

A survey

Predictive Model Based Electrical Consumption

Dhivya V., Diana Paul, R. Kanagaraj

Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India

ABSTRACT

The data mining is the task of analyzing of large quantities of data to derive previously unknown, interesting patterns such as groups of data records, unusual records, and dependencies. It has the ability to turn raw data into useful information. One of the important invention of mankind is electricity. It is considered as a blessing. The availability of this power is very much needed for development and economical stability of the nations. The electricity board in various states and countries perform tasks such as generation, transportation and distribution of electricity to its customers effectively. In this study the various data mining techniques are applied to derive information from the electricity consumption databases.

Keywords: Data Mining, Electricity Consumption

I. INTRODUCTION

Every human in this earth is dependent on electricity. The importance of electricity can be understood during the few minutes of power outages we encounter. The life of people almost stops when power outages. In this modern world one cannot imagine a life without electricity. The electricity is the used in many areas such as Communication, Entertainment, Work, Transportation, Food and Home. Electricity is a constantly developing technology, there would have been less technology in hospitals if electricity was not present, thus the human life is indirectly dependant on electricity. At home, electricity acts as the primary source of energy for water heating or heating, cooking, air conditioning, lighting and various other purposes. The electricity is considered the lifeline of the world due to the reasons mentioned above. The study about the electricity consumption can be done by using techniques and methods of data mining. Data mining is referred as Knowledge Discovery in Data.

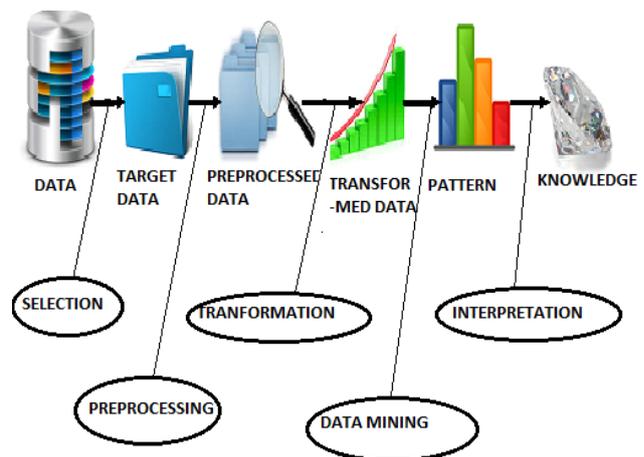


Figure 1. Steps in Knowledge Discovery of Data

Data can be mined from flat files, spreadsheets, database tables, or some other storage format. The most important criteria for the data mining is not the storage format, but its applicability to the problem to be solved. Data mining would not be an easy task, as it has complex algorithms and data will not always available at a single location. It has to be integrated from various heterogeneous data sources. A few widely used techniques include Characterization, Association and Correlation Analysis, Classification, Prediction, Cluster analysis, Outlier analysis, and Evolution analysis.

The data mining techniques are used to forecast the consumption of electricity. Forecasting is an art and science. According to Neils Bohr “Prediction is a very difficult art, especially when it involves the future”. In the present world, an underestimate of electrical consumption may lead to localized brownouts or even blackouts. In case of overestimate could lead to authorization of plants which may not be needed for years. Short term forecasting helps in regulation. The forecasting of data various from simple extrapolation method to time series techniques. The forecasting of electricity using the data mining techniques is the objective of the proposed system. A survey to perform the forecasting of electricity is being done in this paper.

II. LITERARY SURVEY

Ratore et al, proposes that by using data mining techniques the discovery of the electrical consumption pattern at regional level of the city can be determined[1]. The relationship between the electrical consumption and the atmospheric temperature along with geographical patterns such as river, farm, ground and highway are to be mined. The mining techniques included in the system are clustering and association rule mining techniques. The validation of the proposed system are done for a Sangli city. The clustering is the technique which is applied to form groups or clusters of data representing a common property. The association rule mining is used to establish relationship between the different entities. The clustering is done for the electrical consumption data and atmospheric temperature. The k-means algorithm is being used for the clustering. One of the conclusion obtained from clustering is that electrical consumption increases as with the increase of temperature. Thus electrical consumption is directly proportional to atmospheric temperature. The “Apriori” algorithm is used to derive the association rules by using the support and confidence. A set of five association rules have been derived. The data mining model depicted is generalized and can be derived for various other geographical region by analyzing the spatial data of the region and the electrical consumption data.

