A novel grey object matrix incidence clustering model for panel data and its application

Liu Yong, Du Jun-liang, Li Jiao
(School of Business, Jiangnan University, Jiangsu Wuxi 214122, China)

Abstract: In order to fully excavate the information contained in the multi-index panel data, one take decision objects as the research object, and the development state matrix and the development speed matrix of the decision objects are defined by considering the cross-section information and time information of the decision objects, and then the distances among the objects over the indexes are given. Based on grey incidence analysis, the absolute difference and relative difference between the measure value matrices are used to characterize and measure the close degree of the development state matrix and the development level matrix of the decision objects, so that the grey object matrix absolute incidence analysis model is established, and then according to the grey incidence degree between the objects, the objects can be clustered based on hierarchical clustering algorithm. Finally, a clustering problem of regional patent research and development (R&D) efficiency is used to verify the validity and rationality of the proposed model.

Key words: panel data; section information; time information; grey incidence analysis; hierarchical clustering algorithm

1. Introduction

Grey incidence analysis is an important part of grey system theory, which is also the cornerstone of grey system analysis, grey decision and grey clustering. Being different from the statistical methods such as incidence analysis that require large samples, the grey incidence analysis is suitable for the case of small sample size, which is often used to access the connection or influence degree among system variables. Its basic idea is to determine whether the link between different sequences is tight according to the geometry of the sequence curve [1]. Due to its unique advantages, the grey incidence analysis model and method has been becoming a hot issue for domestic and foreign scholars since it was proposed. By combing the related literatures on grey incidence analysis, we can see that they mainly establish points grey incidence analysis model[2], grey absolute (relative) incidence degree model[3], grey similarity and similar incidence analysis model[4], T-type incidence degree model[5], B-type incidence degree model[6], C-type incidence degree model[7], grey entropy incidence degree model[8], and Slope incidence degree model[9], and discuss the normative, consistent and translational properties of grey incidence analysis model, and apply these models and methods to energy [14], supplier selection [15] and other areas. Although these results enrich and develop the research of grey incidence analysis, these methods are suitable for dealing with the problems with cross-sectional data and time series data, while there exist a large number of problems to be analyzed and solved with matrix data and panel data in reality, so that a novel incidence analysis model should be established and researched [16].

According to the existing literatures on grey incidence analysis method to deal with problems with panel data, they are related to five aspects shown as follows. (1)AHP, ANP, TOPSIS, Fuzzy soft sets and other decision-making methods or models are combined with grey incidence analysis to construct some hybrid models [17-21]. These models only utilize grey incidence analysis method as a pretreatment method of data, which restricts the expansion and application space of grey incidence analysis. (2) According to the geometric characteristics of the data, based on the semi-definite characterization of the Hessel matrix and the proximity of the relative concavity and convexity of the incidence sequences, a grey incidence analysis model is established [22], however the model produces the problem of deviation resulting from continuous form to discrete form. (3) By extending from the curves relationship of sequence to surface relationship, based on the similarity of spatial geometry, a three-dimensional grey absolute incidence model is constructed in the form of the family of surfaces[23,24], but this extended model is affected by the presentation of the family of surfaces.(4) Based on the spatio-temporal characteristics of multi-index panel data, from the view of measuring the similarity between the objects from three aspects-the "horizontal" distance, "incremental" distance and "variation" distance of the panel data, a grey matrix incidence analysis model for panel data is established[25], however, the model is difficult for assembling information as a result of comprehensively considering the absolute quantity, incremental index and timing fluctuation of the panel data.(5)To analyze and extract the development law of the things, Cui and Liu constructed grey incidence model for panel data by using the grey incidence analysis method on the basis of the relative degree between reference sequence and comparison sequence[26]; Li et al constructed cumulative generation sequence of the time series of all indexes under the different objects, and then they designed the index incidence analysis model by exploiting the average generating rate of the generation sequence to represent the dynamic change trend of the original sequence and synthesizing the triple different information of deviation, difference and separation[27,28]. Based on the above discussion, although these research results can describe and solve the problem of panel data or high-dimensional data to a certain extent, and enrich and expand the scope of grey relational analysis model, they do not take the cross-sectional dimension and time dimension of the decision object into account to excavate the information contained in the multi-index panel data, so that there are some deviations of the results based on these models. According to the existing problems of grey incidence analysis model, one take decision objects as the research object, and based on the thought of grey incidence analysis, the grey object matrix absolute incidence analysis method is constructed by considering the cross-section information and time information of the decision objects, and then a hierarchical clustering algorithm is designed.

The remainder of this paper is organized as follows: the grey object matrix absolute incidence analysis method based on multi-index panel data is proposed in Section 2. In section 3, a hierarchical clustering algorithm based on grey object matrix absolute incidence analysis model is designed. In section 4, we make the case study to cluster and assess the regional patent R&D efficiency. The paper is concluded with some remarks, which are given in Section 5.

2. A grey object matrix absolute incidence analysis method based on multi-index panel data

Suppose that there exists a multi-index panel data decision information system denoted as $S=\{U,A,V,C\}$, where, $U=\left\{1,2,...,N\right\}$ and $A=\{a_1,a_2,...,a_m\}$ stands for a set of objects and indexes, respectively; $V=\cup v_{ij}^t(i=1,2,...,n;\ j=1,2,...,m;\ t=1,2,...T)$ expresses for the value field of panel data, and v_{ij}^{t} shows the observed value of the object i at time t with respect to the index j.

