

Research Article

Construction and Analysis of Macroeconomic Forecasting Model Based on Biclustering Algorithm

Chao Pang 

School of Economics, Yunnan University of Finance and Economics, Yunnan 650000, China

Correspondence should be addressed to Chao Pang; pangchao@stu.ynufe.edu.cn

Received 8 December 2021; Revised 6 January 2022; Accepted 8 January 2022; Published 7 February 2022

Academic Editor: Miaochao Chen

Copyright © 2022 Chao Pang. 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.

In recent years, with the globalization of information and Internetization, the phenomenon of information overload has appeared in the production of large amounts of data, and data mining has emerged as the times require. Clustering technology is a representative data mining technology. Cluster analysis has been applied to data mining and has achieved significant results. However, with the deepening of people's understanding, it has been found that this either/or classification is increasingly not suitable for fuzzy classification problems. Therefore, fuzzy clustering technology, which combines the strengths of machine learning and fuzzy mathematics, has become the new darling of clustering technology and has achieved outstanding results in clustering accuracy. How to obtain a more accurate division from the vast economic statistics of the Statistical Yearbook has become a difficult problem, especially when there is no prior information. Based on China's macroeconomic statistics, this paper applies the biclustering method to the field of economic zoning for the first time, researches and predicts the economic region division plan of China's provinces and the economic growth model of each province, and combines the results with traditional levels. The results of the class methods are compared. The research results show that the hierarchical clustering algorithm is relatively intuitive and easy to apply to the overall analysis of the national economic divisions. The result of the biclustering algorithm has its unique advantages in mining the commonalities of various provinces under certain attribute sets.

1. Introduction

Macroeconomic zoning is the foundation of regional economic research and is the most basic unit used to analyze regional gaps, conduct regional adjustments, and promote regional development. In the process of regional socioeconomic development, the scientific division of socioeconomic regions is a prerequisite for rational formulation of regional economic development plans. My country is a typical large developing country, with large differences in the economic, social, resource, and environmental aspects of various regions. How to divide my country's overall macroeconomic regions by analyzing macroeconomic statistical data that have both time and spatial distribution characteristics over several years is a problem that this article will focus on.

The division methods of economic regions are generally divided into traditional classification methods and numerical

classification methods. Traditional classification methods are usually based on experience and relevant professional knowledge for qualitative classification. Although it can achieve certain results, the results are relatively general. It is difficult to make a more detailed description of the differences and connections between the research objects, sometimes because of the researcher. The subjective intent of the classification affects the objectivity of the classification. Numerical classification method can weaken the subjectivity and arbitrariness of traditional classification method to a certain extent. It has become more and more widely used in the study of economic zoning. Many scholars at home and abroad have used various numerical analysis methods for national economic zoning and regions. In-depth research work has been carried out on economic zoning. However, the differences between different methods and sample data will affect the final classification results [1–5]. Therefore, when solving specific problems, we need to combine

subjective judgments with objective facts to give a more reasonable analysis. The biclustering algorithm is shown in Figure 1.

The most notable feature of the traditional classical taxonomy is that it has the characteristic of either or the other; that is, the same thing belongs to and only belongs to a certain category, and there cannot be two situations where it does not belong to any category or belongs to more than one category at the same time. It is precisely because the result of this classification is clear and distinct, and there is no ambiguity, so this classification method is also called hard classification [6, 7].

However, in real life, people often use inaccurate but meaningful language in their daily communication, that is, vague language. If computer technology is used to recognize and analyze these vague language and information, it seems very difficult. The famous American cybernetics expert, Professor L.A. Zadeh, fully realized this contradiction and put forward the core idea of fuzzy mathematics, which is to use fuzzy thought to make clear and accurate mathematical explanations, which gave rise to fuzzy mathematics [8, 9].

Since the 1950s, Western countries have begun to standardize regional planning and regional policy work. John analyzed in detail the concept of regional planning on the three levels of country, metropolis, and city. The United States first began to implement standardized regional divisions. In 1969, in order to meet the standardization needs of regional analysis and policy makers, the US Bureau of Economic Analysis (BEA) divided detailed standard economic zones based on county data and the division of metropolitan areas. Its division method became important for other countries. Bongaer S. D. divides regions according to the functional consistency within the economic zone. This division method of functional consistency can better compare the divided regions and solve the injustice caused by the many types of economic divisions in the past for interregional policy making. Davis believes that regional planning should play a central role in coastal management, and the federal guidelines for special regional management plans should support advanced regional planning through coastal management behavior. Bryan proposed a systematic regional planning method, which uses the integer planning method within the framework of multicriteria decision analysis to set priorities for vegetation management and vegetation restoration to achieve the goal of multiple natural resource management. R. I. Chman uses principal component analysis and core principal component prefiltered cluster analysis to regionalize and classify sea level pressure. The results show that the filtered cluster analysis of the core principal component analysis method captures more accurately than the core principal component analysis. When it comes to the nature of the input data, the clustering calculation after the core principal component analysis filter is more efficient [10–14].

After China changed from a planned economy to a market economy, more and more scholars realized the important role of numerical classification in economic regionalization. On the basis of China's existing administrative regions, Liu Dongliang used the clustering method of

