

# The Application Analysis of Data Mining-Based Energy Internet and Big Data

Lina Yuan<sup>1,\*</sup>, Huajun Chen<sup>1</sup>, Bo Tian<sup>1</sup>, and Jing Gong<sup>2</sup>

<sup>1</sup>College of Data Science, Tongren University, Tongren, China

<sup>2</sup>Graduate School, Tongren University, Tongren, China

\* 893422817@qq.com

**Abstract.** The increase of energy utilization efficiency, the improvement of energy efficiency level, and the reduction of pollutant emission are the main targets of China's energy revolution. A series of energy technologies, energy management and energy supplies, such as the production, transformation, configuration and consumption of energy in the background of Energy Internet (EI), are moving in developing more diversified, cleaner, and highly efficient direction. With parallel network of distributed energy and implementation application of various smart watches and energy storage equipment, the amount of data involved in the intelligent terminal will be exponential growth, the workload of EI on data information processing is increasingly growing. This paper focuses on the application analysis of EI and big data (BD) based on data mining methods, and takes the operation model of electric vehicles (EV's) charging and discharging data analysis based on vehicle-cooperation as a classical case.

**Keywords:** energy utilization; Energy Internet; big data; electric vehicles.

## 1. Introduction

Energy Internet (EI) is for the purpose of developing green low carbon energy, integration of Internet information technology, and a novel smart energy system combined “long-distance large-capacity” power transmission technology with new renewable energy, which can effectively leverage its energy resources across the country and around the world, and at the same time support free access and operation of a large number of “the polymer of power generation and utilization”. In this sense, energy revolution as a new starting point and the vigorous promotion of constructing EI is bound to powerfully drive the establishment of the global energy system, in order to prompt the comprehensive construction of a clean and renewable energy pattern. Especially, at present, due to the dispersion, difficult collection, and complicated analysis of relevant data on EI, the centralized management of Big Data (BD) will change this situation. On the basis of abundant data dimension, and the combination of BD technology with the current rapid development and great achievements, it will significantly enhance the analysis of EI application scenario and the effect of assessment and professional simulation.

EI, based on the basic concept of the Internet, integrates their own advantages of information network and energy network, and forms a huge “wide area network (WAN)” consisting of “backbone” and “local area network (LAN)” [3], where backbone is usually the power grid, and LAN is some micro network and a distributed energy network using new energy and clean energy. The distinct feature of EI is open equivalence, not only including various energy and information, but also making energy and information exchange with each other, and both achieve dynamic balance, that is why the distributed energy and equipment, to the maximum extent, can integrate and interconnect with EI like a USB flash drive.

The advantages of EI are far more, and it can also implement hundreds of millions of interconnection among equipment, machines, and systems at the energy end, through high-end network sensors and applications, meanwhile, employing big data mining and analysis techniques, all kinds of data information (such as meteorological data, environment data, traffic data, grid operating data, etc.) are to be integrated into various forms of research reports, to facilitate the scientific understanding of the relevant personnel, thus work efficiency is greatly improved, and energy situation can be early warning and adjustment in time.

## 2. Basic Theories and Algorithms of Data Mining

Data mining refers to automatically extract the data acquisition process of those useful information hidden in the data from the data set. The most commonly used and accepted definition of data mining: extract the implied, unaware in advance, and potentially useful information and knowledge from abundant, incomplete, noisy, fuzzy, and random data in the process of data mining. Data mining is to extract, convert, analyze, and process a great deal of business data, to dig up the valuable, hidden, and potential business opportunities.

### 2.1 The Process of Data Mining

In order to form the unified specification of data mining industry, the European Union and other agencies jointly drafted the standard process of data mining industry, i.e., Cross-industry Standard Process for Data Mining (CRISP-DM), as shown in Fig. 1, where the arrow indicates the most important and frequent attachment relationship between each stage. In practical application, the order of each stage is not absolutely unchanged, and there must be a repetition of each stage.

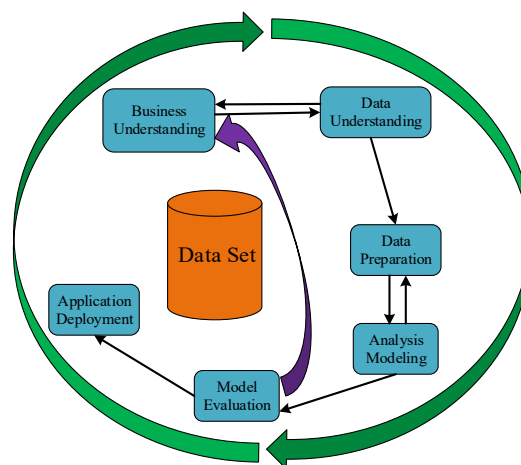


Figure 1. The process of data mining.