- (a) IF (River AND Farm, Near) THEN Electricity consumption (Medium OR Low)
- (b) IF (River OR Farm, Near) THEN Electricity consumption (Medium OR Low)
- (c) IF (Ground OR Highway, Near) THEN Electricity consumption (High OR Medium)
- (d) IF ((River OR Farm, Near) AND (Highway, Near) THEN Electricity consumption (Medium)
- (e) IF (Ground AND Highway, Near) THEN Electricity consumption (High)

Figure 2. Association rule set derived for Sangli city[1]

Fu et al, provides the glossary for the researches on how the current time series data mining deployed and a guideline for the research direction for future investigation[3]. The time series is the collection of data made chronically. The characteristics include the large data, high dimensionality and update regularly. Thus the data mining on time series data need be summarized into a series of steps to derive knowledge from it. The steps include Time series representation and indexing, Similarity measure, Segmentation, Visualization, Mining in time series. The time series representation and indexing includes the reduction of dimension by one of the techniques namely sampling, piecewise aggregate approximation(PAA),adaptive piecewise constant approximation (APCA),linear interpolation, perceptually important points (PIP). The representation of the time series can be from numerical to symbolic form. The efficiency of the indexing in depending on the precision of approximation in reduced dimension space.

The similarity measure is fundamental step for time series analysis and data mining. The similarity measure can be done by whole sequence matching and subsequence matching. The segmentation can be considered as preprocessing step in time series analysis. Visualization is the mechanism of presenting the proposed time series for the future analysis. A few tools used for visualization are cluster and calendar-based visualization tool, Time Searcher tool, spiral visualization tool and VizTree. The mining in time series is done as Pattern discovery and clustering, Classification, Rule discovery and Summarization[3]. Qian et al, proposes a novel data-representation scheme and a framework for clustering algorithms [4]. Clustering is an important tool in data mining. It is used

to discover the grouping structure which is present in a set of objects or data sets. It has applications in area, granular computing, text mining, information retrieval, bioinformatics, customer analysis, Web data mining, and scientific data exploration. The objective of clustering is to group a set of objects into meaningful clusters. The clusters have objects which are similar to each other and other clusters are very different. Various clustering algorithm are to deployed for this purpose.

For the numeric data, the k-means-type algorithms are very representative, that effectively and efficiently organize the objects into several clusters. In k-means algorithms, all objects are depicted by the features with numeric domains. Therefore, the objects described can be considered in a Euclidean space, and the similarity or the dissimilarity between two objects can be measured by a Euclidean distance, a cosine distance, and so on.

```

Pseudocode for k-means algorithm
Make initial guesses for the means  $m_1, m_2, \dots, m_k$ 
Until there are no changes in any mean
Use the estimated means to classify the samples into clusters
For  $i$  from 1 to  $k$ 
Replace  $m_i$  with the mean of all of the samples for cluster  $i$ 
end_for
end_until

```

Figure 3. Pseudocode for k-means algorithm

Feng et al, proposes a soft skill based association rules mining. Association rule is used to capture interesting dependencies between variables in large datasets. The paper offers a detailed insights into soft set based association rule mining. Two case studies are conducted to demonstrate the soft set based association rule mining approach. Soft set theory is a general mathematical tool to deal with uncertainty. The soft sets is founded on the idea of parameterization, that suggests that objects should be perceived from various points of view. These views provide a description of the whole entity with high complexity.

Various authors have introduced concepts such as algebraic operations in soft set theory and examined their basic properties, new operations in soft set theory to further consolidate the algebraic basis of soft sets, congruence relations on soft sets and discussed certain lattice structures, soft set model of equivalence classes

of information systems and applied it to the selection of clustering attributes for categorical datasets, soft set based algorithm for clustering attribute selection and their extensions have been successfully applied to various algebraic structures and decision making problems. In addition to the concepts introduced the improving some existing definitions for mining association rules based on soft sets is done in the paper presented by Feng et al[5].

