

Data Mining Technology and its Applications to Power Systems

Amit Kumar Mishra
Department of Electrical Engineering,
Faculty of Engineering,
Dayalbagh Educational Institute,
Agra-282010

Ajay Kumar Saxena
Department of Electrical Engineering,
Faculty of Engineering,
Dayalbagh Educational Institute,
Agra-282010

ABSTRACT

With increasingly environmental constraints the modern Power and Energy Systems are experiencing huge transformations in many ways. The quantum of data in power systems is growing rapidly due to large database used by power systems engineers for various operations. From the power generation plants the electrical energy is transmitted and distributed to end users. Frequent failure of various equipments and the systems has made it impossible to maintain the continuity of supply. Sometimes these failures are beyond the control of the power system operator. The operation and planning of power systems provide a large amount of data and it is difficult to extract the useful information from this large database that is continuously used by operators. Data mining is a process of extracting interesting and previously unknown knowledge from a set of data. The data mining techniques help power systems planner/operator to have smooth system planning, operation and are useful for extracting useful information from the existing data banks. The paper describes the data mining technology and its applications that would be useful in power systems.

Keywords

Data Mining, Machine learning, Cluster analysis, Neural Networks

1. INTRODUCTION

The engineering design, planning, installation, operation and maintenance tasks are performed by the power industry to facilitate high quality and secure supply of electricity to the consumers. Continuous development and use of recent techniques has also been a challenge for power system engineers.

The deregulation of the Electric Power System has led to several challenges, constraints and complexity of power system operations. Supply of real-time high quality electric power to the consumers by the suppliers has become a difficult task. Depleting conventional power generation sources has led to search for newer and renewable generation sources. The integration of newer power generation sources with the conventional sources of generation leads to systems that are much more complex and poses greater challenges to

power system engineers. Integration of large numbers of smaller and distributed generation sources has created several problems with the grid. Some of these include load forecasting, generation scheduling, system operation, control and risk management. Power systems have complex characteristics due to non linearity. Various problems associated with the wind or solar power integration pose greater technical challenges.

The adoption of data mining techniques has been actuated by the growth of available data in the electric power sector which is an interdisciplinary task and needs coordination of an electrical and computer engineer to interface power system and data mining [1]. Information plays a vital role in the field of Information Technology. The main task of any data mining tool is to collect data from heterogeneous sources, store, maintain and analyze it to extract desired information from that.

One of the main reasons to use data mining in a power system is to find interesting patterns and trends in data which includes various events of system operation. There is a continuous demand to utilize information languishing in this era to make

it effective and efficient due to large quantity of data that are generated from system operations.

So data mining techniques have an important role to play in power system operation and control. These techniques can be applied to load forecasting, fault classification, critical bus identification, load characterization and various other applications. These techniques are also used to find out undiscovered relationships in data that allows making predictions. Data mining is also termed as knowledge discovery in databases (KDD) [2].

This paper elaborates the aspect of data mining and its uses in various fields. Data mining in today scenario is having wide spectrum, which has motivated researcher to understand linear as well as nonlinear relationship. This paper discusses the past, present and future of data mining and its application to power systems. Data mining enables the users to deduce the result by applying the set of patterns (models, relations or examples). Data mining has potential to filter out the most important information from the data warehouse. The techniques that is used in data mining to gather information is called modeling. This technique works directly on the data. The field machine learning, Inductive learning, supervised and unsupervised learning have been drawn from data mining research.

2. DATA MINING PROCESS

2.1 Machine Learning

Machine learning is new generation of data mining. It is combination of statistics and Artificial Intelligence (AI). Machine learning is an evolution of Artificial Intelligence because it combines AI heuristic operations with the advanced statistical analysis. Computer programs learn about the data they study and make decisions based on qualities of study data using statistics for fundamental concepts [3].

2.2 Inductive Learning

Inductive learning is a model based learning where the subject is analyzed on the basis of set patterns and thus classified in class and similar pattern and hence enable user to classify to object into said class by their pattern and rules or trends. This type of learning is dynamic and adaptive. Due to this nature, it is only possible to study a small number of properties of object to classify into classes.

