Research Article

Fusion Analysis of Economic Data of the Medical and Health Industry Based on Blockchain Technology and Two-Way Spectral Cluster Analysis

Runhua Liu,^{1,2} Chengcheng Zhang,³ and Tenglong Feng³

¹Department of Medicine and Health Management, Guizhou Medical University, Guiyang 550025, China

²Department of Culture and Content Between South Korea and China, Hoseo University, Cheonan-si, Chungcheongnam-do 31066, Republic of Korea

³Economic Research Institute, Hoseo University, Cheonan-si, Chungcheongnam-do 31066, Republic of Korea

Correspondence should be addressed to Tenglong Feng; fengtenglong1987@163.com

Received 26 August 2021; Revised 28 October 2021; Accepted 5 November 2021; Published 24 November 2021

Academic Editor: Wenqing Wu

Copyright © 2021 Runhua Liu et al. 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 huge potential in gene expression analysis, which is helpful for disease diagnosis, new drug development, and life science research, the two-way clustering algorithm was proposed and it was widely used in gene expression data research. In order to understand the economic data of medical and health industry, this paper analyzes the economic data of the medical and health industry in different regions of China based on blockchain technology and two-way spectral cluster analysis and makes statistics on the economic data of the medical and health industry in eastern, central, and western regions of China. This paper studies the development status of China's medical and health industry and the factors affecting the agglomeration of medical and health service industry and analyzes them under the blockchain technology and two-way spectral cluster analysis method. The results show that the overall development trend of China's medicine and health is from government-led to government, society, and individual sharing. After the transformation of blockchain technology and two-way spectral cluster analysis, the output value of the pharmaceutical industry increased by about 10%.

1. Introduction

Traditional clustering analysis algorithms mainly deal with static data information. Due to the high-speed, real-time, and continuous characteristics of real-time data streams, traditional clustering analysis algorithms cannot be used. Blockchain technology is essentially a distributed witness technology. The so-called distributed means that the data are not concentrated in a certain data server center, but stored in various nodes in the network. The network members themselves are the storage carriers of the data and directly share, copy, store and synchronize the data. The so-called “witness” is to confirm and notarize the information uploaded in this distributed network. Once the information is uploaded and verified successfully, it cannot be tampered with, achieving the purpose of “witness.” The data is stored in the blockchain instead of a centralized

server, which can protect the data from being tampered with, making the data more credible and reliable. In addition, the permanent preservation of data also prevents the occurrence of denial. Therefore, blockchain technology fundamentally solves the many problems of the current traditional centralized system due to the existence of third parties.

Different from the general clustering method, in two-way clustering, not only the genes must be clustered, but also the changes in experimental conditions must be considered at the same time. The clustering composed of object subsets and attribute subsets identifies gene combinations with consistent expression patterns in the subsets of specific conditions, that is, two-way clustering. The clustering changes found in the dynamic analysis of data flow play an important role in the economic data fusion analysis of medical and health industry.

For the fusion analysis of economic data in the medical and health industry based on blockchain technology and two-way spectral clustering analysis, experts at home and abroad have conducted many studies. Yokoya showed the scientific results of the 2017 Data Fusion Competition organized by the IEEE Geosciences and Remote Sensing Society Image Analysis and Data Fusion Technical Committee. It aims to establish an accurate model (evaluated by the accuracy index of the undisclosed test city reference), and it is computationally feasible (evaluated by a limited-time test phase) [1]. Paola believes that the use of multisensor data fusion technology is essential to effectively merge and analyze heterogeneous data collected by multiple sensors, which are generally deployed in smart environments. A context-aware, self-optimized, and self-adaptive sensor data fusion system based on a three-tier architecture is proposed. The results show that the proposed solution is superior to the static method of context-aware multisensor fusion and achieves a lot of energy saving while maintaining a high degree of inference accuracy [2]. Liu believes that there is an urgent need to develop a data fusion method that can integrate data from multiple sensors to better characterize the randomness of the degradation process. The article developed a method to construct a health index by fusing multiple degradation-based sensor data [3]. In the article, Ghamisi proposed a new framework for using extinction profile (EP) and deep learning to fuse hyperspectral and optical detection and rasterization data derived from ranging. The results are compared with other methods, and the proposed method can achieve accurate classification results. It should be noted that the article uses the concept of deep learning to integrate LiDAR and hyperspectral features for the first time, providing a new opportunity for further research [4]. Chen FC proposed a deep learning framework based on convolutional neural network (CNN) and naive Bayesian data fusion scheme, called NB-CNN, which is used to analyze a single video frame for crack detection. At the same time, a new data fusion scheme is used to aggregate information extracted from each video frame to improve the overall performance and robustness of the system. For this reason, a CNN is proposed to detect crack patches in each video frame, and the proposed data fusion scheme maintains the temporal and spatial consistency of the cracks in the video, and the naive Bayesian decision effectively discards false positives [5]. In order to create a positioning system that provides high-availability attitude estimation, Tao Z also integrates dead reckoning sensors. Then the data fusion problem is expressed as sequential filtering. A reduced-order state-space modeling of the observation problem is proposed to provide an easy-to-implement real-time system. Experimental results show that, in terms of accuracy and consistency, this tightly coupled method performs better than the loosely coupled method that uses GNSS positioning points as input [6]. Beyca integrates multiple in situ sensor signals to detect initial abnormalities in the ultraprecision machining (UPM) process. Through the development of a new supervised learning method, the DP model state estimation is combined with the evidence theory sensor data fusion method to make a cohesive decision about the UPM process