multivariate analysis to discuss China's large-scale economic regionalization at the national level. Through principal component analysis and cluster analysis, Liu Qinqu believes that the level of economic development in Henan Province can be divided into four levels, which are spatially represented as four geographic regions. Liu Zheng used the basic ideas and principles of the AHP model to construct an index system for ecological economic zoning and established a research method for ecological economic zoning, taking Tanghai County's national ecological demonstration area as an example. Zheng Dexiang used the economic indicators reflecting the forest location as a factor and applied a self-organizing competition network to establish a model, and, after continuous learning and testing, the obtained network simulation results were used to carry out economic zoning of the forest land in 68 counties (cities) in Fujian Province. Based on the analysis of traditional regional economic difference analysis methods, Zhang Yanwen proposed a new method of regional economic spatial difference analysis based on spatial clustering, hierarchical maps, and axis analysis and used the data of per capita GDP in Northeast China in 2000 for empirical analysis. Jiang Ling used the MFPT matrix method established based on short- and medium-distance passenger flow to divide economic function zones and tested the effectiveness of the zoning plan. Chen Shuangying applied SPSS to carry out an empirical analysis on the level of circular economy development in various regions of China, focusing on the process and results of clustering using the method of systematic clustering. Peng Ping used the gravity model to study the economic zoning of 91 counties and cities in Jiangxi Province and divided the province into four major economic regions, Nanchang Yijiujiang, Jingdezhen Yiyingtian, Xinyu Yipingxiang, and Ganzhou, and coordinated the development of the economic regions. The question puts forward scientific planning measures. Li Xuemei and Zhang Suqin analyzed the application of cluster analysis technology in data mining and explained the implementation process of cluster analysis technology with an example of regional division in macroeconomics. Based on the principles of ecological economics, Zhang Yongming selected 37 characteristic indicators suitable for the classification of the ecological economic system in Shandong Province and used principal component analysis and systematic clustering to divide the 17 prefecture-level cities in Shandong Province into three major ecological economic categories and 7 subcategories. Lin Aiwen proposed a grey clustering method based on weighted common origin. This method uses segmentation and common origin to calculate the clustering function and, after reasonable weighting, distinguishes each clustering element under its clustering index in order to evaluate the regional natural resources and select Hubei Province for case analysis [15–18].

The rapid development and progress of fuzzy clustering theory has spurred the collaborative development of related fields, especially the computer intelligence of fuzzy clustering technology [19–22].

In this paper, the biclustering method is applied to the field of economic zoning for the first time, and the characteristics of the biclustering method are analyzed in detail and the results are compared with the results of traditional

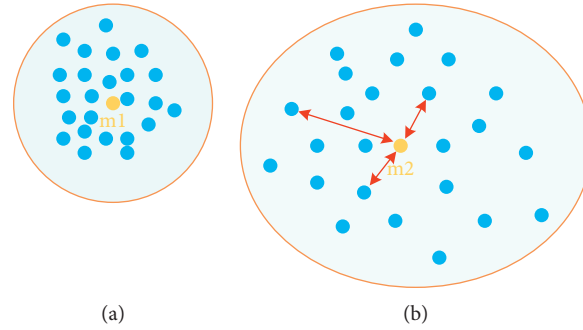


FIGURE 1: Biclustering algorithm. (a) C1. (b) C2.

clustering methods. The study found that the biclustering method has its unique advantages in mining the correlation between provinces, especially the local correlation.

2. Biclustering Algorithm

It can be summarized in this way that the biclustering algorithm can effectively identify the set of objects that show similar behavior patterns in a specific set of attributes. The biclustering algorithm is also widely used in many other different fields, especially the analysis of gene expression data in biological information [23–28].

The hierarchical method is to create a layered structure by decomposing a given set of data objects. According to the formation method of hierarchical decomposition, hierarchical methods can be divided into two types, bottom-up and top-down. The bottom-up aggregation hierarchical clustering method is to initially treat each object (itself) as a cluster and then aggregate these original clusters to construct larger and larger clusters, until all objects are aggregated into one cluster or until certain termination conditions are met. Most of the hierarchical clustering methods belong to this type of method, but they are different in the definition and description of the distance between objects within the cluster. The top-down decomposition hierarchical clustering method [29–31] is the opposite of the bottom-up method. It first regards all objects as the content of a cluster; it is continuously decomposed to make it become smaller and smaller but with more and more small clusters, until all objects constitute a cluster by themselves or satisfy a certain termination condition (such as a threshold of the number of clusters or a threshold of the shortest distance between the two closest clusters). The disadvantage of the hierarchical method is that it cannot be traced back after the (group) decomposition or merging. This feature is also useful because there is no need to consider the combinatorial explosion caused by different options when decomposing or merging. But this feature also makes this method unable to correct its own wrong decisions. When doing clustering analysis of data, this paper adopts the hierarchical clustering method embedded in EisenbergCluster3.0 software [32–34]. The software provides four cores of hierarchical clustering algorithms, namely, centroid clustering, single associative clustering, full associative clustering, and average associative clustering. After comparison, the average associative

clustering algorithm is selected. The distance between class A and class B is defined as follows:

$$d(A, B) = \frac{\sum_{x \in A, y \in B} d(x, y)}{|A||B|}. \quad (1)$$

Given a value with n rows and m columns to represent matrix A , the element a_{ij} is given specific value, which represents the relationship between row i and column j . Such a data matrix A with n rows and m columns is defined by its row set and column set:

$$\begin{aligned} X &= \{x_1, x_2, \dots, x_n\}, \\ Y &= \{y_1, y_2, \dots, y_m\}. \end{aligned} \quad (2)$$

Different biclustering algorithms produce different types of biclustering; they are as follows:

- (1) Constant value biclustering (Figure 2(a))
- (2) Row (column) constant value biclustering (Figures 2(b) and 2(c))
- (3) Biclustering with consistent constant value (Figures 2(d) and 2(e))
- (4) Biclustering with consistent evolution (Figures 2(f), 2(g), 2(h), and 2(i))

The simplest biclustering algorithm identifies those subsets with constant values. An example of constant biclustering is shown in Figure 2(a). Other biclustering methods look for subsets with constant values in rows or columns on the data set. Figures 2(b) and 2(c) are biclusters with constant values in rows and columns, respectively. A more complex biclustering method is to find biclusters with consistent values in rows and columns. Figures 2(d) and 2(e) are two examples of this clustering. Each row (column) can be evolved by adding a constant value (Figure 2(d)) or multiplying by a constant (Figure 2(e)) additional rows (columns).