Table 1. Comparison between Machine and Inductive Learning

Characteristics	Machine Learning	Inductive Learning
Approach	Automated Approach	Model Based Approach
Nature	Concept Based as in Algorithms	Dynamic and adaptive
Coverage	Multiple observation at a time	Small sets at a time

2.3 Supervised Learning

In supervised learning user help the system to construct a model by defining class /set patterns to the object. Thus the system has only to find the common properties in the example. The description of the unseen object is formulated and it forms the class further it is used to find out set patterns and finally it can be used to predict the class of an unseen object.

2.4 Unsupervised Learning

In unsupervised learning no class is defined by the user. In this learning technique, user can make an observation of learning. Thus systems have to identify the pattern set by it. So here environment of learning phase plays a vital role to define the class.

3. DATA MINING APPLICATIONS

This part deals with typical application areas of Data Mining in Power Systems. There are various methods that are used for power system problems. The main machine learning technique used in power system is artificial neural networks, but the other data mining techniques are also used in some power system areas with good results [4]. Huisheng [5] used Fuzzy/ neural network for solving the problem of Fault classification in the transmission lines. Kezunovic [6] used

neural network for detection and diagnosis of transient fault in transmission lines. Mori [7] describes some fundamental concepts of power systems by introducing data mining techniques in power systems and also state the kind of problems that the electric industry tries to solve with data mining technique. Vasilic [8] used the neural network for faults classifications in transmission lines. Tso [9] used statistical analysis for detecting the substations most sensitive to the disturbances in Transmission and distribution systems. Vasilic [10] used the fuzzy / neural network for faults classifications in transmission lines. Dola [11] have used Decision tree and neural network for solving the problem of Faults classification in the distribution system. Figueiredo [12] used decision tree for classification of electric energy consumer in distribution system. Silva [13] used neural network for fault detection and classification in transmission lines. Costa [14] used neural network for fault classification in transmission lines. Chia-Hun et. al. [15] have used adaptive wavelet networks for Power-quality detection in power system disturbances. Hagh [16] used neural network for faults classification and locations in transmission lines. Dash [17] used support vector machine for classification identification of series and compensated transmission lines. Bhende [18] used neural network for detection and classification of power quality disturbance. Pang et. al [19] have used wavelet networking for power quality detection of power system disturbance. Saibal [20] used wavelet networking for classification of transient in distribution system. Ramos et. al. [21] have used decision tree for characterization and classification of consumer in distribution system. A Mosavi [22] used multiple criteria Decision making pre-processing using Data mining tools.

Neural networks have been put in use for fault location quite recently and have gained significant importance since Sobajic et. al. [23] have used neural networks for the prediction of critical clearing time. Wide usage of neural networks started by late eighties and during early nineties. To achieve greater efficiency in power system fault detection, classification and finding its location the neural networks are also generally usually used. A lot of research has been done and a large number of papers have been published for finding fault locations using neural networks [24]. Certain significant techniques and results that have been published are briefly discussed here. A majority of the work mentioned here made use of feed-forward multilayer perceptrons technique. Kulicke et. al. [25] have used neural networks for the detection of faults on transmission lines and also differentiated between arcing and non arcing faults. A new technique for the detection and location of high speed faults using neural networks has been proposed by Rikalo et. al. [26].

Based on the review of literature, the methods and specific problem where data mining is used is represented in Fig 1 and Fig 2

