Analysis of integrated energy customers under the background of energy revolution

Lili Wang¹, Linjuan Zhang¹, Changqing Xu¹, Ruting Huang²*

¹State Grid Henan Economic Research Institute, Zhengzhou, Henan, 450000, China
²China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, Fujian, 361005, China
*Corresponding author’s e-mail: rutinghuang@stu.xmu.edu.cn

Abstract. China’s rapid economic development has brought a series of environmental problems, which seriously threaten people’s production and life. Driving the energy revolution and building clean and safe energy systems have become an important part of China’s sustainable development. As a crucial means to reduce energy cost and improve energy efficiency, integrated energy service has been gradually developed in recent years. In this context, how to manage the integrated energy customers and improve their satisfaction and loyalty has become an urgent problem to be solved. This paper applies the method of high dimensional data clustering (HDDC) to study this issue. Moreover, a case study has been analysed to illustrate the suitability and effectiveness of the proposed framework with the daily electricity consumption data of industrial enterprises. This paper has practical value for the integrated energy service providers.

1. Introduction

Energy is the foundation of economic and social development and unreasonable energy use will bring a series of ecological environment problems (Song et al., 2018). As environmental issues have become more prominent in recent years, how to explore a green and efficient energy structure has become the focus of China. Since president Xi Jinping proposed the energy revolution in 2014, a number of policies have been issued. Among them, integrated energy service is a way to bring economic benefits and environmental benefits, such as carbon dioxide and pollutant emission reduction. Recently, China is constantly promoting the development of integrated energy service. In October 2017, the State Grid issued a document named “Opinions on the Development of Integrated Energy Service Business” which aimed to promote the development of integrated energy service industry.

However, as an emerging business, integrated energy service is still at a premature developmental phase. The competition in the energy market will be becoming increasingly fierce as more and more companies are involved into integrated energy service. Thus the traditional supply-centered service may be increasingly unsuitable for the development trend of integrated energy service market and more attention should be paid to the customer demand. Under this background, how to attract and manage different customers is crucial to the success of integrated energy service providers.

The need for different type of customers differs in reality. As stressed by Wyner (1996), 80% of the company’s sales profits come from 20% of key customers. The undifferentiated service strategy will not only lead to the customer loss, but also the inefficient resource allocation. Thus, how to categorize the customers according to their importance and serve the customers based on their classification has become a problem that urgently needs to be solved. We apply the method of high dimensional data clustering (HDDC) to study this issue. Moreover, a case study has been analyzed to illustrate the suitability and effectiveness of the proposed framework with the daily electricity consumption data of industrial enterprises. This paper has practical value for the integrated energy service providers.
clustering (HDDC) to study this problem. Moreover, a case study has been analysed to illustrate the suitability and effectiveness of the proposed framework with the daily electricity consumption data of industrial enterprises. Our paper has practical value for the integrated energy service providers.

2. Literature review

The concept of customer classification was first proposed by Smith (1956) in the mid-1950s, which is developed from the concept of market segmentation. As a marketing strategy, customer classification refers to dividing customers according to a certain standard, and then providing service accordingly. Our paper is correlated with two streams of literature. One is the research on classification algorithms and the other is the specific application of classification algorithms.

2.1 Improved algorithms of customer classification

Improved algorithms of customer segmentation has attracted much attention. For example, Heilman and Bowman (2002) extended the traditional single-category brand selection logit hybrid model and comprehensively used the consumer purchasing data of multi-category products to differentiate consumers. Kuo et al. (2006) proposed a new clustering method, which combined neural network SOM algorithm and genetic K-means algorithm and they also used the data from the actual freight industry to prove the effectiveness. Cheng and Chen (2009) proposed a method to subdivide customers based on RFM model. Subbalakshmi et al. (2015) proposed a method for finding the optimal number of clusters in a dynamic data set based on fuzzy contours. Chen et al. (2017) proposed a novel consumer clustering algorithm based on the product tree. Munusamy and Murugesan (2020) improved the dynamic fuzzy c-means clustering algorithm and used the improved algorithm to study consumer dynamic segmentation.

2.2 The application of customer classification algorithms

A number of scholars have studied many interesting problems by using the improved classification algorithms. For example, the classic RFM model proposed by Hughes (1996) is widely used for consumer classification. Davies et al. (1996) used the data of ATM users’ views on service to analyze user satisfaction and divide ATM users into four categories based on the neural network analysis model. Vellido et al. (1999) used the unsupervised neural network SOM algorithm to categorize the online shopping market and found the SOM algorithm was a very effective data visualization tool. Kim et al. (2003) used the artificial neural network SOM algorithm to divide the tourist market in Western Australia and verified that SOM and projection functions would make the data more intuitive. Hsu et al. (2012) used transaction data to segment consumers based on the concept hierarchy method. Chen et al. (2015) used the RFM clustering algorithm to investigate the time series data of consumer transactions to predict consumer profitability. Guo et al. (2018) used clustering algorithms to classify the types of household electricity consumption.