The rapid development and progress of fuzzy clustering theory have spurred the collaborative development of related fields, especially the computer intelligence of fuzzy clustering technology. The problem solved by the last biclustering method analyzed here is to find a bicluster with consistent evolution, which is also the most general biclustering model. These methods treat the elements in the matrix as symbolic values and try to find a subset of rows and

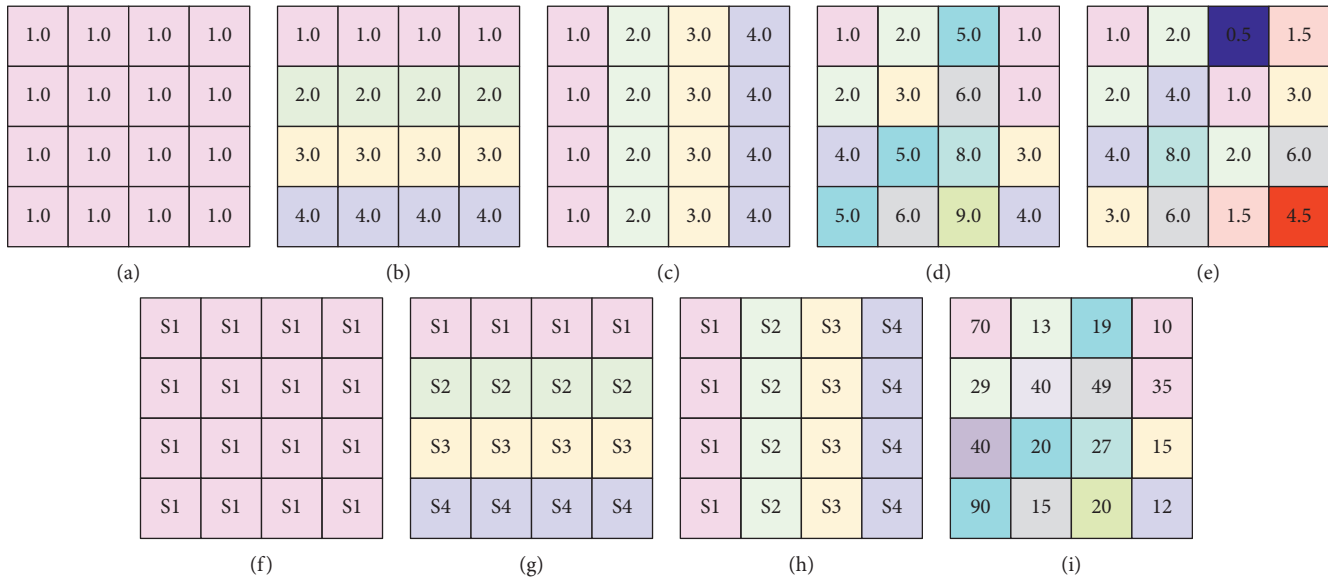


FIGURE 2: Different types of biclustering.

columns that have consistent behavior, regardless of the actual values in the data matrix. The nature of uniform evolution can be observed on the entire bicluster, on the rows and columns of the submatrix (Figure 2(f)), on the rows of the bicluster (Figure 2(g)), or on the columns (Figures 2(h) and 2(i)).

The biclustering algorithm assumes one of the following situations: there is only one bicluster in the matrix (Figure 3(a)), or the matrix contains K biclusters, where K is what we expect to be certain. Although most algorithms assume that there are several biclusters in the matrix, general algorithms only expect to find one bicluster. In fact, although some algorithms can find more than one bicluster, the target bicluster is usually the best one through some index tests.

When the biclustering algorithm considers that there is more than one bicluster, the following biclustering structure can be obtained (Figures 3(b)–3(i)):

- (b) Diagonal matrix biclustering (rows and columns are reordered and form diagonal matrix blocks)
- (c) Nonoverlapping biclustering of checkerboard structure
- (d) Row-specific biclustering
- (e) Column-specific biclustering
- (f) Nonoverlapping double clustering of tree structure
- (g) Nonoverlapping nonspecific biclustering
- (h) Overlapping biclustering with a hierarchical structure
- (i) Overlapping biclusters placed randomly

The biclustering algorithm has two different goals: to identify one or a given number of biclusters. Some algorithms try to identify one bicluster at a time, such as Cheng and Church, and Sheng et al. identify one bicluster each time and repeat the process to finally find other biclusters. Lazzaroni and Owen also tried to find biclustering in an

iterative process to get the lattice model. There are also some algorithms that try to find all biclusters at the same time. FLOC is the method used. First, the data is added to each row or column with independent probability to generate an initial set of biclusters, and then the number of biclusters is increased iteratively.

It has been widely and successfully applied. One of the outstanding advantages of fuzzy theory and technology is that it can be better. Describing, imitating, and the way of thinking of human beings using fuzzy thinking to make those disciplines seem to have nothing to do with mathematics in the past or have little to do with mathematics in the past can use quantitative and clear mathematical descriptions to build models. Considering the complexity of the problem, some heuristic solving algorithms are used to solve this problem. These algorithms can be divided into the five following categories:

- (1) Clustering of iterative rows and columns
- (2) Subsystem method
- (3) Greedy iterative solution
- (4) Exhaustive method
- (5) Identification of distribution parameters

QUBIC is a qualitative biclustering algorithm proposed by Li Guojun and Ma Qin. Compared with other current methods, this algorithm can solve the biclustering problem on a more general model and basically overcomes all the difficulties faced by the current biclustering problem. A core feature of the QUBIC algorithm is that it can identify all statistically significant biclusters. Another important feature is that it can find the most general biclustering model, that is, the biclustering with the scaling mode. At the same time, QUBIC is a very efficient method, which can solve the biclustering of thousands of objects under thousands of conditions within a few minutes (desktop CPLT time). This method has been well applied in the field of bioinformatics.