Definition 1 Assume that $x_{ij}(t)$ respects the dimensionless measure of the index value v_{ij}^{t} of $j(j=1,2,...,m)$ at time $t(t=1,2,...,T)$ of $i(i=1,2,...,N)$, on the premise of $i=1,2,...,n; j=1,2,...,m; t=1,2,...T$, if it satisfies $\mu_{ij}(t)=\dfrac{\Delta x_{ij}(t)}{x_{ij}(t)}$, then

exist a relationship of, then

$$\boldsymbol{x}_i=\begin{bmatrix} x_{i1}(1) & x_{i2}(1) & \cdots & x_{ij}(1) & \cdots & x_{im}(1) \\ x_{i1}(2) & x_{i2}(2) & \cdots & x_{ij}(2) & \cdots & x_{im}(2) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1}(t) & x_{i2}(t) & \cdots & x_{ij}(t) & \cdots & x_{im}(t) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1}(T) & x_{i2}(T) & \cdots & x_{ij}(T) & \cdots & x_{im}(T) \end{bmatrix}, i=1,2,...,N$$

$$\boldsymbol{\mu}_i=\begin{bmatrix} \mu_{i1}(2) & \mu_{i2}(2) & \cdots & \mu_{ij}(2) & \cdots & \mu_{im}(2) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mu_{i1}(t) & \mu_{i2}(t) & \cdots & \mu_{ij}(t) & \cdots & \mu_{im}(t) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mu_{i1}(T) & \mu_{i2}(T) & \cdots & \mu_{ij}(T) & \cdots & \mu_{im}(T) \end{bmatrix}, i=1,2,...,N$$

are called as the development state matrix and development speed matrix under panel data of the object i, respectively.

Where, $x_{ij}(t)$ and $\mu_{ij}(t)$ stands for the absolute quantity and the relative quantity of the index, respectively, and $\Delta x_{ij}(t)=|x_{ij}(t)-x_{ij}(t-1)|$.

For the clustering problem with panel data, the development of the objects with different indexes in different stages take on different states. Each sequence in the matrix of panel data represents the corresponding system behavior index. The columns of the matrix describe the cross-sectional characteristics of each object at a time, while the rows of the matrix reflect the temporal characteristics of the object. It can be said that the object is mainly affected by the cross-section information and time information in two aspects. Therefore, for any two objects, the proximity degree requires the comprehensive and integrated information on these two aspects to measure. In order to accurately characterize the dimension characteristic of panel data, one uses the dissmilarity degree to reverse the similarity between objects, the smaller the dissmilarity degree, the greater the similarity. From the two dimensions of the section and the time, the proximity of the cross-sectional information and the time information is characterized and measured by relative development state level difference and development speed level of different objects' index respectively, so that the grey object matrix absolute incidence analysis model is constructed.

Definition 2 For $\forall i,s\in X, \forall a_j\in A, \forall t\in T$ and $i\neq s$, if $\boldsymbol{x}_i,\boldsymbol{x}_s$ and μ_i,μ_s represent the development state matrix and development speed matrix of object i and s under panel data, respectively, then

$$d_{isj}^{1}=\frac{1}{T}\sum_{t=1}^{T}\Delta_{isj}^{1} \tag{1}$$

$$d_{isj}^{2}=\sqrt{\frac{1}{T-1}\sum_{t=1}^{T-1}(\mu_{ij}(t)-\mu_{sj}(t))^{2}} \tag{2}$$

are called as the development state horizontal distance and development speed horizontal distance between the object i and s over the multi-index panel data, respectively.

Where, $\Delta_{isj}^{1}=|x_{ij}(t)-x_{sj}(t)|$.

Definition 3 For $\forall i,s\in X,\forall a_j\in A,\forall t\in T$ and $i\neq s$, if d_{isj}^{1} and d_{isj}^{2} stands for the development state horizontal distance and development speed horizontal distance between the object i and s, respectively, then

$$\xi_{isj}=\frac{\min_i\min_j d_{isj}^{3}+\rho\max_i\max_j d_{isj}^{3}}{d_{isj}^{3}+\rho\max_i\max_j d_{isj}^{3}} \tag{3}$$

$$d_{is}^{A}=\sum_{j=1}^{m}w_j\xi_{isj} \tag{4}$$

are called as the grey incidence coefficient and incidence degree between the object i and s over the index and the index set under panel data, respectively.

Where, $d_{isj}^{3}=\lambda_1 d_{isj}^{1}+\lambda_2 d_{isj}^{2}$, $0\le\lambda_1\le 1, 0\le\lambda_2\le 1, \lambda_1+\lambda_2=1$, $\rho\in(0,1)$; w_j is the weight of index $j(j=1,2,...,m)$ under the index set, $0\le w_j\le 1,\sum_{j=1}^{m}w_j=1$.

The definition of 3 is the grey object matrix absolute incidence analysis model.

Theorem 1, Grey object matrix absolute incidence analysis model has the following basic properties:

1. Normality, $0<\xi_{is}\le 1$;
2. Symmetry;
3. Proximity;
4. Comparability;
5. Uniqueness;
6. Similarity;
7. Parallelism;
8. Consistency.

Proof: (similar to the proof in the literature 1, slightly)

3. Hierarchical clustering algorithm based on grey object matrix absolute incidence analysis model

The cluster of grey incidences is mainly used to classify factors of the same type in order to simplify complicated systems, and then it classifies the observed objects by setting thresholds on the basis of grey incidence degree between any two observed objects [1].

Definition 4 Assume that d_{is}^{A} is the distance between the decision object i, s on the multi-index set A, and for $\forall i, s \in U$, if $d_{is}^{A} = d_{si}^{A}$ is satisfied, then the matrix

$$d = \begin{bmatrix} d_{11}^{A} & d_{12}^{A} & \dots & d_{1s}^{A} & \dots & d_{1n}^{A} \\ d_{21}^{A} & d_{22}^{A} & \dots & d_{2s}^{A} & \dots & d_{2n}^{A} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{i1}^{A} & d_{i2}^{A} & \dots & d_{is}^{A} & \dots & d_{in}^{A} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{n1}^{A} & d_{n2}^{A} & \dots & d_{ns}^{A} & \dots & d_{nn}^{A} \end{bmatrix}$$

is called as the incidence matrix of the objects over the multi-index.

