Analysis of Regional Differences in Technological Innovation Based on Principal Component Distance Weighting Algorithm and FCM Algorithm

Mingrui Yue
School of Management Engineering, Zhengzhou University, Zhengzhou City, China
Mingruiyue@zzu.edu.cn

Abstract. In this paper, first of all, the optimal FCM algorithm was used to classify the economic development level of each sub-region. Secondly, the principal component analysis method was employed to comprehensively score and rank the economic development level of each sub-region. Thirdly, based on the classification and ranking, the differences in regional economic development were analyzed. Taking the economic disparities of western provinces as an example, the research showed that this algorithm could effectively comprehensively analyze the economic development disparities of western provinces and well reflect the characteristics of economic development of each province. Therefore, it was an improved and feasible method for analyzing the differences in economic development level of sub-regions within a large region.

Keywords: Principal component distance weighting algorithm, FCM algorithm, Scientific and technological innovation, Regional economy

1. Introduction
At present, scientific and technological innovation in China develops rapidly and has become an important driving force to promote economic development. However, there are also many problems in China’s scientific and technological innovation, among which the lack of financial support is the biggest problem. Besides, China’s regional development is unbalanced between the north and the south, and there is a serious unbalance between the scientific and technological innovation capacity and the government’s scientific and technological investment, and the driving force of scientific and technological innovation on economic growth shows a gradual decreasing tendency from the southeast to the northwest, which is expanding year by year, thus resulting in the unbalanced development of China’s regional economic development. Economic imbalances have led to the emergence of other social problems, such as the lack of social equity and the increase of social contradictions and conflicts, which have seriously hindered national development. Therefore, it is of great theoretical significance to study how to promote the healthy and steady development of the regional economy through scientific and technological innovation and investment. China’s economy has been developing rapidly, but the extensive economic development strategy has been confronted with severe challenges, so the sustainable development of China's economy needs a new engine. Based on the historical research on the economic development of various countries in the world, it is found that the contribution of
scientific and technological innovation to the economy of developed countries has exceeded 75%. However, innovation-driven enterprises need strong financial support, but because of risk avoidance and cost control, they have limited innovation-driven capabilities. The government can not only provide scientific and technological innovation support through direct investment but also support competent enterprises through administrative means to conduct technological innovation, so as to provide a steady stream of different driving forces for the sustainable and healthy development of China’s regional economy, and then create a good atmosphere and environment for collective innovation in the whole society. Later, after examining various methods from the perspective of clustering accuracy and regional economic development differences, in this paper, an intelligent hybrid algorithm of regional economic development difference analysis was proposed. It plans to first use the FCM algorithm to classify the economic development level of western provinces, and then adopt principal component analysis (PCA) to comprehensively score their economic development level. To get a better classification effect, the proposed genetic simulated annealing optimization FCM algorithm based on principal component distance weighting was used for clustering.

In this paper, the economic development differences of western provinces are taken as an example, and based on the intelligent hybrid algorithm of regional economic development difference analysis, the economic development differences of western provinces were comprehensively analyzed.

2. Methods for analyzing differences in regional economic development

2.1. FCM algorithm

FCM algorithm is a fuzzy clustering algorithm based on the objective function, and its idea is that this classification minimizes the J value of the objective function and maximizes the similarity between objects in the same cluster [13].

FCM algorithm is expressed as follows:

$$\min J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^\beta d_{ij}^2$$

$$\sum_{j=1}^{c} \mu_{ij} = 1$$

Where U is its similarity classification matrix, V is the clustering center point of each category, c is the number of categories to be divided, d_{ij} is the Euclidean distance between the ith data sample and the center point of the j category, $\mu_{ij} \in [0,1]$ is the membership degree of sample $x_i$ to j category, and $\beta \in [1,\infty)$ is a weighted parameter. When $\beta \in [1.5,2.5]$, the effect is the best.

2.2. Principal component distance weighting algorithm

Due to the different contribution rates of principal component factors, the first principal component has the largest contribution rate, followed by the second principal component, the third principal component, and so on. Their importance in classification is classified by primary and secondary factors. The direct substitution of principal component factors for original data results produces a classification distortion. Therefore, different weights should be given when calculating the distance between samples, in order to achieve the purpose of self-adapting sample data.