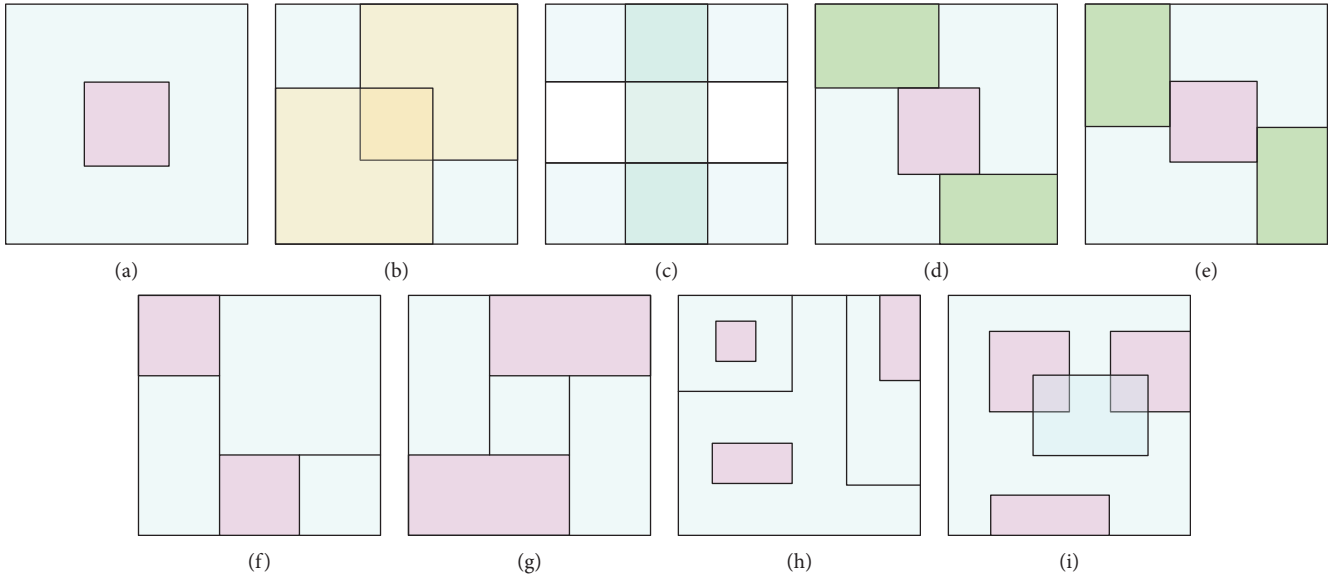


FIGURE 3: Biclustering structure.

There are three main steps of the QUBIC algorithm. The first step is to construct a representation matrix through a discrete method based on outlier thinking. A qualitative method is used to represent the expression value, so that a new matrix can be obtained to represent the object expression data set under multiple attributes. The purpose of this is to effectively construct a biclustering model in the general sense. Since the object of this article is the collection of macroeconomic data of various provinces and the attribute is a macroeconomic indicator, if the original data dispersion method of QUBIC (each row is discrete) is used for clustering, its economic significance cannot be explained. The discrete method in the algorithm has been improved as follows. The following chart is shown in Figure 4.

When constructing the representation matrix for each object, thus

$$c = \frac{m}{2}, \tag{3}$$

$$s - 1 = [m \times q].$$

In the above equation, q is an optional parameter value. Object i is in column j of the attribute initial data matrix (m rows and n columns), and its values are arranged in ascending order as follows:

$$v_{1j} \cdots v_{s-1j} v_{sj} \cdots v_{c-1,j} v_{cj} v_{c+1,j} \cdots v_{m-s+1,j} v_{m-s+2,j} \cdots v_{mj}. \tag{4}$$

The value under j is considered invalid if and only if its expressed value belongs to

$$(v_{ic} - d_i, v_{ic} + d_i). \tag{5}$$

Here

$$d_i = \min\{v_{ic} - v_{is}, v_{i,m-s+1} - v_{ic}\}. \tag{6}$$

The algorithm considers that all values satisfying the following formula are low expression, and they are represented by -1 by default.

$$S \leq v_{ic} - d_i. \tag{7}$$

All the values of the following formulas are highly expressed and are represented by 1 by default.

$$S \geq v_{ic} + d_i. \tag{8}$$

Of course, the data with high (low) expression can be further subdivided according to the size of the value. Use 1 (2) to represent the high (second highest) expression, and use -1 (-2) to represent the low (second low) expression (see below for details) (case analysis).

At present, fuzzy clustering has made outstanding achievements in computer simulation technology and e-commerce and other high-tech aspects. Similarly, fuzzy clustering analysis theory has also been successfully applied in economic management, environmental science, and traditional industries such as biology, environment, agriculture, and medical care and achieved good results. The matrix constructed above is called a representation matrix, in which the expression level of each object under any attribute is represented by an integer value. Two objects under a subset of attributes are considered to have related expression modes if their corresponding integers in the two corresponding rows in the expression matrix are equal. Here, the correlation level between two objects under a specific condition set is defined as the number of attributes that satisfy the condition. In practical applications, I am also very interested in the completely opposite state: the integers in the corresponding column have the same absolute value but the opposite sign. If each pair in a row of a submatrix is either correlated or negatively correlated, it is said to be feasible. The biclustering problem is to find all the local optimal submatrices in a given matrix. The second step is to

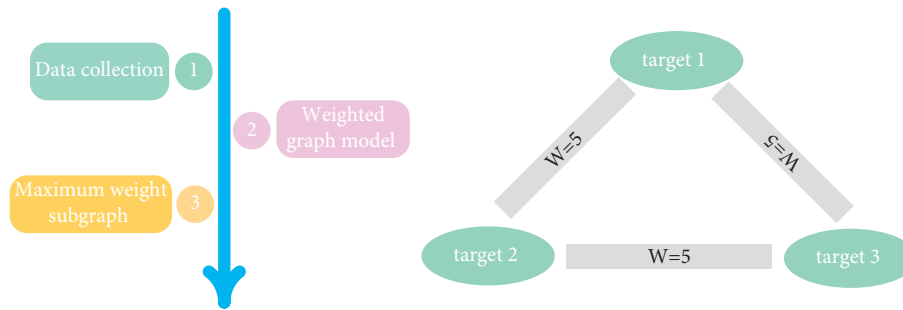


FIGURE 4: Following chart of the method.

