



Sharif University of Technology  
Scientia Iranica  
Transactions E: Industrial Engineering  
<http://scientiairanica.sharif.edu>



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

Y. Liu\*, J.-L. Du, and J. Li

*School of Business, Jiangnan University, Jiangsu Wuxi 214122, China.*

Received 4 September 2017; received in revised form 12 May 2019; accepted 17 June 2019

## KEYWORDS

Panel data;  
Section information;  
Time information;  
Grey incidence  
analysis;  
Hierarchical clustering  
algorithm.

**Abstract.** In order to fully excavate the information contained in multi-index and small sample panel data, one takes decision objects as the research object. 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 indices are given. Based on grey incidence analysis, the absolute difference and relative difference between the measured 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. Subsequently, according to the grey incidence degree between the objects, the objects can be clustered based on the 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.

© 2021 Sharif University of Technology. All rights reserved.

## 1. Introduction

The panel data combines the characteristics of the time data series and cross-section data, and it can describe the dynamic characteristics of the research object. It is widely used in the modeling practice of economic and management problems, and it has an important position in the analysis and research of economic and management problems. In addition, multi-indicator panel data can provide a more important application background and value for providing comprehensive information and data characteristics of research objects. Panel data clustering analysis has been widely used in areas such as marketing [1],

finance [2], energy [3,4], etc., and its importance is self-evident. For instance, if one wanted to measure the level of China's green technology innovation efficiency, its data should include time, evaluation indicators and subjects to be evaluated. This is a typical panel data that contains three evaluation dimensions. Traditional two-dimensional data forms can only express information on any two of the three dimensions, such as time-evaluation indicators, time-evaluation subjects and evaluation indicators-evaluation subjects. Cases similar to this include assessment of air pollution in the environmental field, identification of financial risks in the financial sector and predictions of energy consumption in the energy sector. Panel data clustering requires efficient aggregation of feature information in three dimensions, whereas traditional clustering methods cannot be applied to panel data. In other words, many problems in the real world contain panel data information. How to effectively mine the effective

\*. Corresponding author. Tel.: +8615061889051  
E-mail address: [clly1985528@163.com](mailto:clly1985528@163.com) (Y. Liu)

information contained in the panel data is of great significance for the analysis of clustering results and real problems.

Research into the panel data clustering method is still in its infancy, and the existing methods are mainly based on multivariate statistical theory. Bonzo and Hermosilla (2002) first introduced multivariate statistical methods into panel data analysis and used probabilistic connection functions to improve clustering analysis algorithms for panel data analysis [5]. Subsequent research involves fixed effect [6], time series [2,7], ordinal data [8], parameter estimation [9,10], etc. However, the statistical method performs cluster analysis, which has certain requirements on the number of samples. Melanization of statistics such as the variance and covariance is likely to result in loss of data information. With the development of computer technology and the improvement of computing power, clustering algorithms for panel data are beginning to emerge. These studies are mainly related to the Bayesian approach [11,12], artificial bee colony [13], data mining [14], DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [15] and GMM (Gaussian Mixture Model) [16]. The shortcomings of the above literature is that their clustering ideas mainly focus on the improvement and design of the clustering algorithm, but do not examine the multiple dynamic information features of the data format of the panel data. For the latter, considering the spatiotemporal characteristic of panel data and the uncertainty of the decision-making system, some scholars adopt uncertain system theories and methods such as grey incidence analysis, rough set [17,18] and others [19] to deal with panel data.

Grey incidence analysis is an important part of the grey system theory, which is also the cornerstone of grey system analysis, grey decision, and grey clustering. Being different from the statistical methods that require large samples, the grey incidence analysis is suitable in cases of small size samples, and 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 [20]. Due to its unique advantages, the grey incidence analysis model and method is becoming a hot issue for domestic and foreign scholars since it was proposed. Currently, it is widely used in economic analysis and evaluation [21], engineering management [22,23], supplier selection [24] and other fields.

According to the existing kinds of literature on the grey incidence analysis method to deal with problems with panel data, they are related to five aspects shown as follows.

1. AHP, ANP, TOPSIS, and Fuzzy soft sets. Other

decision-making methods or models are combined with grey incidence analysis to construct some hybrid models [25–29]. These models only utilize the 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 [30]. However, the model produces the problem of deviation resulting from continuous form to discrete form;
3. By extending from the relationship of the curves 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 [25,31], but this extended model is affected by the presentation of the family of surfaces;
4. Based on the spatiotemporal characteristics of multi-index panel data, from the view of measuring the similarity between the objects from three aspects: “horizontal”, “incremental” and “variation” distances of the panel data, a grey matrix incidence analysis model for panel data is established [32]. 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, Li et al. (2015) constructed a cumulative generation sequence of the time series of all indices under the different objects. They subsequently 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 [33,34]. Kuh (1959) used inter-group estimates in the paper and found that the cross-sectional samples may be inconsistent with the time series samples [35]. Due to the two-dimensional nature of the panel data, the correlation between the correlation of the variables in the time series dimension and the cross-sectional dimension may be heterogeneous. This heterogeneity is an important reason for great changes caused by different effect settings in the estimation results [36,37]. Frühwirth-Schnatter (2011) considered continuous and non-continuous longitudinal time series data sets in panel data clustering [7]. Izakian et al. (2015) proposed a Dynamic Time Warping (DTW) distance for clustering panel data [38].

The study of panel data cross-section dimensions mainly includes dependence [39], heteroscedasticity [40], etc. Therefore, dealing with panel data, it is important to consider both cross-sectional and time dimensions.