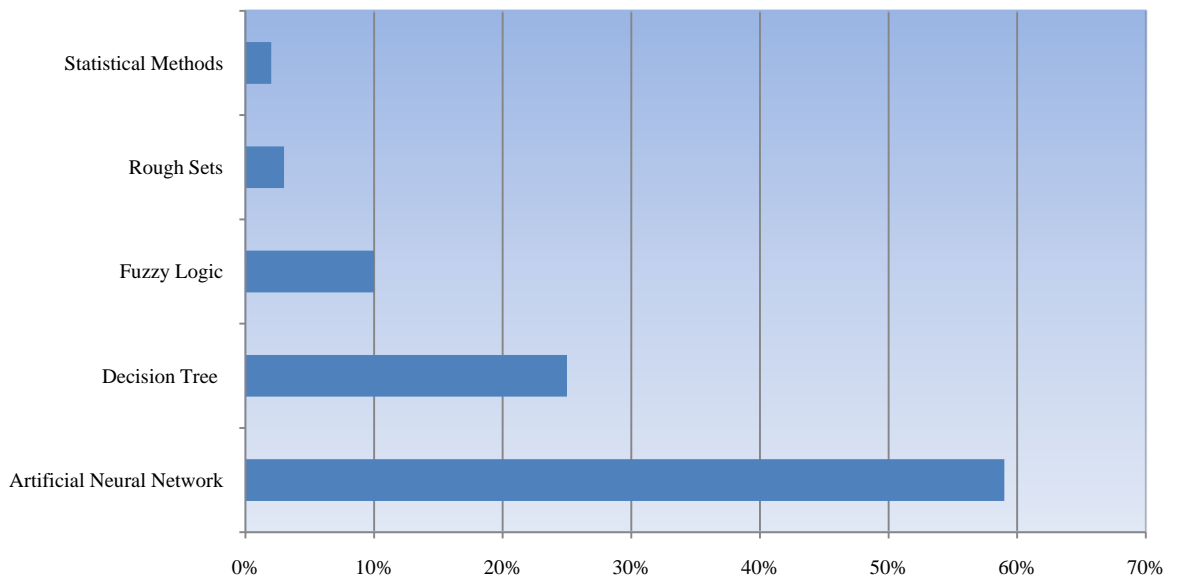


Fig 1: Data mining method applied to power systems

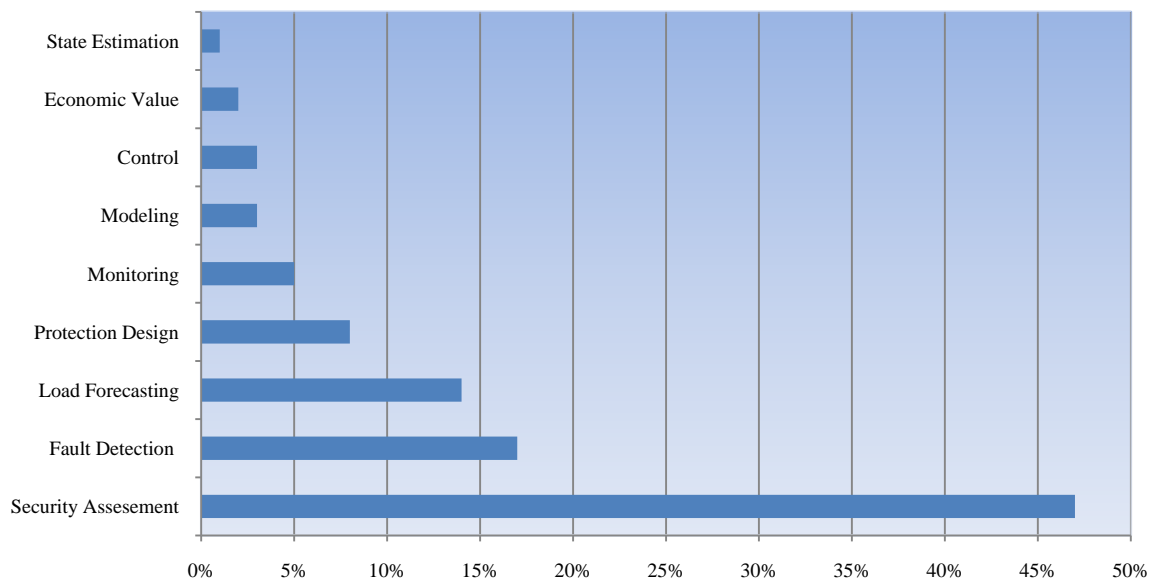


Fig 2: Different types of problems in Power systems

3.1 Classification And Prediction

Data analysis can be performed in two forms: Classification, and Prediction. Classification is a supervised learning technique [27]. It is applied to the areas where the objective or class level is pre-characterized or definitely known. Prediction is a form of data analysis which is applied to data classes to predict future data sets. It is a type of data analysis and

investigation which is connected to the information classes to anticipate future information sets. It is a supervised learning that uses some sort of relationship in the middle of data and yield values so that selected data can be utilized to predict the class of a new object that is not known [27]. Figure 3 shows the prediction process.