build an empowered graph model. For a given representation matrix, construct a weighted graph G , with objects as vertices and edges connecting a pair of objects, and the weight of each edge is the correlation level of the corresponding two objects. The greater the weight, the stronger the correlation between the corresponding two rows. Intuitively, the objects in the biclustering should constitute a subgraph G with extremely large weights, because, in the conditional subset, because these objects have a high degree of correlation, the weight of each involved edge is larger. However, it is worth noting that not all subgraphs with extremely high weights correspond to a bicluster. There is no polynomial algorithm for identifying all subgraphs with maximum weights in a weighted graph, because the problem of identifying the maximum clique in a graph is a special case of this problem, and the maximum clique problem is a well-known NP-C problem. Therefore, in QUBIC's solution, the problem of finding subgraphs in the graph is not directly solved, but a heuristic algorithm is constructed based on the constructed representation matrix to solve the biclustering problem. In the third step, based on the constructed model, a heuristic algorithm is used to find the subgraphs with extremely large weights corresponding to the biclustering. At the beginning, an edge with the largest weight is taken as the seed to construct the initial biclustering, starting from the selected seed, iteratively expanding the biclustering in this matrix. Consider a matrix M with m rows and n columns discussed above, representing the representation level of n objects under m attributes, a corresponding weighted graph G , a set of vertices V , and a set of edges E . The weight of each edge is the number of columns with the same nonzero integer that the two objects have. This algorithm iteration is performed on the edge set S arranged in descending order.

The edges are

$$e = g_i g_j. \quad (9)$$

At least one of g_i and g_j is not in the previously determined bi-cluster. The basic idea is to iteratively expand biclustering in the vertical and horizontal directions according to the selected seed. When it can no longer be expanded, that is, when the following formula reaches the maximum, output the submatrix (I, J) of M found, and I is the row submatrix and J is the column submatrix.

$$\begin{aligned} \min\{|I|, |J|\}, \\ M_{\text{sub}} = (I, J). \end{aligned} \quad (10)$$

This algorithm has some unique and powerful features: (1) It will not miss a meaningful biclustering. If the construction of a significant bicluster is not completed due to some reasons in the algorithm, resulting in the failure to recognize the bicluster, this problem will be corrected later by selecting other edges as seeds. (2) This algorithm can not only find objects related to expressions but also find objects whose expressions are exactly the opposite. (3) Although this is a greedy algorithm, because the algorithm traverses all the seeds, it will not miss the optimal solution.

3. Research on Macroeconomic Zoning Based on Biclustering Algorithm

This paper collects and sorts out the macroeconomic data of 31 provinces, municipalities, and autonomous regions in China for 9 years from 1999 to 2007 and focuses on selecting 17 macroeconomic indicators. The selection of indicators is mainly based on the indicator design that affects the sustainable development of China's economic regions. Certainly, the data for each year comes from the National Statistical Yearbook. Some indicators are missing in some years, but this does not affect the clustering results as a whole.

At present, the commonly used clustering methods can respectively cluster the rows or columns of the data matrix, while the biclustering method is a method of clustering in both the row and column dimensions at the same time. Because the data of macroeconomic analysis has both time and space characteristics, it is necessary to reduce the dimensionality of the data first to make it suitable for biclustering analysis. In this paper, the indicators plus the year mark are used as new indicators, thus reducing the data to the two-dimensional space of the new indicators and provinces. Such a dimensionality reduction method can overcome the lack of information caused by averaging a certain index over a certain period of time. In order to make the data comparable, the data is normalized by the software EisenbergCluster3.0. This paper uses the hierarchical clustering method embedded in Eisenbergcluster3.0 software to analyze the data. It shows the clustering results of various

provinces and cities. The different industries are compared in Figure 5.

In the first category, Beijing, Tianjin, and Shanghai are my country's three municipalities directly under the Central Government (except Chongqing). These three regions are all economically developed regions in my country. Selecting the regional GDP as a representative economic indicator, the average GDP growth rates of Beijing, Tianjin, and Shanghai from 1999 to 2007 were 10.8%, 11.4%, and 9.4%, respectively, which are relatively similar in value, and they have maintained a steady growth rate. Under other indicators (the gross value of the primary industry, the employment population, and the population of the primary industry), the three cities also have great similarities, which is also in line with the actual situation of the three municipalities. Take Beijing as an example. As the national political, economic, and cultural center, agriculture only accounts for a small proportion of the regional economy. Statistics show that, in 2009, the three industrial structures in Beijing were 1 : 23.2 : 75.8, and the tertiary industry had already accounted for more than 75% (Figure 6). This means that the clustering method is looking for the overall optimum, while the biclustering method produces a partial pattern, so it is also looking for a local optimum.

The second category covers nine provinces from south to north in the central and eastern regions of my country. Because of their high values under several attributes such as gross domestic product, secondary industry output value, and local fiscal revenue, this category can be known. It belongs to a collection of economically developed provinces. From 1999 to 2007, the average contribution rate of these nine provinces to the national GDP was 51.77%, accounting for more than half of the national GDP. Through more in-depth comparison and analysis, it can be found that the three provinces of Liaoning, Heilongjiang, and Hubei are closer and can be classified into a subcategory. The average GDP growth rates from 1999 to 2007 were 18.2%, 16%, and 15.5%, respectively. Fujian, Zhejiang, Jiangsu, Guangdong, and Shandong are more similar and can be divided into the second subcategory, with average annual GDP growth rates of 18%, 27.8%, 26%, 29.7%, and 26.5%. Shanxi is a subcategory of its own, with an average annual GDP growth rate of 31.2%. Generally speaking, the provinces under this category have maintained a relatively rapid growth rate, which basically represents the overall speed and level of my country's economic development. The prediction is shown in Figure 7.

