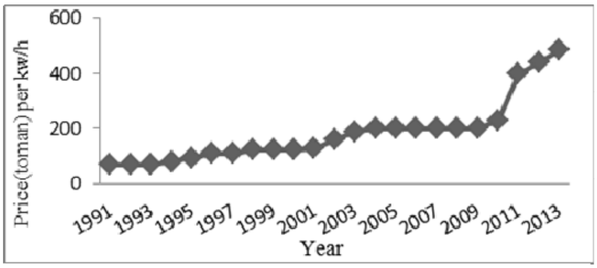

Figure 2. The electricity consumption price from 1991 to 2013

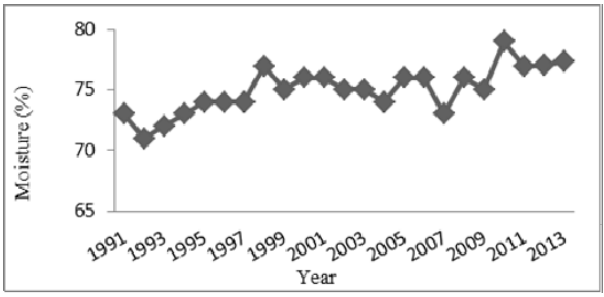

Figure 3. The average moisture from 1991 to 2013

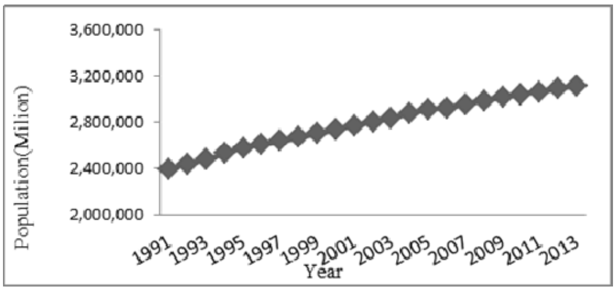

Figure 4. The population from 1991 to 2013

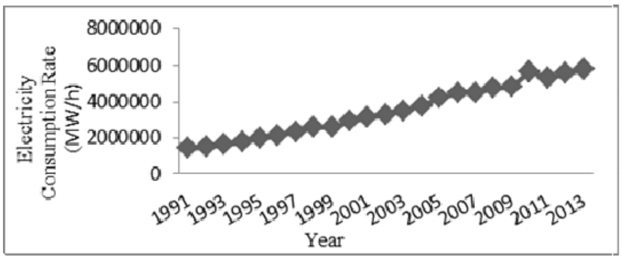

Figure 5. The electricity consumption rate from 1991 to 2013

In table 2, a general description of the dataset is presented. As you can see in this table, the highest consumption rate equals 5787185 MW/h which refers to the year 2013 and the lowest rate is related to the year 1991 which equals 1483678 MW/h. Also, highest temperature equals 18.20 degrees centigrade which refers to the year 2010 and the lowest temperature equals 15.56 degrees centigrade that is related to the year 1992.

TABLE II. Dataset description

Year	maximum temperature	minimum temperature	Maximum electricity consumption rate	Maximum electricity consumption rate
1991-2013	18.20	15.56	5787185	1483678

Further, we apply the regression models, neural network and the SVM on the dataset using the Clementine software and study the obtained results and select the best output for the electricity consumption prediction.

4. Results and Discussion

Using the numerical prediction models, we build the desired models and then select the best model in terms of relative error and correlation rates. In table 3, the results obtained from the prediction models are shown.

TABLE III. Comparison of prediction models

Model	Relative Error (%)	Correlation Coefficient
Regression	0.9	0.996
Neural Net	1.6	0.968
SVM	11.1	0.976

As you can see in table 3, the most suitable model is related to the Regression model. This model has the least error rate and the most correlation rate. In the next stage, the result obtained from the Regression model will be studied. Also, the real data will be compared to the estimated data and the relative error pertinent to each prediction will be estimated. The relative error rate between the real values and the estimated values are shown in table 4.

