Identification of Comprehensive Energy Consumption Feature Based on Rough Set Bayesian Classification Algorithm

Jian Liu¹, Cengkai Ouyang²,³, Shuanghuang Zhao²,³, Zhengqi Tian²,³ and Lihui Wang¹

¹ State Grid Jiangsu Electric Power Co., Ltd Research Institute, Nanjing 211103, China;
² State Grid Jiangsu Electric Power Co., Ltd Marketing Service Center, Nanjing 210019, China;
³ Key Laboratory of Electric Energy Measurement in State Grid Corp, Nanjing 210019, China
Email: 63400481@qq.com

Abstract. Mining of valuable information from user energy consumption data efficiently and accurately has always been a research hotspot in the power industry. Bayesian classification method is one of the important data processing methods in the field of machine learning and data mining research. It has the advantages of simplicity, high efficiency and stable classification effect, and it provides an effective solution to the user's comprehensive energy consumption feature identification. A model describing energy consumption data to predict trends is built, training data set is analyzed, a classification model is constructed, and the data records in the database to a given category are mapped, which can be applied to data prediction. By studying incomplete information systems, expanding rough sets, constructing extended models, Bayesian classification algorithms based on rough sets theory is designed, and user comprehensive energy consumption feature identification is realized. Experiments show that the user comprehensive energy consumption feature identification algorithm based on rough set Bayesian classification can greatly improve the classification accuracy.

1. Introduction
At present, electricity, water, heat, gas and other energy supply enterprises implement the "one household, one meter" policy. However, energy information collection and business processing are independent[1]. There is no resource reuse and sharing of public utility infrastructure, which leads to repeated investment and operation. Problems such as low efficiency of maintenance management and insufficient level of intelligent business development[2]. Breaking through the traditional energy system architecture, developing an integrated energy system, achieving interconnection, coordination and coordination of multiple energy supply networks, and simultaneously carrying out energy measurement, collection, monitoring and analysis are important means to reduce energy consumption in the whole society and improve energy efficiency. Become an inevitable choice in the world's energy field.

The role of the integrated energy system is to improve energy efficiency while reducing energy consumption and emission. The key to the application of the integrated energy system is to build a fully functional integrated energy system energy information collection, monitoring and energy consumption analysis platform[3,4]. This article establishes the relationship between the integrated energy system node index and the comprehensive energy consumption index. Based on the complete...
collection of information on the consumption of electricity, water, gas and other energy by different users such as residents and enterprises, each energy consumption assessment is established for different energy consumption characteristics. Model, analyze the energy consumption data, study the composition and changes of various energy consumption in different energy consumption periods, find out the relevant factors that affect the user's energy consumption, and then propose the energy consumption optimization strategy to adapt to it. Scientific and reasonable energy use theoretically supports the implementation of the energy monitoring and analysis function of the comprehensive energy system.

2. Energy Consumption System
The smart grid has transformed the power system from the traditional supply-side, unidirectional power supply, basically relying on manual management operation mode to the direction of full participation of users, bidirectional flow of flows, and high degree of automation behavior[5,6]. With the unified deployment of electric energy service management platforms, grid companies can obtain rich user-side information and grasp various types of user energy consumption data in real time, so as to conduct a more detailed analysis of user energy consumption behavior in typical scenarios to reduce energy consumption. Emissions and energy consumption provide some theoretical and data support.

The electric energy service management platform needs a reasonable data collection system as a support, with wide coverage, timely communication, and the integration of multiple energy consumption data. At present, most areas of the country have basically covered smart electric energy meters, and the construction of electricity information collection system has a certain scale. In contrast, water, gas, heat and other industries, although the Ministry of Housing and Urban-Rural Development issued the "Technical Conditions for Data Transmission of Household Meters", but the stipulations are not clear, and various industries have not formed a unified and standardized smart metering standard, so that various industries and manufacturers can only carry out remote collection meter reading through their own understanding. A single "one household, one meter" cannot meet the application requirements, and user energy consumption monitoring needs to be conducted for comprehensive energy consumption such as electricity, water, heat, and gas.

On the basis of obtaining comprehensive energy consumption data of customers such as electricity, water, heat, and gas, it is necessary to study its variation rules and extract similar energy consumption behavior characteristics, and comprehensively and systematically propose methods for analyzing customer energy consumption behavior characteristics[7,8]. It can fully reflect the differences in energy consumption behaviors of different categories of users. In typical scenarios, the energy consumption behavior patterns of users 'main energy-consuming devices provide a data basis for users' energy consumption analysis, demand response, and potential replacement of electrical energy. At present, there is a lack of accurate definitions of customer energy-using behaviors, and lack of theoretical research and technical means related to customer energy-using behaviors[9]. Studying customer energy consumption behavior can reasonably guide all types of energy on the user side to participate in peak-shaving and valley-filling of the energy supply network, reduce peak load pressure, and effectively improve operating efficiency and economy, thereby realizing energy saving and emission reduction on the user side and the energy supply side.