Based on the above discussion and comparing the results of Table 1, there are two obvious shortcomings in the existing research. For panel data clustering based on statistical and intelligent algorithm methods, these clustering methods only work on large sample data. Because statistics and big data algorithms are modeled on massive data, it is difficult to get good results when the sample size is small. On the other hand, although data processing methods based on the grey theory can describe and solve the problem of small sample panel data or high-dimensional data to a certain extent, and enrich and expand the scope of the grey relational analysis model, they do not take the cross-sectional and time dimensions of the decision object into account to excavate the information contained in the multi-index panel data. At the same time, the number of clusters is generally given by humans, lacking scientific rationale, which often leads to discrepancies between clustering results and reality. Therefore, considering these two aspects, this paper attempts to mine the cross-section information and time information of the research object, and uses the grey absolute incidence analysis method to construct the grey object matrix absolute incidence clustering method based on multi-index and small sample panel data. On this basis, the hierarchical clustering algorithm is applied to overcome the drawbacks of traditional grey clustering to determine the clustering threshold, so that it is simpler and more reasonable to obtain the clustering number.

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 the grey object matrix absolute incidence analysis model is established. In Section 4, the case is used to cluster and assess the regional patent R&D efficiency. Finally, this study gives some conclusions that have been concluded from the novel grey object matrix incidence clustering model and cases 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  $U = \{1, 2, \dots, N\}$  and  $A = \{a_1, a_2, \dots, a_m\}$  stand for a set of objects and indices, respectively;  $V = \cup v_{ij}^t (i = 1, 2, \dots, n; j = 1, 2, \dots, m; t = 1, 2, \dots, T)$  expresses the value field of panel data, and  $v_{ij}^t$  shows the observed value of the object  $i$  at time  $t$ , with respect to index  $j$ . It can be converted to a two-dimensional table on the plane, as shown in Table 2.

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

$$x_i = \begin{bmatrix} x_{i1}(1) & x_{i2}(1) & \dots & x_{ij}(1) & \dots & x_{im}(1) \\ x_{i1}(2) & x_{i2}(2) & \dots & x_{ij}(2) & \dots & x_{im}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1}(t) & x_{i2}(t) & \dots & x_{ij}(t) & \dots & x_{im}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1}(T) & x_{i2}(T) & \dots & x_{ij}(T) & \dots & x_{im}(T) \end{bmatrix},$$

$i = 1, 2, \dots, N,$

**Table 1.** Comparison between clustering method proposed in this paper and existing methods.

Literature	Multidimensional information	Large sample	Small sample	Number of clusters
[5]	✓	✓		Artificial determination
[6]		✓		Artificial determination
[2,7]	✓	✓		Artificial determination
[9,10]		✓		Artificial determination
[11–15]		✓		Automatic determination
[32]	✓		✓	Artificial determination
[33,34]	✓		✓	Artificial determination
Grey clustering			✓	Artificial determination
This paper	✓		✓	Automatic determination

**Table 2.** Three-dimensional table description method of panel data.

Time	1					...	$t$					...	$T$				
	index					...	index					...	index				
Object	1	...	$j$	...	$m$	...	1	...	$j$	...	$m$	...	1	...	$j$	...	$m$
1	$v_{11}^1$	...	$v_{1j}^1$	...	$v_{1m}^1$	...	$v_{11}^t$	...	$v_{1j}^t$	...	$v_{1m}^t$	...	$v_{11}^T$	...	$v_{1j}^T$	...	$v_{1m}^T$
$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
$i$	$v_{i1}^1$	...	$v_{ij}^1$	...	$v_{im}^1$	...	$v_{i1}^t$	...	$v_{ij}^t$	...	$v_{im}^t$	...	$v_{i1}^T$	...	$v_{ij}^T$	...	$v_{im}^T$
$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
$n$	$v_{n1}^1$	...	$v_{nj}^1$	...	$v_{nm}^1$	...	$v_{n1}^t$	...	$v_{nj}^t$	...	$v_{nm}^t$	...	$v_{n1}^T$	...	$v_{nj}^T$	...	$v_{nm}^T$

$$\mu_i = \begin{bmatrix} \mu_{i1}(2) & \mu_{i2}(2) & \dots & \mu_{ij}(2) & \dots & \mu_{im}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_{i1}(t) & \mu_{i2}(t) & \dots & \mu_{ij}(t) & \dots & \mu_{im}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_{i1}(T) & \mu_{i2}(T) & \dots & \mu_{ij}(T) & \dots & \mu_{im}(T) \end{bmatrix},$$

$$i = 1, 2, \dots, N,$$

are called the development state matrix and development speed matrix under panel data of object  $i$ , respectively, where  $x_{ij}(t)$  and  $\mu_{ij}(t)$  stand for the absolute quantity and 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 objects with different indices in different stages takes on different states. Each sequence of panel data matrix represents the corresponding system behavior index. The columns of the matrix describe the cross-sectional characteristics of each object, 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-sectional information and time information. Therefore, for any two objects, the proximity degree is required to measure the comprehensive and integrated information of these two aspects. In order to accurately characterize the dimension characteristic of panel data, the dissimilarity degree is used to reverse the similarity among objects; the smaller the dissimilarity degree, the greater the similarity. The panel data contains information on two dimensions: section and time. The section information is reflected in the relative development state of different objects. The time information is reflected in the development speed level. Therefore, the relative development level and development speed can be used to measure the similarity of section information and time information. So that the grey object matrix absolute incidence analysis model is constructed.

