Research on potential user identification model for electric energy substitution

Huajian Xia1, Meiling Chen1, Haiying Lin1, Shuo Yang2, Bo Miao3 and Xinshi Zhu2

1Beijing China-power information technology Co., Ltd. Beijing China
2State Grid Hebei Electric Power Company, Hebei China
3China Electric Power Research Institute, Beijing China

*Corresponding author e-mail: xingyue_xia@163.com

Abstract. The implementation of energy substitution plays an important role in promoting the development of energy conservation and emission reduction in China. Energy service management platform of alternative energy users based on the data in the enterprise production value, product output, coal and other energy consumption as a potential evaluation index, using principal component analysis model to simplify the formation of characteristic index, comprehensive index contains the original variables, and using fuzzy clustering model for the same industry user’s flexible classification. The comprehensive index number and user clustering classification based on constructed particle optimization neural network classification model based on the user, user can replace electric potential prediction. The results of an example show that the model can effectively predict the potential of users' energy potential.

1. Introduction
In recent years, the pollution problem caused by the whole society pay close attention to environmental protection. To improve the quality of the environment and adjust the structure of energy consumption, the national Power Grid Corp timely put forward the development strategy of "substituting electricity for coal, replacing electricity with oil and Electricity came from far away", which has promoted the rapid development of electric energy substitution industry. Under this background, how to quickly and efficiently excavate the potential for alternative energy will be an important part of the future power market. The literature [1] quantifies the potential for alternative energy, and constructs an alternative model of electric energy potential to predict the energy substitution in different scenarios; The literature [2] is based on the current situation of energy consumption in Beijing, established the evaluation index system of energy instead of other energy sources, to forecast the future demand of major terminal energy sources in Beijing; The literature [3] takes Gansu Province as the research object, studies the structure and characteristics of coal, oil and other energy sources, and expounds the general direction of Gansu power saving potential users. These theoretical studies have promoted the rapid development of electric energy substitution, and carried out a deep level theoretical study on the potential analysis of energy substitution. However, the research in this field is still in the initial stage in our country. Moreover, there is little research on the methods of replacing typical user potential with electrical energy.
This article relies on the power services management platform big data, select enterprise total output value, product output, coal, oil and natural gas are selected as user characteristic indexes, and using Principal component analysis to reduce the dimensionality of various indexes. By analyzing the index characteristic variance and the cumulative contribution rate, to form the comprehensive indexes including the main characteristics; Using fuzzy clustering analysis model of the same industry users flexible classification, comprehensive index of the number of users and the simplified fuzzy clustering based on categories, the establishment of neural network identification model based on particle swarm optimization, implementation of power user electric energy alternative classification, in order to assess users of alternative energy potential. It can provide some theoretical support for the relevant management departments of power industry to tap the potential energy users, judge the energy, replace the market changes, and formulate corresponding energy replacement decisions.

2. Research on user characteristic analysis model

User characteristics analysis model of customer characteristic index based on analysis of simplified model of a number of indicators using the principal component, form a comprehensive index, the result is more effective and reasonable model, specific steps are as follows:

(1) Index screening

Electric energy substitution refers to "substituting electricity for coal", "substituting electricity for gas" and "substituting electricity for oil", and the greater the consumption of energy, such as coal, the greater the potential for electricity substitution. At the same time, the enterprise product and product output growth have a positive impact on energy consumption, therefore, the model to consume user GDP, the output of products, coal, oil and natural gas as the characteristic index of [4] potential analysis.

