An Algorithm of Analytic Hierarchy Process Model Based on Electric Power Big Data of Smart Grid

Qian Sun  
*State Grid Henan Electric Power Research Institute, Zhengzhou, Henan, China*

Bo Zhu*  
Hangzhou Phantec Technology Co., Ltd., Hangzhou, Zhejiang, China

Qiang Li  
State Grid Henan Electric Power Research Institute, Zhengzhou, Henan, China

Yu Liu & Tingjun Yang  
Hangzhou Phantec Technology Co., Ltd., Hangzhou, Zhejiang, China

ABSTRACT: An algorithm of analytic hierarchy process model based on electric power big data of smart grid is studied in this paper with respect to the analysis on the maturity of regional development. In this algorithm, the author uses data-mining technology to analyze the layout of surrounding functional blocks from aspects of consumers distribution and land usage with consumers of a central urban area as the objects and the data of power consumers as the basis. Meanwhile, the author combines data of supporting facilities like business, cultural entertainment, sports, health care, education and scientific research to calculate the regional development maturity and analyze the data-mining model through the AHP model based on electric power big data of smart grid. And finally a mathematic model of AHP with strong robustness is constructed. The experiment shows that an analytical hierarchy process algorithm is first designed for the realization of the development guidance target of urban function areas in the undertaking that smart grids support smart cities. In the meantime, a maturity analyzing technique based on the advanced and reliable data-mining algorithm is also proposed for regional development level.

Keywords: analytical hierarchy process; maturity of regional development; data mining; smart grid

1 INTRODUCTION

Power data of a residential area is closely related to the regional maturity in terms of requirements. Data-mining technology is able to analyze the corresponding relation between power data and living indicators of a residential area in depth and finally establish a mathematic model for the data-mining calculation. A data-mining model of the overall electricity consumption level of consumers and ages of residents can be established for the analysis. A data-mining model of the weighted analysis of each functional module around a residential area and the infrastructure of hardware facilities can be established for statistics. And a data-mining model of electricity consumption law and habit of a residential area and the population can be established for qualitative calculation.

Power data can be calculated and analyzed through the above-mentioned data-mining technology so as to obtain regional electricity consumption habit and family structure through analyses on power data and family information. In the meantime, association rules like the age and the income of a householder are able to know consumers through electricity consumption information from various aspects. Data-mining on a residential area and the infrastructure indicators is able to carry out a scientific weighted calculation and a qualitative analysis on regional development and fi-
nally to accomplish the target model design of a data-mining model of a residential area and its maturity.

2 ANALYSIS MODEL OF REGIONAL DEVELOPMENT MATURITY BASED ON AHP

2.1 Overview of the analytic hierarchy process

The analytic hierarchy process, AHP for short, was first put forward by Thomas Saaty (T.L. Saaty), an operational researcher of America, in the mid 1970s. It is a systematic and hierarchical analytical method that combines qualitative analyses and quantitative analyses [1-2]. It is soon taken seriously all over the world because of its practicality and efficiency in terms of dealing with complex decision-making problems. It has been widely applied in economic planning and management, energy policy and allocation, behavioral science, military command, transportation, agriculture, education, talents, medical health, environment, etc. AHP has a variety of merits, the most important one of which is concision. It not only can be applied in cases with uncertainties and subjective information but also allows experience, insight and intuition in a logical way. Perhaps the biggest advantage of AHP is that hierarchy is proposed so that users can seriously measure the relative importance of indicators [3-7]. In this paper, the weight value of each influencing factor is calculated on the basis of AHP.

2.2 Optimized design of AHP

With a hierarchical analysis model, the author first classifies electricity consumption habits of consumers, then establishes an analysis model of residential maturity on this basis, and finally analyzes the regional development. The overall framework of the algorithm is presented in Figure 1.

![Figure 1. The whole process of the evaluation system.](image)

Figure 1. The whole process of the evaluation system.

Analysis processes of electricity consumption habits classification and residential maturity are presented in Figure 2 and Figure 3.