#### 2.1.1 Business Understanding.

The first stage of the CRISP-DM, before the start of data mining, is first to understand the current faced problems of data and business. To maximize the value of data mining, target problems should have a specific definition, appropriate data is prepared for data mining, thus the obtained results can be explained.

#### 2.1.2 Data Understanding.

Under the guidance of business analysis target, collect and understand the original data, merge data from different sources to the same data mining library, roughly remove some redundant or irrelevant data, and evaluate the available data.

#### 2.1.3 Data Preparation.

Organize and clean a series of available raw data, by selecting variables, choosing records, create new variables, and transform variables, based on the selected algorithms and tools to work for data processing, in order to achieve modeling requirements.

#### 2.1.4 Analysis Modeling.

For different data mining goals apply data mining tools to set up data analysis model, adopting diverse data mining algorithms, then build a distinct data mining models, and finally dig various types. Meanwhile, for the same data mining tasks, multiple data mining algorithms may be utilized, and the final algorithm selected need to consider many factors, such as the characteristics of the data to be processed, the algorithm principle, implementation mechanisms, data analysts familiar with algorithms, etc. In practical application, usually on the same data set, using a variety of data mining

algorithms, and through comparison and selection of multiple models, ultimately determine the most effective model on the current data set.

### 2.1.5 Model Evaluation.

To evaluate the established model, focus on whether the results are consistent with the objectives of step one. First, validate each process and procedure for the process of data mining, in order to make sure no error occurred for the whole process of data mining. Secondly, evaluate the quality, effectiveness, rationality, etc. of the data mining results. In practical applications, the main aspects of model evaluation for data mining model include: Model classification and prediction accuracy, model performance, model complexity, comparison between models, clustering effectiveness, etc. Common methods of model evaluation contain: Confusion Matrix, Information Gain, visualization approach, etc.

### 2.1.6 Application Deployment.

After establishing and validate data mining model, there are mainly two used methods. On one hand, provide the model to analysts for reference and assist in decision-making. On the other hand, the model is applied to various data sets to carry out data mining for different business objectives.

## 2.2 The Algorithms of Data Mining

Use a few commonly used data mining methods to EI involved in massive data to explore value, mainly including: classification analysis, clustering analysis, association analysis, regression analysis, dimension reduction analysis. The specific included algorithms are shown in Table 1.

Table 1. Summary of Data mining algorithms.

The types of algorithms	The names of algorithms
Classification analysis	Logistic Regression
	Bayesian
	Support Vector Machine (SVM)
	Perceptron
	Neural Network
	Random Forests
	Restricted Boltzman Machines
	Canopy Clustering
	K-means Clustering
	Fuzzy K-means
	Expectation Maximization
Clustering analysis	Mean Shift Clustering
	Hierarchical Clustering
	Dirichlet Process Clustering
	Latent Dirichlet Allocation
	Spectral Clustering
Association analysis	Parallel FP Growth Algorithm
Regression analysis	Locally Weighted Linear Regression
Dimension reduction analysis	Singular Value Decomposition
	Principal Components Analysis
	Independent Component Analysis
	Gaussian Discriminative Analysis

## 3. Data Mining Methods and Simulation Results

Energy Internet (EI) is a novel industry pattern merging “Energy” and “Internet”, which breaks the boundaries between the industry with the most intuitive impact on the traditional power industry, and combines with three mayor industry: information technology, electronic communication, and conventional electric power. Furthermore, the most representative and hierarchical influence on the

traditional industry for EI is the transition of the electricity market competition pattern and the focus of competition.

EI and big data technology is only effective combination and implementation in practical application in practical applications, various data source information, such as weather information, geographic information, environmental conditions, traffic conditions, power grid characteristic, power utilization, etc., can be made screening selection and processing analysis to the maximum limit, which indeed serves the vast masses of people and all kinds of users in need, not only stay in the theoretical level, and establishes the application scenarios of EI and big data technology according to the following steps as shown in Fig. 2.