conditions. It is detected and classified as 90% accuracy [7]. These studies provide a reference for the creation of this paper, but due to certain problems in related research algorithms and insufficient data samples, the results of related research are not consistent.

In this paper, we have two innovations. One is that the paper proposes a method to search for various biclusters using different bicluster quality evaluation indicators. Second, it compares and analyzes the experimental results of the algorithm in this paper and other commonly used biclustering integration methods in expressing data. The comparison results show that the blockchain-based technology and the two-way spectral clustering analysis method proposed in this paper are implemented in indicators, and time performance is better than other methods.

2. Data Fusion Analysis Method

2.1. Blockchain Technology. Blockchain is derived from the underlying support technology of the Bitcoin network. It is a decentralized public ledger facing the world [8]. The block header contains the version number, timestamp, random number, difficulty value, hash value of the previous block, and the root hash value of the Merkle tree, as shown in Figure 1 [9]. The blockchain is built on the entire network, and the extension of the blockchain network is convenient. Any place that has access to the Internet can be connected to the blockchain so that it can realize transactions across borders and supervisions, reduce supervision costs, and improve convenience.

Blockchain is a storage form of blockchain technology. The blockchain is composed of “blocks” connected in chronological order, and corresponding information is recorded in each “block.” Blockchains can be divided into three types, namely, public chains (represented by Bitcoin and Ethereum), alliance chains (represented by R3 alliance, BCOS platform), and private chains. Among them, the public chain is an open platform facing the world. Any individual or organization can freely access and use the services of the public chain and can also withdraw freely [10]. As the underlying chain, the public chain can develop decentralized applications for specific businesses based on the public chain.

The contribution of nodes in the public chain will be rewarded with digital tokens, and nodes participating in the world will jointly maintain the public chain. The public chain has achieved complete decentralization, but it lacks effective supervision, and transaction throughput is relatively low. At present, it cannot fully adapt to commercial applications with large business volumes. The alliance chain is jointly maintained by several organizations to maintain a blockchain, which is mainly used for the blockchain platform as a new way of cooperation between organizations to reduce the cost of business collaboration between alliance members and improve business operation efficiency [11, 12]. The alliance chain can have no token mechanism, nodes are provided by alliance members, the generation of each block is jointly determined by the preselected nodes, and other nodes can participate in verification and transactions. The

alliance chain will provide a supervision interface and even allow the setting of supervision nodes to achieve a kind of semidecentralization. Private chain is a blockchain system managed by an individual or a single organization. It has all the read and write permissions of the blockchain. The transaction throughput is much higher than that of the public chain. It is generally used for the internal business of exchanges and financial institutions. With the help of blockchain, the platform improves the business efficiency within the organization in a low-cost way [13].

The essence of a smart contract is a collection of data (state) and code (business function), which are stored in a specific address on the Ethereum blockchain. They can be triggered by transactions on the blockchain, and this code can be read from the blockchain to get data and write data [14]. Ethereum uses smart contracts to extend the functionality of the blockchain to support developers in building decentralized applications. At present, there are thousands of decentralized applications that are being developed and deployed based on Ethereum, and hundreds of decentralized applications have been running stably on the Ethereum blockchain network [15].

Traditional centralized systems face problems such as high cost, low business operation efficiency, and insecure data storage. From the nature of blockchain, it can be seen that blockchain can provide good solutions [16]. In the blockchain system, there is no need for a trusted third party to do credit endorsement, and the nodes in the network can still carry out normal transactions and business operations in an environment that does not need to trust each other. Data does not need a centralized server for storage and management but is secured by cryptography technology, distributed consensus algorithm, and so on, so that the data cannot be tampered with and can be traced [17].