Overall, it is widely recognized that offering differentiated service strategies according to the customer types is a key component of business success. As an emerging business, the studies on integrated energy service from the perspective of customer clarification still lack. What should be noted is that the division of the customer is always based on multiple indicators with annual data, ignoring the high-frequency information. However, the integrated energy customers always have high-frequency data, such as electricity consumption and power load data. To fully use these data, this paper uses the funHDDC proposed by Bouveyron and Jacques (2011) and Schmutz et al. (2020) to classify the integrated energy customers and provide some references for the integrated energy service providers.

3. Method

As illustrated, we applies the clustering method for time series to make the full use of high frequency data of integrated energy customers. The clustering method, named funHDDC is based on functional latent mixture model. Let we first assume that there are a series of independent multivariate curves \{X_1, X_2, \cdots, X_n\}, such as electricity consumption curves and typical load curves. These curves are the realizations of a \(L_2\) continuous multivariate stochastic process \(X_t = \{X_t(t)\}_{t \in [0, T]}\), which belongs to
In practice, the functions of the observed curves are unknown and we only have the information of the discrete observations $X^j_i(t_j), \cdots, X^j_{N_i}(t_j)$. Thus it is essential to reconstruct the function form from the finite discrete observations. A common way to realize this is to assume that each curve can be approximated with the combination of the basis functions.

$$X^j_i(t) = \sum_{r=1}^{R_j} C_{jr}^j(X^j_i)\phi_r^j(t)$$

funHDDC is a model-based clustering method, which is suitable for univariate and multivariate function data, which evaluate for each curve belonging to the cluster $Z^j_i \in \{1, \cdots, K\}$. This algorithm is dependent on a Karhunen-Loeve decomposition per cluster to propose a fine representation of the curves, having the following assumption on distribution.

$$Z^j_i \sim M(\pi_1, \cdots, \pi_K)$$

$$\delta^j_i \sim N(\mu_k, \Delta_k)$$

The inference of the model parameters $\theta = \{\pi_k, \mu_k, \Delta_k\}$ can be estimated by the EM algorithm, which includes two steps, the expectation (E) and maximization (M) steps (Dempster et al., 1977).

### 4. Case analysis

It should be noted that funHDDC is a model-based clustering algorithm which is appropriate for either univariate or multivariate functional data (Martínez-Álvarez et al., 2019). Thus, the model can be applied to analyze multi-dimensional time series data, such as combining heat, gas and electricity consumption. Due to data limitation and availability, this paper only uses one univariate of daily electricity consumption data of industrial enterprises in 2017. After deleting the enterprises with missing data, 173 industrial enterprises are studied. As for the integrated energy service, the industrial enterprises are valued customers. Thus the analysis and classification of industrial enterprises are meaningful for the integrated energy service market.

Before categorizing integrated energy customers, the optimal number of categories needs to be determined. According to the BIC guidelines, the optimal category is 9.

| $K$ | threshold | complexity | BIC     |
|-----|-----------|------------|---------|
| 1   | 0.2       | 131        | -269,668.70 |
| 2   | 0.2       | 263        | -263,742.52 |
| 3   | 0.2       | 395        | -262,007.24 |
| 4   | 0.2       | 527        | -261,522.37 |
| 5   | 0.2       | 659        | -261,471.44 |
| 6   | 0.2       | 791        | -261,564.73 |
| 7   | 0.2       | 395        | -262,007.24 |
| 8   | 0.2       | 1,118      | -262,253.24 |
| 9   | 0.2       | 1,187      | -258,507.05 |
| 10  | 0.2       | 131        | -269,668.70 |

Figure 2 shows the classification results of the integrated energy customers. We can find that due to the Spring Festival, all industrial enterprises have experienced a significant decline in electricity consumption at the end of January and the beginning of February. Moreover, it can be found that users of the class 1 and class 5 have the highest average electricity consumption while the electricity consumption fluctuates greatly for class 1 customer. Average electricity consumption of users in class 3 and 9 is low, and the electricity consumption is relatively average throughout the year without large fluctuations.
However, there are too many categories selected by the BIC model, which may not be applicable in practice. The model is re-processed by selecting four categories. Figure 2 shows the average electricity consumption of the classification results. For the first type of customers, the electricity consumption is low and the fluctuation is small. For the second type, the power consumption is medium and the fluctuation is large. For the third type, customers have large electricity consumption and relatively large fluctuations. For customers in the fourth category, the average electricity consumption is moderate with low fluctuations.

Integrated energy services have a wide range, which can be summarized as two aspects. One is integrated energy, including the supply of cooling, heating, and electricity. The other is integrated service, which refers to engineering, investment, and operational service. According to different service types,
the customer values may differ even with the same classification results. For example, high energy consumption always means high value in views of energy supply. Thus based on the classification results in figure 2, the customers of class 1, class 2, class 3 and class 4 can be seen as ordinary customers, medium value customers, high value customers and medium and high value customers, respectively. However, for different integrated energy services, the values of different categorized customers may change. Some investment or operational services would pay more attention to the fluctuation of energy consumption.