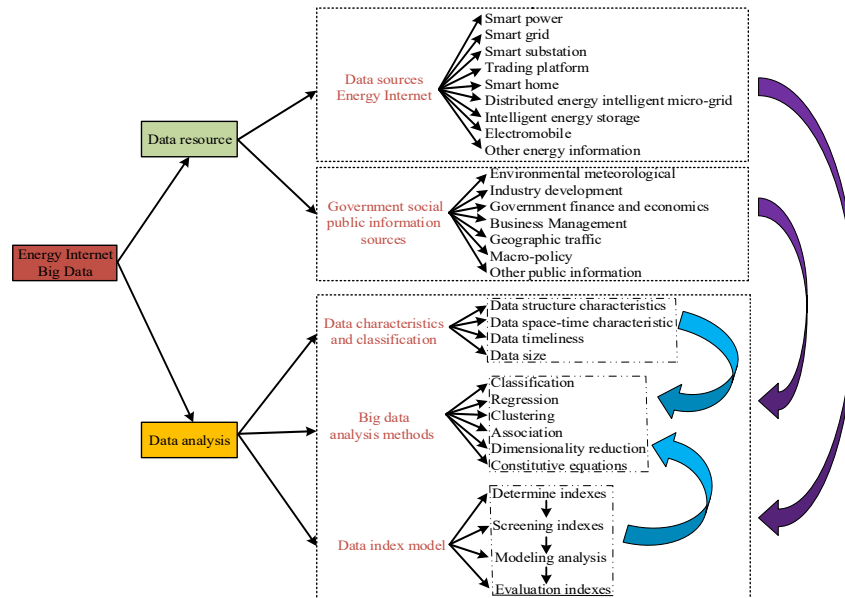


Figure 2. The flow chart of building EI big data.

Firstly, define the construction requirements. EI, especially, has the global EI of wide coverage, environmental diversity, numerous relevant professional fields, geographical range, and long construction cycle. In addition, the construction of the EI involves data types not only numerous but also dispersed. Neither EI nor BD alone will yield the perfect results. Therefore, the requirement must be fully recognized, in order to better research it.

Secondly, develop data platform. Joining various industries and fields, we advocate to establish resources data center of EI and BD technology, and develop data platform, through the platform to provide the presentation of data retrieval, and the analytic capability of computing. The centralized management of data will transform the current situation with the correlation data dispersion, difficult collection, and complicated analysis of EI. Based on abundant data dimension, making great progress of technologies, such as BD, analysis evaluation and professional simulation, etc., will greatly promote the capability of planning and research.

Lastly, carry on practical application. Data platform is the important basic components of EI and BD applications system, is mainly responsible for collecting data used in planning and experimental study to analyze and process. According to its research requirement, EI can collect various data to form data collection channels, build data model, evaluate theoretical calculation, design all kinds of indexes, and build evaluation system. Furthermore, employing the data warehouse, visualization technology, analysis results are shown to form analysis reports on various types of planning studies, really realize the deep integration of EI and BD, and provide the conditions of more convenient data analysis for thousands of households.

#### 4. Classical Case Analysis

In the future, with the rapid expansion of EI, energy storage technology will enter a leapfrog development stage. The application of mass small energy storage equipment, mainly including

charging and discharging piles of electric vehicles, will become the mainstream of energy storage direction. This section explains in detail the operation model of electric vehicles (EV's) charging and discharging data analysis based on vehicle-cooperation. The operators of charging and discharging facilities will operate and manage according to the following seven layers of charging and discharging piles.

First, time-sharing price (TSP). TSP is in the mainstream development direction of price with energy facilities in the future, each charging pile is a small e-commerce company, according to its characteristics, such as different geographical location, different charging time, how much power supply resources, the condition of network congestion, etc. Thus, e-commerce can adjust the price accordingly, to achieve the goal of TSP.

Second, advantages of EVs. In order to ultimately achieve the profit target in the power market, electric car industry can provide service measures, such as demand response, standby application, frequency modulation (FM), peak-shaving, etc., in the development process through the Internet technology. The effective ways to implement this goal is to be good at utilizing the charging and discharging equipment, which will be seen everywhere in the future.

Third, charging load analysis. In order to make the enterprise more understand the energy storage application and load characteristics of EVs, as well as to further speed up the development of enterprise intelligent electricity business, carry out distribution network construction planning. Through the methods, such as data association analysis, linear fitting, etc., in view of the charging equipment load behavior of different EVs, research on the related performance on the characteristics, for instance, time, geography, power, and energy.