Guarín et al, proposes the application of an educational data mining approach to model academic attrition (loss of academic status) at the Universidad Nacional de Colombia. The techniques such as naïve Bayes and a decision tree classifier are used for prediction of the loss of academic status due to low academic performance. The implementation of the prediction models was done using Rapid Miner and Weka for the Bayesian classifier and the decision tree respectively. These methods were selected based on the need for a predictive model that is descriptive at the same time in order to acquire a better understanding of the event of loss of academic status. The model has been evaluated using different scenarios depending on the available information.

Hence based on the need of the information the prediction technique has to be determined. Thus for the implementation of prediction model of electricity, Fuzzy Back Propagation network can be selected.

Devi et al, proposes a survey on forecasting electrical consumption in Tamil Nadu [2]. In this paper the various data mining techniques are discussed. The importance of the forecast of the electricity in Tamil Nadu are being stated. The prediction techniques such as Fuzzy Logic, Particular Neutral network and Expert Systems are highly efficient. The survey details the load prediction in Tamil Nadu. The procedure of data mining is also explained in this paper[2].

III. CONCLUSION

The study has lead to the conclusion that data mining allows the use of different approaches to build different models based on the type of data and objectives. Data mining methods can be predictive and descriptive. Predictive methods are classification techniques and descriptive method covers association rules and

clustering. These techniques can be applied to extract information from electricity consumption data of consumers. Each region in a state has thousands of consumers where electricity is supplied in a hierarchical manner. The mining can be used for short and long term forecasting of electricity demand. The mining techniques such as clustering and rule mining can be used to derive at rule sets.

Thus one of the objectives of the system to be constructed are to derive association rule set by the comparison of electricity consumption and geographical structures. Another objective is to predict the electricity consumption for further years.

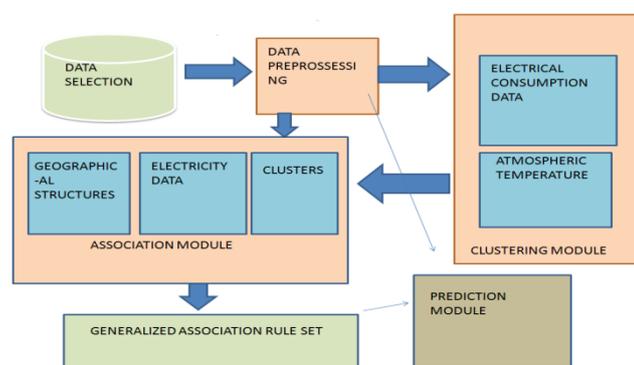


Figure 4. The architectural diagram of the proposed system

The survey study done has led to conclusion and selection of the various data mining that can be used to obtain the objectives mentioned. The data sets are to be collected and after preprocessing the clusters are formed. The clusters are used to derive at the association rule sets after the application of association rule mining. The prediction module uses electrical data of previous years and the association rule sets to obtain the results.

IV. REFERENCES

[1] Ravindra R. Rathod, Rahul Dev Garg, "Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data", *Electrical Power and Energy Systems*, Volume 78, June 2016, Pages 368–374.

[2] Dr. Mrs. M. Renuka Devi, R. Manonmani, "Electricity forecasting using data mining techniques in tamil nadu and other countries– a survey", *International*

Journal of Emerging trends in Engineering and Development, Issue 2, Vol.6, September 2012.

[3] Tak-chung Fu, "A review on time series data mining", *Engineering Applications of Artificial Intelligence*, Volume 24, Issue 1, February 2011, Pages 164–181.

[4] Yuhua Qian, Bing Liu, "Space Structure and Clustering of Categorical Data", *IEEE transactions on neural networks and learning systems*, vol. 27, no. 10, october 2016.

[5] Feng Feng, Junghoo Cho, Witold Pedrycz, Hamido Fujita, Tutut Herawan, "Soft set based association rule mining", Volume 111, 1 November 2016, Pages 268–282.

[6] Camilo Ernesto López Guarín, Elizabeth León Guzmán, and Fabio A. González, "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining", *IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje*, Vol. 10, No. 3, August 2015.