The third category has five provinces: Hebei, Henan, Anhui, Hunan, and Sichuan. These five provinces are basically located in the central region of my country in terms of geographic location. The main feature of this category is that the primary industry accounts for a large proportion of the province's economy, and it is also a province with a large labor force. Under the four attributes of urban per capita disposable income, rural per capita disposable income, urban residents' living consumption expenditures, and rural residents' living consumption expenditures, the values are low, indicating that the living standards of the people in this type of economic region are not high. When using a

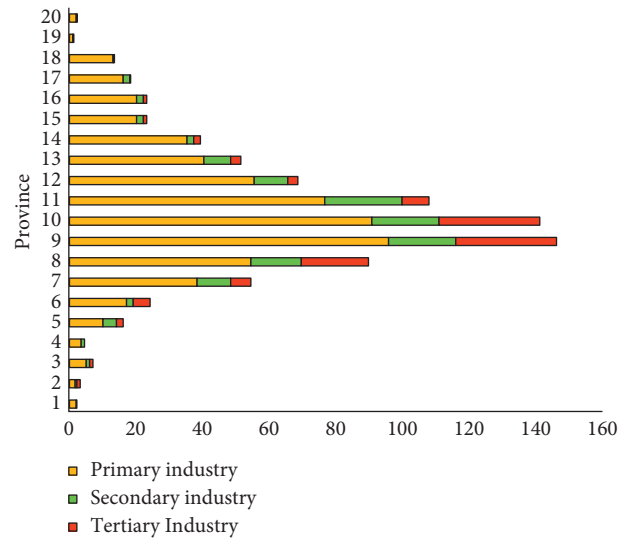


FIGURE 5: Different industry.

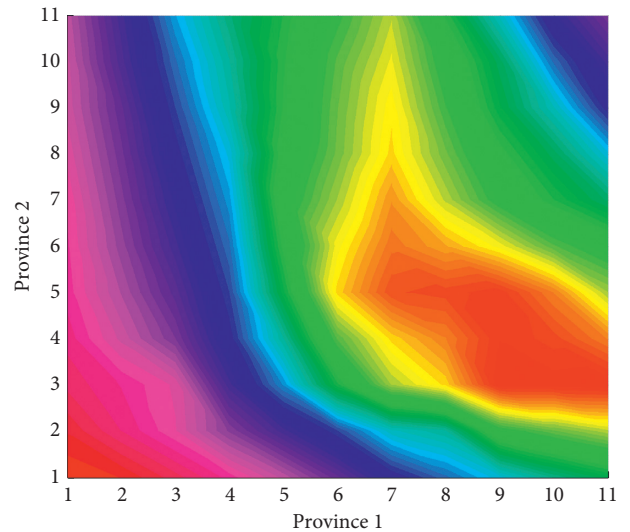


FIGURE 6: Provinces comparison.

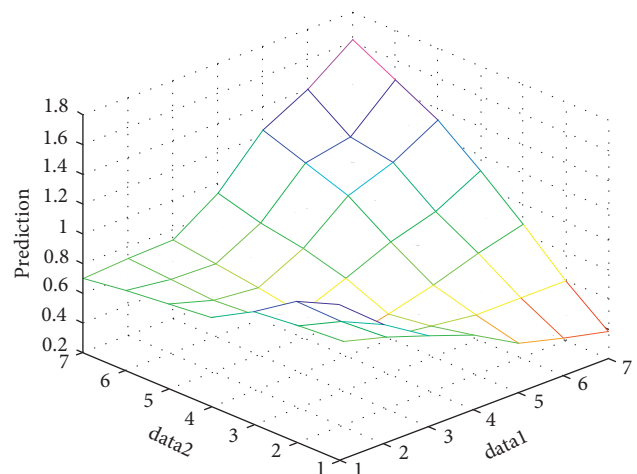


FIGURE 7: Prediction.

clustering algorithm, each object in an object cluster is defined by all attributes, and each attribute in a similar attribute cluster is characterized by the activities of all objects. However, when using the biclustering algorithm, each object in the biclustering is only determined by a certain subset of attributes, and each attribute in the biclustering is also only determined by a certain subset of the objects.

The fourth category is the seven provinces and cities in the central and western regions of my country: Jiangxi, Shaanxi, Guangxi, Guizhou, Chongqing, Yunnan, and Gansu. These provinces and cities belong to several provinces and cities with relatively slow economic development in China. The average contribution rate of these seven provinces to my country's GDP from 1999 to 2007 was only 11.5%, but the average economic growth rate was 20%. It is shown that the economic development of these provinces and cities is in a good state. The data are compared in Figure 8.

The fifth category almost covers the border provinces and regions in the north and southwest of my country, and the economic development speed is relatively slow. Among them, the three provinces of Inner Mongolia, Jilin, and Xinjiang are more similar and belong to provinces with a shortage of labor, and all other attributes have no obvious characteristics. The share of these three provinces in my country's GDP is between 1% and 2%, and their economic conditions are average. The other two provinces, Hainan and Tibet, have lower values under the attributes of gross national product, local fiscal revenue, and employed population, indicating that the level of economic development is still relatively low. These four provinces account for less than 1% of my country's GDP, and they are economically backward areas. Therefore, the purpose of the biclustering algorithm is to find a subset of common objects and a subset of attributes by clustering in the row and column directions at the same time, instead of clustering in these two dimensions separately.

The QUBIC algorithm is a biclustering algorithm, which generates a matrix containing different number attributes and objects through program analysis of the data. The objects within the same matrix have greater similarity under the attributes of the matrix and are clustered in one category. Among them, this method therefore gets rid of the limitation of clustering under the premise that all attributes are involved and can observe economic phenomena from a relatively novel perspective, thereby discovering its regular characteristics. A total of 9 biclusters were obtained by running QUBIC software on standardized data. Due to space limitations, we selected two typical biclusters for analysis. The convergence is shown in Figure 9.

