

Research Article

An Improved Grey Clustering Model with Multiattribute Spatial-Temporal Feature for Panel Data and Its Application

Jian Li  and Wan-ming Chen 

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Correspondence should be addressed to Wan-ming Chen; cwmnuaa@163.com

Received 26 August 2019; Revised 17 December 2019; Accepted 4 January 2020; Published 30 January 2020

Academic Editor: Kai Zhang

Copyright © 2020 Jian Li and Wan-ming Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the complexity and uncertainty of the objective world and the limitation of cognition, it is difficult to extract the information and rules contained in the panel data effectively based on the traditional panel data clustering method. Given this, considering that the absolute amount level, increasing amount level, and volatility level are the main indicators to represent the spatial-temporal feature of the panel data, a novel grey clustering model with the multiattribute spatial-temporal feature of panel data is established, and then it is applied in the regional high-tech industrialization in China. The results show that the proposed model can make full use of the spatial-temporal feature information of the panel data, identify the problems existing in the clustering objects, and make the clustering results more objective and practical.

1. Introduction

Panel data, which combine the feature of time series and cross-sectional data, describe the systematicness and dynamics of the research object in a more detailed way and provide the information and data feature of the research object in a more comprehensive way. It has an important application background and value and has become a hot issue of academic research.

Panel data analysis mainly adopts econometric methods and approaches, focusing on the exploring of panel data modeling and parameter estimation and applying it to the local market effect of production and service trade [1] and other fields. However, due to the complexity of the objective world and uncertainty and the limitation of human cognition, it is difficult to effectively mine the data only by traditional panel data analysis methods and affects the quality and efficiency of decision-making. However, effective clustering results are often decisive in data mining [2]. At present, the research on panel data clustering analysis is still in its infancy, and the traditional clustering analysis method is mainly suitable for cross-sectional data clustering. It is difficult to describe and extract the information and rules contained in panel data. In order to solve the clustering problem of panel data, Bonzo and Hermosilla [3] first pioneered the application of multivariate statistical analysis

theory in panel data analysis and proposed the use of probability theory and random heuristics to optimize the panel data for random clustering analysis. Subsequently, some scholars from different perspectives and perspectives study the model building, statistical selection, and reconstruction of panel data. Specifically, the research focuses on two kinds of problems in panel data clustering: how to measure the similarity of data objects (distance and similarity coefficient) and what clustering method to adopt. Luo et al. [4] proposed to take the polygon area of two adjacent points between the selected scheme and the ideal scheme as the correlation coefficient and constructed the grey correlation degree formula. Cui [5] analyzed the geometric feature similarity of two groups of sequences in the form of spatial vectors and established the calculation formula of a grey relational degree by using Euclidean distance of vectors. Cui and Liu [6] measured the correlation degree by the proximity of the development velocity index and the growth velocity index between the correlation factor matrix and the system characteristic behavior matrix. For the similarity of data objects, there are two main methods to measure the similarity of data objects: distance and similarity coefficient. Among them, Zhu and Chen [7] used multivariate statistical methods to describe and analyze the feature of panel data, designed the similarity distance between objects, and constructed a panel data clustering method based on the similarity of panel data. According to the principle of cluster

analysis, Zheng [8] constructed the distance function and deviation square sum function of multi-index panel data based on the analysis of panel data format and digital feature and introduced the clustering process of multi-index panel data. Xiao et al. [9] used principal component analysis to conduct dimensionality reduction processing on panel data, constructed similarity index of sequence matrix of the comprehensive evaluation function, and then conducted clustering analysis on panel data. Ren [10] defined the comprehensive Euclidean distance by using the horizontal quantity, increment, and increment change rate of the multi-index panel data, measured the similarity by using the comprehensive form of matching measure and Euclidean distance, and proposed a multi-index panel data fusion clustering method based on shape feature similarity. Facing the multi-dimensional feature of panel data, considering the geometric similarity of indexes, Zhang and Liu [11] constructed an extended grey relational analysis model based on the matrix and then used it to study the clustering problem of panel data. And Wu and Liu [12] used the Hessian matrix to define convexity, then used the convexity of data to characterize similarity between samples, proposed a three-dimensional grey convex correlation degree, and established a threshold-based clustering method. However, the previous literature did not consider the problem of distance measurement among samples in three-dimensional space. By replacing the distance in the fuzzy *c*-means algorithm with the surface similarity index, Ma et al. [13] proposed a similarity index calculation method that considers both the distance and set similarity of individuals in three-dimensional space. Aiming at the limitation of similarity measure and distance in the literature, Li and He [14] proposed a comprehensive similarity distance based on the spatiotemporal feature of panel data, taking into comprehensive consideration the features of “absolute index,” “incremental index,” and “temporal fluctuation” of panel data.

According to the existing researches, the clustering methods may be divided into three kinds of partition-based, density-based, hierarchy-based, model-based, and grid-based. Specifically, there are fuzzy *c*-means clustering [14], Ward clustering [14, 15], regression clustering [16–18], correlation matrix hierarchical clustering [19–21], numerical analysis clustering [10, 22, 23], and grey relational clustering [11, 12, 24–28]. On basis of “absolute index,” “incremental index,” and “fluctuation index,” Li et al. [15] reconstructed the distance function and Ward clustering algorithm of similarity measure of panel data and proposed a panel data-adaptive weight clustering algorithm. Nelsen and Dean [16] constructed an adaptive semi-parametric NHPP model from the statistical point of view and used it to analyze the mixed nonhomogeneous Poisson process of longitudinal counting panel data with weak heterogeneity type. Aiming at the binary feature of panel data, Assmann and Hogrefe [17] adopted the Bayesian reasoning method to estimate the model parameters and then proposed the panel data clustering method. Juárez and Steel [18] constructed a new panel data clustering method based on the dynamic feature, equilibrium level, and covariance of the data using the autoregressive model of *T* distribution with the skewed heavy tail. Liu et al. [19] introduced a correlation matrix hierarchical clustering method to extract multiple correlation factors from resting fMRI data, and it can capture spontaneous fMRI signals from