Definition 5 Assume that d stands for the incidence matrix of the objects over the multi-index, for $\forall i, s \in U, A$, the threshold value $\alpha \in [0,1]$, if $d_{is}^{A} \geq \alpha$, then i, s are the same class.

The threshold value α can be determined according to the needs of the actual problem. If the threshold value α is closer to 1, the more detailed classification, and the lower the number objects in each component; if the threshold value α is smaller, then the more coarse the classification is, and each component of the variable is relatively more. Generally, $\alpha = 0.5$.

According to the definition 5, we can find that grey incidence clustering is based on the determined clustering threshold value, which lacks theoretical basis, however, the hierarchical clustering algorithm can overcome the shortcoming. The hierarchical clustering algorithm is a clustering algorithm by repeatedly calculating the similarity between two types of data points and combining the most similar data of two points in all data points to realize the final classification. In summary, the hierarchical clustering algorithm is to determine the similarity between them by calculating the distance between each data point and all data points. The smaller the distance, the higher the similarity; the greater the distance, the lower the similarity; then the final distance from the nearest two data points or categories combined to generate clustering tree. As shown in Figure 1.

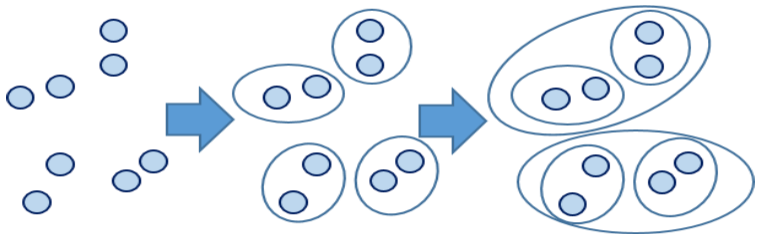

Figure 1 Schematic diagram of hierarchical clustering algorithm

Based on the above analysis, it is well known that the algorithm is a typical greedy algorithm. This algorithm first takes the local optimal and finally optimizes the global optimal. Because the results based on the hierarchical clustering algorithm depend entirely on the choice of the parameters, there is no need of global objective function similar to the K mean method and the problem of local minimum or difficultly selecting the initial point. Meanwhile, the merge operation of the hierarchical clustering algorithm can often one-time implement the final cluster results, which means that once two clusters are combined, they are not revoked. Therefore, the hierarchical clustering algorithm not only avoids the error caused by the uncertainty of human cognition, but also overcomes the problem of setting grey incidence clustering threshold value, so that the cluster results based on is closer to the real situation. In view of this, on the basis of the grey incidence degree between any two objects, we exploit the hierarchical clustering algorithm to substitute the cluster of grey incidences to cluster analysis.

In the the hierarchical clustering algorithm, there are many ways such as Euclidean, Seuclidean, Mahalanobis, Cityblock, Minkowski, Cosine, and Hamming to calculate distance matrix. However, these distance calculation methods focus on the construction of distance, and they do not reflect the actual background of the research problem and cannot totally reflect and excavate the information contained in panel data, so that a novel distance matrix should be established to reflects the information of the panel data and meet the requirements of the hierarchical clustering.

In fact, the greater the grey incidence degree between the two clustering objects, the smaller the distance between the two clustering objects, on the contrary, visa. It can be seen that grey incidence degree between the two clustering objects is inverse relationship with distance, and then the reciprocal of the grey incidence degree can be used to express the distance. Because the distance between any object and its own is 0, the principal diagonal of the distance matrix is all 0, the principal diagonal of the grey incidence degree matrix is all 1, and therefore, the transformation by taking grey incidence degree to be reciprocal and then reduce by 1 can be realized. Assume that d' stands for the distance matrix in this paper, the distance matrix d' can be obtained as follows:

$$d' = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1s} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2s} & \dots & d_{2n} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{i1} & d_{i2} & \dots & d_{is} & \dots & d_{in} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{ns} & \dots & d_{nn} \end{bmatrix}$$

Where, $d_{is} = 1/d_{is}^{A} - 1, i = 1, 2, \cdots, n; s = 1, 2, \cdots, n.$

The algorithm flow of the hierarchical agglomerative clustering based on the grey object matrix absolute incidence analysis can be determined and shown in Figure 2, and its specific steps are as follows:

Step 1: Calculate grey incidence degree between any two objects and turn it into distance;

Step 2: Classify each decision object as a class and get a total of N classes;

Step 3: Merge the closest two classes into one class;

Step 4: Calculate the distance between the new class and all the old classes;

Step5: Repeat the third and fourth steps until the end of the merge into a class (which contains N objects) or to satisfy certain conditions;

Step 6: End.

