

Research Article

Data Mining of Regional Economic Analysis Based on Mobile Sensor Network Technology

Yucong You 

Department of Economy and Trade, Guangzhou College of Business and Technology, Guangzhou, 510000 Guangdong, China

Correspondence should be addressed to Yucong You; yucong_you@sohu.com

Received 10 September 2021; Revised 25 November 2021; Accepted 10 December 2021; Published 6 January 2022

Academic Editor: Mu Zhou

Copyright © 2022 Yucong You. 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.

With the continuous development of regional economy, the difference of regional economy has also aroused the attention of all walks of life. Due to the limitations of the traditional research methods, the research results are relatively simple and unable to conduct a more comprehensive analysis. The traditional methods include the following: (1) analyze the evolution of regional logistics based on the location Gini coefficient and location quotient of GIS, and reflect the situation of industrial agglomeration from the annual change curve of the location Gini coefficient; (2) use SPSS12.0 software to perform multivariate or event factors, and analyze and calculate the factor score to sum up several principal component factors; and (3) the production function analysis method is used to measure the economies of scale and agglomeration. As an extension, the relationship between the estimated total output and the agglomeration index of the factor input to measure the uniform state of the industrial distribution department is an effective measurement method for the agglomeration economy. In order to promote the sustainable development of regional economy, this paper analyzes the regional economy comprehensively based on the emerging mobile sensor network technology and data mining technology. Firstly, this paper analyzes the location technology of mobile sensor networks based on sequential Monte Carlo, selects the C-means clustering method which is suitable for economic large-sample clustering analysis, and constructs a complete data mining model. Then, the model is used to analyze the economic, social, natural, and educational science and technology indicators of a certain region from 2015 to 2019. The results show that the first principal component weight of economic indicators is the highest proportion of fiscal revenue, which is 0.986. This shows that the role of fiscal revenue in economic indicators is greater. The main index of urban consumption is 72.0, which is the highest. This shows that the population growth rate and the average consumption of urban households in social indicators play a greater role. The first principal component of natural index has the highest weight of pollution emission, which is 0.47, while the second principal component has the highest weight of total energy consumption, which is 0.48. This shows that the pollution emissions and total energy consumption in the natural indicators play a greater role. In the educational science and technology index, the first principal component weight is the highest, which is 0.61. This shows that the education funds play an important role in educational science and technology indicators. Therefore, the data mining model based on mobile sensor network technology can comprehensively and accurately analyze various indicators of regional economy.

1. Introduction

1.1. Background Significance. The regional economy is closely related to the local economic revitalization, which has an important reference role for the government to make a decision. If the economic development of a certain area is good, then it means that the local government's decision is correct, and the government will maintain the original decision; but if there are signs of economic recession, it means

that there is a problem with the government's decision and it will push the government to renew decision-making. Mobile sensor network technology can monitor and collect all kinds of information in real time and has superior information acquisition ability [1, 2]. Data mining technology is a collection of machine learning, artificial intelligence, and statistics-related results and is widely used in various fields [3]. The combination of mobile sensor network technology and data mining technology can improve the efficiency and

quality of data acquisition and mining and obtain effective information quickly. The application of this synthetic technology to the analysis of regional economy does not only provide a new idea for the analysis of regional economy but also can transform the research results into productivity and promote the development of regional economy.

1.2. Related Work. Wireless sensor networks have been used in many mobile applications, such as wildlife tracking and participatory urban sensing. Zhu et al. introduce the impact of industrial agglomeration on the regional economy in a simulated intelligent environment based on machine learning and propose a method to detect the industrial agglomeration index to analyze industrial agglomeration. By establishing an industrial agglomeration index system, the level of manufacturing industry integration is objectively analyzed and the impact of manufacturing industry integration is empirically tested [4]. Lu et al. solve the environmental monitoring problem by developing an event-triggered limited time control scheme for mobile sensor networks. The proposed control scheme can be executed independently by each sensor node and consists of two parts: one part is a finite-time consensus algorithm, and the other part is an event trigger rule. The consensus algorithm is used to enable the position and speed of sensor nodes to quickly track the position and speed of the virtual leader within a limited time. The event trigger rule is used to reduce the update frequency of the controller to save the computing resources of the sensor node. Under the fixed communication topology and the switched communication topology, some stability conditions are derived for the mobile sensor network using the proposed control scheme [2].