anesthetized mice. Stanila et al. [20] firstly divided the EU countries into two groups by the hierarchical clustering method, then used panel data to estimate the impact factors of each group on employment rate, and finally constructed the employment rate prediction model of EU member states. In order to accurately measure the parameters of the panel clustering model, Wang et al. [22] proposed the Stein form estimation for the linear panel data model and constructed the asymptotic distribution of the Stein form estimation. The results show that within the local asymptotic framework, the asymptotic risk estimated by Stein is strictly smaller than that estimated by the fixed effect. Considering that the sample is affected by other viewpoints or clustering behavior and the panel data are nonlinear, George et al. [23] proposed a nonlinear panel data model which can produce endogenous “strong” and “weak” cross-sectional dependence and used the relevant approximation theory to estimate and infer the model, clustering the objects. Falletta and Sauer [29] started from boundary integral to discretization, considered numerical solution of wave equation in two-dimensional space, and generated loosely approximated sequences by panel clustering and boundary integral operators to improve the validity and accuracy of panel clustering; Li et al. [24] constructed the cumulative generating sequence of time series with different targets. The dynamic trend of the original sequence was characterized by the average generating rate of the generating sequence, and then a mean-AGRA grey incidence correlation clustering algorithm is proposed under panel data. According to the spatiotemporal characteristics of panel data, from the three dimensions of absolute amount level, increasing amount level, and volatility level, Liu et al. [30] defined the conception of the comprehensive distance between decision objects, then proposed a grey incidence analysis clustering approach for panel data, and discussed its computing mechanism of threshold value by exploiting the thought and method of three-way decisions.

According to the above discussion and analysis of the clustering analysis of panel data, the existing methods mostly use traditional econometric models and clustering methods to deal with clustering problems according to the similarity among objects. However, due to the complexity and uncertainty of the objective world and panel data, it is difficult to extract the information and rules contained in panel data effectively based on traditional panel data clustering methods. In view of this, considering that the grey clustering method has unique advantages in dealing with “small sample, poor information” clustering problem, according to the spatial-temporal feature of panel data, using grey relational analysis and grey target decision-making method, a grey clustering model with multiattribute spatial-temporal feature is constructed, in order to expand the application field of the grey decision-making method and provide method support for panel data clustering problem.

2. A Grey Clustering Decision-Making Model

The grey clustering method mainly uses the grey correlation or the whitening weight function method to classify the evaluation index or the evaluation object according to the actual needs. According to the classification method of the

grey system, the grey clustering method mainly includes grey correlation clustering and grey whitening weight function clustering. Since Professor Deng found the theory of grey system, the grey clustering method has become an important research topic because of its unique advantages in dealing with the clustering problem of “small sample, poor information”. Grey clustering methods mainly include grey correlation clustering, grey fixed weight clustering, grey variable weight clustering, grey optimal clustering, grey trend correlation clustering, grey entropy weight clustering, hybrid grey clustering, and optimization and extension of grey clustering evaluation model. These grey clustering models can only solve the clustering problem of cross-sectional data, and it is difficult to deal with the economic and social problems with panel data effectively. In view of this, according to the spatial-temporal feature of panel data, the grey clustering model of the multiattribute spatial-temporal feature is constructed by using the grey decision method.

There is a multi-index panel data decision information system $S = \{U, A, V, C\}$, among, $U = \{1, 2, \dots, N\}$ representing the clustering object collection; $A = \{a_1, a_2, \dots, a_m\}$ is the index collection; $V = \cup v_{ij}^t (i = 1, 2, \dots, n; j = 1, 2, \dots, m; t = 1, 2, \dots, T)$ is the panel data

range, where v_{ij}^t is the observed value of the index of the clustering object i about index j at the time t ; and $C = \{c_1, c_2, \dots, c_l, \dots, c_q\} (l = 1, 2, \dots, q)$ represents the spatial-temporal feature set of objects.

For multi-index panel data, it can not only describe the spatial feature of the object at a certain point but also reflect the dynamic evolution of the object in a certain time domain. Multi-index panel data can reflect three aspects of the information of the described object: (1) the absolute level of the development of indicators in a certain period; (2) the incremental level of the indicators of a specific individual with time; and (3) the fluctuation level of the indicators of a specific individual. The absolute level, incremental level, and fluctuation level are substantially the main indicators to characterize the spatial-temporal feature of the panel data. Accordingly, the absolute level, incremental level, and fluctuation level of the object can be defined according to the spatial-temporal feature of the panel data.

Definition 1. Let $x_{ij}(t)$ be the dimensionless measure value of the index $j (j = 1, 2, \dots, m)$ of the time $t (t = 1, 2, \dots, T)$ object $i (i = 1, 2, \dots, N)$, for $i = 1, 2, \dots, n; j = 1, 2, \dots, m$; and $t = 1, 2, \dots, T$, if $\mu_{ij}(t) = \Delta x_{ij}(t) / x_{ij}(t)$ and $\eta_i(t) = \bar{x}_i(t) / S_i(t)$, separately called matrices:

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

$$\mu_i(t) = \begin{bmatrix} \mu_{i1}(2) & \mu_{i2}(2) & \dots & \mu_{ij}(2) & \dots & \mu_{im}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_{i1}(t) & \mu_{i2}(t) & \dots & \mu_{ij}(t) & \dots & \mu_{im}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_{i1}(T) & \mu_{i2}(T) & \dots & \mu_{ij}(T) & \dots & \mu_{im}(T) \end{bmatrix}, \quad i = 1, 2, \dots, N. \quad (2)$$

$$\eta_i(t) = [\eta_i(1), \dots, \eta_i(t), \dots, \eta_i(T)]^T, \quad i = 1, 2, \dots, N. \quad (3)$$

where $x_i(t)$, $\mu_i(t)$, and $\eta_i(t)$ are the absolute level matrix, increment level matrix, and fluctuation level matrix of objects, respectively, under panel data.