2.2. Two-Way Spectral Cluster Analysis Method. With the rapid development of science and technology in the world today, the development of mankind has produced a large amount of data. How to quickly and fully utilize these data and find useful information is a major challenge [18, 19]. Data mining is to mine and analyze the original data from a large number of data sources to obtain effective knowledge information so as to make guidance and decision-making. Under normal circumstances, the process of data mining mainly has the following steps, as shown in Figure 2.

The two-way clustering algorithm is fundamentally different from the traditional clustering algorithm. The traditional clustering algorithm is only one-way clustering of rows or columns, while the two-way clustering algorithm considers the whole matrix at the same time; that is, at the same time, it performs cluster analysis on rows and columns to detect the local information of the matrix [20]. However, based on the two-way clustering algorithm of gene expression data, a gene or a sample can belong to different “clusters” at the same time; of course, it can also not exist in any “clusters”; that is, it can be between “clusters” and “clusters.” The overlapping part is shown in Figure 3. Rows represent genes, and columns represent the edges of two

adjacent conditions, that is, the direction of the gene expression level of a gene under two adjacent conditions.

The algorithm can exclude extra rows and columns from the two-way clustering results so as to achieve the purpose of shielding the rows and columns contained in the previous two-way clustering results so that the algorithm can produce different results through continuous iteration. Two-way cluster analysis plays an important role in gene expression profile data, which is mainly manifested in the following two aspects. In drug research, the results of two-way cluster analysis based on gene expression data are useful for the research of drug mechanism, drug development, the judgment of drug efficacy, and the detection of drug targets has played a great role. In terms of disease diagnosis, cancer heterogeneity is the biggest difficulty facing current cancer diagnosis and treatment. However, we can use the two-way data of gene chip cluster analysis that is used to identify cancer subtypes, thereby developing personalized treatment approaches. It can also be used to detect new tumor markers for early diagnosis and corresponding treatment.

Most of the two-group analysis algorithms are currently based on either Jewish or metaheuristic optimization methods, so these algorithms require some quality evaluation indicators to calculate the quality of the search and the direction of the search [21]. In fact, the research process of biclustering is the process of proposing a large number of biclustering indicators. The quality of biclustering evaluation indicators directly determines the efficiency and benefit of biclustering analysis algorithm.

In the two-way clustering, the two-way clustering set with the smallest average mean square residual is determined, and it is saved as the contemporary optimal two-way clustering set. Otherwise, the iteration is terminated, and the contemporary optimal two-way clustering set is output as a result. The mean square residual of a bicluster $B(I, J)$ is defined as

$$MSR(B) = \frac{1}{VolB} \sum_{i=1}^k \sum_{j=1}^l (b_{ij} - b_{i.} - b_{.j})^2. \quad (1)$$

Its related index is expressed as

$$RI(B) = \sum_{j=1}^l \frac{R_{Ij}}{l} \sum_{j=1}^l \frac{(1 - \sigma_{Ij}^2 / \sigma_j^2)}{l}. \quad (2)$$

Among them, R_{Ij} is the correlation index of the j -th column in the bicluster and σ_{Ij}^2 is the local variance of all elements in the j -th column in the bicluster B , but σ_j^2 is the global variance of all the elements in the j -th column in the entire gene expression data A .

For a bicluster $k * l$ of size $B(I, J)$, $k * (l - 1)$ do the following transformations to obtain a matrix M of size, and each element of b'_{ij} is defined as follows:

$$b'_{ij} = \begin{cases} 1, & \text{if } b_{ij} \leq b_i(j+1), \\ -1, & \text{if } b_{ij} \geq b_i(j+1), \\ 0, & \text{if } b_{ij} = b_i(j+1). \end{cases} \quad (3)$$

FIGURE 1: Block structure diagram.

FIGURE 2: Data mining process.

FIGURE 3: Two-way clustering analysis algorithm. (a) Single cluster structure. (b) Diagonal structure. (c) Checkerboard structure. (d) Row mutually exclusive structure.

Among them, $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, l - 1$. Then, the corresponding similarity times N of any two gene sums in double cluster $B(I, J)$ are defined as follows:

$$N = \max_{i' \in I, i' \neq i} N. \quad (4)$$

Among them, when the value $\delta(x) = 1$ of x is true, $\delta(x) = 0$. Based on the formula biclustering $B(I, J)$, the maximum number $\max N$ of similarities of gene i is defined as follows:

$$\max N = \sum j \in J \delta(b'_{ij} = b'_{ij}). \quad (5)$$

Any three genes of bicluster $B(I, J)$ are defined as follows:

$$\text{CSI}(i_1, i_2, i_3) = \frac{\sum j \in J \delta(b'_{i_1 j} = b'_{i_2 j} = b'_{i_3 j})}{\text{Max}N}. \quad (6)$$

Each data point has neighbor points.

$$Z_j = B^T(x_{ij} - d) \in R^m. \quad (7)$$

This transformation should be reversible, where Z_j is the X_{ij} mapping result and Z_j can also be obtained by X_{ij} doing the inverse transformation.