(2) Building data matrix

N sample data of the same industry are selected based on the electric energy service management platform. Suppose each sample has l variables, l=1, 2… n. Then you can form a data matrix [5]-[7].

\[
A = \begin{pmatrix}
    a_{11}, & a_{12}, & \ldots, & a_{1l} \\
    a_{21}, & a_{22}, & \ldots, & a_{2l} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1}, & a_{n2}, & \ldots, & a_{nl}
\end{pmatrix}
\]  

(1)

(3) Data standardization

Due to the different dimensions of the feature indexes, the samples are standardized. Then you can get the following data formula.

\[
R_{ij} = (x_{ij} - \bar{x}_j)/s_j \\
\bar{x}_j = \frac{\sum_{i=1}^{n}(x_{ij})}{n} \\
s_j^2 = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2/(n-1)}{n}
\]  

(2)

(4) Eigenvalues and eigenvectors are computed

Calculate the correlation coefficients \(r_{jk}\) between the index j and the index k in the original data:

\[
r_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2/S_j[\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2/S_k]
\]  

(3)

According to the formula mentioned above, the correlation coefficient matrix B can be obtained:

\[
B = \begin{pmatrix}
    b_{11}, & b_{12}, & \ldots, & b_{1k} \\
    b_{21}, & b_{22}, & \ldots, & b_{2k} \\
    \vdots & \vdots & \ddots & \vdots \\
    b_{n1}, & b_{n2}, & \ldots, & b_{nk}
\end{pmatrix}
\]  

(4)
(5) Determine the number of principal components
It is assumed that $T_i$ is the sum of the variance contribution rates of the $m (m \leq l)$ principal component components [8], and the formula is as follows:

$$T_i = \frac{\sum_{k=1}^{m} \lambda_k}{\sum_{k=1}^{l} \lambda_k} \tag{5}$$

(6) Solving rotation matrix
The coordinate system is orthogonally rotated at an angle $\theta$, so that the coordinate axis $y_1$ of the ellipse axis is taken, and the coordinate $y_2$ is taken in the direction of the elliptical short axis, and the rotation formula can be obtained [8]:

$$\begin{cases} y_{1j} = l_{1j} \cos \theta + l_{1j} \sin \theta \\ y_{2j} = -l_{2j} \sin \theta + l_{2j} \cos \theta \end{cases} \tag{6}$$

Suppose the $U$ represents the coordinate rotation matrix.

$$U = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \tag{7}$$

(7) Determine comprehensive index function
Taking the variance contribution rate of the principal component factor as the weight, and through the linear combination of the factor and the weight, the comprehensive evaluation index function of the power energy substitution can be obtained:

$$F = \frac{\omega_1 z_1 + \cdots + \omega_m z_m}{\omega_1 + \cdots + \omega_m} \tag{8}$$

3. Research on user identification model of electric energy substitution potential

3.1. Research on clustering model of similar users
The comprehensive index is determined using the above model, the use of alternative energy to fuzzy C means algorithm implementation of fuzzy clustering of large data users, according to the replacement battery user evaluation implementation of user alternative potential, the specific steps of the model are as follows:

(1) Determine membership matrix
The sample space matrix is $X$.

$$X = \begin{pmatrix} x_{11}, & x_{12}, & \ldots, & x_{1p} \\ x_{21}, & x_{22}, & \ldots, & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}, & x_{n2}, & \ldots, & x_{np} \end{pmatrix} \tag{9}$$

If $n$ samples are divided into $c (2 \leq c \leq n)$, Make $U = (u_{ik})_{c \times n}$ indicate the membership matrix, $u_{ik}$ denotes the membership degree of the first $k$ samples. The membership degree of the same data set is equal to 1 by normalizing the membership matrix.

$$\sum_{i=1}^{c} u_{ik} = 1 \quad k = 1, 2, \ldots, n \tag{10}$$
(2) Defining objective functions
Set \( V = \{v_1, v_2, ..., v_c\} \) stands for clustering centers, \( v_i = \{v_{i1}, v_{i2}, ..., v_{ip}\} \). Define the objective function \( V \):

\[
V(U, W) = \sum_{i=1}^{c} \sum_{j=1}^{p} u_{ij}^p d_{ik}^2
\]

(11)

Among them, \( d_{ik} = \|x_k - v_i\| \)
When the \( u_{ij} \) satisfies the following formula, can ensure that the target function \( V \) is defined and the minimum value is obtained.

\[
u_{ij} = \frac{1}{\sum_{k=1}^{c}(d_{ij}/d_{kj})^2/(m-1)}
\]

(12)

In order to further verify the effectiveness of the clustering results, reference [12] method is used to analyze the clustering results from three indexes, such as discrete degree, center scatter degree and average membership degree.