Among,

$$\begin{aligned} \Delta x_{ij}(t) &= x_{ij}(t) - x_{ij}(t-1), \\ \bar{x}_i(t) &= \frac{\sum_{j=1}^m x_{ij}(t)}{m}, \\ S_i(t) &= \frac{\sum_{j=1}^m (x_{ij}(t) - \bar{x}_i(t))^2}{m-1}. \end{aligned} \quad (4)$$

According to the spatial-temporal feature of the panel data, this paper sets the spatial and temporal features of the subjects as absolute quantity level, incremental level, and fluctuation level and marks them as c_1, c_2 , and c_3 , then $l = 1, 2, 3$ and $q = 3$. Let $\gamma_{ij_l}(t)$ be a measure of the spatial-temporal feature c_l of the index j at the time t for the object i , which represents the spatial-temporal feature attribute values of absolute quantity level, incremental level, and fluctuation level. Correspondingly, the temporal and spatial matrices of objects can be defined.

Definition 2. Assume $\gamma_{ij_l}(t)$ represents the measured value of the object i at the time t about the indicator j spatial-temporal feature l , for $\forall i \in U$, $\forall a_j \in A$, and $c_l \in C$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$; $t = 1, 2, \dots, T$, if $\gamma_{i_l}^{\max}(t) = \max_i\{\gamma_{ij_l}(t)\}$, or $\gamma_{i_l}^{\min}(t) = \min_i\{\gamma_{ij_l}(t)\}$, then the matrices

$$\gamma_{i_l}(t) = \begin{bmatrix} \gamma_{i_{1_l}}(1) & \gamma_{i_{2_l}}(1) & \dots & \gamma_{i_{j_l}}(1) & \dots & \gamma_{i_{m_l}}(1) \\ \gamma_{i_{1_l}}(2) & \gamma_{i_{2_l}}(2) & \dots & \gamma_{i_{j_l}}(2) & \dots & \gamma_{i_{m_l}}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \gamma_{i_{1_l}}(t) & \gamma_{i_{2_l}}(t) & \dots & \gamma_{i_{j_l}}(t) & \dots & \gamma_{i_{m_l}}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \gamma_{i_{1_l}}(T) & \gamma_{i_{2_l}}(T) & \dots & \gamma_{i_{j_l}}(T) & \dots & \gamma_{i_{m_l}}(T) \end{bmatrix},$$

$$\gamma_{i_l}^{\max}(t) = \begin{bmatrix} \gamma_{1_l}^{\max}(1) & \gamma_{2_l}^{\max}(1) & \dots & \gamma_{j_l}^{\max}(1) & \dots & \gamma_{m_l}^{\max}(1) \\ \gamma_{1_l}^{\max}(2) & \gamma_{2_l}^{\max}(2) & \dots & \gamma_{j_l}^{\max}(2) & \dots & \gamma_{m_l}^{\max}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \gamma_{1_l}^{\max}(t) & \gamma_{2_l}^{\max}(t) & \dots & \gamma_{j_l}^{\max}(t) & \dots & \gamma_{m_l}^{\max}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \gamma_{1_l}^{\max}(T) & \gamma_{2_l}^{\max}(T) & \dots & \gamma_{j_l}^{\max}(T) & \dots & \gamma_{m_l}^{\max}(T) \end{bmatrix},$$

(5)

are the spatial-temporal feature l measures of the object i and positive ideal object at a time t with respect to the index j .

According to Definitions 1 and 2, we can obtain $\gamma_{i_1}(t) = x_i(t)$, $\gamma_{i_2}(t) = \mu_i(t)$, and $\gamma_{i_3}(t) = \eta_i(t)$. For the spatial-temporal feature attribute c_1 , the index j_1 satisfies $j_1 = 1, 2, \dots, m_1$ and $m_1 = m$. For the spatial-temporal feature attribute c_2 , the index j_2 satisfies $j_2 = 1, 2, \dots, m_2$ and $m_2 = m$. For the spatial-temporal feature attribute c_2 , the index j_3 satisfies $j_3 = 1, 2, \dots, m_3$ and $m_3 = 1$. Since the larger the absolute quantity and increment and the smaller the fluctuation level, the better, according to Definition 2, we can see that

$$\gamma_{i_1}^{\max}(t) = \begin{bmatrix} x_1^{\max}(1) & x_2^{\max}(1) & \dots & x_{j_1}^{\max}(1) & \dots & x_m^{\max}(1) \\ x_1^{\max}(2) & x_2^{\max}(2) & \dots & x_{j_1}^{\max}(2) & \dots & x_m^{\max}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_1^{\max}(t) & x_2^{\max}(t) & \dots & x_{j_1}^{\max}(t) & \dots & x_m^{\max}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_1^{\max}(T) & x_2^{\max}(T) & \dots & x_{j_1}^{\max}(T) & \dots & x_m^{\max}(T) \end{bmatrix},$$

$$\gamma_{i_2}^{\max}(t) = \begin{bmatrix} \mu_1^{\max}(2) & \mu_2^{\max}(2) & \dots & \mu_{j_2}^{\max}(2) & \dots & \mu_m^{\max}(2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_1^{\max}(t) & \mu_2^{\max}(t) & \dots & \mu_{j_2}^{\max}(t) & \dots & \mu_m^{\max}(t) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \mu_1^{\max}(T) & \mu_2^{\max}(T) & \dots & \mu_{j_2}^{\max}(T) & \dots & \mu_m^{\max}(T) \end{bmatrix},$$

$$\gamma_{i_3}^{\max}(t) = [\eta_i^{\min}(1), \dots, \eta_i^{\min}(t), \dots, \eta_i^{\min}(T)]^T,$$

