Robust Big Data Analytics For Electricity Price Forecasting In The Smart Grid

Mano S[1], Mrs. J . Kalaivani[2]
[2] ASST. PROF
[1, 2] Krishnasamy College Of Engineering and Technology, Cuddalore.

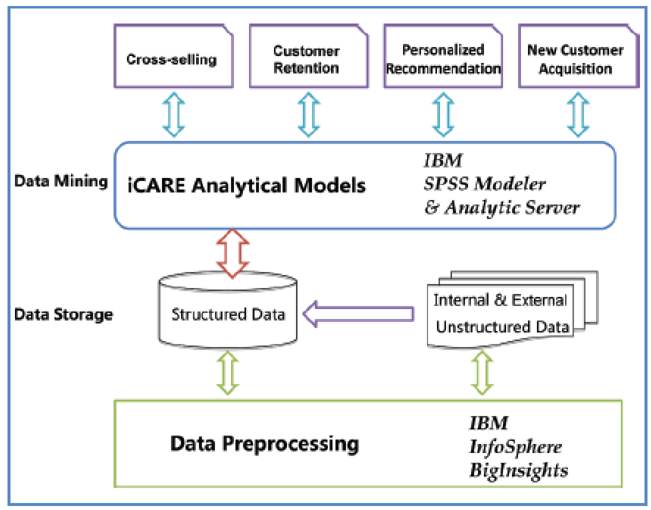
Abstract- *Electricity price forecasting is a significant part of smart grid because it makes smart grid cost efficient. Nevertheless, existing methods for price forecasting may be difficult to handle with huge price data in the grid, since the redundancy from feature selection cannot be averted and an integrated infrastructure is also lacked for coordinating the procedures in electricity price forecasting. To solve such a problem, a novel electricity price forecasting model is developed. Specifically, three modules are integrated in the proposed model. First, by merging of Random Forest (RF) and Relief-F algorithm, we propose a hybrid feature selector based on Grey Correlation Analysis (GCA) to eliminate the feature redundancy. Second, an integration of Kernel function and Principle Component Analysis (KPCA) is used in feature extraction process to realize the dimensionality reduction. Finally, to forecast price classification, we put forward a differential evolution (DE) based Support Vector Machine (SVM) classifier. Our proposed electricity price forecasting model is realized via these three parts. Numerical results show that our proposal has superior performance than other methods.*

Keywords- Data processing, Support vector machine, Big data, Prediction, Correlation analysis,

I. INTRODUCTION

Countries around the world have set aggressive goals for the restructuring of monopolistic power system towards liberalized markets especially on the demand side. In a competitive retail market, load serving entities (LSEs) will be developed in great numbers. Having a better understanding of electricity consumption patterns and realizing personalized power managements are effective ways to enhance the competitiveness of LSEs. Meanwhile, smart grids have been revolutionizing the electrical generation and consumption through a two-way flow of power and information

II. SYSTEM ARCHITECTURE



III. LITERATURE SURVEY

1. Big Data Analytics for Electricity Price Forecast:

Electricity Price forecast is a major task in smart grid operation. There is a massive amount of data flowing in the power system including the data collection by control systems, sensors, etc. In addition, there are many data points which are not captured and processed by the energy market operators and electricity network operators including gross domestic product, the price of fuel, government policy and incentives for renewable and green energy sectors as well as impacts on new technologies such as battery technology advancement and electric vehicles. In this study, data points from 2001 to 2017 were collected and 78 data points are considered for analyses to select the highly-correlated features which could potentially affect the electricity price. In the first step, a comprehensive correlation method using Pearson Correlation Coefficient is applied to find the points which are correlated with the electricity price. In the next step, the correlated data is fed to the machine learning algorithm for price forecast. The algorithm results were tested in the historical data in California and the outcomes were satisfactory for the three years forecast. The combination of featured selection and machine learning is giving superior outcomes than the traditional methods.

2. Data analytics in the electricity sector – A quantitative and qualitative literature review

The rapid transformation of the electricity sector increases both the opportunities and the need for Data Analytics. In recent years, various new methods and fields of application have been emerging. As research is growing and becoming more diverse and specialized, it is essential to integrate and structure the fragmented body of scientific work. We therefore conduct a systematic review of studies concerned with developing and applying Data Analytics methods in the context of the electricity value chain. First, we provide a quantitative high-level overview of the status quo of Data Analytics research, and show historical literature growth, leading countries in the field and the most intensive international collaborations. Then, we qualitatively review over 200 high-impact studies to present an in-depth analysis of the most prominent applications of Data Analytics in each of the electricity sector's areas: generation, trading, transmission, distribution, and consumption. For each area, we review the state-of-the-art Data Analytics applications and methods. In addition, we discuss used data sets, feature selection methods, benchmark methods, evaluation metrics, and model complexity and run time. Summarizing the findings from the different areas, we identify best practices and what researchers in one area can learn from other areas. Finally, we highlight potential for future research.

3. Big data analytics in smart grids: a review

Data analytics are now playing a more important role in the modern industrial systems. Driven by the development of information and communication technology, an information layer is now added to the conventional electricity transmission and distribution network for data collection, storage and analysis with the help of wide installation of smart meters and sensors. This paper introduces the big data analytics and corresponding applications in smart grids. The characterizations of big data, smart grids as well as huge amount of data collection are firstly discussed as a prelude to illustrating the motivation and potential advantages of implementing advanced data analytics in smart grids. Basic concepts and the procedures of the typical data analytics for general problems are also discussed. The advanced applications of different data analytics in smart grids are addressed as the main part of this paper. By dealing with huge amount of data from electricity network, meteorological information system, geographical information system etc., many benefits can be brought to the existing power system and improve the customer service as well as the social welfare in the era of big data. However, to advance the applications of the big data analytics in real smart grids, many issues such as techniques, awareness, synergies, etc., have to be overcome.