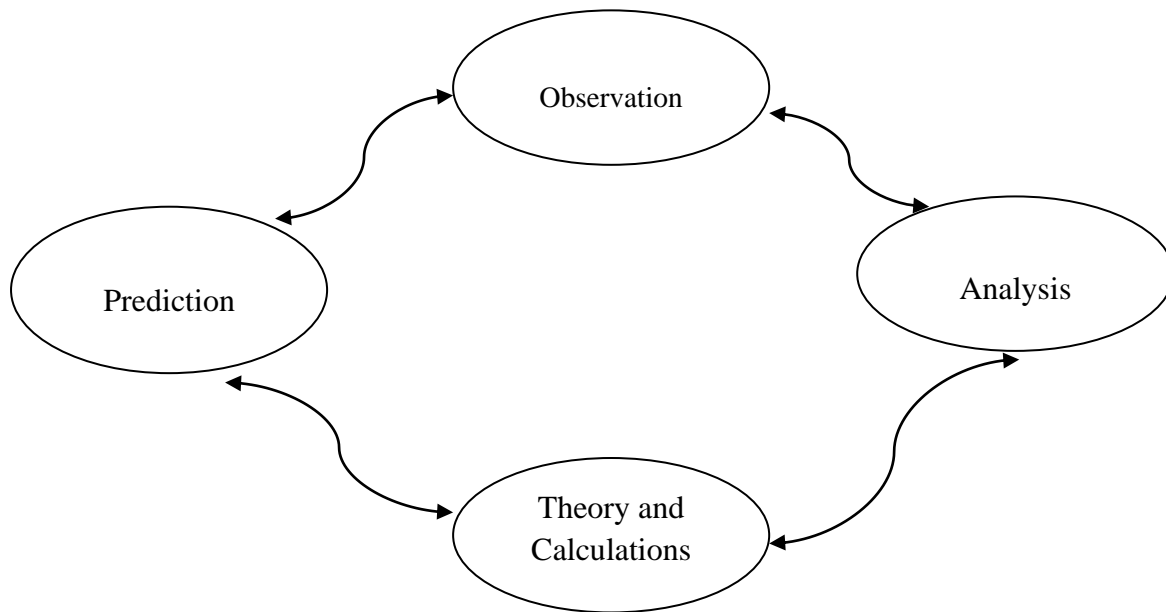


Fig 3: Prediction Process (Observation, Analysis, Theory & Calculations and Prediction)

3.2 Genetic Algorithms

Genetic Algorithm is an area for exchanging of views between Computer Science and Biology. It is a Meta learning strategy which tells that this method would be used by individual with the combination of other learning techniques. To get possible optimized solution for a problem, a genetic algorithm is used by building a population of chromosomes. “Genetic algorithms are adaptive heuristic search algorithms and are based on the evolutionary ideas of natural selection process and genetics” [28].

Inside generation of a population, the chromosomes are arbitrarily adjusted with expectations of making new chromosomes that have better assessment scores. The next generation population of chromosomes is randomly chosen from the present generation with selection probabilities based on the assessment score of every chromosome. Numerous adjustments are conceivable that can upgrade the execution for a given application, and some little improvements have been made to upgrade the execution of this benchmark.

The basic structure of Genetic algorithm is shown below [29]:

Basic Genetic Algorithm ()

```

{
    Initialization;
    Evaluation;
    While termination criterion not reached
    {
        Selection_and_Reproduction;
        Crossover;
        Mutation;
        Evaluation;
    }
}
  
```

It uses the combination of selection, recombination and mutation to evolve a solution to a problem.

3.3 Artificial Neural Networks

Genetic Algorithms derive their inspiration from Biology while Artificial Neural Networks are modeled on the human brain. The human brain can be depicted as biological neural system. It is an interconnected web of neurons transmitting intricate patterns of electrical signals. The input signals received by dendrites and based on these inputs signal it produced an output signal via an axon [30].

Artificial neural networks (ANNs), represents computational standards in biological metaphor and are rapidly gaining popularity among power system researchers. Due to improved and efficient performance of ANNs for solving complex problems, the power system engineers are also using this technique for solving various kinds of power system problems [31].