(6)

among,

$$\gamma_{j_1}^{\max}(t) = x_j^{\max}(t) = \max_i\{x_{ij}(t)\},$$

$$\gamma_{i_2}^{\max}(t) = \mu_j^{\max}(t) = \max_i\{\mu_{ij}(t)\}, \quad (7)$$

$$\gamma_{i_3}^{\max}(t) = \eta_i^{\min}(t) = \min_i\{\eta_i(t)\}.$$

According to the matrix of the measured values of the object i and the positive ideal object, the distances between the object i and the positive ideal object with respect to the space-time feature l are

$$d_i^l = \left[\sum_{t=1}^T \sum_{j=1}^{m_l} (\gamma_{ij_l}(t) - \gamma_{j_l}^{\max}(t))^2 \right]^{1/2}. \quad (8)$$

According to the spatial-temporal feature of the panel data, d_i^1 , d_i^2 , and d_i^3 , respectively, represent the distance between the object i and the positive ideal object regarding the absolute quantity level attribute, increment level, and fluctuation level. The smaller the value of d_i^1 is, the greater the absolute amount of object i 's development, and the better its development degree will be. However, d_i^2 depicts the trend difference of indicator value increment between object i and positive ideal object over time. If the object i and positive ideal objects both present the same direction change over time, the closer the object i and positive ideal object are, the smaller the distance is, and vice versa. d_i^3 represents the degree of fluctuation of the index value of the object i and positive ideal object over time. The greater the individual similarity, the smaller the distance between the object i and the positive ideal object. According to formula (8), the expression of d_i^1 , d_i^2 , and d_i^3 is

$$d_i^1 = \left[\sum_{t=1}^T \sum_{j=1}^m (x_{ij}(t) - x_j^{\max}(t))^2 \right]^{1/2},$$

$$d_i^2 = \left[\sum_{t=1}^T \sum_{j=1}^m (\mu_{ij}(t) - \mu_j^{\max}(t))^2 \right]^{1/2}, \quad (9)$$

$$d_i^3 = \left[\sum_{t=1}^T \sum_{j=1}^1 (\eta_i(t) - \eta_i^{\min}(t))^2 \right]^{1/2}$$

$$= \left[\sum_{t=1}^T (\eta_i(t) - \eta_i^{\min}(t))^2 \right]^{1/2}.$$

Definition 3. Assume that $\gamma_{ij_l}(t)$ represents the measured value of the object i at the moment of t about the indicator j spatial-temporal feature l , $\forall i \in U$, $\forall a_j \in A$, and $c_l \in C$, if $d_i^{+l} = \max_i\{d_i^l\}$, $d_i^{-l} = \min_i\{d_i^l\}$, then we called vectors d_i^+ and d_i^-

$$d_i^+ = (d_i^{+1}, d_i^{+2}, \dots, d_i^{+l}, \dots, d_i^{+q}), \quad (10)$$

$$d_i^- = (d_i^{-1}, d_i^{-2}, \dots, d_i^{-l}, \dots, d_i^{-q}),$$

are the object positive and negative ideal target centers, respectively.

In order to better reflect the spatial and temporal features of cluster objects, research objects can be divided into s category according to practical problems and evaluation needs. Accordingly, each spatial-temporal feature attribute value can be divided into s category. In order to determine the classification of each object, we try to use the extreme value difference equalization method to determine the target of each category. As the distance between the research object and the positive ideal object is smaller, if the ideal grey target is smaller, the grey class is better. Correspondingly, $\forall c_l \in C$ and $\forall i \in U$, the method of determining the ideal grey class using the idea of the mean value of the extreme value difference is centered on the extreme point d_i^{+l} of the spatial-temporal feature attribute c_l , and the mean value $(d_i^{+l} - d_i^{-l}) / (s_l - 1)$ ($s_l \geq 2$) of the extreme value difference is taken as the extraction distance. Search and extract corresponding feature values and positions along with each time-space feature until the maximum value under each time-space feature attribute dimension is searched (the situation centered on the extreme point d_i^{+l} is similar). Based on the above method, we can define the ideal objects of each grey class under the combination or difference of spatial-temporal feature attributes. Based on the above method, the definition of the ideal objects of each grey class can be given.

Definition 4. Set panel data decision information system $S = \{U, A, V, C\}$, for cluster object U , $\forall c_l \in C$, $k = 1, 2, \dots, s$. If the following condition is met,

$$d_i^{+lk} = d_i^{+l} - \frac{k(d_i^{+l} - d_i^{-l})}{s - 1}, \quad (11)$$

then

$$d_i^{+k} = (d_i^{+1k}, d_i^{+2k}, \dots, d_i^{+lk}, \dots, d_i^{+qk}) \quad (12)$$

is called the ideal object of k the grey class under the set of space-time feature attributes.