The application of data mining technology in the economic field is also a lot of research. Based on data warehouse, data mining, and OLAP technology, Min and Rui propose a modular design decision support system and describe its structure and key technologies [5]. The system test results show that data mining has a good prediction effect on economic prediction. Arends-Kuenning et al. forecast according to the industry that the change of Mexico-US tariff will lead to regional economic growth, and trade liberalization seems to benefit manufacturing industry to a large extent [6]. The increase in national tariffs leads to a decrease in external demand, which also affects prices on the international market, causing them to fall, and the increase in tariffs results in the internal market being the same as the international market after the levy, so in this case, the tariff on the levy has to be shared between internal consumers and exports over, which can affect the regional economy. Their research only analyzes the impact of trade on regional economic development, without considering other factors.

1.3. Innovative Points in This Paper. In order to have a more comprehensive and thorough understanding of the regional economy, provide data support for the relevant government work, and promote the development of regional economy, this paper conducts an in-depth study on regional economic analysis based on mobile sensor network technology and data mining technology. The innovation points of this study

are as follows: (1) based on mobile sensor network positioning technology, this paper selects the appropriate C-means clustering method as the basic model of data mining and constructs a data mining model which can be used to analyze the relevant indicators of regional economy; (2) using this model, the economic, social, natural, educational, and scientific and technological indicators of a certain region are cluster analyzed; and (3) it is concluded that in the development of the regional economy, the financial revenue, the average consumption level of urban households, dye emissions, total energy consumption, and education funds play an important role. If fiscal revenue increases, then the corresponding government investment will increase, which will promote the development of the regional economy; the increase in consumption will further stimulate consumption, and consumption will stimulate regional economic growth; education is the cornerstone of the nation and will cultivate local talents. This will also boost the regional economy. Therefore, in order to promote the development of regional economy, we must start with these factors.

2. Mobile Sensor Network Technology and Data Mining Technology

2.1. Mobile Sensor Network Technology

2.1.1. Routing Protocol. Wireless sensor networks have limited computing, storage, wireless communication, and power supply capabilities. Good routing protocols can reduce energy consumption as much as possible and prolong the network lifetime. There are two functions of routing protocol. The first is to find the optimal path between the source node and the destination node efficiently and accurately. The second is to group the obtained data and forward the grouped data according to its optimized path [7, 8]. For the application scenario with a small number of nodes moving, the routing protocol can be simplified.

According to the network topology, routing protocols can be divided into two types: plane and cluster. Plane routing protocol has the advantages of simple and easy to expand, while cluster routing protocol has the advantages of low communication cost, low energy consumption, and convenient management [9]. Routing is divided into static routing and dynamic routing. Dynamic routing protocols can be divided into distance vector routing protocols and link state routing protocols. The distance vector routing protocol is based on the Bellman-Ford algorithm, mainly including RIP and IGRP; the link state routing protocol is based on the very famous Dijkstra algorithm in graph theory, that is, the shortest first path algorithm. In the distance vector routing protocol, the router transmits part or all of the routing table to its neighboring routers; in the link state routing protocol, the router transmits link state information to all routers in the same area. According to the router's position in the autonomous system, routing protocols can be divided into interior gateway protocols and exterior gateway protocols. There are two types of interdomain routing protocols: exterior gateway protocol and border gateway protocol. EGP is designed for a simple tree topology. It has

obvious shortcomings when dealing with routing cycles and setting routing strategies. It has been replaced by BGP at present. EIGRP is a proprietary protocol of Cisco, which is a hybrid protocol. It not only has the characteristics of the distance vector routing protocol but also inherits the advantages of the link state routing protocol. Various routing protocols have their own characteristics and are suitable for different types of networks. Low-power adaptive cluster layering (LEACH) protocol uses a fully distributed mechanism to generate cluster heads [10]. Each node produces a random number between [0,1]. If the number is less than the threshold $Y(n)$, then the node will not be the cluster head node for a long time. The calculation of threshold is shown in the following formula:

$$Y(n) = \begin{cases} \frac{r}{1 - r * t[\text{mod}(1/r)]}, & n \in J, \\ 0, & \text{other,} \end{cases} \quad (1)$$

where r is the percentage of ordinary nodes becoming cluster heads in each round, t is the number of current rounds, and J is the set of nodes without selected cluster heads in the last $1/r$ round. When $t=0$, the threshold is equal to r . When $t=1$, the threshold is shown in the following formula:

$$Y(n) = \frac{r}{1-r}. \quad (2)$$

The energy consumption formula in the LEACH protocol is the first-order radio mode. The calculation of energy consumption of sending and receiving 1 bit byte by sensor node is shown in formula (3) and formula (4), respectively:

$$E_{Tx}(i, l) = iE_{elec} + i\epsilon l^\alpha, \quad (3)$$

$$E_{Rx}(i) = iE_{elec}, \quad (4)$$

where ϵ, α are the amplification factor and constant of the signal amplifier, respectively. E_{elec}, l is energy consumption and transmission distance.