The particularity of the user's comprehensive energy consumption structure determines that it will be limited by various external factors. Under the combined effect of many influencing factors, the user's energy consumption will show a very wonderful fluctuation. Faced with the growing number of user groups such as electricity, water, heat, and gas, and the user's energy consumption method is more and more flexible and efficient, cluster analysis, correlation analysis and other technologies are applied to energy consumption data in a big data environment, combined with neural Networks, support vector machines and other machine learning algorithms have become increasingly necessary to mine the potential value of user energy consumption data. Considering the strong correlation between user energy consumption and meteorological factors, each change in meteorological indicators will more or less cause load changes, and this change will be affected by regional factors[10,11]. Therefore, analyzing the correlation between typical user energy consumption and influencing factors can effectively help us analyze the user's energy consumption characteristics, and
then be able to predict the user's energy consumption and tap the demand response and energy saving potential of various energy consumption.

In this paper, fuzzy comprehensive evaluation method (AHP-FCE) is used to evaluate user energy consumption. As a relatively mature comprehensive evaluation method, this method has been widely used in statistical analysis and comprehensive index system evaluation in various fields. This method combines the fuzzy comprehensive evaluation method and the analytic hierarchy process. The system comprehensively considers various factors that affect the evaluation results, which can not only reduce the drawbacks caused by personal subjective assumptions, but also fully reflect the evaluation factors and the ambiguity of the evaluation process. Together improve the effectiveness and reliability of the evaluation.

The research of integrated energy metering integrated acquisition architecture system includes the acquisition system business demand analysis, business operation model research, technical architecture system research, design index system and function research. This paper is based on the SOA architecture to carry out the integrated collection system architecture design; based on Spark + R technology to build a distributed data computing and analysis application engine; based on the public information model to build a unified interface architecture. Among them, SOA is an advanced IT infrastructure, which will also become the “mainstream” architecture of power enterprise information integration applications; Spark + R technology is oriented to the application requirements of complex data and variable forms. The analysis application engine can quickly respond to data quality monitoring and data. Repair, abnormal analysis and other requirements; the public information model is an abstract model that can describe all the main objects of the power enterprise, by providing a standard method for representing power system resources with object classes and attributes and their relationships.

3. Bayes Classification Algorithm Based on Rough Set

Study the user's comprehensive energy consumption model under typical energy consumption scenarios, and construct a typical user's energy consumption impact analysis model. A typical energy consumption scenario refers to the energy consumption environment composed of various influencing factors that can trigger changes in energy consumption, including the comprehensive energy consumption composition, day type, season, and orderly electricity arrangement. Different users' comprehensive energy consumption has distinctive energy consumption pattern characteristics. The same user may adopt different energy consumption behaviors in different energy consumption scenarios. To study the influencing factors and influencing mechanism of user's comprehensive energy consumption behavior is mainly to dig out the internal connection of these influencing factors, which can use different classification methods, such as: Bayesian classification, decision tree classification, artificial neural network algorithm, etc. Bayesian taxonomy data mining method extracts the main influencing factors of user behavior patterns and analyzes the influencing mechanism. The principal component analysis method is used to eliminate redundant factors firstly. The specific steps are as follows: (1) Set up n observation sample matrices to be analyzed, and normalize the transformation of each variable of the matrix; (2) Find the correlation matrix to solve this characteristic equation The characteristic root of (3) takes the top p factors to be analyzed with large variance contribution rate as the main components, that is, extracts the main influencing factors of the comprehensive energy consumption of p users, that is, the characteristic quantity of user energy consumption behavior. Then, a simple Bayesian classification algorithm based on JS divergence feature weighting is used to select a certain number of user comprehensive energy consumption feature quantities as training samples, and the internal connections of these feature vectors are mined to analyze the closest dependence on user energy consumption behavior.

Fuzzy comprehensive evaluation method is an effective multi-factor decision-making method for comprehensive evaluation of things affected by various factors. In the fuzzy comprehensive decision, let \( U = \{u_1, u_2, \ldots, u_n\} \) be n kinds of factors, and \( V = \{v_1, v_2, \ldots, v_m\} \) be m kinds of judgments. The status of various factors is different, and their role is also Not the same, of course, the weights are also different. Man judgments on species are not absolutely positive or negative, the comprehensive judgment should be a fuzzy subset on \( V \).
where \( F(v) \) is the set of all fuzzy set subsets on the domain \( v \), \( B_j = (1, 2, \ldots, m) \) is the membership of \( v_j \) to the fuzzy set \( B \).