The algorithm flow of the optimized AHP proposed in this paper is presented below:

**Step1: establish a hierarchical structure model.** On the basis of in-depth analyses of practical problems, relevant factors are decomposed into multiple layers from top to bottom according to different attributes. Factors of the same layer not only are subject to and have influence on factors of the former layer but also dominate and are influenced by factors of the next layer. The top layer is the objective layer, having only one factor. The lowest layer is usually the project layer or the object layer. It has one or multiple layers in between, known as the criterion layer or the indicator layer. The criterion layer should be further decomposed into sub-criterion layers when there are excessive criteria (for example, more than 9 criteria).

![Figure 2. Analysis on habits of consumers.](image)

![Figure 3. Analysis on residential maturity.](image)
Step 2: Construct paired comparison matrix. From the second layer of the hierarchical structure model, construct paired comparison matrices for factors of the same layer subject to (or influencing) each factor of the former layer with the paired comparison method and 1-9 comparison measures all the way down to the lowest layer.

Step 3: Calculate weight vectors and carry out consistency check. Calculate the largest characteristic root and the corresponding eigenvector of each paired comparison matrix. Carry out consistency checks with consistency indicators, randomly consistency indicators and consistency ratio. If it is qualified after the check, the eigenvector (after normalization) is the weight vector; otherwise, a new paired comparison matrix needs to be constructed.

Step 4: Calculate combined weight vectors and carry out combined consistency check. Calculate the combined weight vector of the objective of the lowest layer and carry out combined consistency check according to the formula. If it is qualified after the check, decisions can be made in line with results represented by combined weight vectors. Otherwise, the model needs to be reconsidered or paired comparison matrices with larger consistency ratio need to be reconstructed.

2.3 The optimal number of clusters based on data characteristics and prior data

In previous studies, the optimal number of clusters are mainly determined in two ways, data characteristics and prior data:

(1) Data characteristics

The number of clusters is determined by specific data characteristics of data to be processed. Normally the size of a characteristic quantity is represented by information entropy, which is largely dependent on data. The calculation of information entropy requires the basic vector collection of data in advance, namely the basic vector space of data. But it is quite difficult to find such a vector space exactly in flow field cluster. So the optimal cluster number of the streamline cannot be determined directly in this way.

(2) Prior data

The number of clusters is determined by the preset clustering standard. This method is similar to the classification algorithm. This standard is normally obtained through empirical data. In the general algorithm of streamline clustering, it is impossible to be fully applicable to all data. Therefore, this algorithm fails to meet the requirements.

Based on similar matrixes to be clustered, the author combines the dimensionality reduction effect of spectral clustering on similar matrixes. The optimal clustering result of the same eigenvector matrix V corresponds to the optimal clustering result of the original similar matrix A.

If there is a matrix shaped like \( \bar{A} = \begin{bmatrix} C_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & C_i & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & C_n \end{bmatrix} \)

the size of its sub-matrix \( C_i \) is m*k and the streamline of \( C_i \) belongs to category i. Now the corresponding eigenvector of \( \bar{A} \) is the same with A, which means that the optimal clustering result of \( \bar{A} \) is the optimal clustering result of A. The determination of the optimal cluster number is converted to the problem of how to transform A to \( \bar{A} \). It is usually difficult for an arbitrary symmetric matrix to find a transformation matrix shaped like \( \bar{A} \). But it could be similar to \( \bar{A} \) as much as possible. The transformation matrix can be similar to \( \bar{A} \) only by setting a corresponding cost function. The cost function used in this paper is shown in the formula below:

\[
J = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 / M_k^2
\]

Here, \( a_{ij} \) is the similarity of streamline i and j; \( M_k \) is the maximum value of line i, standing for the largest distance between a streamline and other streamlines.

The above cost function eliminates the numerical effect. It can be seen that the value of the diagonal matrix J decreases. The process of searching the optimal classification is converted to the process of finding the minimum value of J.