The LEACH protocol has the advantages of low energy consumption and long network life cycle. However, the signal between multiple nodes will produce interference, which will affect the communication quality and connectivity of the whole network. In addition, the number of cluster head nodes will directly affect the network life cycle.

2.1.2. Positioning Technology. At present, wireless sensor network positioning technology can be divided into two categories: static and mobile. Static wireless sensor network positioning based on distance measurement uses the received signal strength ranging mechanism to calculate the distance between receiving and receiving nodes by using the radio signal transmission model [11, 12]. The localization technology of mobile wireless sensor network nodes is currently a major research hotspot in wireless sensor networks. In mobile wireless sensor networks, nodes are mobile, and nodes with unknown location information in the network can use a certain positioning mechanism and predic-

tion mechanism to determine their own current and underlying coordinates. The existing wireless sensor network node's own positioning algorithm has relatively low positioning accuracy and poor positioning real-time performance, and some algorithms have high complexity, and the ranging error is greatly affected by the environment, which is not suitable for locating mobile nodes. Aiming at the above-mentioned problems, this paper begins to study the positioning technology of mobile wireless sensor network nodes, as shown in the following formula:

$$P(l)[dB] = P(l_0) - 10n \lg\left(\frac{l}{l_0}\right) + R_\sigma, \quad (5)$$

where $P(l), P(l_0)$ are the received power at a distance of l, l_0 , n is the channel attenuation factor, and R_σ is a Gaussian random variable. When the signal transmission time is used for ranging, as shown in the following formula:

$$l = (t - t_0) \times v, \quad (6)$$

where t, t_0 are the time of transmitting and receiving signals, respectively; v is the propagation speed of signals; and l stands for distance. When the time difference of signal arrival is used for ranging, the calculation is shown in the following formula:

$$l = [(t_2 - t) - (t_1 - t_0)] \frac{v_1 - v_2}{v_1 v_2}. \quad (7)$$

Among them, t, t_0 are the time of receiving and transmitting RF signal and t_2, t_1 are the time of receiving and transmitting ultrasonic signal, respectively. They are the propagation speed of RF signal and ultrasonic signal.

Although the accuracy of static wireless sensor network positioning without ranging is low, it does not need hardware support, so the cost and energy consumption are low.

MCL algorithm for mobile wireless sensor network localization based on sequence Monte Carlo improves the accuracy and reduces the cost of localization by using the mobility of nodes, but the storage is very limited [13]. MCB algorithm improves the sampling power and reduces the amount of calculation. The DRL algorithm of nonstatistical mobile wireless sensor network localization can reduce the information flooding between nodes and improve the positioning accuracy [14]. CDL algorithm belongs to centralized location and is easier to implement.

2.1.3. MIMO Technology. Multiple-input multiple-output (MIMO) technology can improve transmission rate and expand coverage without increasing bandwidth and transmitting power [15]. The channel capacity of the MIMO system is shown in the following formula:

$$C = [\min(f, r)]B \log_2\left(\frac{\omega}{2}\right), \quad (8)$$

where f, r are the number of transmitting and receiving antennas, respectively; B, ω represent the signal bandwidth

and the average signal-to-noise ratio of the receiving end, respectively; and C represents channel capacity.

MIMO technology has two functional forms: diversity and multiplexing [16]. The space-time code gain is reduced by space-time coding. The transmitter of spatial multiplexing divides the data into several substreams and sends them at the same time. After the receiver extracts and recovers the signals, it merges them into the original data stream.

Cooperative virtual MIMO technology in wireless sensor networks mainly studies energy consumption characteristics and transmission strategy. The energy consumption characteristics are studied based on the point-to-point transmission model, and the transmission strategy is to explore the scenario of data collection nodes in the detection area.

2.2. Regional Economic Analysis Model

2.2.1. Regression Analysis Model. Regression analysis model uses statistical methods to analyze the causal relationship between variables. If there is only one independent variable and dependent variable in the regression analysis and the relationship is linear, then the analysis model is univariate linear regression [17]. If the independent variables are greater than or equal to two and the relationship is still linear, then the analysis model is multiple linear regression [18]. The univariate linear regression model is shown in the following formula:

$$y = ax + b + \sigma, \quad (9)$$

where a , b are regression coefficients and σ is random disturbance term, which is generally assumed to obey normal distribution. The standard equations solved by multiple linear regression model are shown in the following formula:

$$\begin{cases} \sum y = n\alpha_0 + \alpha_1 \sum x_1 + \alpha_2 \sum x_2, \\ \sum x_1 y = \alpha_0 \sum x_1 + \alpha_1 \sum x_1^2 + \alpha_2 \sum x_1 x_2, \\ \sum x_2 y = \alpha_0 \sum x_2 + \alpha_1 \sum x_1 x_2 + \alpha_2 \sum x_2^2. \end{cases} \quad (10)$$

When using the regression analysis model, it is assumed that dependent variables are affected by independent variables, and their relationship is expressed by regression mathematical model. Then, get the required data, use statistical methods to determine the relationship between dependent variables and independent variables, and finally, get the regression equation, and use the equation to predict the analyzed object. For example, if the goods on the market are within the normal fluctuation range, then the sales of the goods are also within the normal range. If the value of the goods increases, then the sales of the goods will decrease. Based on this change, a regression can be drawn to predict sales.