TABLE IV. Comparison between real consumption with estimated consumption

Year	Real Consumption(MW/H)	estimated Consumption(MW/H)	Relative Error (%)
2009	4833174	5009938	3.66
2010	5675742	5490331	3.7
2011	5311477	5413243	1.92
2012	5576999	5716601	2.50
2013	5787185	5824856	0.65
Mean Relative Error (%)	2.48		

Figure 6 is a graph that shows the difference between the real electricity consumption and the estimated electricity consumption using the regression prediction model between the years 1991 and 2013.

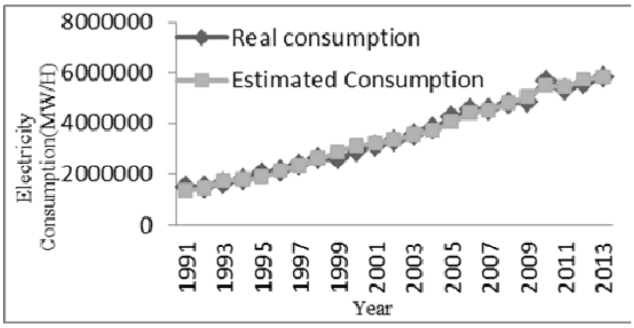

Figure 6. A comparison between the real electricity consumption and the estimated electricity consumption

As it was mentioned earlier, four independent or predictive variables were employed in this research for prediction. These variables include population, temperature, moisture, and electricity consumption price. The study of the variables influence ability rate on the target variable can be very important. In table 5, the variable importance rate on the prediction is shown.

TABLE V. Variable importance rate on the prediction

Variables	*Importance Rate (%)*
Population	43.4
electricity consumption price	26.7
Temperature	21.5
Moisture	8.4

The population field has the most effect on the electricity consumption rate. However, the effects of the electricity consumption price and the temperature should not be ignored. As it was shown in table 3, the model that was built using the regression method has less error rate and therefore better performance compared to other methods. The electricity consumption rate related to the years 2009 to 2013 is considered as the prediction test. The results showed a 2.48% average relative error. Further, the electricity consumption rate related to the years 2014 to 2020 will be predicted. First, the independent variables values related to the datasets that include temperature, moisture, population and electricity consumption price, will be calculated using the growth coefficient. Table 6 shows the estimated values related to this data.

TABLE VI. The estimated values related to Independent Variables

Year	Temperature (Centigrade)	Population (million)	Moisture (%)	electricity consumption price(Toman)
۲۰۱۴	18.08	3141748	77.757	532
۲۰۱۵	18.22	3168416	78.127	586
۲۰۱۶	18.37	3195083	78.523	644
۲۰۱۷	18.51	3221750	78.893	709
۲۰۱۸	18.66	3248418	79.29	780
۲۰۱۹	18.81	3275085	79.686	858
۲۰۲۰	19.06	3301753	80.582	944

Using the estimated values in table 6 along with the regression model built by Clementine, we proceed to calculate power consumption rate. The results obtained from this prediction are presented in table 7 and figures 7.

TABLE VII. Predicting electricity consumption Rate from 2014 to 2020

Year	*Rate(MW/H)Predicting electricity consumption*
۲۰۱۴	5991434
۲۰۱۵	6186759
۲۰۱۶	6376851
۲۰۱۷	6555836
۲۰۱۸	6726603
۲۰۱۹	6886974
۲۰۲۰	7076796

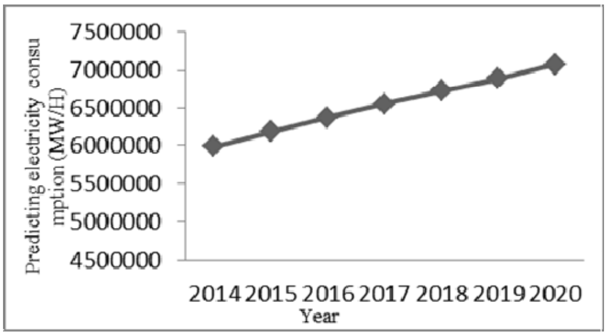

Figure 7. The Chart Predicting electricity consumption Rate from 2014 to 2020

Based on the prediction made from the years 2014 to 2020, the electricity consumption rate will increase yearly by around 3.2%. This increase for the years 2007 to 2013 and 2000 to 2006 was 4% and 10%, respectively.

Despite of the previous references which are used only one type of consumption, in this paper we study on the several type of power consumption includes the domestic, commercial, industrial and agricultural consumptions.