For  $\forall i, s \in X, \forall a_j \in A, \forall t \in T$  and  $i \neq s$ , if  $\mathbf{x}_i, \mathbf{x}_s$  and  $\mu_i, \mu_s$  represent the development state matrix and

development speed matrix of an object and under panel data, respectively. By referencing the panel data mean and variance processing method from the pieces of literature [41,42], one can obtain as follows:

$$d_{isj}^1 = \frac{1}{T} \sum_{t=1}^T \Delta_{isj}^1, \quad (1)$$

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

which are the development state horizontal distance and development speed horizontal distance between object  $i$  and  $s$  over the multi-index panel data, respectively. The proof of Theorem 1 in Appendix A can be used to prove that the settings of Eqs. (1) and (2) are in accordance with the grey sequence operation, where,  $\Delta_{isj}^1 = |x_{ij}(t) - x_{sj}(t)|$ .

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

$$\xi_{isj} = \frac{\min_i \min_j d_{isj}^3 + \rho \max_i \max_j d_{isj}^3}{d_{isj}^3 + \rho \max_i \max_j d_{isj}^3}, \quad (3)$$

$$d_{is}^A = \xi_{is} = \sum_{j=1}^m w_j \xi_{isj}, \quad (4)$$

are called the grey incidence coefficient and incidence degree between object  $i$  and  $s$  over the index and the index set under panel data, respectively, where,  $d_{isj}^3 = \lambda_1 d_{isj}^1 + \lambda_2 d_{isj}^2$ ,  $0 \leq \lambda_1 \leq 1, 0 \leq \lambda_2 \leq 1, \lambda_1 + \lambda_2 = 1$ ,  $\rho \in (0, 1)$ ,  $w_j$  is the weight of the index  $j$  ( $j = 1, 2, \dots, m$ ) under the index set,  $0 \leq w_j \leq 1, \sum_{j=1}^m w_j = 1$ . Then, the value of  $\lambda_1$  and  $\lambda_2$  are determined based on the

specific problem. It can be obtained through expert investigation, analytic hierarchy, the entropy weight coefficient method, the variance contribution method and so on. The principle is to make the clustering results more in line with the actual situation, if  $\lambda_1 > \lambda_2$ , it shows that the cross-sectional features of the panel data affect the clustering results more than the timing characteristics, and vice versa.

Definition 3 is the grey object matrix absolute incidence analysis model.

**Theorem 1.** The grey object matrix absolute incidence analysis model has the following basic properties:

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

**Proof.** the specific proof process is detailed in Appendix A.

There are many methods for measuring the similarity between decision objects in existing literature, and the typical ones mainly involve the following:

1. Minkowski distance.

$$\text{dist}(\vec{X}, \vec{Y}) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}. \quad (5)$$

2. Cosine similarity [33,34].

$$\cos(\theta) = \frac{\vec{X}^T \vec{Y}}{|\vec{X}| \cdot |\vec{Y}|}. \quad (6)$$

3. K-L distance (relative entropy)<sup>[43]</sup>.

$$D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}. \quad (7)$$

The Minkowski distance does not take into account the distribution of individual components (expectation, variance, etc.), which may be different. It exaggerates the role of small variables and is unstable. The cosine similarity requires the integrity of the data for each dimension and is not sensitive to absolute quantities. The K-L distance is difficult to deal with panel data with small samples and unclear probability distribution. Therefore, the grey object matrix absolute incidence proposed in this paper can make up for any lack in the above methods, and fully excavate the information contained in the multi-index panel data.

### 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 classifies the observed objects by setting thresholds on the basis of grey incidence degree between any two observed objects [1].

**Definition 3.** It is assumed that  $d_{is}^A$  is the distance between the decision object  $i$ , and  $s$  on the multi-index set  $A$ . For  $\forall i, s \in U$ , if  $d_{is}^A = d_{si}^A$  is satisfied, then, the matrix:

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

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

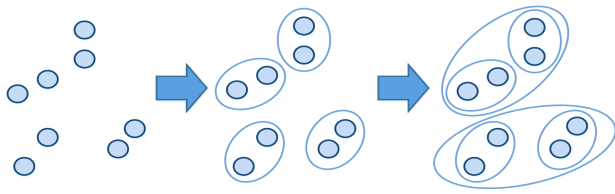
**Definition 4.** It is assumed that  $d$  stands for the incidence matrix of the objects over the multi-index. For  $\forall i, s \in U, A$ , the threshold value  $\alpha \in [0, 1]$ . If  $d_{is}^A \geq \alpha$ , then  $i$ , and  $s$  are the same class.

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

According to Definitions 3 and 4, one can find that grey incidence clustering is based on the determined clustering threshold value, which lacks a theoretical basis. However, the hierarchical clustering algorithm can overcome the shortcoming. The hierarchical clustering algorithm repeatedly calculates the similarity between two types of data points and combines the most similar data of two points in all data points to realize the final classification. In summary, the hierarchical clustering algorithm determines 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 larger the distance, the lower the similarity. Then, the final distance from the nearest two data points or categories are combined to generate a clustering tree, as shown in Figure 1.

Based on the above analysis, it is well known that the algorithm is a typical greedy algorithm. This algorithm firstly forms the local optimum and finally



**Figure 1.** Schematic diagram of the hierarchical clustering algorithm.

achieves the global optimum. Because the results based on the hierarchical clustering algorithm depend entirely on the choice of the parameters, there is no need to make a global objective function and a local minimum in the K mean method and how to select the initial point can be ignored. Meanwhile, the merge operation of the hierarchical clustering algorithm can get the final cluster results directly, 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 a grey incidence clustering threshold value, so that the cluster results from the hierarchical clustering algorithm are closer to the real situation. In view of this, on the basis of the grey incidence degree between any two objects, the cluster of grey incidences are substituted to cluster analysis by the method of the hierarchical clustering algorithm.