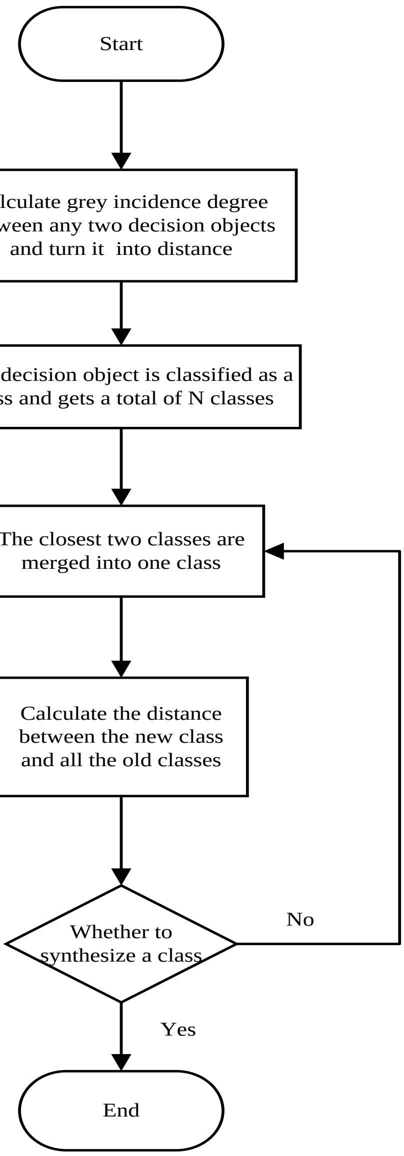

Fig. 2 Algorithm flowchart of the hierarchical clustering based on the grey object matrix absolute incidence analysis

4 Case applications

Patent is an important index to measure the technological innovation capability of regional high-tech industry, and it is a weapon to obtain scientific and technological competitiveness and occupy the high ground of science and technology. After more than 30 years of reform and opening-up, china's scientific and technological innovation activities greatly increased. According to the statistical date of the National Bureau of Statistics 2008-2016, the patent R&D investment (R&D staff full-time equivalent and R&D expenditure) and output (patent application volume, technical market turnover) increase from 173.6 million person-years, 3710.2 billion yuan RMB, 573 thousand, 87.6 billion yuan RMB in 2007 to 375.9 million person-years, 1416.99 billion yuan RMB, 27985001, 986.6 billion yuan RMB in 2015, respectively, and the growth rate is much faster. However, through analysis, we can see that technology R&D in China is still in the initial stage, and there are some problems such as uneven regional development and low efficiency. Compared with the efficiency of patent R&D in developed countries, the driving force of patent R&D in our country has not yet appeared, and there exist a variety of waste in the factors of scientific and technological development, and the “fragmentation” of patent input and output is more serious, which directly affects the utilization efficiency of China's scientific and technological resources, as well as strategic deployment of the transformation of economic development mode and the construction of an innovative country. To evaluate and analyze the development status of regional patent R&D in china, and grasp and clarify the differences in the status of patent development in various regions, and formulate policies and measures for R&D in different regions, according to the statistical data of “China Statistical Yearbook” (2012-2016), “ China Statistical Yearbook of science and technology” (2012-2016) and SIPO (People's Republic of China Intellectual Property Office), one takes the data of the efficiency of patent R&D in 30 provinces or regions (provinces, municipalities and autonomous regions of China, and The Tibet region is not included because of lacking some data) as a research object, and uses the proposed model to evaluate the efficiency of regional patent R & D in China.

The efficiency evaluation of provincial patent R&D is a complex systematic project, and it is necessary to establish a comprehensive and objective evaluation index system. According to the principles of scientific, comprehensive and data availability, on the basis of the relevant research[29], from the angle of patent innovation input and output, an evaluation index system of patent R&D efficiency is designed. In terms of input, macro environment and R&D input should be focused on. Generally speaking, macro environment can be measured and characterized by the per capita GDP and the number of people with a college degree or above, and the R&D input is mainly measured by two indexes of internal expenditure on R&D and full-time equivalent of R&D personnel. In terms of output, it mainly describes the level and capability of patent R&D output from three dimensions of output capability, output quality and industrialization level, and the output capability is generally measured by the number of patent applications accepted, the output quality is measured by two indexes of the number of invention patent application accepted and technology market turnover, and the industrialization level is measured by high tech industry profits and export volume. The specific evaluation indexes are shown in Table 1.

Table 1 The index system of patent R&D efficiency in China's provinces and cities

First level index	Second level index	Third level index
Input	Macro environment	Per capita GDP
		the number of people with a college degree or above
	R & D input	internal expenditure on R&D
		full-time equivalent of R&D personnel
Output	output capability	the number of patent applications accepted
	output quality	the number of invention patent application accepted
		technology market turnover
	industrialization level	high tech industry profits
		export volume

According to the preliminary research and the evaluation information from the experts invited by us, the weight of each index, the development status level and development speed can be determined, respectively, and be shown as follows:

$$w_1 = 0.2, w_2 = 0.08, w_3 = 0.1, w_4 = 0.12, w_5 = 0.06, w_6 = 0.15, w_7 = 0.11, w_8 = 0.1, w_9 = 0.08$$

$\lambda_1 = 0.7, \lambda_2 = 0.3$.

Based on the proposed model in this paper, grey incidence degree between regions and incidence matrix under the index system of the patent R&D efficiency can be calculated and shown in Table 2, and then the hierarchical clustering distance matrix between any two regions can be determined by using the designed method in this paper, and shown in Table 3. According to the clustering object distance matrix of 30 regions in China under the index system of patent R&D efficiency, the “Ward” method is used to cluster 30 regions, and then the specific clustering categories can be obtained and shown in Fig. 3.