2.2.2. Principal Component Analysis Model. Principal component analysis will reflect the sample characteristics of multiple index variables through the clustering of principal components into a small number of comprehensive variables [19]. Because a large number of indicators will be involved in

the regional economic analysis, these indicators are not completely independent but are interrelated and jointly reflect the economic development of the region [20]. A large number of related indicators will increase the workload and increase the difficulty of analysis. But simply reducing the indicators will reduce the credibility of the results.

The original indexes in the index matrix describing the comprehensive characteristics of a region are dimensionless.

$$Y_i' = \frac{Y_i - \bar{Y}}{S}, \quad (11)$$

where Y_i , \bar{Y} are the i th original index and sample mean value, S is the standard deviation. Then, calculate the correlation coefficient matrix of the index, as shown in the following formula:

$$R_{zp} = \frac{1}{n} \sum_{i=1}^n Y_{iz} Y_{ip}. \quad (12)$$

The third step is to calculate the eigenvalue and eigenvector and solve the characteristic equation as shown in the following formula:

$$|R - \omega P| = 0. \quad (13)$$

The fourth step is to calculate the contribution rate of principal components, as shown in the following formula (14):

$$Q_p = \frac{\omega_p}{\sum_{i=1}^p \omega_i}. \quad (14)$$

Finally, the score of the principal component is calculated, as shown in formula (15). After calculation, the principal component is arranged in the order of score.

$$F = v_1 y_1 + v_2 y_2 + \dots + v_n y_n, \quad (15)$$

where v_n is the common factor weight.

2.2.3. Prediction Analysis Model. The prediction analysis model can explain the relationship between the prediction object and the related factors. The main prediction and analysis model used is the grey prediction model, which analyzes the mathematical relationship between the factors themselves and the factors according to the behavior characteristic data and information of the specific gray system [21]. The grey prediction is based on the grey system theory and uses the differential equation model to predict the change of the eigenvalue of the system and estimate the occurrence time of the abnormal value of the eigenvalue [22, 23].

Sequence grey prediction predicts the eigenvalues of variables in a future period according to the existing eigenvalues. According to the occurrence time of the current abnormal value, the catastrophe grey prediction can predict the occurrence time of the future abnormal value. The grey prediction of seasonal disaster is very special, and the abnormal value of its eigenvalue often appears in a specific period

of time. Topological grey prediction can be used to predict a group of data without changing rules. However, the grey prediction of the system is different from the above-mentioned prediction methods, which is aimed at the overall prediction of the system [24].

2.3. Data Mining Technology and Process

2.3.1. Data Mining Technology. Neural network method uses a large number of neurons, which are connected into a network to achieve a large number of parallel computing [25]. The neural network algorithm needs to construct a threshold object. If the sum of a set of logical unit variables is not less than a given threshold, a value will be output [26]. Suppose that the input value is x_1, x_2, \dots, x_n and the weighted coefficient of the input value is g_1, g_2, \dots, g_n , the variable summation is shown in the following formula:

$$M = \sum_{i=1}^n (x_i * g_i). \quad (16)$$

Naive Bayes method is also one of the common data mining methods. In order to use naive Bayes classification, the assumption of independence must be satisfied. The expression of naive Bayes classifier is shown in the following formula:

$$C = \arg \max_{C_j \in C} P(C_j) \cdot \prod_{i=1}^n P(t_i | C_j), \quad (17)$$

where C, C_j represent the set of categories and category j , respectively; $P(C_j)$ denotes the probability of occurrence of category j ; and $P(t_i | C_j)$ represents the probability of occurrence of term t_i in category C_j .

K -nearest neighbor algorithm finds k -nearest neighbor documents which are closest to a given test document in the training set and then takes the categories of adjacent documents as candidate categories of the test documents and scores them according to the classification. The decision rule of scoring is shown in the following formula:

$$P(\bar{c}_t, l_j) = \sum_{u_j \in \text{KNN}} \text{sim}(\bar{c}_t, \bar{u}_j) y(\bar{u}_j, l_j) - a, \quad (18)$$

where $y(\bar{u}_j, l_j)$ is 1 or 0; $\text{sim}(\bar{c}_t, \bar{u}_j)$ is the similarity between test document \bar{c}_t and training document \bar{u}_j ; and a is the threshold of binary decision.