In the hierarchical clustering algorithm, there are many ways such as Euclidean, Mahalanobis, Cityblock, Minkowski, Cosine, and Hamming to calculate the 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. They cannot totally reflect and excavate the information contained in the panel data,

so, a novel distance matrix should be established to reflect 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, and vice versa. It can be seen that the grey incidence degree between the two clustering objects is an inverse relationship with distance, and then the reciprocal of the grey incidence degree can be used to express the distance. Because the distance between an 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 the grey incidence degree to be reciprocal and then reducing 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:

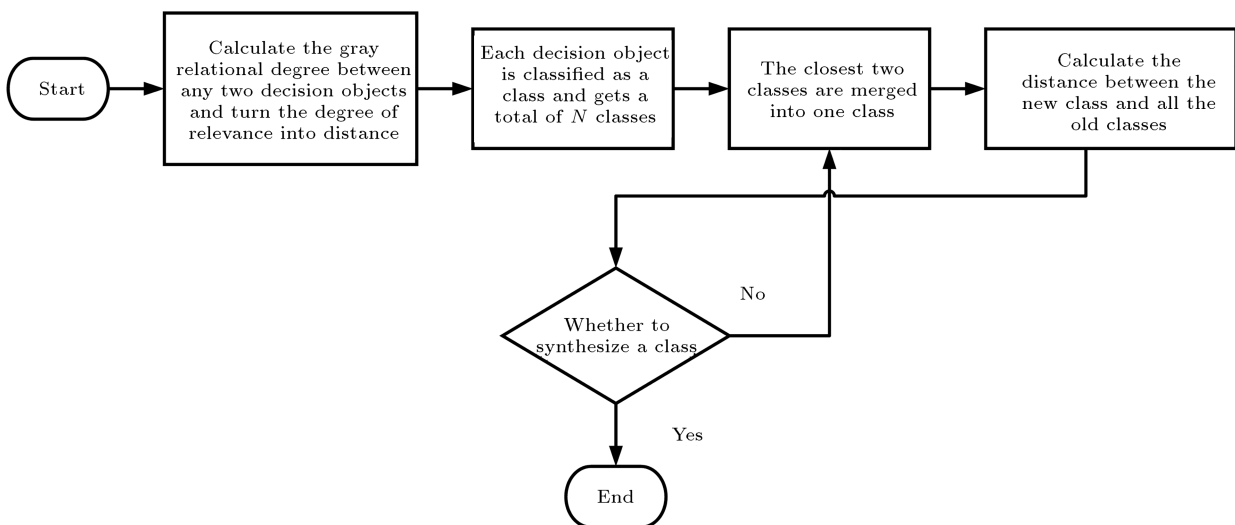
$$d' = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1s} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2s} & \dots & d_{2n} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{i1} & d_{i2} & \dots & d_{is} & \dots & d_{in} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ d_{n1} & d_{n2}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

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

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

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

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



**Figure 2.** Algorithm flowchart of the hierarchical clustering.

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

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

**Step 5:** 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.

#### 4. Case applications

The patent is an important index to measure the technological innovation capability of regional high-tech industry. It is necessary for gaining scientific and technological competitiveness and occupies the high ground of science and technology. After more than 30 years of reform and opening-up, China's scientific and technological innovation activities have greatly increased. According to the statistical data 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, it can be seen that technology R&D in China is still at 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 China country has not yet appeared, and there exists a variety of waste in the factors of scientific and technological development. 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 a strategic deployment of the transformation of economic development mode and the construction of an innovative country. By evaluating and analyzing the development status of regional patent R&D in China, one can 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 the “China Statistical Yearbook” (2012–2016), the “China Statistical Yearbook of Science and Technology” (2012–2016) and CNIPA (China National Intellectual Property Administration), one takes the data of the efficiency of patent R&D in 30 provinces or regions (provinces, municipalities and autonomous regions of China), (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 relevant research [44], 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, the macro environment can be measured and characterized by the per capita GDP and the number of people with a college degree or above. The R&D input is mainly measured by two indices of internal expenditure on R&D and the 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. The output capability is generally measured by the number of patent applications accepted. The output quality is measured by two indices of the number of invention patent applications accepted and the technology market turnover, and the industrialization level is measured by high tech industry profits and export volume. The specific evaluation indices are shown in Table 3, and the standard and well-developed dataset of panel data are shown in Table 4.

According to preliminary research and the evaluation information from invited experts, the weight of each index can be determined, respectively, and it is shown as follows:

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

Based on the adaptive algorithm from the literature [45], the weights of cross-sectional information and time dimension information can be calculated as follows:

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

Based on the proposed model in this paper, the grey incidence degree among regions and the incidence matrix, the index system of the patent R&D efficiency can be calculated as shown in Table B.1 (Appendix B). Then, the hierarchical clustering distance matrix between any two regions can be determined using the designed method in this paper, as shown in Table B.2 (Appendix B). 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, as shown in Figure 3.

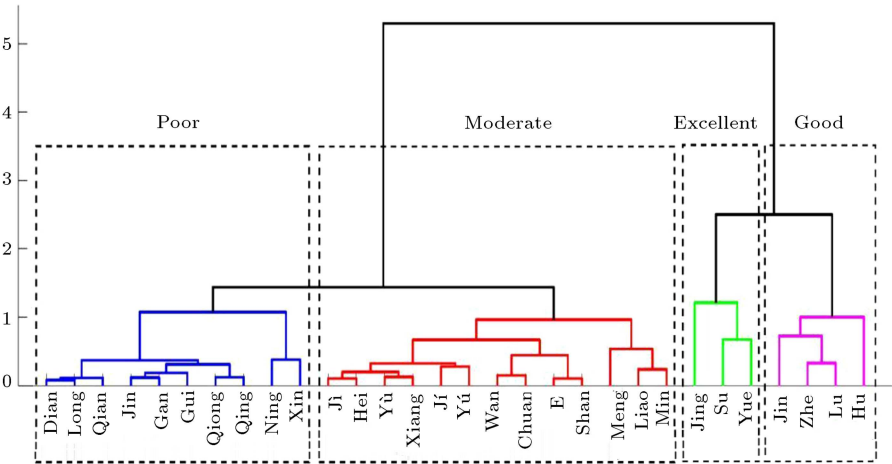
From Figure 3, we can see that 30 provinces and cities in China mainland can be clustered into the four categories of excellent, good, medium and