3.4 Decision Trees

A decision tree can be utilized as a model for a sequential decision problem under uncertainty. A decision tree uses graphical representations for decision which has to be made, for events that can occur. The probability is allocated to the events, and values associated with it are calculated for each outcome. An objective of the analysis is to determine the best decisions. Decision tree construct classification or regression models in the form of a tree type structure. It separates a dataset into smaller and smaller subsets while in the meantime a related choice tree is incrementally created. The final results of a tree have decision nodes and leaf nodes. A decision node has two or more branches where as leaf node denotes a classification or decision. The root node is the topmost decision node in a tree which corresponds to the best predictor. Decision trees can hold both categorical and numerical data [32].

So decision tree algorithms scale up very well for large data sets. This algorithm produced best result then the naïve prediction on the data sets.

3.5 Clustering

Clustering is an example of unsupervised learning technique. It is a powerful method to extract useful information from the data set. Clustering is defined as the process of organizing objects into groups whose elements are homogenous in some way and that is called segment or cluster. A cluster is a collection of objects which are similar between them and are different to the objects belonging to other clusters [33]. The object that is not part of any cluster is called outlier. It is a classification of objects from data in which classes are not pre labeled

4. CONCLUSIONS

Modern power system is a very large and interconnected complex system. Due to its complex nature and reliability issues a large amount of data is to be collected, analyzed and processed for smooth operation of the system. In order to find the desired information with greater accuracy in minimum time, data mining techniques are used in power systems also. In this paper various methods used for fault detection and security of the power system employing data mining technique has been discussed in brief.