According to the measurement distance of the object i and k grey object ideal object in the spatial-temporal feature attribute c_l , the distance between the object i and the k ($k = 1, 2, \dots, s$) object ideal object with respect to the spatial-temporal feature attribute c_l can be calculated as

$$d_i^{lk} = |d_i^l - d_i^{+lk}|. \quad (13)$$

For the research object i , the closer to the ideal object of the k grey class, the higher the degree of similarity, the more likely it belongs to the grey class. Considering that the grey relational analysis is based on the similarity degree of the geometric shape of the sequence curve to determine whether it is closely related or not, it can be used to calculate the grey relational degree of the ideal grey object k and study object i . Let ε_i^{lk} be the correlation coefficient between object i and grey ideal object k on the space-time feature attribute l , and it is

$$\varepsilon_i^{lk} = \frac{\min_i \min_l d_i^{lk} + \rho \max_i \max_l d_i^{lk}}{d_i^{lk} + \rho \max_i \max_l d_i^{lk}}. \quad (14)$$

Correspondingly, the grey correlation degree of the object i and k grey object ideal objects are obtained:

$$\varepsilon_i^k = \sum_{l=1}^q w_l \varepsilon_i^{lk}. \quad (15)$$

where $0 \leq w_l \leq 1$ and $\sum_{l=1}^q w_l = 1$.

According to the correlation degree of cluster object i , the larger the ε_i^k is, the more the cluster object i belongs to the grey class k , and by $\max_{1 \leq k \leq s} \{\varepsilon_i^k\} = \varepsilon_i^{*k}$, it can be judged that cluster object i belongs to the grey class k .

3. Case Analysis

The regional high-tech industry is an important part of the national high-tech industry. Its industrialization determines the level of regional economic development, the product's competitiveness, and even the future of the development of regional economic. The major developed countries and regions in the world regard the high-tech industries as the country's top strategic task. They formulate policy measures and adjust resource inputs to command the world economy and enhance international competitiveness. After more than 20 years of development, China's high-tech industry has made a remarkable achievement. However, due to the differences in the level of economic development and the uneven distribution of resources in various regions, the development of high-tech industries in various regions is seriously unbalanced. How to judge the development status of regional high-tech industrialization comprehensively and accurately, stimulate innovation practice in various regions, and improve the level of regional high-tech industrialization have become urgent problems for national and local governments. This study aims to objectively, comprehensively, and truly reflect the development status of China's regional high-tech industrialization and to comprehensively evaluate the development differences in various regions.

Therefore, considering regional development status, development level, and development stability, this paper sets out a regional high-tech industrialization evaluation index system from the two aspects of high-tech industrialization level and the benefits of high-tech industrialization (the data come from the compilation of Chinese science and technology statistics). Then, this study tries to use the model constructed in this paper to measure and evaluate the development status of regional high-tech industries. The central government and the national macroeconomic departments can use the indicators constructed in this study to timely grasp the development trend of high-tech industrialization in various regions and formulate suitable policies for the development of high-tech industrialization. This policy will enable China to get rid of the phenomenon of regional convergence in economic development, promote industrial restructuring and upgrading, and accelerate the development of economic.

Referring to the compilation of China's scientific and technological statistics, the regional high-tech industrialization assessment is mainly divided into two aspects of high-tech industrialization level and efficiency. The indicators for

measuring the level of high-tech industrialization primarily include the added value of high-tech industries accounts for the proportion of industrial added value, the added value of knowledge-intensive service industry accounts for the proportion of GDP, the export value of high-tech products accounts for the proportion of merchandise exports, and the sales revenue of new products accounts for the proportion of main business income. The indicators for measuring the benefits of high-tech industrialization are mainly high-tech industry labor productivity, high-tech industry value-added rate, and knowledge-intensive service industry labor productivity. Correspondingly, the regional high-tech industrialization development evaluation index system is composed of the above seven indicators, which are recorded separately a_1, a_2, \dots, a_7 . In consideration of the spatial-temporal feature of China's provincial high-tech industrialization development, it is more reasonable to evaluate the development status of high-tech industrialization from the perspective of absolute development level and coordination level. Due to the incomplete statistics of Tibet, according to the compilation data of statistical compilation of science and technology in China from 2007 to 2014, 30 provinces and cities in mainland China were selected to conduct cluster analysis on the panel data of high and new technology industrialization with the model constructed in this paper. First, dimensionless processing is performed on the values of high-tech industrialization development indicators in 30 provinces in mainland China from 2007 to 2014. Then, according to Definition 1, the absolute quantity level matrix, the incremental level matrix, and the wave level matrix under the spatial-temporal feature of each province can be calculated. It is a good state that when absolute quantity level and the increment level of the spatial-temporal feature attribute are as large as possible, and the fluctuation level is as small as possible. According to the absolute quantity level matrix, the incremental level matrix, and the fluctuation level matrix, the spatial-temporal feature matrix of the positive ideal object can be determined, and the distance matrix about the 30 provinces and the positive ideal object can be obtained.

In order to know the level of each province in China, it is assumed that the province can be divided into three grey categories. Each province can be divided into three levels: high, medium, and low, from the perspective of the dimension of the absolute quantity level spatial-temporal characteristic attribute, the incremental level spatial-temporal characteristic attribute, and the coordination level spatial-temporal characteristic attribute. According to the distance matrix of the spatial-temporal feature attributes of 30 provinces and positive ideal objects (provinces), three grey class ideal objects in the spatial-temporal feature attribute set can be determined by using the extreme value difference equalization contraction method. Three grey class ideal objects are as follows:

$$\begin{aligned} d_i^{+1} &= (d_i^{+11}, d_i^{+21}, d_i^{+31}) = (0.1057, 0.0936, 0.0798), \\ d_i^{+2} &= (d_i^{+12}, d_i^{+22}, d_i^{+32}) = (0.3418, 0.2844, 0.3189), \\ d_i^{+3} &= (d_i^{+13}, d_i^{+23}, d_i^{+33}) = (0.5779, 0.5452, 0.5580). \end{aligned} \quad (16)$$