Comprehensive evaluation of \( B \) depends on the weight of each factor, that is, the weight is a fuzzy subset on \( U \), \( A = (a_1, a_2, \ldots, a_m) \in F(U) \) and \( \sum_{i=1}^{m} a_i = 1 \), where \( a_i \) represents the weight of the \( i \)-th factor. Given the weight \( A \), a comprehensive evaluation \( B \) is obtained accordingly. Fuzzy comprehensive evaluation is based on fuzzy mathematics membership degree theory to transform qualitative evaluation into quantitative evaluation, that is, to use fuzzy mathematics to make an overall evaluation of things or objects restricted by various factors.

Combined with actual field research and user energy consumption influencing factors, a user energy consumption factor system was established. Set first-level indicators: energy-saving equipment application \( U_1 \), new technology application \( U_2 \), smart home system \( U_3 \), new energy technology \( U_4 \).

Energy efficiency factor set.

\[
U = \{u_1, u_2, u_3, u_4\} \tag{2}
\]

where \( U_1 = \{u_{11}, u_{12}, u_{13}\} \), \( U_2 = \{u_{21}, u_{22}, u_{23}\} \), \( U_3 = \{u_{31}, u_{32}, u_{33}\} \), \( U_4 = \{u_{41}, u_{42}, u_{43}\} \).

The fuzzy mathematical formula is given by

\[
f : U \rightarrow F(V) \tag{3}
\]

\[
\mathbf{u}_i \rightarrow f(\mathbf{u}_i) = (r_{i1}, r_{i2}, r_{i3}, \ldots, r_{im}) \in F(V) \tag{4}
\]

Fuzzy mapping induces a fuzzy relationship, which can be expressed by a fuzzy matrix as

\[
R = \begin{bmatrix}
  r_{11} & r_{12} & \ldots & r_{1m} \\
  r_{21} & r_{22} & \ldots & r_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{n1} & r_{n2} & \ldots & r_{nm}
\end{bmatrix} \tag{5}
\]

where \( r_{ij} \) is the membership of the \( i \)-th energy efficiency factor set to the \( j \)-th evaluation set, which is determined by the membership function. \( r_{ij} \) is given by

\[
r_{ij} = f_{ij} = \sum_{j=1}^{n} f_{ij} \tag{6}
\]

where \( f_{ij} \) is the number of times the \( i \)-th judging factor \( u_i \) is rated as the \( j \)-th judging set \( v_j \).

After determining the order relationship, the ratio of the importance degree \( S_k \) regarding energy efficiency factor concentration is obtained by

\[
S_k = \frac{a_{k-1}}{a_k} \tag{7}
\]

\[
a_n = (1 + \sum_{k=2}^{n} a_k) \prod_{i=k}^{n} (S_i)^{-1} \tag{8}
\]

\[
a_{k-1} = S_n a_k (k = n, n-1, \ldots, 2) \tag{9}
\]
By performing a hierarchical comprehensive evaluation of user energy consumption, and the comprehensive evaluation result of each sub-factor set is obtained by

\[ B_j = A_j \cdot R_i = (b_{j1}, b_{j2}, \ldots, b_{jm}) \]  
(10)

\[ b_{ij} = \sum_{k=1}^{e} (a_{ik}r_{kj}) \quad 1 \leq j \leq m \]  
(11)

The fuzzy evaluation space \((U, V, R)\) is evaluated, and the evaluation matrix \(R\) of the factor set \(U\) is comprehensively evaluated by the second-level factor \(B_j\) constitutes. User energy consumption estimation factor set final comprehensive evaluation system is obtained by

\[
R = \begin{bmatrix}
B_1 \\
B_2 \\
\vdots \\
B_L
\end{bmatrix} \cdot \begin{bmatrix}
A_1 & R_1 \\
A_2 & R_1 \\
\vdots & \vdots \\
A_L & R_1
\end{bmatrix}
\]  
(12)

The specific steps of user comprehensive energy consumption impact analysis based on Bayesian classification are as follows: (1) Preparation stage: use correlation analysis to identify whether any two given feature quantities are statistically relevant, eliminate redundant information, and obtain principal features \(A\) prediction tuple is formed according to the predicted user feature measurement value, and the observed user feature measurement value is divided into multiple training tuples, so that the training tuple and the associated class labels form a training set. (2) Training stage: use JS divergence to represent the amount of information that feature items can provide, and address the shortcomings of JS divergence from word frequency and text frequency within and between categories and inverse categories modified with coefficients of variation Consider the three aspects of frequency, adjust and correct the JS divergence, and finally calculate the weight of each feature item, and bring the weight into the formula of Naive Bayes; calculate the prior probability of each class Probability density \(P\) of each feature quantity in each category, and training; (3) Application stage: calculate the \(P\) value for each category for the data collected in real time, and determine the user's comprehensive energy consumption impact based on the maximum value.