The cluster is a matrix with a scale of 81 (9×9). The 9 provinces and cities are classified under the conditions of a subset of all attributes, that is, the urban per capita disposable income of individual years, the living consumption expenditure of urban residents, and local fiscal revenue. The expression values are relatively high and they are selected to be clustered into one category, indicating that the living standards of the people in these provinces and cities are better. In the clustering results of the aforementioned

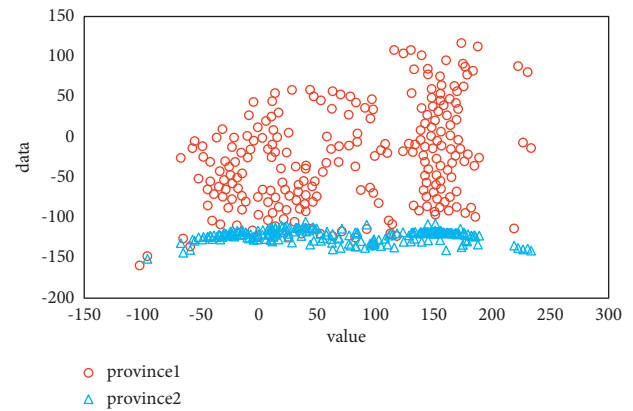


FIGURE 8: Data in different provinces.

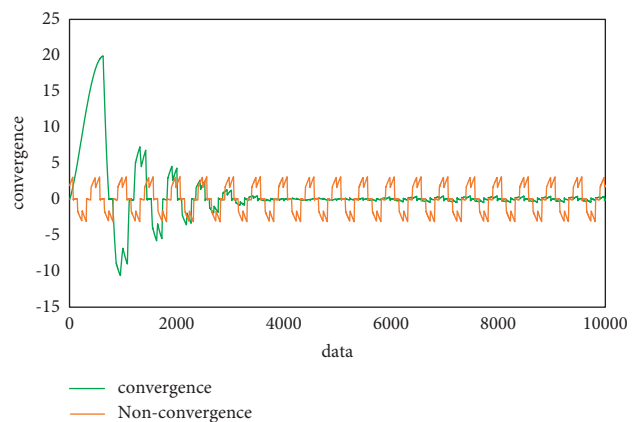


FIGURE 9: Convergence.

cluster, the nine provinces and cities belong to different categories. In comparison, the biclustering algorithm can find out those parts of provinces and municipalities that are similar under more specific attributes, thereby discovering some details hidden behind general economic phenomena.

The cluster will be simultaneously selected from the provinces and cities with high and low expression values under the two attributes of tertiary industry output value and government fiscal expenditure. Statistics show that Jiangsu, Shandong, and Guangdong contributed 8.6%, 8.1%, and 10.8%, respectively, of the output value of the tertiary industry to the output value of the tertiary industry in the country from 1999 to 2007. They are the three provinces with very developed tertiary industry, while Guizhou, Tibet, Hainan, Qinghai, and Ningxia, from 1999 to 2007, had average contribution rates of 0.9%, 0.17%, 0.5%, 0.7%, and 0.4%, respectively, to the output value of the tertiary industry, which are relatively backward provinces. Therefore, the clustering result conforms to the status quo. Such double clustering allows us to compare similar and opposite situations conveniently, which is also a highlight of the QUBIC algorithm.

Since the result of the previous running parameter only screens out those objects with high expression values, it is rarely involved in provinces and municipalities whose expression values are not significantly high or low. Therefore,

after we adjust the parameters, we can adjust the expression values to the second most significant ones. The objects are filtered out. Considering more detailed division when the data is discrete, QUBIC software is run to get a total of 16 biclusters. We select two of them for analysis. Under the conditions of primary industry output value, industrial output value, and construction industry output value, Anhui, Hunan, Sichuan, Fujian, Hubei, Heilongjiang, and Shanghai are the better regions, while Inner Mongolia, Gansu, and Xinjiang are the better regions.

Under this parameter, we can also find clusters with both low expression and sublow expression in an attribute set. For example, double cluster 4, Shanxi, Guizhou, and Gansu are in the secondary industry output value and local fiscal expenditure attributes. The next low expression is the three provinces with low expression under the consumption expenditure of rural residents.

Through the analysis of the operating results of the above two parameters, we can see that biclustering has a high degree of flexibility in the study of economic zoning, and the scale of clusters and the classification of clusters can be controlled by adjusting the parameters.

4. Conclusion

In summary, this article uses clustering and biclustering algorithms to analyze my country's macroeconomic data and obtains some meaningful results. First of all, in the research method, this paper uses the hierarchical clustering method in the clustering algorithm to do cluster analysis. On the other hand, this paper uses the QUBIC algorithm to do biclustering analysis, which breaks through the restriction that all clustering conditions in the clustering method must participate. This also solves the problem of "either or the other" in the classification of objects. The same object can belong to different categories under different conditions. At the same time, this method can find objects with completely opposite expressions at the same time, which is of great significance in economic analysis. Secondly, the data processing method in this article is also different from the past. Most studies on economic zoning use the method of segmented averaging to process data over a period of time, and the results are rough. The data dimensionality reduction method in this paper makes the analysis results specific to each year, making the results of economic zoning more detailed. Finally, based on the empirical analysis results, the following conclusions can be drawn:

- (1) The result of hierarchical cluster analysis gives a division of my country's overall economic region.
- (2) QUBIC got different clustering results from the previous method.
- (3) The development mode of high expression areas is low, a good template for expressing regions.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] Y. Yu, C. Yang, Q. Deng, T. Nyima, S. Liang, and C. Zhou, "Memristive network-based genetic algorithm and its application to image edge detection," *Journal of Systems Engineering and Electronics*, vol. 32, no. 5, pp. 1–9, 2021.
- [2] Y. Ishida and S. Hashimoto, "Asymmetric characterization of diversity in symmetric stable marriage problems: an example of agent evacuation," *Procedia Computer Science*, vol. 60, no. 1, pp. 1472–1481, 2015.
- [3] P. Zoha and R. Kaushik, "Image edge detection based on swarm intelligence using memristive networks," *IEEE Trans. on CAD of Integrated Circuits and Systems*, vol. 37, no. 9, pp. 1774–1787, 2018.
- [4] J. Pais, "Random matching in the college admissions problem," *Economic Theory*, vol. 35, no. 1, pp. 99–116, 2018.
- [5] J. J. Jung and G. S. Jo, "Brokerage between buyer and seller agents using constraint satisfaction problem models," *Decision Support Systems*, vol. 28, no. 4, pp. 291–384, 2020.
- [6] Y. Liu and K. W. Li, "A two-sided matching decision method for supply and demand of technological knowledge," *Journal of Knowledge Management*, vol. 21, no. 3, 2017.
- [7] J. Byun and S. Jang, "Effective destination advertising: matching effect between advertising language and destination type," *Tourism Management*, vol. 50, no. 10, pp. 31–40, 2015.
- [8] A. N. Nagamani, S. N. Anuktha, N. Nanditha, and V. K. Agrawal, "A genetic algorithm-based heuristic method for test set generation in reversible circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 2, pp. 324–336, 2018.
- [9] C. Koch and S. P. Penczynski, "The winner's curse: conditional reasoning and belief formation," *Journal of Economic Theory*, vol. 174, pp. 57–102, 2018.
- [10] D. Ettinger and F. Michelucci, "Creating a winner's curse via jump bids," *Review of Economic Design*, vol. 20, no. 3, pp. 173–186, 2016.
- [11] J. A. Brander and E. J. Egan, "The winner's curse in acquisitions of privately-held firms," *The Quarterly Review of Economics and Finance*, vol. 65, pp. 249–262, 2017.
- [12] Z. Palmowski, "A note on var for the winner's curse," *Economics/Ekonomia*, vol. 15, no. 3, pp. 124–134, 2017.
- [13] B. R. Routledge and S. E. Zin, "Model uncertainty and liquidity," *Review of Economic Dynamics*, vol. 12, no. 4, pp. 543–566, 2009.
- [14] D. Easley and M. O'Hara, "Ambiguity and nonparticipation: the role of regulation," *Review of Financial Studies*, vol. 22, no. 5, pp. 1817–1843, 2019.
- [15] P. Klibano, M. Marinacci, and S. Mukerji, "A smooth model of decision making under ambiguity," *Econometrica*, vol. 73, no. 6, pp. 1849–1892, 2005.
- [16] Y. Halevy, "Ellsberg revisited: an experimental study," *Econometrica*, vol. 75, no. 2, pp. 503–536, 2017.
- [17] D. Ahn, S. Choi, D. Gale, and S. Kariv, "Estimating ambiguity aversion in a portfolio choice experiment," *Working paper*, vol. 5, no. 2, pp. 195–223, 2019.
- [18] T. Hayashi and R. Wada, "Choice with imprecise information: an experimental approach," *Theory and Decision*, vol. 69, no. 3, pp. 355–373, 2010.

- [19] K. Zima, E. Plebankiewicz, and D. Wiczorek, "A SWOT analysis of the use of BIM technology in the polish construction industry," *Buildings*, vol. 10, no. 1, 16 pages, 2020.
- [20] P. Sun, B. Liu, and T. Sun, "Injury status and strategies of female 7-a-side rugby players in Anhui Province," *Sports Boutique*, vol. 38, no. 3, pp. 72–74, 2019.
- [21] P. Guild, M. R. Lininger, and M. Warren, "The association between the single leg hop test and lower-extremity injuries in female athletes: a critically appraised topic," *Journal of Sport Rehabilitation*, vol. 30, no. 2, pp. 1–7, 2020.
- [22] U. G. Inyang, E. E. Akpan, and O. C. Akinyokun, "A hybrid machine learning approach for flood risk assessment and classification," *International Journal of Computational Intelligence and Applications*, vol. 19, no. 2, Article ID 2050012, 2020.
- [23] Q. Liu, S. Du, B. Wyk, and Y. Sun, "Double-layer-clustering differential evolution multimodal optimization by speciation and self-adaptive strategies," *Information Sciences*, vol. 545, no. 1, pp. 465–486, 2021.
- [24] H. R. Medeiros, F. D. Oliveira, H. F. Bassani, and A. Araujo, "Dynamic topology and relevance learning SOM-based algorithm for image clustering tasks," *Computer Vision and Image Understanding*, vol. 179, pp. 19–30, 2019.
- [25] Y. Deng, D. Huang, S. Du, G. Li, and J. Lv, "A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis," *Computers in Industry*, vol. 127, Article ID 103399, 2021.
- [26] J. J. Chan, K. K. Chen, S. Sarker et al., "Epidemiology of Achilles tendon injuries in collegiate level athletes in the United States," *International Orthopaedics*, vol. 44, no. 3, pp. 585–594, 2020.
- [27] W. Li, G. G. Wang, and A. H. Gandomi, "A survey of learning-based intelligent optimization algorithms," *Archives of Computational Methods in Engineering*, vol. 28, pp. 1–19, 2021.
- [28] G. G. Wang, A. H. Gandomi, A. H. Alavi, and D. Gong, "A comprehensive review of krill herd algorithm: variants, hybrids and applications," *Artificial Intelligence Review*, vol. 51, no. 1, pp. 119–148, 2019.
- [29] A. Onan, S. Korukoglu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.
- [30] M. Jiang, Z. Zhang, J. Jiang, Q. Wang, and Z. Pei, "A collaborative filtering recommendation algorithm based on information theory and bi-clustering," *Neural Computing & Applications*, vol. 31, no. 12, pp. 8279–8287, 2019.
- [31] A. Onan, "Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering," *IEEE Access*, vol. 7, pp. 145614–145633, 2018.
- [32] B. H. Hedi, M. Imed, and G. Abdessalem, "The dynamic spillover effects of macroeconomic and financial uncertainty on commodity markets Uncertainties," *Economies*, vol. 9, no. 2, pp. 91–101, 2021.
- [33] B. Mikkil, H. Eric, and K. S. Jan, "Modeling, forecasting, and nowcasting U.S. CO2 emissions using many macroeconomic predictors," *Energy Economics*, vol. 96, pp. 1–9, 2021.
- [34] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2019.