In linear algebra, the Eigen value and the eigenvector can be solved in a relatively traditional way of solving equations, which are normally solved through numerical analyses in practical projects however. Jacobi iterative method is frequently used to solve the Eigen value and the eigenvector of a symmetric matrix through the matrix iteration. It solves the product of Jacobi rotation matrix \( J_k \), the basic process of which is shown below:

\[
A^{(k+1)} = J_k^T A^{(k)} J_k, \quad k = 0, 1, ...
\]

Here, \( J_k \) is an orthogonal matrix.

With the increase of k , \( A^{(k)} \) converges gradually to a diagonal matrix. Normally \( J_k \) is the Givens transformation matrix. The form is presented in the formula:

\[
J_k = G(i_k, j_k, \theta_k) = \begin{bmatrix} 1 & \cdots & \cos \theta_k & \sin(-\theta_k) \\ \vdots & \ddots & \sin \theta_k & \cos \theta_k \\ \cos \theta_k & \sin \theta_k & 1 & \cdots \\ \vdots & \ddots & \cdots & \cdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}
\]

Because A is a symmetric matrix, \( A^{(k)}(i,j) \) and \( A^{(k)}(j,i) \) will gradually become 0 if an appropriate \( \theta_k \) is selected. The fastest convergence is \( \theta_{max} \), the value of which is shown in the formula below:
\[ \tan 2\theta_{\max} = \frac{2A^{(k)}(p,q)}{A^{(k)}(q,q) - A^{(k)}(p,p)} \]

In the above formula, \( A^{(k)}(p,q) \) is the element with the largest absolute value in the matrix \( A^{(k)} \) except for the main diagonal element. \( p,q \) are determinant numbers.

The minimizing process of \( J \) is to select a proper \( \theta_k \) so as to gradually reduce \( J \), \( A^{(k)}(i,j) \) and \( A^{(k)}(j,i) \). The derivation of \( J \) is presented below:

\[
\frac{\partial J}{\partial \theta_k} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial a_{ij}}{\partial \theta_k} M_k^2
\]

The above formula decreases progressively in the process of iteration, the transformation matrix of which is set as \( R_k \). The process of proof is not elaborated here due to the limited length of the paper.

In conclusion, the clustering process is as follows:

1. A rough referential range of cluster number is defined as \( C = [C_1 \ldots C_m] \);
2. Repeat (3), (4), (5) for each \( C_i \);
3. Initialize \( A^{(0)} = A \);
4. As for the iterative transformation \( k \), calculate \( R_k \) and \( A^{(k)} \) according to \( A^{(k-1)} \).
5. Stop iteration according to the parameter \( C_i \), calculate the corresponding \( J \), and renew the minimum \( J_{\min} \).
6. The obtained classification result of \( J_{\min} \) is the optimal classification result.

On account of the fact that the probability that values with large deviation like 1 and \( n \) do not belong to the optimal cluster number is quite small, the calculation time can be effectively saved if a certain referential range of the optimal cluster number is predefined.

2.4 Evaluation system of regional maturity

A conventional statistical method generally carries out a statistical analysis on factors of a residential area. The mean value of factors is taken as the standard value. But the precondition is that there is no correlation between factors. Once there is a correlation, the referential standard value will have a corresponding deviation, leading to inaccurate evaluations. If all object lines are divided into subcategories through the clustering analysis that is similar to the analysis of consumers habits in the above section, the mean value of factors of each category can be obtained. The category accounting for the largest proportion is selected as the classification standard. This processing method filters the impact of a deal of abnormal data on the one hand and avoids data interference caused by factors correlation on the other hand because clustering deals with the data dimension as a whole [8-10].

Eight dimensions, including proportions of five different electricity consumption habits, the number of supermarkets in 2Km, the number of bus stations in 1Km and the number of hospital in 3Km, are taken as the analysis objects. Each data dimension is processed with normalization, the principle of which is in line with the formula below:

\[ \text{Value}_{\text{norm}} = \frac{W \cdot (\text{Value}_{\text{real}} - \text{Value}_{\text{min}})}{(\text{Value}_{\text{max}} - \text{Value}_{\text{min}})} \]

Here, \( \text{Value}_{\text{norm}} \) is the normalized value; \( \text{Value}_{\text{real}} \) is the actual value; \( \text{Value}_{\text{min}} \) and \( \text{Value}_{\text{max}} \) are respectively the minimum value and the maximum value; \( W \) is the normalized weight, representing the attribute proportion.