Fourth, technical classification. There are obvious differences between direct current (DC) fast charging pile and ordinary alternative current (AC) charging pile from the technical perspective. The validity of the charging stations in the public area is extremely high, and the ordinary AC charging pile is difficult to meet its needs, while the DC fast pile can be fully charged and suitable for the public charging station. For normal car owners, the charging performance of ordinary AC charging pile can basically satisfy the needs of private cars, and can solve the problems in residential areas.

Fifth, charging mode. The car charge is a common charging mode, and in addition to this method, it is also very effective to charge alone battery to power battery using charging battery after the body and battery are opened. Whatever the mode, the ultimate goal of charging is to provide fast, high efficient, safe, and reasonable power supplies for the battery, and through this separation of charging technology, the battery replacement device can quickly remove the power cell from the vehicles and then directly replace the other, which is far more efficient than the vehicle charging mode.

Sixth, functional service. In view of the current demand of power users, charging pile not only needs the ability to supply at any time, based on this, but also realize the function of real-time time, meter, and amount charged, to act as the power user's purchasing equipment. In addition, as the transfer facilities continue to be deep in all walks of life and mass life, the charging pile should also have the function of mobile payment. As a result, the charging pile is not only better than the measurement of electricity, to make the data more accurate, but also updates and updates more reliable communication capabilities to meet the growing demand for communications.

Seventh, charging behavior analysis. For electric power enterprises, the specific location distribution of the power customers is an important basis for the customer construction of the best payment channels. To get this information, the enterprise can make statistical and analysis of the charging address and customer category of the electric car charging facility. Meanwhile, with increasing open of future power market, in order to make the power price system more diverse and more able to attract customers, analyze the customer's credit rating, combining data such as payment time, payment, payment timeliness, the amount of the payment, and residential property of the users.



## 5. Conclusion

In a word, leveraging the techniques of BD and data mining to analyze and process various data involved in the power field, can help the members of the grid area find the problems, optimize the operation, and improve the benefits in the environment of EI.

## Acknowledgments

This work was financially supported by the science and technology cooperation program project in Guizhou in 2015 (Project Number: [2015]7251 Marked by the word LH in Guizhou Province), the project of platform talent in Guizhou in 2016 (Project Number: [2016]5611), and partly supported by the National Natural Science Foundation of China (NO.61562703).

## References

- [1]. A. Q. Huang, M. L. Crow, G. T. Heydt, J. P. Zheng, S. J. Dale, The Future Renewable Electric Energy Delivery and Management (FREEDM) System, The Energy Internet, Proceedings of the IEEE, 2011, 99(1): 133-148.
- [2]. Peng Xiaosheng, Deng Diyuan, Cheng Shijie, et al., Key Technologies of Electric Power Big Data and Its Application Prospects in Smart Grid, Proceedings of the CSEE, 2015, 35(5): 503-511 (In Chinese).
- [3]. Liu Ruizhi, Evaluation Model of Market Transaction in Energy Internet Based on Big Data, North China Electric Power University, March 2017 (In Chinese).
- [4]. Chen Juan, Research on Planning of Regional Distributed Energy System under Energy Internet, North China Electric Power University, June 2017 (In Chinese).
- [5]. Wa Kun, Li Huining, Feng Yixiong, Tian Guangdong, Big Data Analytics for System Stability Evaluation Strategy in the Energy Internet, IEEE Transactions on Industrial Informatics, 2017, 13(4): 1969-1978.
- [6]. Wang Guilan, Zhou Guoliang, Zhao Hongshan, Liu Hongyang, Real-time big data technologies of energy internet platform, 2016 IEEE International Conference on Power System Technology (POWERCON), Wollongong, NSW, Australia, 28 Sept.-1 Oct. 2016: 1-6.
- [7]. A. K. Idrees, W. L. Al-Yaseen, M. A. Taam, O. Zahwe, Distributed Data Aggregation based Modified K-means technique for energy conservation in periodic wireless sensor networks, 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Jounieh, Lebanon, Lebanon, April 2018: 1-6.
- [8]. Tai-Yeon Ku, Wan-Ki Park, Hoon Choi, Energy Peak Reduction Mechanism with Prediction of Demand and PV Generation on Big Data, 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, Jan. 2018: 713-716.
- [9]. Rui Fu, Feng Gao, Rong Zeng, Jun Hu, Yi Luo, Lu Qu, Big data and cloud computing platform for energy Internet, 2017 China International Electrical and Energy Conference (CIEEC), Beijing, China, Oct. 2017: 681-686.