5. REFERENCES

- [1] J. Morais, Y. Pires, C. Cardoso, and A. Klautau, 2009. "An Overview of Data Mining Techniques Applied to Power Systems".
- [2] Han, J. & Kamber, M., 2001. "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers.
- [3] J.Furnkranz, D.Gamberger, N.Lavrac, 2012. "Foundation of Rule Mining".
- [4] Z. A. Vale, C. Ramos, S. Ramos, and T. Pinto, 2009. "Data mining applications in power systems: Case-studies and future trends," Transmission and Distribution. Conf. Expo. Asia Pacific, pp. 1–4.
- [5] Huisheng Wang, Keerthipala, W.W.L. 1998. Fuzzy Neuro approach to fault classification for transmission line protection. IEEE Transactions Power Delivery, Vol.13, No. 4, 1093-1104, ISSN 0885-8977.
- [6] Kezunovic, M., Vasilic, S., Gul-Bagriyanik, F. 2002. Advanced Approaches for Detecting and Diagnosing Transients and Faults.
- [7] Mori, H., Kosemura, N., Kondo, T., Numa, K., 2002. Data mining for short-term load forecasting. Power Engineering Society Winter Meeting. pp. 623 – 624, ISBN 0-7803-7322-7.
- [8] Vasilic, S., Kezunovic, M., 2002. An Improved Neural Network Algorithm for Classifying the Transmission Line Faults. IEEE Power Engineering Society Winter Meeting, pp. 918 – 923. ISBN 0-7803-7322-7.
- [9] Tso, S.K., Lin, J.K., Ho, H.K., Mak, C.M., Yung, K.M., Ho, Y.K., 2004. Data mining for detection of sensitive buses and influential buses in a power system subjected to disturbances. IEEE Transactions on Power Systems, Vol. 19, No.1, 563 – 568, ISSN 1558-0679.
- [10] Vasilic, S., Kezunovic, M. 2005. Fuzzy ART Neural Network Algorithm for Classifying the Power System Faults. IEEE Transactions on Power Delivery, Vol. 20, No. 2, 1306-1314, ISSN 0885-8977.
- [11] Dola, H.M., Chowdhury, B.H. 2005. Data mining for distribution system fault classification. Power Symposium, 2005. Proceedings of the 37th Annual North American, pp. 457 – 462, ISBN 0-7803-9255-8.
- [12] Figueredo, V., Rodrigues F., Vale, Z., Gouveia, J. B. 2005. An electric energy Consumer characterization Framework based on data mining techniques. IEEE Transactions Power Systems, Vol. 20, No. 2, 596- 60), ISSN 1558-0679.
- [13] Silva, K. M., Souza, B. A., Brito, N. S. D.2006. Fault detection and classification in transmission lines based wavelet transform and ANN. IEEE Transaction on Power Delivery, Vol 21, No. 4, 2058-2063, ISSN 0885-8977.
- [14] Costa, F. B., Silva, K. M., Souza,B. A., Dantas, K. M. C., Brito, N. S. D. 2006. A method for fault classification in transmission lines bases on ANN and wavelet coefficients energy. International Joint Conference Neural Networks. pp. 3700 – 3705, ISBN 0-7803-9490-9.
- [15] Chia-Hung Lin, Chia-Hao Wang.2006.Adaptive wavelet networks for power-quality detection and discrimination in a power system, IEEE Transactions on Power Delivery, Vol 21, No. 3, 1106 - 1113, ISSN: 0885-897.
- [16] Hagh, M.T.; Razi, K., Taghizadeh, H. 2007. Fault classification and location of power transmission lines using artificial neural network, International Power Engineering Conference, pp. 1109 – 1114, ISBN 978-981-05-9423-7.
- [17] Dash, P.K., Samantaray, S.R., Panda, G. 2007. Fault Classification and Section Identification of an Advanced Series-Compensated Transmission Line Using Support Vector Machine. IEEE Transactions on Power Delivery, Vol 22. , No. 22, 67 – 73, ISSN 0885-8977.
- [18] Bhende C. N., Mishras S., Panigrahi B. K. 2008. Detection and classification of power quality disturbances using S-transform and modular neural network. Electric Power Systems Research, Vol. 78, 122-128.
- [19] Pang, P., Ding, G. 2008. Power quality detection and discrimination in distributed power system based on wavelet transform. 27th Chinese Control Conference (CCC 2008), pp. 635 – 638, ISBN 978-7-900719-70-6.
- [20] Saibal Chatterjee, Sivaji Chakravorti, Chinmoy Kanti Roy and Debangshu Dey. 2008.Wavelet network-based classification of transients using dominant frequency signature, Electric Power Systems Research, Vol. 78, No. 1, 21-29.
- [21] Ramos, S., Vale, Z. 2008. Data mining techniques application in power distribution utilities. Transmission and Distribution Conference and Exposition, pp. 1-8. ISBN. 978-1-4244-1903-6.
- [22] Mosavi,A.2010.Multiple Criteria Decision –Making Preprocessing Using Data Mining Tools.Journal of Computer Science,7(2),26-34.
- [23] Pao YH, Sobajic DJ, 1988. "Autonomous Feature Discovery of Clearing time assessment", Symposium of Expert System Applications to Power Systems, pp. 5.22-

5.27. 32.

- [24] S. B. Ayyagari, 2011. “Artificial Neural Network Based Fault Location for Transmission Lines”.
- [25] Kulicke B, Dalstein T, 1995. “Neural network approach to fault classification for high speed protective relaying”. IEEE Transactions on Power Delivery, vol. 4, pp. 1002 – 1009. 33.
- [26] Kezunovic M, Rikalo I, Sobajic DJ, 1996. “Real-time and Off-line Transmission Line Faulty Classification Using Neural Networks”, Engineering Intelligent Systems, vol. 10, pp. 57-63.
- [27] S. Sumathi and S .N .Srivinandam, 2006. Introduction to Data Mining and its Applications, Vol 29.
- [28] GeneticAlgorithms,2005.http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/hmw/article1.html
- [29] E. W. J.Lebak, A .Reuther, 2005.“Polymorphous Computing Architecture (PCA) Kernel-Level Benchmarks,” ReVision.
- [30] D.Shiffman, 2012. “The Nature of Code”.
- [31] Vidya Sagar S,Vankayala,Nutakki D.Rao,2003. Artificial Neural Networks and their applications to power systems.
- [32] http://www.saedsayad.com/decision_tree.htm
- [33] S. Kumar, M. Kathuria, A. K. Gupta, and M. Rani, 2012. “Fuzzy clustering of web documents using equivalence relations and fuzzy hierarchical clustering” .CSI Sixth Int. Conf. Software. Eng., pp. 1–5.