Based on the minimum Euclidean distance, the C -means clustering method divides the point set (x_1, x_2, \dots, x_n) into m clusters, calculates the Euclidean distance between the center of the set and each point, and generates the distance set L . Find m points (c_1, c_2, \dots, c_m) in the distance set as the original center point; then, the distance from other vector points to the M original center points is shown in the following formula:

$$l(x_i, c_j) = \sqrt{\sum_{m=1}^n (x_{im} - c_{jm})^2}. \quad (19)$$

Then, adjust the cluster center according to the following formula:

$$c_{jm} = \frac{\sum_{x_{jm} \in c_j} x_{jm}}{N_j}. \quad (20)$$

2.3.2. The Basic Process of Data Mining. Data mining process generally includes three main stages: data preparation, data mining, and result expression and interpretation [27]. Data preparation stage can be subdivided into data integration, selection, and preprocessing. According to the specific needs of the search-related raw data, select the targeted data. Data preprocessing uses cleaning and transformation methods to process noisy and random data, which can improve the quality of data mining [28].

In the stage of data mining, it is necessary to select appropriate tools for data mining and then to determine the appropriate tools for data mining. When selecting data mining tools, we need to consider the characteristics of data and the requirements of the actual running system.

According to the final decision-making purpose, the information mined needs to be interpreted and evaluated by experts in related fields. The process of knowledge discovery in database needs to be evaluated according to the initial requirements and then optimized after analysis. Finally, the most valuable information is visualized and conveyed to decision-makers.

2.3.3. System Structure of Data Mining. The structure of data mining system usually includes data source and its server, data mining engine, pattern evaluation, graphical user interface, and knowledge base [29]. Data source mainly provides data needed by data mining, which can be database, data warehouse, or other types. The data source server can filter and clean the data in the data source according to the user requirements and integrate the required target data sets.

Data mining engine is the core of the system, which has the analysis functions of characterization, association mining, classification, and clustering. Pattern evaluation can remove useless patterns and retain interesting patterns according to interest threshold. GUI presents interesting patterns to users in a more intuitive way, that is, visualization of mining results. The role of knowledge base is embodied in storing knowledge, guiding data mining, pattern interpretation, and evaluation.

3. Experiments on Data Mining of Regional Economic Analysis

3.1. Mobile Sensor Network Localization Based on Sequential Monte Carlo

3.1.1. Prediction and Filtering. The MCL algorithm is divided into two stages: prediction and filtering, and the node motion in the prediction stage obeys the random model. The maximum velocity of node motion is v_{\max} . At moment e , the circle is drawn with k_{e-1} as the center and the maximum velocity as the radius.

$$p(k_e|k_{e-1}) = \begin{cases} \frac{1}{(\pi v_{\max}^2)}, & \text{if } d(k_e, k_{e-1}) < v_{\max}, \\ 0, & \text{if } d(k_e, k_{e-1}) \geq v_{\max}, \end{cases} \quad (21)$$

where $d(k_e, k_{e-1})$ is the Euclidean distance between k_e, k_{e-1} . The greater the speed of the node, the larger the predicted area.

In the filtering stage, the node will filter out the sample points which do not conform to the actual situation according to the new observation value. If the number of remaining sample points after filtering is too small, it is necessary to repeat the prediction stage and the correction stage until the number meets the requirements.

3.1.2. Important Sampling. MCL algorithm uses importance sampling, assuming that all samples are obtained from the normalized important function π independent sampling. After measuring the weight of each sample, the posterior probability distribution is estimated. In order to simplify the calculation, the prior distribution is selected as the recommended distribution. The calculation of important functions is shown in the following formula:

$$\pi(k_e|q_0, q_1, \dots, q_e) = p(k_0) \prod_{n=1}^e p(k_n|k_{n-1}). \quad (22)$$

3.1.3. Resampling. With the increase of the variance of important weights, the important samples will degenerate, so it is necessary to resample to eliminate the small normalized full-time sample points. In MCL algorithm, the weight is either 1 or 0, so it is necessary to keep enough effective samples to ensure that the effective sampling scale is not less than the total number of samples for repeated sampling.

In the process of resampling, the sample points collected in each round are saved to the same set until the requirements are met. In this way, the computational complexity of resampling is reduced and the robustness is avoided. In the sampling process, in order to reduce the computational difficulty, we divide the weight of the MCL into 1 or 0, compare the weight of the sample with it, classify the sample, and divide it into two sets.

3.2. Establishment of Regional Economic Data Mining Model

3.2.1. Application of C-Means Clustering Method in Regional Economy. The C-means clustering method is based on the minimum Euclidean distance, and its specific calculation process has been described in detail in Section 2, and will not be repeated here. C-means clustering method has the advantages of less calculation, stronger robustness, and faster processing speed, which is suitable for clustering analysis of economic samples.