At that time, $j = k$ and $\delta_{jk} = 1$. Otherwise, $\delta_{jk} = 0$.

$$X'_{ij} = d + BB^T(X_{ij} - d) = d + BZ_j. \quad (8)$$

The formula is inverse transformation. In actual operation, due to the influence of noise data or different transformation methods, there are errors between X'_{ij} and X_{ij} , as shown in the following:

$$\varepsilon_j = X_{ij} - X'_{ij} = X_{ij} - d - BB^T(X_{ij} - d). \quad (9)$$

Perform objective optimization operations on B and d . The specific formula is as follows:

$$\min \left\{ \sum_{j=1}^k \omega_j \|\varepsilon_j\|^2 \right\} = \min \left\{ \sum_{j=1}^k \omega_j \|X - d1^T - BZ\|^2 \right\}, \quad (10)$$

where ω_j is the weight of the error ε_j . According to the feature decomposition, the minimum weighted mean square value of B is obtained. S is the weighted covariance matrix of neighboring points.

The customization of data is based on the premise of which distribution it conforms to, and then training and analysis are carried out according to the hypothetical distribution model. Therefore, learning the distribution of feature data according to the energy model can solve all the above problems. Then,

$$E(v, h | \theta) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j. \quad (11)$$

Among them, θ is the parameter model, a_i is the bias of the visible layer unit, b_j is the bias of the hidden layer unit, and W_{ij} is the connection weight between the visible layer and the hidden layer. The joint probability distribution that

can be obtained according to the energy function is as follows:

$$Z\theta = \sum_{v,h} e^{-E(v,h|\theta)}, \quad (12)$$

where $Z(\theta)$ represents the normalization factor which was in the calculation of joint probability. The likelihood function is solved through specific calculations, and the formula can be expressed as

$$p(v | \theta) = \frac{\sum_h e^{-E(v,h|\theta)}}{Z(\theta)}. \quad (13)$$

According to the state of the hidden layer unit, the formula for obtaining the visible layer unit in reverse is

$$P(v_i = 1 | h, \theta) = \sigma \left(a_i + \sum_j W_{ij} h_j \right). \quad (14)$$

The specific solution algorithm $P(v|\theta)$ of the function is to use the contrast divergence algorithm and then calculate the minimum mean square value of the translation vector d :

$$d = \frac{\sum_{j=1}^k w_j X_{ij}}{\sum_{j=1}^k w_j}, \quad (15)$$

$$S = \frac{1}{k} \sum_{j=1}^k w_j (X_{ij} - d)(X_{ij} - d)^T.$$

When the above formula is transformed, size ω is related. The weight of the sample point reflects the possibility that the point is noisy data. If the error is large, it means that the point is likely to be noise; otherwise, the point is less likely to be noise. The following functional relationship is satisfied between the weight and the error:

$$\omega(\varepsilon) = \begin{cases} 1, & |\varepsilon| \leq \mu, \\ \exp \left[\frac{(\varepsilon - \mu)^2}{2\sigma^2} \right], & \text{otherwise} \end{cases} \quad (16)$$

Calculate the cost function. If the cost function ε is less than a certain threshold or the change of the cost function during two iterations ε is less than a certain threshold, the algorithm stops, and the cost function is

$$J(U, v_1, \dots, v) = \sum_{i=1}^c J_i = \sum_{i=1}^n u_{ij}^m d^2(x_i, v_j). \quad (17)$$

Update the membership matrix U , and then return to the step:

$$u_{it} = \sum_{j=1}^c \left(\frac{d_{it}}{d_{jt}} \right)^{-2/(m-1)}. \quad (18)$$

For the membership matrix output by the algorithm, no human intervention is required in the algorithm implementation process. In order to avoid the possible misjudgment of this method, based on the cosine similarity, the

cosine value of the angle between the point and the cluster center is used to weigh the Euclidean distance. Then,

$$d_v(x_t, v_j) = \text{sim}(x_t^{(j)}, v_j) * \sqrt{(x_t^{(j)}, v_j)^T (x_t^{(j)}, v_j)}. \quad (19)$$

Among them, $t = |v_j|$, $|v_j|$ represents the number of samples v_j in a cluster that is the cluster center and represents all sample points in the cluster where the cluster center is located.