3.2. Combination identification model based on particle swarm optimization and neural network
BP neural network is a multilayer feed forward network, with a high degree of self-learning and adaptive ability, but the slow convergence and easy to fall into local minimum; particle swarm optimization algorithm with complex and multi peak data effectively and quickly find the advantages of the global optimal solution, it can make up the deficiency of BP neural network. In order to enhance the overall performance of the neural network identification model.

(1) Principle of neural network recognition model
The BP neural network is composed of input layer, hidden layer and output layer. The network structure is shown in figure 1, \( W_{ij} \) and \( W_{jk} \) respectively represent the weights of the input layer I to the hidden layer J, and the weights of the hidden layer j to the input layer k [14].

![Neural network identification model structure](image)

Figure 1. Neural network identification model structure

The model uses \( a = \{a_1, a_2, ..., a_n\} \), which is extracted from the feature model, as the model input. Output layer node number \( y = \{y_1, y_2, ..., y_m\} \). According to the number of user clustering classes, the output is represented as a value between [0, 1], and the network output mode is as follows:
\[
\begin{align*}
\gamma_{jm} &= 1, \quad x \in j \\
\gamma_{jm} &= 0, \quad x \notin j 
\end{align*}
\] (13)

(2) Neural network model optimization

Suppose that the hidden layer activation threshold of BP neural network is \( \delta_j \) and the output layer activation threshold is \( \delta_k \). The weights and thresholds of the neural network model are used to form a community \( U = [W_{ij}, \theta_{ij}, W_{jk}, \theta_{jk}, \ldots], i = 1, \ldots, n; j = 1, \ldots; q = 1, \ldots, m] \), the dimension of the search space in the community by particle swarm optimization:

\[
S = (q \cdot n + q) + (m \cdot q + m)
\] (14)

There are \( N \) particles in the particle community, so that the space of the \( i \) particle is \( \eta_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iS}) \). The best location of particles is \( \xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iS}) \), the Particle velocity is \( v_i = (v_{i1}, v_{i2}, \ldots, v_{iS}) \), The whole community was searched for the best location for \( \tau_i = (\tau_{i1}, \tau_{i2}, \ldots, \tau_{iS}) \),

\[
\begin{align*}
\theta_{ik}^{t+1} &= \omega v_{ik}^{t} + c_1 \xi_{ik}^{t} - \eta_{ik}^{t} \\
&\quad + c_2 \eta_{ik}^{t} - \eta_{ik}^{t} \\
\eta_{ik}^{t+1} &= \eta_{ik}^{t} + v_{ik}^{t+1}
\end{align*}
\] (15)

\( K \) stands for Iteration number; \( \omega \) stands for Inertia weight; \( c \) stands for Acceleration factor; \( r \) Means the random numbers uniformly distributed in the interval \([0, 1]\).

4. Example analysis

4.1. User profile analysis

Electric service platform and related research based on the data from 200 alternative energy users to the same user as an example, GDP, the output of products, coal, petroleum, natural gas, electricity consumption data as the characteristic index, through the user characteristic analysis model, can get the following results:

| component | Initial eigenvalue | Rotation sum of squares loading |
|-----------|--------------------|----------------------------------|
|           | Total Variance (%) | Cumulative (%)                   | Total Variance (%) | Cumulative (%) |
| 1         | 2.55               | 51.08                            | 2.08               | 41.51          | 41.51          |
| 2         | 1.23               | 24.70                            | 1.20               | 24.09          | 65.6           |
| 3         | 0.63               | 12.66                            | 1.14               | 22.82          | 88.42          |
| 4         | 0.44               | 8.81                             |                     |                |                |
| 5         | 0.14               | 2.76                             |                     |                |                |