However, there are also other types of consumptions (e.g. public consumption, street lightings, and free consumptions) that we disregard due to their trivial share in the total electricity consumption rate. Figure 8 shows the electricity consumption rate based on the type of consumption.

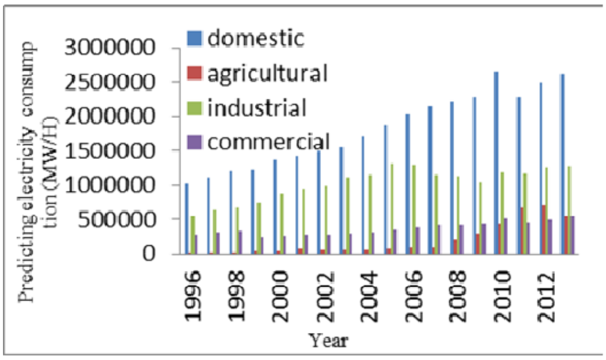

Figure 8. The electricity consumption rate based on the type of consumption from 1996 to 2013

As you can see in this figure, one of the prominent factors in the total electricity consumption rate is the domestic consumption type. This type on average makes up about 48.96% of the total consumption rate from 1996 to 2013. This rate for the industrial, commercial and agricultural types of consumption out of the total consumption rate is 27.02, 9.50 and 3.16, respectively.

5. Conclusions and Suggestions

In this research, based on the conducted analyses, the regression model generated less error rate on the tentative data compared to other models towards the electricity consumption prediction. For this reason, the regression method was employed for electricity consumption prediction. The predictive variables in this research include population, temperature, moisture, and electricity consumption price. Based on the estimated values of table 7, the electricity consumption growth rate in the year 2020 will increase by 22.28% compared to the year 2013. While the electricity consumption growth rate between the years 2006 to 2013 and the years 1999 to 2006 equaled 28.41 and 73.53, respectively. Moreover, the results obtained from this prediction indicate that the electricity consumption on average will have a 3.2% increase per year. As it was mentioned in table 5, the most prominent variable in electricity consumption is population. Also, one of the reasons for the decrease in electricity consumption growth rate compared to two decades ago is the decrease in population growth rate. Another reason for the decrease in electricity consumption growth rate compared to two decades ago is the sharp increase in electricity consumption price in two decades ago. In the conducted analysis, population is the most important factor participating in electricity consumption and has a great effect on the increase in electricity consumption rate.

Using the monthly or daily data can help make more accurate predictions. For example, having a 3-year period electricity consumption rate, we can have more data and therefore it helps us in making predictions with less error rate.

References

[1] C. J. Chang and C. D. Li and C. C. Chen and W. C. Chen, Forecasting short term electricity consumption using the adaptive grey-based approach—An Asian case, OMEGA journal, 40, 2012, 767-772.

[2] Z. Al-Garni and S. M. Zubair and J. S. Nizami, A regression model for electric energy consumption forecasting in Eastern Saudi Arabia. Energy, 19, 1994, 1043–1051.

[3] S. K. Kalogirou and M. Bojic, Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy, 25, 2000, 479–487.

[4] S. Ahmad and M. Y. Hassan and M. P. Abdollah and H. A. Rahman and F. Hussin and H. Abdollah and R. A. Saidur, Review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 33, 2014, 102-108.

[5] K. F. Geoffrey and K. W. Kelvin, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy journal, 32, 2007, 1761-1767.

[6] J. L. Deng, Control problems of grey systems. Systems and Control Letters, 1, 1982; 288–381.

[7] Kusiak, Data mining for prediction of wind farm power ramp rates. IEEE International Conference on Sustainable Energy Technologies, 2008, 1099 – 1103.

[8] V. Bianco and O. Manca and S. Nardini, Electricity consumption forecasting in Italy using linear regression models. Energy journal, 34, 2009, 1413–1421.

[9] Ekonomou, Greek long-term energy consumption prediction using artificial neural networks. Energy, 35, 2010, 512–517.

[10] K. Kavaklioglu and H. Ceylan and H. Ozturk and O. Canyurt, Modeling and prediction of Turkey's electricity consumption using Artificial Neural Networks. Energy Conversion and Management, 50, 2009, 2719–2725.