2.3. Data Fusion. The data integration process includes information retrieval, data processing, data integration, and result analysis [22]. Due to the variability of data, in the process of multisensor data integration, data must be integrated systematically, and data integration is divided into two levels according to function. All-round data connection with data preview, location recognition, and tracking was functions. High-resolution data integration is important for the analysis of trends and errors as a process to obtain the overall integration results [23]. Data fusion plays an important processing and coordination role in multi-information sources, multiplatforms, and multiuser systems, ensuring the connectivity and timely communication between the units of the data processing system and the collection center.

We use Figure 4 to illustrate the data-level fusion method. Data-level fusion is based on the raw data collected by each sensor to directly perform sublevel fusion; that is, data compilation and analysis are performed before the raw data collected by each sensor is processed [24]. Data-level fusion can retain the effective information in the original data as much as possible, but its disadvantage is that when the sensor data value is too large, the statistical accuracy will be improved, and the original data will be incompletely verified.

The biggest advantage of data-level fusion is that the original information is rich because the processed object is the most original data set, without any preprocessing, the loss of information is negligible, it can provide a large amount of detailed original information, and the accuracy of the fusion result is high. The disadvantage is that the amount of data that needs to be processed is extremely large, and the computer capacity and performance requirements are high. At the same time, the entire fusion process takes a long time, which will directly affect the real-time performance of the system; the original data is easily interfered with by external data, and the system should have good fault tolerance. Commonly used methods include weighted average algorithm, wavelet transform, and other algorithms [25].

In order to solve the shortcomings of data fusion, this paper is aimed at detecting the characteristics of a certain ambiguity in the data set and using fuzzy logic methods to identify and classify the detected data sets. Fuzzy set theory is essentially a kind of multivalued logic. In the process of fusion, a number between 0 and 1 is set for each data to express the credibility in the fusion process, and then the multivalued value is used. Logical reasoning method merges data to realize data fusion [26, 27].

3. Data Fusion Experiment and Results

3.1. Economic Status of the Medical and Health Industry. Data analysis is carried out on sites where Chinese medical institutions and health centers are focused on consulting data. In the eastern part of China, the medical centers of Beijing, Tianjin, Hainan, and Shanghai are higher than 1; Hebei and Shandong are nearly three. The strength of agglomeration in the year is higher than 1. The agglomeration index of Jiangsu, Zhejiang, Fujian, and Guangdong is below 1, as shown in Figure 5.

It can be seen that the four eastern cities of Beijing, Tianjin, Shanghai, and Hainan are densely populated and well-developed, stimulating an infinite demand for medical services, and they have relative agglomeration advantages from the perspective of demand. From the perspective of supply, the average number of health personnel in each medical institution in the above three cities is higher than the average in the eastern region, while the average in the four provinces, including Jiangsu, is lower than the overall level in the eastern region. The four provinces of Jiangsu, Zhejiang, Fujian, and Guangdong are the eastern coastal areas, and the large population base is the key factor for the formation of agglomeration levels lower than the levels of other eastern regions.

The results of statistics on the average personnel of medical institutions in the eastern region are shown in Table 1.

Through statistics on the medical industry in the western region, the results of the study found that the concentration of medical service industries in Ningxia, Inner Mongolia, and Xinjiang in the western region are all above 1. Qinghai and Shaanxi are basically above 1. The agglomeration level is below 1, and only one year exceeds 1 a year; the agglomeration levels in Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, and Gansu are all below 1. The result is shown in Figure 6.

It can be seen that the average level of health personnel in each medical institution in Ningxia, Inner Mongolia, and Xinjiang is above the western average. The level of health personnel in each medical institution in Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, and Gansu is almost less than or close to the western average, as shown in Table 2.

The study found that the medical service industry concentration levels in the four provinces of Anhui, Jiangxi, Henan, and Hunan were below 1; the concentration levels of the five provinces of Hubei, Shanxi, Liaoning, Jilin, and Heilongjiang were all above 1. The average number of health personnel in each medical institution cannot reflect the concentration of the medical service industry in the region from the perspective of supply. The results of the research are shown in Figure 7.

For the central region, from the effective analysis of Theil index and local health expenditures, the expenditures in the region are relatively consistent, which meets the needs of the local population and fully integrates into the government's health expenditures. There are significant differences among different provinces in China. The

FIGURE 4: Data-level fusion.

FIGURE 5: Medical and health agglomeration value in different cities.

TABLE 1: Average personnel in eastern region.

	2015	2016	2017	2018	2019	2020
BJ	30.72	33.08	28.37	28.80	20.64	22.65
TJ	30.26	32.11	33.84	29.74	21.02	20.29
SH	20.68	23.67	57.07	56.35	37.90	35.41
HN	13.84	15.24	16.46	18.12	9.60	10.03
JS	19.07	18.40	17.50	26.01	13.18	13.72
ZJ	16.45	16.81	16.37	17.75	10.06	10.68
FJ	13.75	11.83	23.8	26.63	6.74	6.27
GD	20.23	23.01	26.31	29.22	11.43	12.10

average ranking of the central region is 9. Compared with the eastern and western regions, the overall assessment of the local health expenditure in the central region is carried out. The average personnel of each institution in the central region are shown in Table 3.