3.2.2. Selection of Samples and Indicators. The principles of comprehensiveness, objectivity, periodicity, and independence should be followed in the selection of indicators. The cluster index of regional economy must be able to reflect the comprehensive content of sustainable development of

regional economy, so it should include not only economic indicators but also social, natural, educational, and technological indicators. Sample independence means that the samples are independent of each other, and there is no pairing relationship. Both samples are required to be from a normal distribution, and the mean is required to be a descriptive statistic that is meaningful for testing, such as the comparison of the wages of boys and girls. Objectivity means that the selection conditions of these samples are fixed and cannot be regarded as transfer; periodicity means that the indicators have certain rules; the comprehensive principle means that the basic information reflected by the samples is more comprehensive.

In the economic indicators, we need to consider the per capita income and consumption, the proportion of the secondary and tertiary industries, and the proportion of fiscal revenue. In the social indicators, we need to consider the average consumption of urban and rural households, population growth, aging, and social infrastructure construction. The natural indicators include pollution emissions and treatment costs, the consumption of various types of energy, and the storage of various resources. The indicators of educational science and technology include educational funds, the number of students and teachers, scientific research funds, and the number of patent applications.

In addition, because the data in the indicators change greatly, the variables with small absolute values may not show their role, and the clustering effect will also be affected. In order to avoid this kind of situation as much as possible, each data is dimensionless, and the processing method is shown in formula (11).

3.2.3. Modeling. Based on the mobile sensor network positioning technology, this paper models the factors under the four indicators of a region from 2015 to 2019 and analyzes the impact of economy, society, nature, education, and science and technology on the regional economic development. There are two types of hierarchy-based clustering methods: (1) aggregate hierarchical clustering. Initially, each object is a cluster, and then, these atomic clusters are merged according to their similarity to each other. Most hierarchical methods fall into this category, the main difference between them being the different definitions of similarity between clusters; (2) divided hierarchical clustering, which is the opposite of the above process. The processing flow of the data mining model is as follows.

As shown in Figure 1, input all the data of each attribute index, preprocess it, and then, put it into storage. The C-means clustering method and association analysis are used for data mining, and the mining results are visualized and compared.

4. Discussion on Data Mining Results of Regional Economic Analysis

4.1. Cluster Analysis Results of Economic Indicators. The data of per capita income and consumption, the proportion of the secondary and tertiary industries, and the proportion of fiscal revenue in the region from 2015 to 2019 are input into

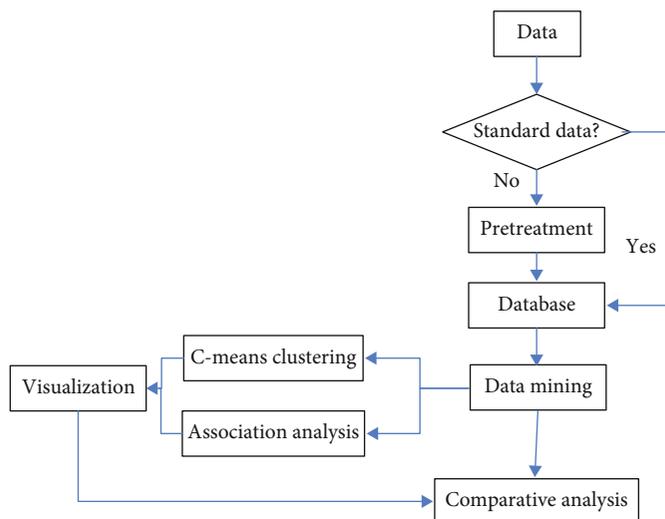


FIGURE 1: Processing flow chart of the data mining model.

TABLE 1: Economic index data.

Index	2015	2016	2017	2018	2019
Per capita income (yuan)	15936	18225	21740	23819	25084
Per capita consumption (yuan)	6798	8976	10035	11652	13447
Proportion of secondary industry (%)	42.3	47.1	50.8	53.6	55.3
Proportion of tertiary industry (%)	26.2	28.9	32.4	35.9	38.6
Proportion of fiscal revenue (%)	13.8	14.3	15.4	15.8	16.1