5. Conclusions
Under the background of the energy revolution, the integrated energy business has entered a period of rapid development. Integrated energy service providers are facing pressure to build new systems and business models. Thus in a complex and changing market environment, how to manage the integrated energy customers and design the business strategy to improve customer satisfaction and loyalty has become an urgent problem to be solved. This paper applies the method of high dimensional data clustering (HDDC) to study the problem. Moreover, we conduct a case analysis with the daily electricity consumption data of 173 industrial enterprises to illustrate the suitability and effectiveness of the proposed framework.

As mentioned, values of different categorized customers may change for different types of integrated energy service. However, strategies for the categorized customers with the same value may be similar. For the high value customers, which contributes to the most of revenue with a small proportion, one-to-one personalized and specialized service can be adopted. For middle and high value customers, it may not be suitable to take one-to-one special service, but understanding their feedback and needs is necessary. For medium value users, which account for the largest proportion, it is essential to formulate reasonable market pricing for them. For the ordinary or sub-prime customers, the main focus is on their creditworthiness and risk.

Acknowledgements
This work is supported by the Science and Technology Project of State Grid Corporation of China entitled “The Adaptivity and Business Mode of Campus Integrated Energy Systems Based on the Dynamic Portrait Analysis”.

References
[1] Bouveyron, C., Jacques, J. (2011). Model-based clustering of time series in group-specific functional subspaces. Advanced Data Analysis and Classification, 5(4), 281-300.
[2] Chen D., Guo K., Ubakanma G. (2015). Predicting customer profitability over time based on RFM time series. International Journal of Business Forecasting & Marketing Intelligence, 2(1):1.
[3] Chen, X., Fang, Y., Yang, M., Nie, F., Zhao, Z., Huang, J.Z. (2017). Purtreeclust: A clustering algorithm for customer segmentation from massive customer transaction data. IEEE Transactions on Knowledge and Data Engineering, 30(3), 559-572.
[4] Cheng C.H., Chen Y. S. (2009), Classifying the Segmentation of Customer Value via RFM model and RS Theory. expert systems with applications, 36(3):4176-4184.
[5] Davies F., Moutinho L., Curby B. (1996). ATM user attitudes: a neural network analysis. Marketing Intelligence & Planning, 14(2):26-32.
[6] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the royal statistical society series b-methodological, 39(1), 1-22.
[7] Guo, Z., Zhou, K., Zhang, X., Yang, S., Shao, Z. (2018). Data mining based framework for exploring household electricity consumption patterns: A case study in China context. Journal of Cleaner Production, 195, 773-785.
[8] Heilman, C.M., Bowman, D. (2002). Segmenting consumers using multiple-category purchase data. International Journal of Research in Marketing, 19(3), 225-252.
[9] Hsu F.M., Lu L.P., and Lin C.M. (2012). Segmenting customers by transaction data with concept hierarchy. Expert Systems with Applications, 39(6):6221-6228.

[10] Hughes, A. M. (1996). Boosting response with RFM. Marketing Tools, 4-8.

[11] Kim J., Wei S., Ruys H. (2003). Segmenting the market of West Australian senior tourists using an artificial neural network. Tourism Management, 24(1):25-34.

[12] Kuo, R.J., An, Y.L., Wang, H.S., Chung, W. J. (2006). Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation. Expert systems with applications, 30(2), 313-324.

[13] Martínez-Álvarez, F., Schmutz, A., Asencio-Cortés, G., Jacques, J. (2019). A novel hybrid algorithm to forecast functional time series based on pattern sequence similarity with application to electricity demand. Energies, 12(1), 94.

[14] Munusamy, S., Murugesan, P. (2020). Modified dynamic fuzzy c-means clustering algorithm—Application in dynamic customer segmentation. Applied Intelligence, 1-21.

[15] Schmutz, A., Jacques, J., Bouveyron, C., Cheze, L., Martin, P. (2020). Clustering multivariate functional data in group-specific functional subspaces. Computational Statistics, 1-31.

[16] Smith, W.R. (1956). Product differentiation and market segmentation as alternative marketing strategies. Journal of marketing, 21(1), 3-8.

[17] Song, M., Peng, J., Wang, J., and Zhao, J. (2018). Environmental efficiency and economic growth of China: A Ray slack-based model analysis. European Journal of Operational Research, 269(1), 51-63.

[18] Subbalakshmi, C., Krishna, G.R., Rao, S.K. M., Rao, P.V. (2015). A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set. Procedia Computer Science, 46, 346-353.

[19] Vellido A., Lisboa, P.J.G., Meehan, K. (1999). Segmentation of the on-line shopping market using neural networks. Expert Systems with Applications, 17(4):303-314.

[20] Wyner G. A. (1996). Customer valuation: Linking behavior and economics. Marketing Research, 8(2):36-38.