The evaluation formula is shown below after the standard reference value is obtained:

\[ \text{Rank} = \sum \frac{W_k \cdot \text{Dev}_{\text{real}}}{\text{Dev}_{\text{standard}}} \]

In the formula, \( \text{Dev}_{\text{real}} \) is the actual coefficient; \( \text{Dev}_{\text{standard}} \) is the standard coefficient; The standard score is \( \sum W_k \).

In the above formula, the evaluation value increases with the attribute value. That is to say, the larger the attribute value is, the higher the evaluation score will be.

3 EXPERIMENT AND ANALYSIS

3.1 Analysis on electricity consumption habits

Electricity consumption data of 2000 households from four sampled residential areas in 30 days are analyzed in this paper. In the case that these data are divided into five categories, the mean value of each category is taken as a data curve. The overall classification is presented in Figure 4.

Proportions of different habits are listed in the table below:

<table>
<thead>
<tr>
<th>Habits</th>
<th>Habit 1</th>
<th>Habit 2</th>
<th>Habit 3</th>
<th>Habit 4</th>
<th>Habit 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportions</td>
<td>21.1%</td>
<td>36.7%</td>
<td>6.2%</td>
<td>2.5%</td>
<td>33.5%</td>
</tr>
</tbody>
</table>

It can be known from the above data that the overall electricity consumption level of consumers of habit 1 is relatively low. Consumers of this kind are the elderly, who have obvious daily routine. Compared with habit 1, the overall electricity consumption level of consumers of habit 2 is higher. Consumers of this kind are office workers, who are middle-aged. Proportions of habit 3 and habit 4 are small but electricity consumptions and fluctuations are relatively large. It means that consumers of this two kinds have no regular daily routine and most of them may be unmarried young people. The most obvious feature of consumers of habit 5 is that the electricity consumption in late night is quite small yet the overall consumption level is moderate. Besides, the number of consumers of this kind is large. So, this electricity consumption habit is
the most common, which represents the electricity consumption per capita.

It is easy to obtain the association rule between each habit and family information like age and income of a householder through the data clustering result and the analysis on family information. Thus, it is able to know all consumers from electricity consumption information in different ways.

3.2 Analysis on residential area maturity

Take 20 residential areas of city A as examples. Analyze and evaluate data of residential areas, calculate weights through the analytic hierarchy process, and construct a discrimination matrix:

\[ W = \begin{bmatrix}
1 & 4 & 1 & 1/2 & 1/2 & 1 & 1/5 & 1/3 & 1/3 \\
1/4 & 1 & 1 & 1 & 1 & 1/7 & 1/5 & 1/5 \\
1 & 1 & 1 & 1 & 1 & 1/5 & 1/3 & 1/3 \\
2 & 1 & 1 & 1 & 1 & 1/3 & 1/2 & 1/2 \\
2 & 1 & 1 & 1 & 1 & 1/3 & 1/2 & 1/2 \\
2 & 1 & 1 & 1 & 1 & 1/3 & 1/2 & 1/2 \\
5 & 7 & 5 & 3 & 3 & 3 & 1 & 1/2 & 1/2 \\
3 & 5 & 3 & 2 & 2 & 2 & 2 & 1 & 1 \\
3 & 5 & 3 & 2 & 2 & 2 & 2 & 1 & 1
\end{bmatrix} \]

The maximum characteristic root of the normalized eigenvector of the above matrix is \( \lambda_{\text{max}} = 9.59 \). According to

\[ CI = \frac{\lambda_{\text{max}} - n}{n - 1} \]

the consistency indicator \( CI = 0.0733 \) can be calculated.