Supposing $G_1,G_2,\ldots,G_d$ are the column vectors of each principal component factor extracted; $I_1,I_2,\ldots,I_n$ are the new row vectors after extraction of the principal component, and the corresponding contribution rates are the number that $\alpha_1,\alpha_2,\ldots,\alpha_d$ are new attributes, then the distance between sample $I_i$ and sample $I_j$ is expressed as:
\[ d_{ij} = \left( \sum_{m=1}^{d} \alpha_m (I_{im} - I_{jm})^2 \right)^{\frac{1}{2}} \]  

2.3. FCM algorithm based on principal component distance weighting

According to Formula (2), the FCM algorithm was improved, and then the new FCM clustering algorithm based on principal component distance weighting was obtained, as follows:

\[
\begin{aligned}
\min J(U,V) &= \sum_{i=1}^{n} \sum_{j=1}^{c} \left( \mu_{ij} \right)^{\beta} \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{jm} \right)^2 + \sum_{i=1}^{n} \tau_i \left( \sum_{j=1}^{c} \mu_{ij} - 1 \right) \\
\sum_{j=1}^{c} \mu_{ij} &= 1 
\end{aligned}
\]  

(3)

Where \( d \) is the number of new attributes.

According to the Lagrange multiplier method, the above algorithm can be translated into the following formula:

\[
\begin{aligned}
\min J(U,V) &= \sum_{i=1}^{n} \sum_{j=1}^{c} \left( \mu_{ij} \right)^{\beta} \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{jm} \right)^2 + \sum_{i=1}^{n} \tau_i \left( \sum_{j=1}^{c} \mu_{ij} - 1 \right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{c} \left( \mu_{ij} \right)^{\beta-1} \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{jm} \right)^2 + \tau_i = 0 
\end{aligned}
\]  

(4)

Where \( \tau_i \) is the Lagrangian multiplier.

Calculating the extremum of the objective function in Formula (4), the partial derivative of parameters \( \tau_i, \mu_{ij} \) could be obtained, with a value of 0, and then the following formulas were obtained:

\[
\frac{\partial J}{\partial \tau_i} = \sum_{j=1}^{c} \mu_{ij} - 1 = 0 
\]  

(5)

\[
\frac{\partial J}{\partial \mu_{ij}} = \beta \left( \mu_{ij} \right)^{\beta-1} \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{jm} \right)^2 + \tau_i = 0 
\]  

(6)

The value of membership \( \mu_{ij} \) was obtained as follows:

\[
\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{km} \right)^2 \right)^{\frac{1}{2-1}}} \left( \sum_{m=1}^{d} \alpha_m \left( I_{im} - v_{jm} \right)^2 \right)^{\frac{1}{2-1}} 
\]  

(7)

Formula (7) is a sufficient condition for minimizing the objective function, and an iterative expression of the membership degree \( \mu_{ij} \) value that minimizes the objective function.

The iterative expression of the clustering center value is:
Formulas (7) and (8) were used to iteratively modify the membership degree and clustering center of data. When the algorithm converged, the clustering center and the membership degree of each sample for each category were determined to complete the classification.

2.4. How to use the algorithm to analyze the regional economic development differences

(1) The FCM algorithm optimized in this paper was used to cluster various economies of regional economy and analyze the characteristics of different categories of economies.

(2) The principal component analysis method was adopted to comprehensively score the economic development level of each economy, and then the economic development ranking of each economy was obtained.

(3) The regional economic development differences were studied and analyzed according to classification and scoring.

3. Empirical analysis on the economic development differences of western provinces

3.1. Determine principal component and variance contribution rate

Based on the principal component analysis, it could be concluded that the variance contribution rate of the first three principal components was 88.1%>85%, as shown in Table 1. Each principal component was a linear combination of original indexes, and the new comprehensive indexes were set as F1, F2 and F3. The principal axis (corresponding to principal component) information table is Table 1.

| Principal axis | Contribution rate% | x1   | x2   | x3   | x4   | x5   | x6   | x7   | x8   | x9   | x10  |
|----------------|--------------------|------|------|------|------|------|------|------|------|------|------|
| F1             | 56.4               | -0.294 | -0.282 | -0.37 | -0.388 | -0.256 | -0.333 | -0.273 | -0.356 | 0.259 | -0.319 |
| F2             | 19.4               | -0.367 | -0.424 | 0.182 | 0.189 | 0.386 | 0.347 | -0.42 | 0.049 | -0.227 | 0.341 |
| F3             | 12.3               | 0.382 | -0.3   | -0.121 | -0.08 | -0.37 | -0.213 | 0.413 | -0.229 | -0.571 | 0.108 |

3.2. Selection of experimental parameters

Parameter settings of the algorithm are shown in Table 2.

| Parameter                        | Setting                           |
|----------------------------------|-----------------------------------|
| Weighted parameters              | 2                                 |
| Maximum iterations               | 32                                |
| Termination tolerance of objective function | 0.00005                          |
| Selection range of optimal cluster number | [2, 6]                           |
| Maximum number of genetic algorithm evolution | 88                               |
| Crossover probability            | 0.01                              |
| Variation digits                 | 12                                |
| Variation number                 | 15                                |
| Chain length                     | 35                                |
| Initial temperature              | 110                               |
| Temperature attenuation coefficient K | 0.75                            |
| Tolerance                        | 0.003                             |
| Seed number                      | 3                                 |
| End temperature                  | 0.1                               |
Experimental description:
Weighted parameter $\beta$ was selected as 1.5, 2 and 2.5 for experiment, and the results showed that for the relevant data in 2015, when the weighted parameter $\beta$ was selected as 2, the sample could be classified more clearly.