**Table 3.** The index system of patent Research and Development (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
		The full-time equivalent of R&D personnel
	Output capability	The number of patent applications accepted
Output	Output quality	The number of invention patent Application accepted
		Technology market turnover
	Industrialization level	High technology industry profits
		Export volume

**Table 4.** Three-dimensional table description method of panel data in China's provinces and cities.

Time	1					...	$t$					...	5				
	Index					...	Index					...	Index				
Object	1	...	$j$	...	9	...	1	...	$j$	...	9	...	1	...	$j$	...	9
1	$v_{11}^1$	...	$v_{1j}^1$	...	$v_{19}^1$	...	$v_{11}^t$	...	$v_{1j}^t$	...	$v_{19}^t$	...	$v_{11}^T$	...	$v_{1j}^T$	...	$v_{19}^T$
$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
$i$	$v_{i1}^1$	...	$v_{ij}^1$	...	$v_{i9}^1$	...	$v_{i1}^t$	...	$v_{ij}^t$	...	$v_{i9}^t$	...	$v_{i1}^T$	...	$v_{ij}^T$	...	$v_{i9}^T$
$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
30	$v_{301}^1$	...	$v_{30j}^1$	...	$v_{309}^1$	...	$v_{301}^t$	...	$v_{30j}^t$	...	$v_{309}^t$	...	$v_{301}^5$	...	$v_{30j}^5$	...	$v_{309}^5$



**Figure 3.** Tree cluster diagram of patent Research and Development (R&D) efficiency of provinces and cities in mainland China.

poor. For the category of excellent, the efficiency of patent R&D is relatively high, and the patent R&D is at the mature stage, which is embodied in the high input and output of patent R&D. These provinces and cities include Beijing (Jing), Guangdong (Yue) and

Jiangsu (Su). Those provinces and cities with good efficiency of patent R&D are Tianjin (Jin), Zhejiang (Zhe), Shandong (Lu) and Shanghai (Hu), respectively. Those provinces and cities with general efficiency of patent R&D (at medium level) cover Hebei (Jí), Hei-



longjiang (Hei), Henan (Yú), Hunan (Xiang), Jilin (Jí), Chongqing (Yú), Anhui (Wan), Sichuan (Chuan), Hubei (E), Shaanxi (Shan), Inner Mongolia (Meng), Liaoning (Liao), Fujian (Min). Those provinces and cities with low efficiency of patent R&D (at poor level), include Yunnan (Dian), Gansu (Long), Guizhou (Qian), Shanxi (Jín), Jiangxi (Gan), Guangxi (Gui), Hainan (Qiong), Qinghai (Qing), Ningxia (Ning), Xinjiang (Xin). From the clustering results, one can see that the efficiency of the regional patent R&D is imbalanced, and the efficiency of the eastern provinces and cities are significantly better than those of the central and western provinces and cities. The efficiency of the central provinces and cities are better than that of western provinces and cities, while the efficiency of the western provinces and cities are the worst. Furthermore, there are few provinces and cities with higher efficiency, and many more regions are general and poor, which indicates that the scale effect of China's patent R&D has not yet been highlighted, and the efficiency of patent R&D is not as high as a whole.

For those regions with relatively high efficiency, they rely on strong scientific research institutions, government policies and adequate financial support. Thus, they develop early science and technology, so that their market internationalization, information level and transformation rate of patent achievements are 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, and take on responsibility to occupy the best of international science and technology in order to develop patent R&D and realize the rise of China's manufacturing industry. Some measures, such as making full use of international resources for innovation to accelerate the internationalization level and expand the vision of globalization, adjusting the patent structure to reduce development costs, increasing the number of high-tech industry patents to break the international patent barriers, and participating in international competition should be taken.

For these regions with relatively good efficiency of patent R&D, their economy develops well, their science and technology power is strong, and thus their patent input is at 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 make effective use of the scientific and technological market as the first direction in patent R&D, broaden the scope of patent R&D and form a reasonable patent structure.

Realising the inherent advantages of patent R&D, these provinces and cities must continue to maintain and consolidate a leading position, and strive to seek a path of internationalization and globalization development.

For those 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 mainly 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, and 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 will not be 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, local governments should 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 invalid patents of , low value and low efficiency. Based on the clustering results, the authors management insights are as follows:

1. The category of "Excellent". The patent R&D efficiency of such provinces and cities is unlikely to have a greater breakthrough in the short term. The future development direction of these provinces and cities should be tuned into the international market. It is necessary to take the initiative to assume the height of international technology. They should make full use of international innovation resources, accelerate the level of internationalization, expand the horizon of globalization; adjust the patent R&D structure to reduce R&D costs, increase the number of patents in high-tech industries, break through international patent barriers, and participate in international competition;
2. The category of "Good". To improve the level of patent R&D in such provinces, it is necessary to improve the utilization efficiency of innovative resources. The R&D of patents should be based on the effective demand of the technology market. The scope of patent R&D should be broadened and a reasonable patent structure formed. China must continue to maintain and consolidate a leading position and seek the path of international and global development;
3. The category of "Moderate". Such areas must maintain the healthy, stable, green and sustainable development of patent R&D. In terms of