The average health staff of each medical institution in the central region of our country is slightly lower than that of the eastern region but higher than that of the

western region. The values of Anhui, Hubei, Jilin, and Heilongjiang are all above the average. Henan has tended to be below the average in recent years; Jiangxi, Hunan, Shanxi, and Liaoning are all below the average. With the level of medical service, industry was agglomeration. The reason is that the number of medical institutions in Shanxi and Liaoning is higher than that in Jilin and Heilongjiang.

FIGURE 6: Agglomeration level of the medical industry in the western region.

TABLE 2: Average personnel in medical institutions.

	2015	2016	2017	2018	2019	2020
NX	17.57	16.82	19.02	18.27	8.08	8.51
NMG	14.65	14.04	12.8	17.21	6.73	6.37
XJ	13.8	13.74	16.6	18.21	9.21	8.82
GX	15.3	15.20	16.62	17.13	6.48	7.02
CQ	13.54	13.52	15.35	16.	8.74	8.04
SC	10.72	10.8	13.45	14.54	5	5.14
GZ	12.61	14.45	15.86	17.02	5.71	6.06
YN	12.41	13.42	14.30	15.31	7.7	8.06
GS	7.03	7.16	7.36	8.78	4.9	4.05

We have made statistics on the number of for-profit medical institutions and nonprofit medical institutions in various regions, and the results are shown in Table 4.

From the above table, the number of nonprofit organizations in the eastern region has increased year by year since 2015, and the number of for-profit organizations has been basically stable except for a few years; the number of nonprofit organizations in the central region has increased year by year, and the for-profit organizations have been basically stable, while in the western region, nonprofit organizations have increased year by year, and profit-making organizations have remained stable.

3.2. Data Fusion Changes. We have made statistics on the expenditures of the medical and health industry as a percentage of GDP and structure, and the results are shown in Figure 8.

It can be seen that the overall development trend of our country's medicine and health is a government-led transformation into a form of sharing by the government, society, and individuals. Government expenditures are declining gradually, and personal and social expenditures are increasing year by year, finally reaching a balance.

We made statistics on the changes in the medical and health industry under the blockchain technology and the two-way spectral clustering analysis method, and the results are shown in Figure 9.

We can see from Figure 9 that, after the change of blockchain technology and the two-way spectrum analysis system, the medical and healthcare industry has greatly improved. Among them, the result of the pharmaceutical industry has increased by about 10%, the cost has fallen, and an increase in the population and a reduction in bed rest time have led to a significant improvement of the medical and healthcare industry. We use examples to count different variables and analyze the fusion results. The results are shown in Figure 10.

4. Discussion

4.1. Algorithm Discussion. As a new type of technical means in the field of data mining, the bidirectional clustering algorithm successfully overcomes the shortcomings of traditional clustering algorithms. It can cluster both the gene direction and the conditional direction at the same time. That is, while retaining the global information, the local information of the gene expression matrix can still be mined.

FIGURE 7: Medical clusters in the central region.

TABLE 3: Average staff of each institution in the central region.

	2015	2016	2017	2018	2019	2020
AH	19.63	21.01	24.07	28.01	11.21	12.34
JX	12.13	12.86	15.10	19.36	5.46	5.67
HN	23.41	24.52	31.26	32.8	6.45	6.71
HN	15.52	15.07	17.49	18.36	5.25	5.13
HB	26.44	25.28	24.07	26.53	9.17	9.16
SX	17.41	17.21	16.89	19.16	5.51	5.60
LN	17.01	16.11	17.39	17.68	7.82	8.09
JN	17.34	15.54	15.59	15.79	8.64	8.54
HLJ	21.79	22.35	22.56	54.56	10.02	10.89

TABLE 4: Number of medical institutions by region.

Years	East		Central		West	
	Profit	Nonprofit	Profit	Nonprofit	Profit	Nonprofit
2015	4788	1011	6046	845	4606	982
2016	4849	1224	5927	1063	4517	1055
2017	5033	1426	5847	1038	4546	1223
2018	5083	1407	5805	1072	4449	1226
2019	5187	1531	5788	1228	4416	1451
2020	4320	2761	4432	2101	4042	2063

The traditional clustering analysis algorithm and data stream clustering analysis algorithm are researched and analyzed, and the data stream clustering analysis algorithm based on density grid is mainly discussed, and it is analyzed and summarized, and the improvement ideas are put forward. Combining the data stream clustering analysis algorithm based on density grid with fast processing speed and strong real-time characteristics, the DSG-stream algorithm is proposed on the problem of its insufficiency of cluster boundary processing and the uniform division of single-mode grids. The grid is divided into different thicknesses and granularity. The concepts of boundary grid and internal grid

are introduced, and the grid influence factor is combined for clustering processing. The algorithm is based on a two-stage processing framework: online stage maintenance of grid feature vectors and dynamic processing of the internal grids form microclusters, and the boundary grids are fine-grained clustering in the offline stage to obtain the clustering results.