Due to the space limitation, the correlation coefficients for calculating the spatial-temporal feature attributes of 30 provincial and three grey ideal objects are no longer listed. According to the preference of the decision-maker or the government in the actual evaluation process, the weights of the space-time feature attributes are assumed to be $w_1 = 0.5$, $w_2 = 0.3$, and $w_3 = 0.2$. Correspondingly, the degree of association between the 30 provinces and the three grey class ideal objects with respect to the spatial-temporal feature attributes can be obtained. According to the degree of grey correlation, the grey class of each province can be determined. The objects belonging to the "low class" are as follows: Hainan, Guizhou, Yunnan, Gansu, Qinghai, Ningxia, Xinjiang, and Guangxi; the objects belonging to the "medium class" are as follows: Hebei, Jilin, Heilongjiang, Anhui, Fujian, Henan, Hunan, Sichuan, Shaanxi, Liaoning, Hubei, Chongqing, Shanxi, Inner Mongolia, and Jiangxi; and the objects belonging to the "high class" are as follows: Beijing, Tianjin, Shanghai, Jiangsu, Zhejiang, Shandong, and Guangdong. The calculation results are shown in Table 1.

Specifically, in terms of the "low-class" category, it includes Hainan, Guizhou, Yunnan, and other regions. The province domain high-tech industry development is relatively backward in this grey category. Compared with the advantageous regions, these regions still have much room for improvement in economic development and industrial scale. In addition, the infrastructure development of the high-tech industry is relatively weak, and there are few enterprises with high-tech content. To develop and promote the transformation and development of high and new technology industries, we should take the characteristic and advantageous industries of this region as the breakthrough point and combine the development of strategic emerging industries. And combining with the characteristics and current situation of local industrial development, the construction of innovation-driven leading demonstration zone should be completed to provide experience and path for regional high-tech construction.

Specifically, in terms of the "medium-class" category, it includes Hebei, Jilin, Heilongjiang, Anhui, Fujian, and other regions. In this grey category, the development of the high-tech industry in the province is at a medium level. We can see that these regions have a certain foundation of high-tech industry, but there are certain barriers between its transformation, market service orientation, and the development of science and technology and economic services. Aiming at the development of high-tech industries in this province, it is necessary to improve policy guidance and talent introduction measures and strengthen the construction of advantageous and characteristic high-tech industries. With high-tech zone and industrial base as the carrier, the formation of regional characteristic industrial clusters is promoted, and the development of the modern industrial system is accelerated. In important industries, basic research and high-tech research strive to maintain the position of national leading development.

Specifically, in terms of the "high-class" category, it includes Beijing, Tianjin, Shanghai, Jiangsu, Zhejiang, Shandong, and Guangdong. In this grey category, the

TABLE 1: The grey correlation degree between the development of high-tech industrialization in each province and the ideal grey class in spatial-temporal feature set.

Region	Low	Medium	High
Beijing	0.4572	0.6438	0.8286
Tianjin	0.4831	0.5873	0.7726
Hebei	0.4558	0.6695	0.3807
Shanxi	0.4462	0.5963	0.3987
Inner Mongolia	0.5951	0.6248	0.3855
Liaoning	0.4931	0.5963	0.3983
Jilin	0.3595	0.7125	0.3813
Heilongjiang	0.3945	0.6529	0.4057
Shanghai	0.3857	0.6404	0.8135
Jiangsu	0.3633	0.5399	0.8875
Zhejiang	0.4536	0.5857	0.7932
Anhui	0.3898	0.7568	0.5943
Fujian	0.4338	0.6052	0.5266
Jiangxi	0.5485	0.5989	0.3456
Shandong	0.5832	0.6043	0.6657
Henan	0.5782	0.6807	0.4439
Hubei	0.3763	0.7008	0.5057
Hunan	0.4655	0.6887	0.5645
Guangdong	0.3944	0.4698	0.8138
Guangxi	0.6187	0.5641	0.3942
Hainan	0.8331	0.4969	0.3412
Chongqing	0.4054	0.6690	0.4037
Sichuan	0.3765	0.6837	0.5055
Guizhou	0.7809	0.5344	0.3466
Yunnan	0.7284	0.5642	0.3553
Shaanxi	0.3894	0.7057	0.5562
Gansu	0.6267	0.4859	0.3491
Qinghai	0.8573	0.4829	0.3699
Ningxia	0.8498	0.4792	0.3787
Xinjiang	0.8644	0.4436	0.3459

development of the high-tech industry in the province shows a good level, and its high-tech industry has initially formed its own characteristics and advantages. In these regions, the development of the high-tech industry has played a strong role in promoting regional economic construction and social development and has formed a relatively complete industrial chain. At the same time, it also has the foundation to participate in international competition and seize the opportunity. Therefore, the development and improvement of the industrialization level and efficiency of the new and high technologies in these provinces need to focus on improving the original innovation ability and sustainable development ability, broaden the global vision, enhance the ability to integrate and utilize global innovation resources, focus on cultivating new competitive advantages of industries, and accelerate the development of high-end advanced manufacturing industry.

According to the clustering results, it is necessary to fully consider the feature and differences of the type-free regions, to scientifically and rationally formulate the development objections of high-tech industrialization and to make targeted and effective policies and measures. Compared with the existing high-tech industrialization evaluation methods, the grey clustering method in the multiattribute spatial-temporal feature constructed in this paper fully considers the

dynamic development, absolute development level, and volatility level of the province and objectively assesses the development status of the province's high-tech industrialization, instead of setting the threshold based on the magnitude of the measured value to determine the categories that each province and city belongs to.