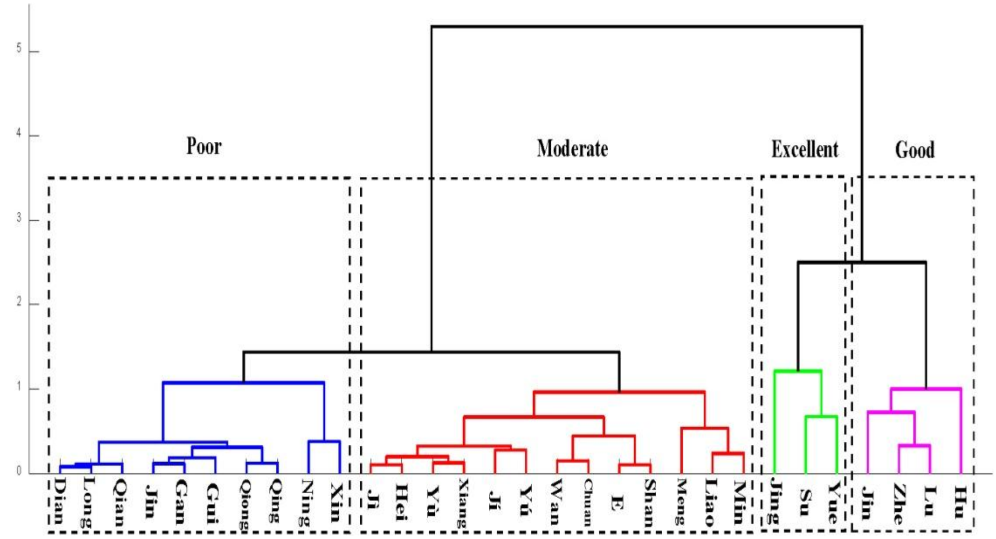

Fig. 3 Tree cluster diagram of patent R&D efficiency of provinces and cities in mainland China

From the Fig. 3, we can see that 30 provinces and cities in China mainland can be clustered into the four categories of excellent, good, medium and poor. For the category of excellent, the efficiency of patent R&D is relatively high, and the patent R&D is in the mature stage, which is embodied in high input and high putout of patent R&D, while these provinces and cities include Beijing(Jing), Guangdong(Yue) and Jiangsu(Su); these provinces and cities with good efficiency of patent R&D are Tianjin(Jin), Zhejiang(Zhe), Shandong(Lu) and Shanghai(Hu), respectively; These provinces and cities with general efficiency of patent R&D (at medium level) cover Hebei(Jì), Heilongjiang(Hei), Henan(Yù), Hunan(Xiang), Jilin(Jí), Chongqing(Yú), Anhui(Wan), Sichuan(Chuan), Hubei(E), Shaanxi(Shan), Inner Mongolia(Meng), Liaoning(Liao), Fujian(Min); These provinces and cities with low efficiency of patent R&D(at poor level): Yunnan(Dian), Gansu(Long), Guizhou(Qian), Shanxi(Jin), Jiangxi(Gan), Guangxi(Gui), Hainan(Qiong), Qinghai(Qing), Ningxia(Ning), Xinjiang(Xin). From the clustering results, we can see that the efficiency of the regional patent R&D takes on imbalance, and the efficiency of the eastern provinces and cities are significantly better than those of the central and western provinces and cities, while the efficiency of the central provinces and cities are better than that of western provinces and cities, and the efficiency of the western provinces and cities are the worst. What’s more, there are few provinces and cities with higher efficiency, and much more regions are general and poor, which indicate that the scale effect of China's patent R&D has not yet been highlighted, and the efficiency of patent R&D is not high as a whole.

For these regions with relatively high efficiency, they rely on a strong scientific research institutions, and government policies and adequate financial support, and then they develop early science and technology, so that their market internationalization, information level and transformation rate of patent achievements is high. However, it is difficult for the efficiency of these regions to have a greater breakthrough in a short time. Therefore, these regions should pay much more attention to international development direction, and take responsibility to occupy the commanding heights of the international science and technology in order to develop patent R&D and realize Chinese manufacturing go out. Some measures such as making full use of international resources for innovation to accelerate the internationalization level and expand the vision of globalization, adjusting patent structure to reduce the development cost, increasing the number of high-tech industry patent to break the international patent barriers, and participate in international competition should be given.

For these regions with relatively good efficiency of patent R&D, their economic is good, and their science and technology strength is strong, and then their patent input are in a higher level. However, the ability of these regions′ output ability and input is obviously insufficient, and their market orientation of patents needs to be improved; the efficiency of patent R&D is not fully excavated and improved. In order to improve the efficiency level of these regions′ patent R&D, it is necessary to improve the efficiency of the usage of innovative resources, and take effective demand of the scientific and technological market as the first direction during the patent R&D, and broaden the scope of patent R&D and form a reasonable patent structure. In the inherent advantages of patent R&D, these provinces and cities must continue to maintain and consolidate the leading position, and strive to seek internationalization and globalization development path.

For these regions with general efficiency of patent R&D, their patent R&D is at a stage of rapid development, but their patent R&D input and output are at a moderate level. These regions are main the central provinces, which are in a stage of rapid economic development. However, these regions cannot blindly pursue the quantity and speed of patent R&D, but ignore the quality. Otherwise, the consequences will be a lot of useless waste of scientific and technological resources, and the efficiency of patent R&D is not high. If these regions want to maintain the healthy, stable, green and sustainable development of patent R&D, they should source flow in terms of input, meanwhile these local governments should well guide the direction, based on the existing characteristics of these regions′ resources and advantages, and increase the input of patent R&D funds, personnel and policy and make reasonable plans for patent R&D according to local conditions. In the aspect of output, these regions should adhere to the market orientation, improve the scientific and technological content and quality of patent R&D, and avoid the patent of invalid, low value and low efficiency.

For these regions with poor efficiency of patent R&D, they lack the infrastructure and technological innovation environment, and their economic development fall behind, and their development level of scientific and technology is low, and high tech enterprises and scientific research universities are fewer in quantity and poorer in quality. At the same time, these local governments have not paid enough attention to them, so that the input and output of patents are at a poor level. Patent R&D of these regions are in the initial stage, and then these governments should strengthen the policy tilt and funding support, and recommend high quality high-tech enterprises and research institutes in order to develop these provinces. Meanwhile, these local governments should improve the scientific and technological innovation infrastructure and the level of economic development in these regions, strengthen cooperation with science and technology projects in the eastern region and utilize regional resource advantages to develop frontier high-tech industry trade.