Through the table analysis, we can see that the eigenvalue of the first factors is 2.55, and the variance contribution rate is 51.08%; The second factor eigenvalue is 1.23, the cumulative variance contribution rate is 75.78%; third factor eigenvalue is 0.63, the cumulative variance contribution rate was 88.44%; the accumulative variance contribution rate was 88.44% higher than or equal to 85%, so you can determine the number of components is 3. According to the above formula, the principal component factor rotation load matrix can be obtained.
Table 2. Principal component factor rotation load matrix

| Original index                  | Comprehensive index B1 | Comprehensive index B2 | Comprehensive index B3 |
|--------------------------------|------------------------|------------------------|------------------------|
| Gross domestic product A1      | 0.161                  | 0.289                  | -0.088                 |
| Product output A2              | 0.447                  | -0.128                 | -0.018                 |
| Coal energy consumption A3     | 0.586                  | -0.378                 | 0.054                  |
| Oil consumption A4             | -0.359                 | 0.892                  | 0.039                  |
| Energy consumption of natural gas A5 | -0.012                  | 0.013                  | 0.886                  |

The factor function of the composite index can be derived from table 2:

\[
\begin{align*}
    f_{B1} &= 0.161 \times A1 + 0.447 \times A2 + 0.586 \times A3 - 0.359 \times A4 - 0.012 \times A5 \\
    f_{B2} &= 0.289 \times A1 - 0.128 \times A2 + 0.589 \times A3 + 0.892 \times A4 + 0.013 \times A5 \\
    f_{B3} &= -0.088 \times A1 - 0.018 \times A2 + 0.054 \times A3 + 0.039 \times A4 + 0.886 \times A5
\end{align*}
\]

(16)

4.2. Potential energy substitution, user analysis

By using the fuzzy clustering model of power users instead of users, the fuzzy C means clustering method is used to select and obtain the users. The reference [12] method can be used to obtain the validity index of the user clustering as the 3-6 class.

Table 3. Validity index

| index | 3   | 4   | 5   |
|-------|-----|-----|-----|
| G1    | 0.69| 0.70| 0.56|
| G2    | 0.51| 0.84| 0.49|
| G3    | 0.72| 0.83| 0.78|
| G     | 0.46| 0.51| 0.47|

As shown in Table 1, the maximum validity of clustering is 0.51, that is, clustering is divided into 4 classes. Through the above analysis, we can see that the number of comprehensive index of potential users is 3, and the number of clustering classes is 4. That is, the input number of BP neural network is 3, the output number is 5, and the network activation function adopts Logsig function. MATLAB tools are used to verify and analyze the relevant data and compare it with the traditional BP neural network model. The simulation results are shown in Figure 2 and figure 3.

![Figure 2. Standard BP neural network training results](image)
According to figure 2 and figure 3 shows that the optimized BP neural network relative to the standard BP, fast convergence speed, and the oscillation is smaller, the training effect is better than the standard BP, can improve the recognition speed, reduce the network output error.

In the process of potential user identification, the clustering classification group is represented by A-D. Because the users in the clustering classification are all electrical users instead of the implemented users, the average replacement power is chosen as the potential of identifying users. 50 users are used to validate the potential identification model, and the following conclusions can be obtained:

### Table 4. Model recognition result analysis

| Clustering classification | A   | B   | C   | D   | E   |
|---------------------------|-----|-----|-----|-----|-----|
| Potential capacity        | 0.35| 15.8| 1.34| 0.83| 0.35|
| (ten thousand kWh)        |     |     |     |     |     |
| Potential user            | 15  | 7   | 13  | 15  | 15  |

The following conclusions can be obtained from table 3:

1. Class A and class D potential users have 15 users, the potential of alternative power were 3 thousand and 500 kWh, 0.83 kWh; B users have a total of 7, the potential users of the replacement battery is 158 thousand kWh; the C user has a total of 13 users, the potential consumption of 13 thousand and 400 kWh;

2. The model can effectively identify the potential users of the electrical energy, and has the characteristics of fast convergence and strong learning ability.