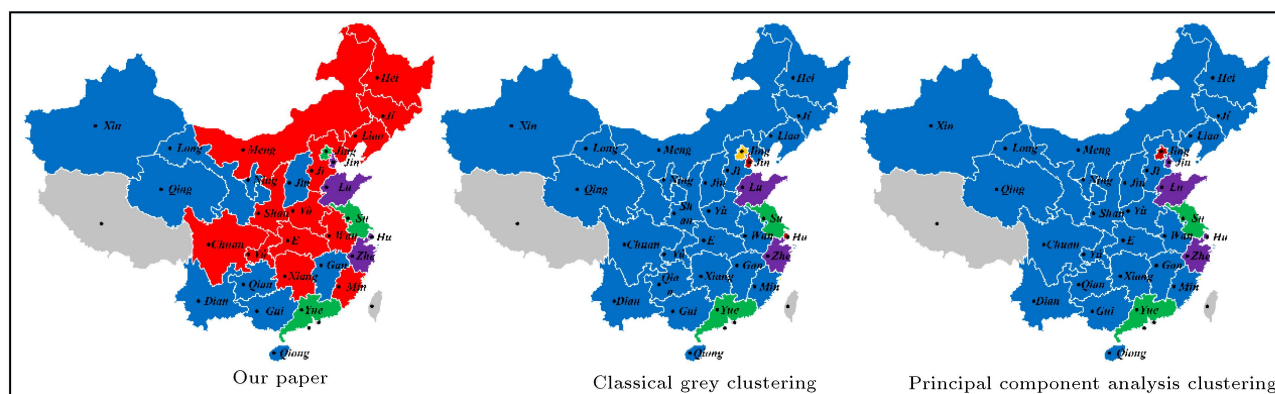


Figure 4. Cluster maps based on different methods.

investment, sources should be opened and expenditures reduced. The government should guide the direction, relying on the existing resources and advantages of the region. It should increase investment in patent R&D funds, talents and policies, and formulate reasonable patent R&D development plans according to local conditions. In terms of output, China must adhere to the market-oriented, improve the scientific and technological content and quality of patent R&D, and avoid invalid, low-value, low-efficiency patents;

4. The category of “Poor”. The R&D of patent in such provinces and cities is in its infancy. To improve the patent R&D of such provinces, the government should strengthen policy support and capital investment support, and introduce high-quality, high-tech enterprises and research institutions. At the same time, the government needs to improve the science and technology innovation infrastructure to improve the level of economic development in the region; strengthen cooperation with science and technology projects in the central and eastern regions, and use regional resources to develop an over the border high-tech industry trade.

In order to better demonstrate the advantages and disadvantages of the proposed method, the clustering results of this paper were compared with classical grey clustering and principal component analysis clustering and the clustering results were presented using maps, as shown in Figure 4.

It can be seen from the clustering map that compared with the method proposed in this paper, classical grey clustering and principal component analysis clustering do not give ideal divisions to the central and western provinces, and they are collectively grouped into one category. The reason may be that the classical grey clustering and principal component analysis clustering cannot simultaneously mine the cross-section dimension and time dimension information of the panel data, resulting in incomplete characteristic extraction

of decision objects. Although the absolute amount of patent input and output in the central and western provinces is relatively low, from the observation period, its patent R&D efficiency is increasing, which has a large area for improvement. In addition, the determination of the clustering threshold is too subjective and lacks a reasonable scientific basis, which will lead to deviation of clustering results to some extent.

Through the above analysis, the proposed model in this paper can make full use of existing information, and mine related information and rules contained in the panel data, and properly deal with the problem with the cluster evaluation of panel data. Thus, the results calculated by the model are consistent with the actual situation.

## 5. Conclusions

In order to fully grasp the information contained in the panel data, under the guidance of the idea and method of the grey incidence analysis model and hierarchical clustering algorithm, this study establishes a novel grey object matrix absolute incidence clustering model based on multi-index panel data, which enriches and develops the application area and range of the 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 a hysteresis effect and periodic fluctuation. Considering this, the proposed model still needs further research.

## Acknowledgments

This work is partially funded by the National Natural Science Foundation of China (71503103); the Humanities and Social Sciences of Education Ministry

(17YJC640233); the Natural Science Foundation of Jiangsu Province (BK20150157); the Social Science Foundation of Jiangsu Province (14GLC008); the Soft Science Foundation of Jiangsu Province (BR2018005); the Jiangsu Province University Philosophy and Social Sciences for Key Research Program (2017ZDIXM034); the Fundamental Research Funds for the Central Universities (2019JDZD06). Even so, this work does not involve any conflict of interest.