Through the above analysis, the proposed model in this paper can make full use of existing information, and mine the related information and rules contained in the panel data, and properly deal with the problem with the cluster evaluation of panel data, which makes the evaluation results be with the actual.

5 Conclusions

In order to fully grasp the information contained in the panel data, based on the thought and method of the grey incidence analysis model and hierarchical clustering algorithm, to a novel grey object matrix absolute incidence clustering model based on multi-index panel data is established, which enriches and develops the application area and range of grey incidence analysis method. Through model and case analysis, the results show that the proposed model in this paper can effectively deal with the clustering decision problem with panel data, and extract the information about the development status and development speed of the objects under panel data. However, the proposed model in this paper cannot efficiently deal with the decision problem which contains hysteresis effect and periodic fluctuation, which is also open to make research.

Acknowledgements

This work is partially funded by the National Natural Science Foundation of China (71503103), Natural Science Foundation of Jiangsu Province (BK20150157); Social Science Foundation of Jiangsu Province (14GLC008); Jiangsu Province University Philosophy and Social Sciences for Key Research Program(2017ZDIXM034); the Fundamental Research Funds for the Central Universities (2017JDZD06).

References

[1]Liu S F, Yang Y J, Lin Y. Grey data analysis: method, models and applications, Springer-Verlag, 2016.

[2]Deng J L. Spread of grey relational space. Journal of Grey Systems, 1995, 7(3): 96-100.

[3]Liu Y, Liu S F, Forrest J. A new grey absolute degree of grey incidence model and application. Chinese Journal of Management Science, 2012, 20 (5): 173-177.

[4]Liu S F, Xie N M, Forrest J. On new models of grey incidence analysis based on visual angle of similarity and nearness. Systems Engineering-Theory & Practice, 2010, 30(5):881-886.

[5]Sun YG, Dang Y G. Improvement on grey T's correlation degree. Systems Engineering-Theory & Practice, 2008, 28(4):135-139.

[6]Wang Q Y. The grey relational analysis of B-mode. J. Huazhong Univ. of Sci. & Tech, 1989, 17(6):77-81.

[7]Wang Q Y, Zhao X H. The relational analysis of c-mode. J. Huazhong Univ. of Sci. & Tech, 1999, 27 (3) : 75-77.

[8]Dang Y G, Liu S F, Liu B, et al. Improvement on Degree of Grey Slope Incidence. Engineering Science, 2004, 6(3):23-26.

[9]Zhang Q S, Guo S J, Deng J L. Grey relation entropy method of grey relation analysis. Systems Engineering-Theory & Practice, 1996, 18(8) : 7-11.

[10]Xiao X P, Li W Z. The relative degree analysis method and sensitivity analysis for multiple attribute decision making with time series. Systems Engineering and Electronics, 1995:36-43.

[11]He W Z, Guo P. Some problems about the grey correlation degree. Mathematical Statistics and Management, 1999,18 (3): 25-30.

[12]Xie N M, Liu S F. The parallel and uniform properties of several relational models. Systems

Engineering, 2007, 25(8):98 -103.
[13]Cui J, Dang Y G, Liu S F. Novel properties of some grey relational analysis models. Systems Engineering, 2009, 27(4):65-70.
[14]Lu I J, Lin S J, Lewis C. Grey relation analysis of motor vehicular energy consumption in Taiwan. Energy Policy, 2008,(36):2556-2561.
[15]Lee Y L, Tsai F C, Liu S F, el al. A scale development of industrial designer ability index through quality function deployment and grey relational analysis methods. Advances in Mechanical Engineering, 2016, Vol. 8(12) 1–11.
[16]Liu S F, Cai H, Yang Y J, et al. Advance in grey incidence analysis modelling. Systems Engineering-Theory & Practice, 2013, 33(8):2041-2046.
[17]Zhu J J, Zhang S T, Chen Y, el al. A hierarchical clustering approach based on three-dimensional gray relational analysis for clustering a large group of decision makers with double information. Group Decision and Negotiation, 2016,25(2):325-354.
[18]Hashemi A H, Karimi A, Tavana M. An integrated green supplier selection approach with analytic network process and improved Grey relational analysis. Int. J. Production Economics, 2015, (59):178–191.
[19]Rajesh R, Ravi V. Supplier selection in resilient supply chains: a grey relational analysis approach. Journal of Cleaner Production, 2015, (86): 343-359.
[20]Wang P, Zhua Z Q, Wang Y H. A novel hybrid MCDM model combining the SAW,TOPSIS and GRA methods based on experimental design. Information Sciences, 2016, (345): 27–45.
[21]Xie N X, Han Y, Li Z W. A novel approach to fuzzy soft sets in decision making based on grey relational analysis and MYCIN certainty factor. International Journal of Computational Intelligence Systems, 2015, Vol. 8, No. 5: 959-976.
[22]Wu L F, Liu S F. Panel data clustering method based on grey convex relation and its application [J]. Control and Decision, 2013, 28(7): 1033-1037.
[23]Zhang K, Liu S F. Extended clusters of grey incidences for panel data and its application. Systems Engineering-Theory & Practice, 2010, 30(7): 1253-1259.
[24]Liu Z, Dang Y G, Qian W Y, et al. Grey grid incidence model based on panel data. Systems Engineering -Theory & Practice, 2014, 34(4): 991-996.
[25]Qian W Y, Wang Y H, Dang Y G. Grey matrix relational modeling and its application based on multivariate panel data[J]. Systems Engineering, 2013, 31(10):70-74.
[26]Cui L Z, Liu S F. Grey matrix similar incidence model for panel data and its application. Chinese Journal of Management Science, 2015,23(11):171-176.
[27]Li X M, Dang Y G, Wang J J. Grey relational clustering model for panel data clustering on indexs and its application. Control and Decision, 2015,30(8):1447-1452.
[28]Li X M, Hipel, K W, Dang Y G. An improved grey relational analysis approach for panel data clustering. Expert Systems with Applications, 2015, 42(23):9105-9116.
[29]Zhao H F, Li W W, Xu S, et al. The efficiency of patent innovation in the different regions of China. Chinese Journal of Management Science, 2008,16(SI):387-392.