### 5. Conclusion

The use of alternative energy potential model of the user identification, users can be divided into different categories of potential objects, narrowing the scope, to focus on the development of alternative energy, at the same time, can provide a theoretical basis for the alternative marketing strategies for the relevant departments to formulate the corresponding power.
Acknowledgments
This work was financially supported by Research and application of marketing analysis and decision making technology based on power service management platform (YD71-16-014) fund.

References
[1] Sun Yi, Zhou Shuang, Shan Baoguo et al. Alternative potential analysis of electric energy under sentimental scenes [J]. Power grid technology, 2017 (1): 118-124.
[2] Yin Hang. Research on the evaluation methods of electric energy instead of other energy resources under the environment of energy saving and emission reduction [D]. Beijing: Master Thesis of North China Electric Power University 2013.3.
[3] Liu Jun, Xu Hui, Liu Yaping et al. Gansu province can use analysis of alternative power characteristics and potential users [J]. Power demand side management, 2016, 18 (1): 44-49.
[4] Zheng Haiyan, Jin Nong, Ji Cong, et al. Power user data analysis technology and typical scene application [J]. Power grid technology, 2015, 39 (11): 3147-3153.
[5] Li Yan, who Zhen xiang, Zhang Fujian. Application of principal component analysis in multi index comprehensive evaluation method [J]. Journal of Hebei University of Technology, 1999, 1 (28): 95-96.
[6] Wu Jinhao, Yang Xiuyuan, Sun Jun. Short term combined forecasting of wind power based on principal component analysis [J]. Electrical technology, 2016, 17 (7): 41-47.
[7] Zhang Yongfeng, Hu Rong. Multi index comprehensive evaluation method of principal component model [J]. Journal of Southwest University for Nationalities (NATURAL SCIENCE EDITION), 2013, 39 (3): 362-365.
[8] He Youshi, Xu Wenqin. Application of factor analysis in comprehensive evaluation of economic benefits of industrial enterprises [J]. Mathematical statistics and theory, 2003, 22 (1): 19-22.
[9] Zhang Hongyan. Fuzzy C means clustering algorithm and its application [J]. Science and technology information, 2014 (5): 178-179.
[10] Mao Hanping, Zhang Yancheng, Hu Bo. Study on segmentation method of crop disease leaf images based on fuzzy C means clustering [J]. Journal of Agricultural Engineering, 2008, 24 (9): 136-140.
[11] Li Lei, Luo Hongqi, Ding Yali. An optimized fuzzy C means clustering algorithm [J]. Computer technology and development, 2009, 19 (12): 71-73.
[12] Zhang Lizi, Cai Xuewen, Lu Yu. For the typical user peak potential analysis Screening [J]. Power system protection and control, 2013 (11): 146-151.
[13] Pan Runqiu, Ma Xiaosong, Liu Jun. A mathematical model for calculating natural quality of cultivated land based on particle swarm optimization and BP neural network [J]. Geography and Geographic Information Science, 2014, 30 (5): 78-82.
[14] He Yong, Xiang Liguo. Research and application of BP neural network model based on Fuzzy Clustering [J]. System engineering theory and Practice, 2004 (02): 79-83.
[15] Yang Fei, Wang Kunming, Ma Xin, et al. Identification of traffic signs using BP neural network classifier [J]. Computer Engineering, 2003, 29 (10): 120-121.
[16] Gao Yanxia, Li Yuisheng. Application of BP neural network based on particle swarm optimization algorithm in image recognition [J]. J Journal of Wuhan Light Industry University, 2006, 25 (4): 35-38.
[17] Yu Hua, Huang Chengwei, Jin Yun, et al. Speech emotion recognition based on particle swarm optimization neural network [J]. Data acquisition and processing, 2011, 26 (1): 57-62.