4. Experiments
In the system experiment, the evaluation set is applicable to each evaluation factor, and the establishment of the evaluation set does not affect the user's energy consumption evaluation. The user energy consumption evaluation level is divided into six levels as shown in Table 1. With the identification of the user's comprehensive energy consumption characteristics, the mixed energy supply operation mode of the end user node is optimized, and the comprehensive energy utilization rate is increased by 12%.

| Index level                  | Features                                      |
|------------------------------|------------------------------------------------|
| First-level indicators V1    | Low energy efficiency, unreasonable electricity consumption strongly |
| Second-level indicators V2   | Low energy efficiency, unreasonable power consumption relatively |
| Third-level indicators V3    | Low energy efficiency, unreasonable power consumption |
| Fourthly-level indicators V4 | Average energy efficiency, reasonable electricity consumption |
| Fifthly-level indicators V5  | Good energy efficiency index, reasonable electricity consumption |
| Sixthly-level indicators V6  | The energy efficiency index is very good and the electricity consumption is reasonable |
5. Conclusions
The analysis methods of the influence quantity of power user energy consumption behavior mode are complicated, and they are not suitable for cluster calculation of large data amount. Building a comprehensive energy consumption impact analysis model based on user energy consumption characteristics is the key to studying household energy consumption behavior. To accurately analyze the user's comprehensive energy consumption impact, different types of users should have different impact analysis models at different times. Even if the same model class is established, its model parameters are also very discrete. The traditional naive Bayesian classification algorithm does not divide its importance according to different feature items, which makes the classification results inaccurate. To address this problem, roughness is introduced, and roughness is used to represent the amount of information that feature items can provide. Considering the feature frequency within and between categories, the roughness is adjusted by weights to modify the Bayesian algorithm. The algorithm is simple and efficient, with strong predictive classification capabilities, and high accuracy.

6. Acknowledgments
The project is supported by the following funds: Science and Technology Project of State Grid Jiangsu Electrical Power Company Research Institute, China (Assessment modeling and energy consumption optimization method design for State Grid Jiangsu Electric Power Research Institute user’s energy consumption.)

7. References
[1] Yun Puyuan, Liu Qiang, Huang Wei, Qu Bo, Liu Guoli. Terminal low–carbon operation strategy of energy efficiency evaluation in integrated energy system [J]. Power Demand Side Management, 2020,21 (4): 54–58.
[2] JIA Hongjie, WANG Dan, XU Xiandong, et al. Study on some key problems related to integrated energy systems [J]. Automation of Electric Power Systems, 2015, 39 (7): 198-207.
[3] Gao Ciwei, Luo Haiming, Zhu Lulu, et al. The energy–saving assessment and optimization of energy storage and electricity utilization technology based on the energy efficiency evaluation of power system [J]. Of China Electrotechnical Society, 2016, 31 (11): 140-148.
[4] Xue Yixun, Guo Qinglai, Sun Hongbin, et al. Comprehensive energy utilization rate for park–level integrated energy system [J]. Electric Power Automation Equipment, 2017, 37(6): 117-123.
[5] Oman. Research and Application of Naive Bayes Classification Algorithm [D]. Master Thesis of Dalian University of Technology, 2004.
[6] Guo Yaguang. Research on classification problems based on rough set and naive Bayes model [D]. Master thesis of Hefei University of Technology, 2005.
[7] Wu Xun, Chen Tefang, Cheng Shu, et al. An online diagnostic method based on output current locus for open-circuit faults of inverters [J]. Transactions of Electrical Engineering, 2017, 32 (Suppl 2): 87-95.
[8] Yang Tingfang, Zhang Hang, Huang Libin, et al. Incipient fault diagnosis based on improved principal component analysis for power transformer [J]. Electric Power Automation Equipment, 2015, 35 (6): 149-153.
[9] Sun Yinshan, Zhang Wentao, Zhang Yiming, et al. Study on the feature extraction and fault discrimination methods of the current signal of the disconnecting coil of high voltage circuit breaker [J]. High Voltage Apparatus, 2015, 51 (9): 134-139.
[10] Zhang Peng, Li Hongbin. A harmonic analysis method based on discrete wavelet transform [J]. Journal of Electrical Engineering and Technology, 2012, (03): 252-259.
[11] Fang Guozhi, Yang Chao, Zhao Hong. Power system harmonic detection method based on FFT and wavelet packet transform [J]. Power System Protection and Control, 2012, 40(05): 779-783.