Table 2 grey object incidence degree matrix under the index system of patent R&D efficiency

	Jing	Jin	J ì	Jin	Nei	Liao	J í	Hei	Hu	Su	Zhe	Wan	Min	Gan	Lu	Y ù	E	Xiang	Yue	Gui	Qiong	Y ú	Chuan	Qin	Tian	Shan	Long	Qing	Ning	Xin
Jing	1.000	0.744	0.607	0.600	0.623	0.658	0.610	0.610	0.770	0.600	0.691	0.638	0.652	0.600	0.704	0.621	0.644	0.616	0.585	0.600	0.588	0.623	0.645	0.587	0.592	0.643	0.592	0.588	0.526	0.542
Jin		1.000	0.767	0.749	0.755	0.831	0.761	0.777	0.845	0.654	0.766	0.762	0.806	0.755	0.740	0.790	0.810	0.785	0.607	0.756	0.725	0.791	0.793	0.739	0.732	0.798	0.735	0.725	0.675	0.667
J ì			1.000	0.936	0.826	0.838	0.918	0.951	0.685	0.610	0.750	0.878	0.868	0.940	0.772	0.928	0.879	0.946	0.634	0.911	0.914	0.898	0.877	0.901	0.904	0.874	0.887	0.911	0.862	0.842
Jin				1.000	0.846	0.820	0.925	0.936	0.672	0.607	0.741	0.879	0.845	0.944	0.761	0.911	0.856	0.923	0.624	0.919	0.922	0.877	0.880	0.908	0.925	0.853	0.922	0.926	0.863	0.856
Nei					1.000	0.841	0.867	0.826	0.689	0.659	0.796	0.763	0.849	0.827	0.757	0.793	0.790	0.811	0.651	0.806	0.839	0.813	0.763	0.816	0.831	0.783	0.837	0.849	0.794	0.789
Liao						1.000	0.850	0.853	0.755	0.677	0.829	0.820	0.899	0.822	0.829	0.838	0.866	0.850	0.696	0.813	0.793	0.851	0.834	0.801	0.792	0.862	0.797	0.796	0.750	0.741
J í							1.000	0.920	0.686	0.628	0.765	0.839	0.877	0.909	0.792	0.876	0.880	0.900	0.651	0.881	0.896	0.901	0.839	0.885	0.894	0.869	0.895	0.908	0.861	0.840
Hei								1.000	0.693	0.625	0.762	0.887	0.865	0.925	0.778	0.923	0.891	0.945	0.636	0.906	0.906	0.905	0.888	0.892	0.887	0.886	0.894	0.913	0.862	0.850
Hu									1.000	0.704	0.764	0.712	0.731	0.676	0.731	0.703	0.744	0.698	0.641	0.673	0.654	0.708	0.733	0.661	0.657	0.748	0.658	0.652	0.599	0.600
Su										1.000	0.743	0.625	0.665	0.605	0.712	0.616	0.662	0.619	0.708	0.597	0.590	0.632	0.632	0.590	0.586	0.658	0.593	0.599	0.543	0.557
Zhe											1.000	0.787	0.827	0.743	0.858	0.761	0.781	0.762	0.724	0.736	0.715	0.777	0.775	0.721	0.717	0.763	0.722	0.726	0.672	0.674
Wan												1.000	0.820	0.907	0.807	0.905	0.859	0.901	0.655	0.902	0.859	0.867	0.932	0.862	0.859	0.855	0.861	0.850	0.788	0.776
Min													1.000	0.852	0.852	0.870	0.872	0.880	0.708	0.837	0.820	0.883	0.844	0.830	0.823	0.855	0.820	0.823	0.774	0.761
Gan														1.000	0.762	0.929	0.855	0.931	0.623	0.936	0.926	0.878	0.902	0.928	0.935	0.849	0.921	0.916	0.850	0.840
Lu															1.000	0.784	0.812	0.787	0.777	0.752	0.739	0.799	0.808	0.735	0.732	0.796	0.741	0.744	0.696	0.682
Y ù																1.000	0.884	0.947	0.638	0.919	0.894	0.910	0.923	0.888	0.884	0.876	0.872	0.887	0.821	0.813
E																	1.000	0.900	0.672	0.853	0.837	0.916	0.890	0.821	0.818	0.950	0.821	0.839	0.800	0.785
Xiang																		1.000	0.645	0.912	0.910	0.910	0.907	0.890	0.890	0.893	0.883	0.907	0.847	0.837
Yue																			1.000	0.618	0.613	0.657	0.650	0.603	0.603	0.669	0.608	0.617	0.560	0.573
Gui																				1.000	0.901	0.879	0.896	0.915	0.917	0.840	0.900	0.891	0.831	0.814
Qiong																					1.000	0.862	0.855	0.892	0.899	0.833	0.889	0.932	0.868	0.850
Y ú																						1.000	0.881	0.850	0.843	0.891	0.844	0.862	0.819	0.794
Chuan																							1.000	0.853	0.854	0.887	0.855	0.851	0.784	0.782
Qian																								1.000	0.950	0.816	0.939	0.887	0.828	0.811
Tian																									1.000	0.815	0.962	0.901	0.826	0.821
Shan																										1.000	0.817	0.837	0.798	0.781
Long																											1.000	0.902	0.829	0.827
Qing																												1.000	0.874	0.875
Ning																													1.000	0.799
Xin																														1.000