In the algorithm, the grid cluster and density threshold are adjusted dynamically, which reflects the real-time changes of data and detects and processes isolated grids, which improves the efficiency of the algorithm and proposes a localized algorithm based on the distributed environment. The node-global node processing model further improves

FIGURE 8: Expenditure and cost composition of the medical and health industry.

FIGURE 9: Changes in the medical and health industry.

the processing speed of the algorithm. Experiments and comparisons verify the clustering accuracy and the operating efficiency of the algorithm, as well as the processing efficiency of the algorithm in a distributed environment.

4.2. Pharmaceutical Industry. With the continuous reform of the medical system and the continuous expansion of medical service marketing, the impact of private medical companies on our country's medical service market for-profit first changed the behavior of our country's general medical companies. In areas where profitable hospitals are

relatively dense, nonprofit hospitals will actually be affected by the hospital's profit-making medical behaviors, and it is changing the market value and service quality of our country's medical service products.

Our country's medicine, especially in the field of traditional medicine, has major problems such as multiple facilities, small scale, hardly control of equipment, information asymmetry, low efficiency, high cost, and confusion. For a long time, pharmaceutical companies have not had their own foundation, and modern medical statistics are no exception to the cost reduction and efficiency improvement of the pharmaceutical industry. The integration of social

FIGURE 10: Different variable changes.

medicine logistics and the integration of the modern medicine and medicine movement system is the primary task of reducing the complexity of the medicine industry.

In the process of implementation, the situation in the past was that primary hospitals were unwilling to be trusted by banks due to their weak financial strength, while large hospitals had strong financial strength and did not require banks to be trusted. It is now required that all grassroots hospitals and community hospitals implement “two lines of revenue and expenditure,” that is, all hospital expenditures are included in financial budget management and all revenues are turned over to special financial accounts.

5. Conclusions

This paper considers the degree of difference between two-way clustering and the degree of fusion between clusters and believes that the optimal number of clusters determined by two-way clustering is better than other algorithms. We compared the accuracy of the two-dimensional clustering algorithm. For high-density genetic data, the two-way clustering algorithm can better extract local information while retaining all this information. It can be seen that the two-way clustering algorithm is better than other clustering algorithms. Of course, there are some problems in the research of this paper. Compared with the clustering ensemble algorithm, the biclustering ensemble algorithm has an extra step of reconstructing the bicluster, and usually, heuristic algorithms that are easy to fall into local optimality are used to solve the problem. However, at present, there is no relevant literature on how to reconstruct the biclustering to obtain the global optimum, and it is still in a blank state of research, which points out a clear direction for the next step of research. How to obtain useful information from these

data and solve some new problems is the focus of current research in the big data industry. Therefore, the related research ideas of gene expression data biclustering analysis are extended to other data, which opens up a new direction for the next step of research for discovering or solving new problems.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors state that this paper has no conflicts of interest.

Acknowledgments

This work was supported by the theoretical Innovation Project of the Federation of Social Sciences in Guizhou (no. GZLCLH-2021-179) and the Scientific Research Fund of Guizhou Medical University (no. YJ2020-BK068).

References

- [1] N. Yokoya, P. Ghamisi, and J. Xia, “Open data for global multimodal land use classification: outcome of the 2017 IEEE GRSS data fusion contest,” *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, no. 5, pp. 1–15, 2018.
- [2] A. De Paola, P. Ferraro, S. Gaglio, G. L. Re, and S. K. Das, “An adaptive bayesian system for context-aware data fusion in smart environments,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1502–1515, 2017.
- [3] K. Liu and S. Huang, “Integration of data fusion methodology and degradation modeling process to improve prognostics,”

- IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 1–11, 2016.
- [4] P. Ghamisi, B. Hofle, and X. X. Zhu, “Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 3011–3024, 2017.
 - [5] F. C. Chen and R. Jahanshahi, “NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 99, pp. 4392–4400, 2018.
 - [6] Z. Tao and P. Bonnifait, “Sequential data fusion of GNSS pseudoranges and Dopplers with map-based vision systems,” *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 3, pp. 254–265, 2017.
 - [7] O. F. Beyca, P. K. Rao, Z. Kong, S. T. S. Bukkapatnam, and R. Komanduri, “Heterogeneous sensor data fusion approach for real-time monitoring in ultraprecision machining (UPM) process using non-parametric bayesian clustering and evidence theory,” *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 1033–1044, 2016.
 - [8] W. Liao, X. Huang, and F. V. Coillie, “Processing of multi-resolution thermal hyperspectral and digital color data: outcome of the 2014 IEEE GRSS data fusion contest,” *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2017.
 - [9] P. Ivan, G. Nuno, and P. Nuno, “From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices,” *Sensors*, vol. 16, no. 2, pp. 184–186, 2016.
 - [10] K. M. Nunes, M. V. O. Andrade, A. M. P. Santos Filho, M. C. Lasmar, and M. M. Sena, “Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy,” *Food Chemistry*, vol. 205, no. Aug.15, pp. 14–22, 2016.
 - [11] N. Yokoya, C. Grohnfeldt, and J. Chanussot, “Hyperspectral and Multispectral Data Fusion: a comparative review of the recent literature,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.
 - [12] M. Schmitt and X. X. Zhu, “Data fusion and Remote sensing: an ever-growing relationship,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 4, pp. 6–23, 2016.
 - [13] L. Mou, X. Zhu, M. Vakalopoulou et al., “Multitemporal very high resolution from space: outcome of the 2016 IEEE GRSS data fusion contest,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3435–3447, 2017.
 - [14] D. Tuia, G. Moser, B. Le Saux, B. Bechtel, and L. See, “2017 IEEE GRSS data fusion contest: open data for global multimodal land use classification [technical committees],” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 70–73, 2017.
 - [15] B. Pradhan, M. N. Jebur, H. Z. M. Shafri, and M. S. Tehrani, “Data fusion technique using wavelet transform and taguchi methods for automatic landslide detection from airborne laser scanning data and QuickBird satellite imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1610–1622, 2016.
 - [16] W. Sun, X. Zhang, Z. Zhang, and R. Zhu, “Data fusion of near-infrared and mid-infrared spectra for identification of rhu-barb,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 171, no. Complete, pp. 72–79, 2017.
 - [17] A. Majumder, L. Behera, and V. K. Subramanian, “Automatic facial expression recognition system using deep network-based data fusion,” *IEEE Transactions on Cybernetics*, vol. 48, no. 99, pp. 103–114, 2017.
 - [18] D. Zhou, K. Zhang, A. Ravey, F. Gao, and A. Miraoui, “Online estimation of lithium polymer batteries state-of-charge using particle filter-based data fusion with multimodels approach,” *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 2582–2595, 2016.
 - [19] L. Sun, M. C. Anderson, F. Gao et al., “Investigating water use over the C hoptank R iver W atershed using a multisatellite data fusion approach,” *Water Resources Research*, vol. 53, no. 7, pp. 5298–5319, 2017.
 - [20] J. Radak, B. Ducourthial, V. Cherfaoui, and S. Bonnet, “Detecting road events using distributed data fusion: experimental evaluation for the icy roads case,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 184–194, 2016.
 - [21] A. Rosa, F. Leone, and C. Scattareggia, “Botanical origin identification of Sicilian honeys based on artificial senses and multi-sensor data fusion,” *European Food Research and Technology*, vol. 244, no. 2, pp. 1–9, 2017.
 - [22] V. H. Jaramillo, J. R. Ottewill, R. Dudek, D. Lepiarczyk, and P. Pawlik, “Condition monitoring of distributed systems using two-stage Bayesian inference data fusion,” *Mechanical Systems and Signal Processing*, vol. 87, pp. 91–110, 2017.
 - [23] J. Xiao, R. Stolkin, M. Oussalah, and A. Leonardis, “Continuously adaptive data fusion and model relearning for particle filter tracking with multiple features,” *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2639–2649, 2016.
 - [24] Weibin, Zhang, and Jinjun, “Hybrid short-term prediction of traffic volume at ferry terminal based on data fusion,” *IET Intelligent Transport Systems*, vol. 10, no. 8, pp. 524–534, 2016.
 - [25] Y. Zhang and S. Prasad, “Multisource geospatial data fusion via local joint sparse representation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3265–3276, 2016.
 - [26] S. B. Tsai, M. F. Chien, Y. Xue et al., “Using the fuzzy dematel to determine environmental performance: a case of printed circuit board industry in taiwan,” *PLoS One*, vol. 10, no. 6, Article ID e0129153, 2015.
 - [27] G. Khatwani and P. R. Srivastava, “Impact of information technology on information search channel selection for consumers,” *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 63–80, 2018.