## References

- Nie, G., Chen, Y.B., Zhang, L.L., and Guo, Y. "Credit card customer analysis based on panel data clustering", *Procedia Computer Science*, **1**(1), pp. 2489–2497 (2010).
- Ando, T. and Bai, J.S. "Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures", *J. Am. Stat. Assoc.*, **112**(519), pp. 1182–1198 (2017).
- Wang, M. and Feng, C. "Decomposition of energy-related CO<sub>2</sub> emissions in China: an empirical analysis based on provincial panel data of three sectors", *Appl. Energy*, **190**, pp. 772–787 (2017).
- Munshi, A.A. and Yasser, A.R.M. "Photovoltaic power pattern clustering based on conventional and swarm clustering methods", *Sol. Energy*, **124**, pp. 39–56 (2016).
- Bonzo, D.C. and Hermosilla, A.Y. "Clustering panel data via perturbed adaptive simulated annealing and genetic algorithms", *Adv. Complex Syst.*, **5**(04), pp. 339–360 (2002).
- Moody, C.E. and Marvell, T.B. "Clustering and standard error bias in fixed effects panel data regressions", *J. Quant. Criminol.*, **26**(2), pp. 347–369 (2020).
- Frühwirth-Schnatter, S. "Panel data analysis: a survey on model-based clustering of time series", *Adv. Data Anal. Classif.*, **5**(4), pp. 251–280 (2011).
- Jacques, J. and Biernacki, C. "Model-based co-clustering for ordinal data", *Comput. Stat. Data Anal.*, **123**, pp. 101–115 (2018).
- Alaeddini, A., Ghazanfari, M., and Nayeri, M.A. "A hybrid fuzzy-statistical clustering approach for estimating the time of changes in fixed and variable sampling control charts", *Inf. Sci.*, **179**(11), pp. 1769–1784 (2009).
- Ghazanfari, M., Alaeddini, A., Niaki, S.T.A., and Aryanezhad, M.B. "A clustering approach to identify the time of a step change in Shewhart control charts", *Qual. Reliab. Eng. Int.*, **24**(7), pp. 765–778 (2008).
- Aßmann, C. and Boysen-Hogrefe, J. "A Bayesian approach to model-based clustering for binary panel probit models", *Comput. Stat. Data Anal.*, **55**(1), pp. 261–279 (2011).
- Yamazaki, K. "Effects of additional data on Bayesian clustering", *Neural Netw.*, **94**, pp. 86–95 (2017).
- Banharnsakun, A. "A MapReduce-based artificial bee colony for large-scale data clustering", *Pattern Recognit. Lett.*, **93**, pp. 78–84 (2017).
- Alizadeh, S., Ghazanfari, M., and Fathian, M. "Using data mining for learning and clustering FCM", *International Journal of Computational Intelligence*, **4**(2), pp. 118–125 (2008).
- Azzalini, A. and Menardi, G. "Density-based clustering with non-continuous data", *Comput. Stat.*, **31**(2), pp. 771–798 (2016).
- Malsiner-Walli, G., Frühwirth-Schnatter, S., and Grün, B. "Model-based clustering based on sparse finite Gaussian mixtures", *Stat. Comput.*, **26**(1–2), pp. 303–324 (2016).
- Peters, G. "Rough clustering utilizing the principle of indifference", *Inf. Sci.*, **277**, pp. 358–374 (2014).
- Saltos, R. and Weber, R. "A Rough-Fuzzy approach for support vector clustering", *Inf. Sci.*, **339**, pp. 353–368 (2016).
- Zhou, J., Lai, Z., Miao, D., Gao, C., and Yue, X. "Multigranulation rough-fuzzy clustering based on shadowed sets", *Inf. Sci.*, **507**, pp. 553–573 (2020).
- Liu, S.F., Yang, Y.J., and Forrest, J., *Grey Data Analysis*, Springer, Singapore (2017).
- Wu, L.F., Liu, S.F., Yao, L.G., and Yu, L. "Fractional order grey relational analysis and its application", *Sci. Iran.*, **22**(3), p. 1171 (2015).
- Zavadskas, E.K., Antucheviciene, J., Turskis, Z., and Adeli, H. "Hybrid multiple-criteria decision-making methods: A review of applications in engineering", *Sci. Iran.*, **23**(1), pp. 1–20 (2016).
- Kolahan, F. and Moghaddam, M.A. "The use of Taguchi method with grey relational analysis to optimize the EDM process parameters with multiple quality characteristics", *Sci. Iran.*, **22**(2), pp. 530–538 (2015).
- Lee, Y.L., Tsai, F.C., Liu, S.F., and Hsu, Y.C. "A scale development of industrial designer ability index through quality function deployment and grey relational analysis methods", *Adv. Mech. Eng.*, **8**(12), pp. 1–11 (2016).
- Zhu, J.J., Zhang, S.T., Chen, Y., and Zhang, L. "A hierarchical clustering approach based on three-dimensional gray relational analysis for clustering a large group of decision makers with double information", *Group Decis. Negot.*, **25**(2), pp. 325–354 (2016).
- Hashemi, S.H., Karimi, A., and Tavana, M. "An integrated green supplier selection approach with analytic network process and improved Grey relational analysis", *Int. J. Prod. Econ.*, **159**, pp. 178–191 (2015).
- Rajesh, R. and Ravi, V. "Supplier selection in resilient supply chains: a grey relational analysis approach", *J. Clean Prod.*, **86**, pp. 343–359 (2015).
- Wang, P., Zhu, Z., and Wang, Y. "A novel hybrid MCDM model combining the SAW, TOPSIS and GRA methods based on experimental design", *Inf. Sci.*, **345**, pp. 27–45 (2016).