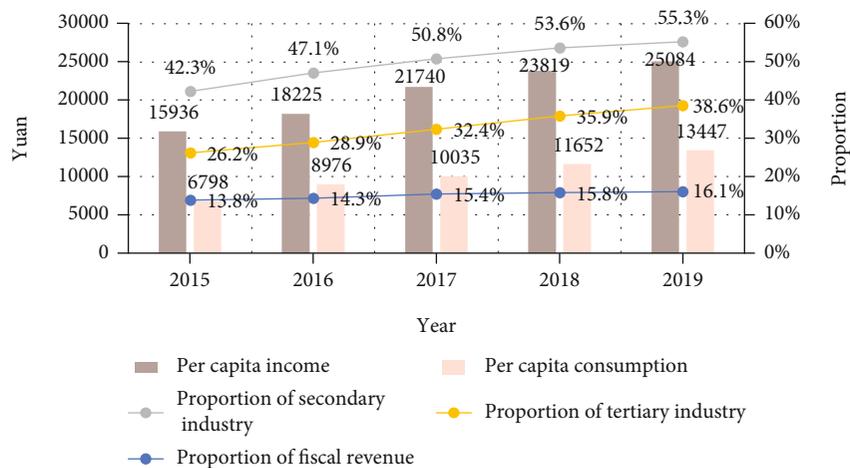


FIGURE 2: Changes in economic indicators.

the data mining model based on mobile sensor network positioning technology, and cluster analysis is carried out.

As shown in Table 1, the per capita income and consumption, the proportion of the secondary and tertiary industries, and the proportion of fiscal revenue in the region from 2015 to 2019 are increasing year by year.

As shown in Figure 2, the per capita consumption level in the region has increased by 97.81% and the per capita income level has increased by 57.4%. The proportion of the secondary industry increased by 30.73%, that of the tertiary industry increased by 47.33%, and the proportion of fiscal

TABLE 2: Calculation results of principal components of economic indicators.

Principal component	Characteristic value	Variance contribution rate (%)	Cumulative variance contribution rate (%)
1	6.387	63.872	63.872
2	2.883	28.829	92.701
3	0.539	5.394	98.095
4	0.186	1.859	99.954
5	0.005	0.046	100.000

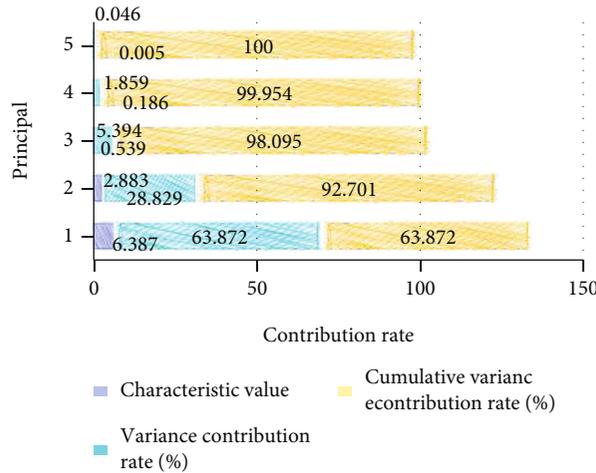


FIGURE 3: Contribution rate of principal component variance of economic indicators.

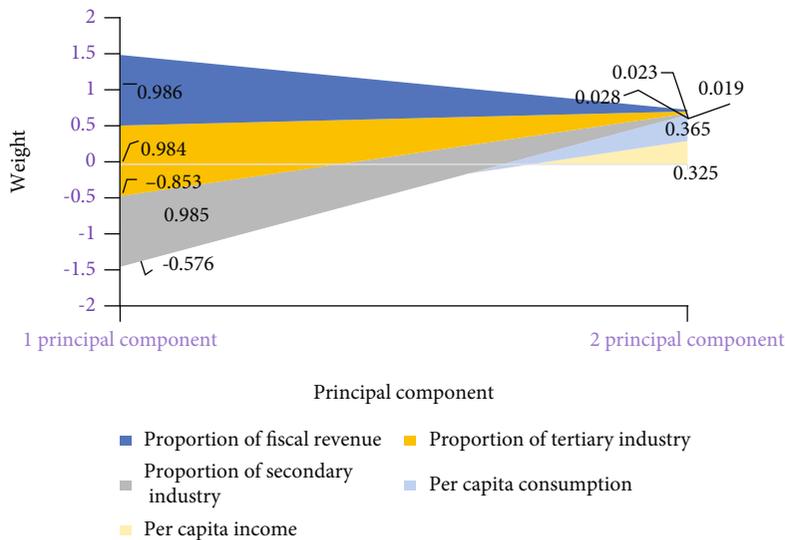


FIGURE 4: Weight values of main components of economic indicators.

revenue increased by 16.67%. This shows that the regional economy presents a stable and sustained growth trend.

The principal component analysis model is used to analyze the data, and the eigenvalues, variance contribution rate, and cumulative variance contribution rate of the indicators such as per capita income and consumption, the proportion of the secondary and tertiary industries, and the proportion of fiscal revenue are calculated.