3.3. Clustering analysis
When the partition fuzziness is smaller, the sample set is more clearly divided and the classification is better. Therefore, for a given weighted parameter $\beta$, the best classification can be obtained if the partition fuzziness is the smallest.

The definition of partition fuzziness is as follows:

$$PF_{\beta}(U,c) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij} - \left( u_{ij} \right)_{0.5}$$

Where $n$ is the number of samples, is the number of clustering, $U$ is the fuzzy partition matrix,

\[
\begin{cases}
1, u_{ij} \geq 0.5 \\
0, u_{ij} < 0.5
\end{cases}
\]

After clustering with the FCM algorithm optimized in this paper, when the the sample was divided into 6 categories, the partition fuzziness $PF_{\beta}(U,c)$ was the minimum value. The target value was always stable in repeated runs, and the partition fuzziness $PF_{\beta}(U,c)$ was always the minimum, so the best classification was 6 categories.

3.4. Results analysis
In 2015, the economy of most provinces in China slowed down due to the global economic downturn, but Chongqing still ranked top one with the same GDP growth as Tibet (11%). Chongqing is the only municipality directly under the central government in Central and Western China, and it is not only an important strategic fulcrum for “China Western Development” but also an important junction point of “One Belt And One Road” and the Yangtze River Economic Belt. Under the background of China’s macro-economic downward pressure, Chongqing takes the lead with the GDP growth rate of more than 10% in successive 3 years, and becomes one of the models of China’s rapid economic transformation development, comprehensive strength. Its comprehensive strength ranks first in the western region, and it is an independent category in the western provinces. In 2015, the economic aggregate of Inner Mongolia ranked third and its per capita GDP ranked first in the western region. However, the resource-based economic development path taken by Inner Mongolia was obviously different from that of other provinces, so it was classified as a separate category. Shaanxi is a province with good economic development in Northwest China, and its economic aggregate ranks second in the west of China, and its economic structure is similar to that of Sichuan and Guangxi, so it was classified into one category together with Sichuan and Guangxi. Although the economic aggregate of Guangxi ranked the 17th in 2015 and the 18th in 2016, it was classified into the same category as Sichuan and Shaanxi, which are important growth poles in the west of China, indicating that Guangxi has similar economic structure characteristics with Sichuan and Shaanxi, and has great development potentials, so it is expected to become an important growth pole in the west of China. Tibet has always been divided into a separate category. Due to historical and geographical reasons, Tibet has inconvenient transportation, few local talents, and a weak economic foundation, which greatly limited its development. Therefore, it is the province with the smallest economic aggregate in China. The GDP growth rate of Tibet took the lead in the whole country in 2015, and its GDP growth rate in 2016 (10%) ranked third in the whole country. However, it does not mean that Tibet has the endogenous driving force for rapid economic growth, but it is more due to its backwardness and the government’s
increased investment projects. Qinghai and Ningxia lack regional policy support, such as state-level new areas, urban agglomerations, and border support, so their economic development is slow and they are in danger of being marginalized.

4. Conclusion

Based on the provincial panel data, the spatial lag panel data model was established by defining geographical proximity spatial weight matrix and geographical distance spatial weight matrix. Through this model, the influence of scientific and technological innovation and governmental input of science and technology on regional economic development was analyzed. According to the applied analysis of economic development differences in the western provinces, it could be seen the proposed intelligent hybrid algorithm of regional economic development difference analysis had a good application. At the same time, the effectiveness of applying new technologies such as data mining in economic research was explored and tested, so it has a broad application prospect in the analysis of the regional economic development differences.

References

[1] Zhang Jinming, Guo Liping, & Zhang Xiaodan. (2012). Influence of interpolation parameters on DEM Interpolation Error in inverse distance weighted interpolation algorithm. Journal of Surveying and Mapping Science and Technology (01), 51-56
[2] Wang Jun An, & Wei Sheng min. (2001). Distance weighted quadratic error measure mesh simplification algorithm. Acta computer aided design and graphics, 13 (002), 189-192
[3] Lu Xiaoyan, Luo Limin, and Li Xiangsheng. "Improvement of FCM algorithm and simulation experiment research." Computer Engineering and application 45.20 (2009): 144-146
[4] Shi Hongbo, Yu Jian, and Huang Houkuan. "An effective implementation of FCM algorithm." Acta Sinica Sinica 025.001 (2003): 63-67
[5] Fang Xuan. Technological innovation and regional economic development [M]. Guangdong Economic Publishing House, 2002
[6] Shen Zhengping, Liu Haijun, and Jiang Tao. "Research on industrial clusters and regional economic development." China soft science, 000.002 (2004): 120-124