29. Xie, N.M., Han, Y., and Li, Z. "A novel approach to fuzzy soft sets in decision making based on grey relational analysis and MYCIN certainty factor", *Int. J. Comput. Intell. Syst.*, **8**(5), pp. 959–976 (2015).
30. Wu, L.F. and Liu, S.F. "Panel data clustering method based on grey convex relation and its application", *Control and Decision*, **28**(7), pp. 1033–1037 (2013).
31. Wu, L.F., Liu, S.F., Yao, L.G., and Yan, S.L. "Grey convex relational degree and its application to evaluate regional economic sustainability", *Sci. Iran.*, **20**(1), pp. 44–49 (2013).
32. Wang, Y.H. and Zuo, W.C. "Research on multi-index gray clustering approach based on adaptive weight for panel data", *J. Grey Syst.*, **30**(3), pp. 13–32 (2018).
33. Li, X.M., Dang, Y.G., and Wang, J.J. "Grey relational clustering model for panel data clustering on indexes and its application", *Control and Decision*, **30**(8), pp. 1447–1452 (2015).
34. Li, X.M., Hipel, K.W., and Dang, Y.G. "An improved grey relational analysis approach for panel data clustering", *Expert Syst. Appl.*, **42**(23), pp. 9105–9116 (2015).
35. Kuh, E. "The validity of cross-sectionally estimated behavior equations in time series applications", *Econometrica*, **27**(2), pp. 197–214 (1959).
36. Bell, A. and Jones, K. "Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data", *Political Science Research and Methods*, **3**(1), pp. 133–153 (2015).
37. Pesaran, M.H., *Time Series and Panel Data Econometrics*, Oxford University Press, UK (2015).
38. Izakian, H., Pedrycz, W., and Jamal, I. "Fuzzy clustering of time series data using dynamic time warping distance", *Eng. Appl. Artif. Intell.*, **39**, pp. 235–244 (2015).
39. Pesaran, M.H. "Testing weak cross-sectional dependence in large panels", *Econom. Rev.*, **34**(6–10), pp. 1089–1117 (2015).
40. Hayakawa, K. and Pesaran, M.H. "Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroscedasticity", *J. Econom.*, **188**(1), pp. 111–134 (2015).
41. Hsiao, C., *Analysis of Panel Data*, Cambridge university press, UK (2014).
42. Liu, Y., Quan, B.T., and Du, J.L. "Object-oriented spatiotemporal characteristic attribute differences clustering approach for panel data and its application", *J. Grey Syst.*, **30**(2), pp. 84–96 (2018).
43. Zarinbal, M., Zarandi, M.H.F., and Turksen, I.B. "Relative entropy fuzzy c-means clustering", *Inf. Sci.*, **260**, pp. 74–97 (2014).
44. Zhao, H.F., Li, W.W., and Xu, S. "The efficiency of patent innovation in the different regions of China",

*Chinese Journal of Management Science*, **16**(SI), pp. 387–392 (2008).

45. Li, Y.G., Dai, Y., and He, X.Q. "Clustering method for Panel data base on adaption weighting", *Systems Engineering-Theory & Practice*, **33**(2), pp. 388–395 (2013).

## Appendix A

### Proof of Theorem 1:

1. **Normality.** Known by Definitions 1 and 2,  $0 < d_{isj}^1, d_{isj}^2$ , and  $0 \leq \lambda_1 \leq 1, 0 \leq \lambda_2 \leq 1, \lambda_1 + \lambda_2 = 1$ , then  $0 < d_{isj}^3 = \lambda_1 d_{isj}^1 + \lambda_2 d_{isj}^2$ . If  $\min_i \min_j d_{isj}^3 = d_{isj}^3$ , then  $\xi_{isj} = 1$ ; if  $\min_i \min_j d_{isj}^3 \neq d_{isj}^3$ , then  $d_{isj}^3 > \min_i \min_j d_{isj}^3$ , thus we can get  $\min_i \min_j d_{isj}^3 + \rho \max_i \max_j d_{isj}^3 < d_{isj}^3 + \rho \max_i \max_j d_{isj}^3$ , then  $\xi_{isj} < 1$ . Obviously,  $0 < \xi_{isj}$ . Therefore,  $0 < \xi_{isj} \leq 1$ . In addition,  $0 \leq w_j \leq 1, \sum_{j=1}^m w_j = 1$ , that is  $0 < \xi_{is} \leq 1$ , the grey incidence degree satisfies the normality.
2. **Symmetry.** If  $U = \{i, s\}$ , then  $|x_{ij}(t) - x_{sj}(t)| = |x_{sj}(t) - x_{ij}(t)|$ ,  $[\mu_{ij}(t) - \mu_{sj}(t)]^2 = [\mu_{sj}(t) - \mu_{ij}(t)]^2$ . In view of this,  $d_{isj}^1 = d_{sij}^1$ ,  $d_{isj}^2 = d_{sij}^2$ , that is  $d_{isj}^3 = d_{sij}^3$ . Obviously, one can get  $\min_i \min_j d_{isj}^3 + \rho \max_i \max_j d_{isj}^3 = \min_s \min_j d_{sij}^3 + \rho \max_s \max_j d_{sij}^3 = d_{isj}^3 + \rho \max_s \max_j d_{sij}^3$ .

Thus, when  $\xi_{isj} = \xi_{sij}$ , that is  $\xi_{is} = \xi_{si}$ , the grey incidence degree satisfies the symmetry.

Since the dimensions of different indicators in the original sample matrix are different, the different dimensions of the indicators will have an impact on the modeling quality and the system analysis results. Therefore, the original data matrix needs to be processed to eliminate the influence of the dimension. For the object behavior matrix in which the multiplication transformation occurs, the correlation coefficient of each point is unchanged after the elements are standardized, so the order-multiplicative transformation consistency and the number-multiplication transformation order-satisfaction are satisfied. Therefore, the grey incidence degree satisfies the parallelism and consistency.

## Appendix B

The index system of the patent R&D efficiency can be calculated as shown in Table B.1. Also, the hierarchical clustering distance matrix between any two regions can be determined using the designed method in this paper, as shown in Table B.2.





## Biographies

**Yong Liu** was born in 1985, and obtained his BS, MS, and PhD degrees in Mechanical Engineering, and System Engineering and Management Science and Engineering from NUAA. Currently, he serves the School of Business at Jiangnan University. He has published over 30 research papers, and is mainly engaged in conflict analysis, soft computing, and grey system.

**Jun-liang Du** was born in 1994, and obtained his

BS degree from the School of Business at Jiangnan University, where he is currently a graduate student. He is mainly engaged in the research of conflict analysis and grey systems.

**Jiao Li** was born in 1994. She obtained her BS degree in Logistics Management from Yangtze University and is currently a graduate student at the School of Business in Jiangnan University. She is mainly engaged in the research of conflict coordination and incentive mechanism of supply chain channels and grey systems.