IV. MODULE DESCRIPTION

1. DATA PROCESSING MODULE:

The proposed technique having six steps to implement the system. In the very first step we are preprocessing the input data. This step includes the data providing and moralization of Usage curve.

2. SYMBOLIC AGGREGATE APPROXIMATION MODULE:

In the second step we are applying the, symbolic aggregate approximation technique to the data's and using that we are obtaining the reduced set of scaled data.

3. HMM Module:

In the third step we are applying the hidden morkove model for the derivation of Usage curve for the individual customer Usage as well as time based Usage history model.

4. PRICE PREDICTION MODULE:

The K-L distance algorithm used here to measure the Usage distance from the any two customer profile curve and based on that it will obtain the matrix for the customer Usage. In fifth step we are we are applying here the modified clustering algorithm to cluster the customers into different profiles based on their dynamic Usage curve.

Finally in our sixth stage we are applying the customer Behaviour driver algorithm to drive the target customers for the variable available business products.

V. PROPOSED METHODOLOGY

In this paper, we investigate the electricity price forecasting issue. Our objective is to predict the electricity price accurately by using the big data from grid. To overcome this challenging obstacle, we propose a Support Vector Machine (SVM) underpinned framework that can predict the price efficiently.

SVM is a classifier that try to find a hyperplane which can divide data into the correct classes. The support vector is a part of data that could help to determine the hyperplane.

We propose a parallelized electricity forecasting framework, called Hybrid Selection, Extraction and Classification (HSEC). The three components of HSEC are parallelized Hybrid Feature Selector (HFS) based on Grey Correlation Analysis (GCA), feature extraction process based on Kernel Principle Component Analysis (KPCA) and differential evolution (DE) based SVM classifier. The HSEC performs feature engineering by selecting features corresponding to the time sequence and the dimensional reduction of electricity price data features. The HFS uses the fusion of two feature selectors based on GCA rather than one to give an appropriate selection of features.

VI. CONCLUSION

Global energy needs are continuously growing. The conventional methods for producing more energy to meet the demand pose a great threat to the environment. Among other solutions, energy efficiency has become a major tool for minimizing the need for producing more energy to cater for the growing demand. Inherently, the cause of improving energy efficiency relies on understanding the usage patterns, identifying the problematic areas, establishing good energy consumption practices and to rectify the faults to reduce energy leakages. The advancement in sensors, ubiquitous computing and communication technologies has provided the basis for effectively collecting the usage data to understand energy usage. The collected data needs to be processed to generate leads for improving energy efficiency. The quality of insights generated from data improves if we consider the current data in context to historic data. This means that data volume for processing will keep on increasing. There can be multiple sources of data so the data formats can also vary. On the use case basis, data processing requires flexibility for customization and variation in speed of data processing. All of these data features refers to application of Big Data technologies for energy efficiency. Distributed parallel computing programming models like MapReduce provide the basic environment for handling Big Data. We leveraged on the power of Map Reduce using Apache Hadoop ecosystem tools to present an end-to end Big Data analytics tool. Hadoop supports scalability to meet large volumes of data sets while the there are other tools that can integrate with Hadoop to process complexity in data. We used the model platform to process real life energy data and generated insights that can be used to improve energy efficiency. The proposed model provides a 'plug and play' environment for many other analytic tools to integrate on a need basis. It is based entirely on open source software components and can be deployed using general purpose hardware or any cloud based model.

REFERENCES

[1] Arundel, A., and Kemp, R. Measuring eco-innovation. United Nations University Working Paper Series, 2009/017 (2009), 1{40.

[2] Balijepalli, V. M., Pradhan, V., Khaparde, S., and Shereef, R. Review of demand response under smart grid paradigm. In Innovative Smart Grid Technologies-India (ISGT India), 2011 IEEE PES (2011), IEEE, pp. 236{243.

[3] Boehm, B. A spiral model of software development and enhancement. SIG-SOFT Softw. Eng. Notes 11, 4 (Aug. 1986), 14{24.

[4] Box, G. E., and Jenkins, G. M. Time series analysis: forecasting and control, revised ed. Holden-Day, 1976.

[5] Commission, F. E. R., et al. Assessment of demand response and ad-vanced metering.

[6] Dean, J., and Ghemawat, S. Mapreduce: simpli ed data processing on large clusters. Communications of the ACM 51, 1 (2008), 107{113.

[7] Dr. Chryssis Georgiou, U. o. C. Reference guide: Acm style. http:// www.cs.ucy.ac.cy/~chryssis/specs/ACMrefguide.pdf. Accessed: 2014-06-09.

[8] Farhangi, H. The path of the smart grid. Power and Energy Magazine, IEEE 8, 1 (2010), 18{28.

[9] Forgy, E. W. Cluster analysis of multivariate data: efficiency versus interpretability of classi cations. Biometrics 21 (1965), 768{769.

[10] Forsstrom,• J., Lahti, P., Pursiheimo, E., Ram•a,• M., Shemeikka, J., Sipila,• K., Tuominen, P., and Wahlgren, I. Measuring energy efficiency.

[11] Grijalva, S., and Tariq, M. U. Prosumer-based smart grid architecture enables a at, sustainable electricity industry. In Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES (2011), IEEE, pp. 1 [12]Hanrahan, P. Vizql: a language for query, analysis and visualization.