Table 3 hierarchical clustering distance matrix under the index system of patent R&D efficiency

	Jing	Jin	J ì	Jin	Nei	Liao	J í	Hei	Hu	Su	Zhe	Wan	Min	Gan	Lu	Y ù	E	Xiang	Yue	Gui	Qiong	Y ú	Chuan	Qin	Tian	Shan	Long	Qing	Ning	Xin
Jing	0.000	0.344	0.647	0.666	0.606	0.521	0.639	0.640	0.298	0.666	0.448	0.567	0.533	0.667	0.420	0.611	0.553	0.624	0.709	0.666	0.702	0.604	0.550	0.702	0.690	0.554	0.690	0.699	0.901	0.845
Jin		0.000	0.304	0.336	0.324	0.203	0.314	0.288	0.184	0.529	0.305	0.313	0.241	0.324	0.351	0.265	0.235	0.274	0.648	0.322	0.380	0.264	0.261	0.353	0.365	0.254	0.360	0.380	0.481	0.500
J ì			0.000	0.069	0.211	0.193	0.089	0.051	0.460	0.640	0.333	0.139	0.152	0.063	0.295	0.077	0.138	0.057	0.577	0.097	0.094	0.114	0.140	0.110	0.106	0.144	0.127	0.097	0.160	0.188
Jin				0.000	0.182	0.220	0.081	0.068	0.488	0.646	0.350	0.137	0.183	0.060	0.314	0.097	0.168	0.084	0.602	0.088	0.084	0.140	0.137	0.101	0.081	0.172	0.085	0.080	0.159	0.168
Nei					0.000	0.189	0.153	0.211	0.452	0.517	0.257	0.311	0.177	0.209	0.321	0.261	0.266	0.234	0.536	0.240	0.191	0.231	0.310	0.226	0.204	0.277	0.195	0.178	0.259	0.267
Liao						0.000	0.177	0.173	0.325	0.478	0.207	0.220	0.113	0.217	0.206	0.194	0.155	0.176	0.437	0.230	0.261	0.175	0.199	0.249	0.263	0.160	0.254	0.256	0.334	0.350
J í							0.000	0.087	0.458	0.591	0.307	0.192	0.141	0.100	0.263	0.142	0.137	0.111	0.535	0.135	0.116	0.109	0.192	0.130	0.119	0.150	0.118	0.101	0.161	0.190
Hei								0.000	0.444	0.601	0.313	0.127	0.157	0.081	0.285	0.083	0.123	0.059	0.573	0.103	0.104	0.105	0.125	0.121	0.128	0.129	0.119	0.095	0.161	0.177
Hu									0.000	0.421	0.309	0.404	0.367	0.479	0.368	0.423	0.343	0.432	0.560	0.486	0.529	0.412	0.364	0.512	0.523	0.336	0.521	0.533	0.670	0.667
Su										0.000	0.345	0.600	0.504	0.652	0.405	0.622	0.511	0.615	0.412	0.676	0.696	0.582	0.582	0.696	0.706	0.519	0.685	0.669	0.841	0.796
Zhe											0.000	0.270	0.209	0.347	0.166	0.315	0.281	0.312	0.380	0.358	0.399	0.287	0.290	0.386	0.394	0.310	0.386	0.378	0.489	0.483
Wan												0.000	0.219	0.103	0.239	0.105	0.164	0.110	0.527	0.108	0.164	0.154	0.073	0.161	0.164	0.169	0.162	0.176	0.270	0.289
Min													0.000	0.174	0.174	0.149	0.147	0.136	0.413	0.194	0.219	0.133	0.185	0.205	0.215	0.170	0.220	0.214	0.292	0.314
Gan														0.000	0.313	0.076	0.170	0.074	0.604	0.069	0.080	0.140	0.108	0.078	0.070	0.177	0.086	0.092	0.177	0.191
Lu															0.000	0.276	0.232	0.271	0.287	0.329	0.353	0.251	0.238	0.361	0.367	0.256	0.350	0.343	0.437	0.467
Y ù																0.000	0.131	0.056	0.567	0.088	0.119	0.098	0.083	0.126	0.131	0.142	0.147	0.128	0.219	0.230
E																	0.000	0.111	0.488	0.172	0.195	0.092	0.124	0.217	0.223	0.052	0.217	0.192	0.251	0.273
Xiang																		0.000	0.551	0.096	0.099	0.098	0.102	0.124	0.124	0.120	0.132	0.103	0.180	0.195
Yue																			0.000	0.617	0.632	0.522	0.538	0.659	0.657	0.494	0.646	0.621	0.785	0.744
Gui																				0.000	0.110	0.137	0.116	0.093	0.090	0.190	0.111	0.122	0.204	0.229
Qiong																					0.000	0.161	0.169	0.121	0.113	0.200	0.125	0.073	0.151	0.176
Y ú																						0.000	0.136	0.177	0.186	0.123	0.185	0.160	0.221	0.259
Chuan																							0.000	0.173	0.172	0.127	0.169	0.175	0.275	0.278
Qian																								0.000	0.053	0.225	0.065	0.128	0.207	0.233
Tian																									0.000	0.227	0.039	0.110	0.211	0.218
Shan																										0.000	0.224	0.194	0.253	0.281
Long																											0.000	0.108	0.206	0.209
Qing																												0.000	0.144	0.143
Ning																													0.000	0.251
Xin																														0.000