Compared with the traditional grey clustering method based on the difference of object attributes [26], the better grey classes are basically the same, but in terms of general grey classes and poor grey classes, the traditional grey clustering method based on the difference of object attributes is different from our grey clustering method based on object multi-attribute spatial-temporal feature. Due to space limitations, we list the clustering results under the incremental level of traditional attribute differences, as shown in Table 2.

The traditional grey clustering method based on the difference of object attributes is used to classify the development of high-tech industrialization in various provinces at the incremental level, and it is found that the "loss-class" categories include the following: Hainan, Guizhou, Yunnan, Tibet, Gansu, Qinghai, Ningxia, and Xinjiang. Inferior grey levels in terms of absolute quantity include the following: Hainan, Guizhou, Yunnan, Tibet, Gansu, Qinghai, Ningxia, Xinjiang, Shanxi, Inner Mongolia, Jiangxi, and Guangxi. And our grey clustering model based on multiple spatio-temporal feature attributes, and considering from the absolute level, incremental level, and fluctuation level, the "low-class" categories in the final results include Hainan, Guizhou, Yunnan, Gansu, Qinghai, Ningxia, Xinjiang, and Guangxi, excluding Inner Mongolia and Jiangxi. The reason for this result is that the traditional grey clustering method based on the difference of object attributes mainly mine information from the absolute development level and dynamic development, ignoring the dynamic development and fluctuation levels of the province. Our model takes these factors into consideration and evaluates the development status and level of high-tech industrialization in the province from the attributes of incremental, absolute, and coordination levels. Relevant data also show that Inner Mongolia and Jiangxi have achieved continuous growth in the benefits and scale of the high-tech industrialization in recent years. For example, Inner Mongolia ranked 5th in the country for high-tech industrialization benefits in 2015, and the value-added rate of the high-tech industries ranked 4th in the country (http://www.nmg.gov.cn/art/2017/7/4/art_1686_137736.html). In the first three quarters of 2017, the high-tech industry in Jiangxi province achieved steady development, and the cumulative added value was 163.824 billion yuan, a year-on-year increase of 11.1% (<http://jxstc.gov.cn/html/1386/2018-06-05/content-9914.html>). Once again, it shows that the current development of high-tech industrialization in Inner Mongolia, Jiangxi, and other regions is in a good situation, and the results of the grey clustering based on the multiattribute spatial-temporal feature are consistent and convincing. According to the above case analysis, it can be known that the constructed model can effectively describe the development trend or future behavior of the research object and achieve effective clustering of the research object.

TABLE 2: The grey correlation degree between the development of high-tech industrialization in each province and the ideal grey class in spatial-temporal feature set.

Region	Low	Medium	High
Beijing	0.1366	0.1924	0.6731
Tianjin	0.3111	0.2577	0.4315
Hebei	0.2656	0.5048	0.2296
Shanxi	0.1911	0.5038	0.2785
Inner Mongolia	0.4032	0.4665	0.1312
Liaoning	0.0418	0.5357	0.4275
Jilin	0.3212	0.5735	0.1053
Heilongjiang	0.2344	0.5546	0.2116
Shanghai	0.1312	0.2843	0.5854
Jiangsu	0.0645	0.3359	0.6003
Zhejiang	0.0743	0.4453	0.4804
Anhui	0.3703	0.4776	0.1552
Fujian	0.1451	0.7009	0.1544
Jiangxi	0.3832	0.5708	0.0461
Shandong	0.1301	0.1962	0.6737
Henan	0.2216	0.4401	0.3402
Hubei	0.1751	0.5033	0.3217
Hunan	0.1986	0.5502	0.2512
Guangdong	0.0509	0.2851	0.6644
Guangxi	0.3818	0.5701	0.0511
Hainan	0.7017	0.1762	0.1231
Chongqing	0.3622	0.4875	0.1513
Sichuan	0.3211	0.3992	0.2801
Guizhou	0.6398	0.2701	0.0902
Yunnan	0.4941	0.3728	0.1332
Shaanxi	0.2316	0.7002	0.0702
Gansu	0.6788	0.3109	0.0111
Qinghai	0.7449	0.2316	0.0255
Ningxia	0.6917	0.2228	0.0865
Xinjiang	0.5105	0.4588	0.0327

4. Conclusions

In order to effectively extract the information and rules contained in the panel data, a grey clustering model of a multiattribute spatial-temporal feature is constructed by using a grey correlation analysis method. Through the model and case analysis, the results show that the model we built has the following characteristics and advantages:

- (1) The model we constructed can more effectively utilize the spatial-temporal feature information of panel data, extract the development law of the panel data, and realize the effective mining of the information of the cluster objects.
- (2) The spatial-temporal feature of the research object, such as the absolute quantity level, the increment level, and the fluctuation level, can make more effective use of the spatial-temporal feature information of the panel data to identify the problems existing in the clustering object. In addition, the results of clustering are more objective and practical, thus expanding the application field of grey relational clustering.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] E. C. Liu and W. X. Li, "The home-market effects of Chinese and American producer services trade: an empirical anal spatial and temporal feature of panel data ysis based on gravity model and WIOD panel data," *China Soft Science*, vol. 4, no. 4, pp. 166–175, 2016.
- [2] S. Y. Wang, Q. W. Wang, and Z. Cao, "The clustering method based on the mutil-dimensional panel data and its application," *Journal of Applied Statistics and Management*, vol. 33, no. 1, pp. 42–49, 2014.
- [3] D. C. Bonzo and A. Y. Hermosilla, "Clustering panel data via perturbed adaptive simulated annealing and genetic algorithms," *Advances in Complex Systems*, vol. 05, no. 04, pp. 339–360, 2002.
- [4] D. Luo, L. L. Ye, B. L. Wei et al., "Grey matrix incidence model for panel data and its application in drought vulnerability risk," *Control and Decision*, vol. 33, no. 11, pp. 2051–2056, 2018.
- [5] X. M. Cui, "Evaluation of core competitiveness of mass entrepreneurship space based on improved grey relation model," *Statistics and Decision*, vol. 7, no. 7, pp. 177–180, 2019.
- [6] L. Z. Cui and S. F. Liu, "Grey matrix similar Incidence model for panel data and its application," *China Journal Of Management Science*, vol. 23, no. 11, pp. 171–176, 2015.
- [7] J. P. Zhu and M. K. Chen, "The cluster analysis of panel data and its application," *Statistical Research*, vol. 24, no. 4, pp. 11–14, 2007.
- [8] B. Y. Zheng, "The clustering analysis multi-variable panel data and its application," *Application of Statistics and Management*, vol. 27, no. 2, pp. 265–270, 2008.
- [9] Z. L. Xiao, B. Y. Li, and S. F. Liu, "The discussion on the clustering way based on the -dimensional panel data and empirical analysis," *Application of Statistics and Management*, vol. 28, no. 5, pp. 831–838, 2009.
- [10] J. Ren, "Fusion clustering analysis of multivariate panel data," *Journal of Applied Statistics and Management*, vol. 32, no. 1, pp. 57–67, 2013.
- [11] K. Zhang and S. F. Liu, "Extended clusters of grey incidences for panel data and its application," *Systems Engineering-Theory & Practice*, vol. 30, no. 7, pp. 1253–1259, 2010.
- [12] L. F. Wu and S. F. Liu, "Panel data clustering method based on grey convex relation and its application," *Control and Decision*, vol. 28, no. 7, pp. 1033–1037, 2013.
- [13] J. J. Ma, G. F. Ma, F. Y. Bo et al., "Chinese provincial patent product analysis based on fuzzy C-mean clustering method for panel data," *Systems Engineering-Theory & Practice*, vol. 09, pp. 2304–2314, 2015.
- [14] Y. G. Li and X. Q. He, "Panel data clustering method and application," *Statistical Research*, vol. 27, no. 9, pp. 73–77, 2010.
- [15] Y. G. Li, Y. Dai, and X. Q. He, "Clustering method for Panel data base on adaption weighting," *Systems Engineering-Theory & Practice*, vol. 33, no. 2, pp. 388–395, 2013.
- [16] J. D. Nelsen and C. B. Dean, "Adaptive functional mixed NHPP models for the analysis of recurrent event panel data," *Computational Statistics and Data Analysis*, vol. 52, no. 7, pp. 3670–3685, 2008.
- [17] C. Assmann and J. B. Hogrefe, "A Bayesian approach to model-based clustering for binary panel probit models,"

- Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 261–279, 2011.
- [18] M. A. Juárez and M. F. J. Steel, “Model-based clustering of non-Gaussian panel data based on skew-t Distributions,” *Journal of Business & Economic Statistics*, vol. 28, no. 1, pp. 52–66, 2010.
- [19] X. Liu, X.-H. Zhu, P. Qiu et al., “A correlation-matrix-based hierarchical clustering method for functional connectivity analysis,” *Journal of Neuroscience Methods*, vol. 211, no. 1, pp. 94–102, 2012.
- [20] L. Stanila, M. E. Andreica, and A. Cristescu, “Econometric analysis of the employment rate for the E. U. Countries,” *Procedia—Social and Behavioral Sciences*, vol. 109, pp. 178–182, 2014.
- [21] Y. Liu, J.-L. Du, R.-S. Zhang, and J. Y.-L. Forrest, “Three way decisions based grey incidence analysis clustering approach for panel data and its application,” *Kybernetes*, vol. 48, no. 9, pp. 2117–2137, 2019.
- [22] Y. Wang, Y. Zhang, and Q. Zhou, “A Stein-like estimator for linear panel data models,” *Economics Letters*, vol. 141, pp. 156–161, 2016.
- [23] K. George, M. James, and S. Yongcheo, “A nonlinear panel data model of cross-sectional dependence,” *Journal of Econometrics*, vol. 179, no. 2, pp. 134–157, 2014.
- [24] X. M. Li, Y. G. Dang, and J. J. Wang, “Grey relational clustering model for panel data clustering on indicators and its application,” *Control and Decision*, vol. 30, no. 8, pp. 1447–1452, 2015.
- [25] X. Li, K. W. Hipel, and Y. Dang, “An improved grey relational analysis approach for panel data clustering,” *Expert Systems with Applications*, vol. 42, no. 23, pp. 9105–9116, 2015.
- [26] Y. Liu, T. Zhou, B. T. Quan et al., “Grey development clustering method and application based on multi-attribute difference of objects,” *Operations Research and Management Science*, vol. 27, no. 12, pp. 57–63, 2018.
- [27] T. B. Jia and S. F. Liu, “Organization power configuration model based grey relational clustering,” *Mathematics in Practice and Theory*, vol. 46, no. 7, pp. 31–37, 2016.
- [28] W. C. Tao and Y. G. Dang, “Collaborative filtering recommendation algorithm based on grey incidence clustering,” *Operations Research and Management Science*, vol. 27, no. 1, pp. 84–88, 2018.
- [29] S. Falletta and S. A. Sauter, “The panel-clustering method for the wave equation in two spatial dimensions,” *Journal of Computational Physics*, vol. 305, pp. 217–243, 2016.
- [30] Y. Liu, J.-L. Du, and Y.-H. Wang, “An improved grey group decision-making approach,” *Applied Soft Computing*, vol. 76, no. 3, pp. 78–88, 2019.