Inquire the average random consistency indicator \( RI \). Values of \( RI \) when \( n = 1 \ldots 9 \) are presented in the following table:

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>RI</td>
<td>0</td>
<td>0</td>
<td>0.58</td>
<td>0.90</td>
<td>1.12</td>
<td>1.24</td>
<td>1.32</td>
<td>1.41</td>
<td>1.45</td>
</tr>
</tbody>
</table>

This is a group of standard indexed generated randomly. Calculate the consistency ratio \( CR \)

\[ CR = \frac{CI}{RI} \]

It can be calculated that \( CR = 0.0506 < 0.1 \). Now the consistency of the matrix is acceptable.

In conclusion, weights of the above 9 influencing factors are:

<table>
<thead>
<tr>
<th>Factors</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electricity</td>
<td>0.0625</td>
</tr>
<tr>
<td>Habit 1</td>
<td>0.0478</td>
</tr>
<tr>
<td>Habit 2</td>
<td>0.0594</td>
</tr>
<tr>
<td>Habit 3</td>
<td>0.0757</td>
</tr>
<tr>
<td>Habit 4</td>
<td>0.0757</td>
</tr>
<tr>
<td>Habit 5</td>
<td>0.0757</td>
</tr>
<tr>
<td>Supermarket</td>
<td>0.2121</td>
</tr>
<tr>
<td>Bus station</td>
<td>0.1955</td>
</tr>
<tr>
<td>Hospital</td>
<td>0.1955</td>
</tr>
</tbody>
</table>

The distribution of data in a parallel coordinate system is shown in Figure 5.

Normalized data samples can be obtained through the normalization of the above data in line with the above-mentioned principle. The clustering analysis is carried out on all data of transformed samples. The counter normalization of the classification results of each clustering center when classification numbers are respectively 3 and 5 is shown in the Table 1.

Cluster each category when classification numbers are respectively 3 and 5. Establish a similarity matrix of the above data and calculate the optimal clustering number of the similarity matrix, which is 5.

Data of different level can be obtained according to the above data distribution. Carry out data fitting through normal distribution in line with the number of clusters. According to sample proportions and the feature of normal distribution, the total number of the data with the largest proportion reaches 50%. Data of the two are basically identical. Standard sample data used in this paper are presented as follows:
Evaluation scores of the data of the residential area mentioned below can be obtained through the above formula:

\[ \text{Electricity consumption} \times \text{Location} \times \text{Education} \times \text{Supermarket} \times \text{Bus station} \times \text{Hospital} \]

<table>
<thead>
<tr>
<th>Habit 1</th>
<th>Habit 2</th>
<th>Habit 3</th>
<th>Habit 4</th>
<th>Habit 5</th>
<th>Supermarket</th>
<th>Bus station</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.15</td>
<td>0.21</td>
<td>0.25</td>
<td>0.28</td>
<td>0.11</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0.17</td>
<td>0.26</td>
<td>0.28</td>
<td>0.16</td>
<td>0.13</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.10</td>
<td>0.11</td>
<td>0.23</td>
<td>0.29</td>
<td>0.21</td>
<td>0.16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.16</td>
<td>0.11</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0.19</td>
<td>0.21</td>
<td>0.29</td>
<td>0.20</td>
<td>0.11</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>0.25</td>
<td>0.18</td>
<td>0.24</td>
<td>0.20</td>
<td>0.13</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

The comprehensive evaluation score is \( \text{Rank} = 4.38 \), which is slightly higher than the standard value.

4 CONCLUSION

Based on data of power consumers, the author not only studies an analyzing algorithm of regional development maturity based on AHP with data-mining technology but also designs and verifies the data-mining model of residential area and regional maturity. Data-mining technology based on power data is able to establish effectively a variety of mathematical models like scientific electricity consumption habit and electricity consumption classification so as to provide original data support for detailed analyses on residents and obtain rational information like age and income of consumers. It guides the livelihood and development of residents in all respects, such as guiding electricity consumption, comparison and analysis of strong and weak development and problems to be solved for the development, so as to carry out comprehensive analyses on the general situation of urban areas distribution and the matching degree of urban areas and functional areas and finally realize the residential area planning guidance target based on power data mining in the undertaking that smart grids support smart cities.
REFERENCES