As shown in Table 2, the eigenvalue of the first principal component is 6.387, and its variance contribution rate is 63.872%. The eigenvalue of the second principal component is 2.883, and its variance contribution rate is 28.829%. The eigenvalue of the third component is 0.539, and its variance contribution is 5.394; the eigenvalue of the fourth component is 0.186, and its variance contribution is 1.859; the eigenvalue of the fifth component is 0.005, and its variance contribution is 0.046.

As shown in Figure 3, the cumulative variance contribution rate of the first two principal components reaches

92.701%, which indicates that only the numerical changes of the first two principal components need to be analyzed.

As shown in Figure 4, the index with the highest weight in the first principal component is the proportion of fiscal revenue, which is 0.986. The second is the proportion of the secondary industry and the tertiary industry. In the second principal component, the index of per capita consumption with the highest weight is 0.365. This shows that the financial revenue and the proportion of the secondary and tertiary industries play a greater role in the economic indicators affecting the regional economic development.

4.2. Results of Cluster Analysis of Social Indicators. The data of average consumption of urban and rural households, social infrastructure construction projects, population growth, and aging population proportion in the region from 2015 to 2019 are input into the data mining model based on mobile sensor network positioning technology, which is created in this paper, for clustering analysis.

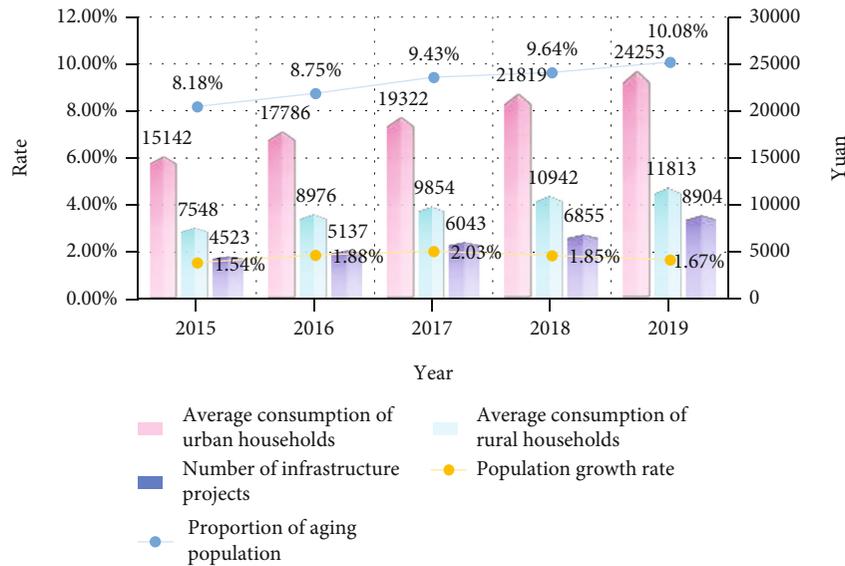


FIGURE 5: Changes in social indicators.

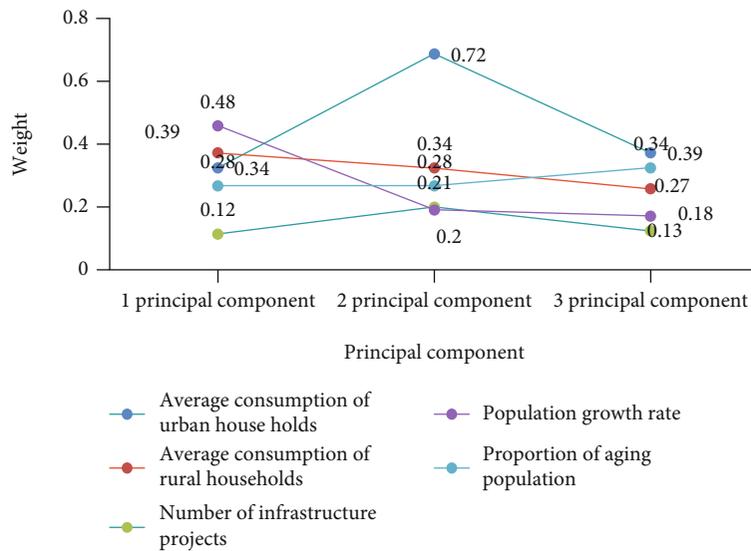


FIGURE 6: The weight value of main components of social indicators.

As shown in Figure 5, the average consumption of urban and rural households and the number of social infrastructure construction projects in the region are increasing year by year, with a cumulative growth of 60.17%, 56.51%, and 96.86%, respectively, in the five years. The population growth rate showed a trend of first growth and then decline, with a cumulative growth rate of 8.44% over the past five years. The proportion of the aging population shows a trend of increasing year by year, with a cumulative increase of 23.23% in the past five years.

According to the clustering results of social indicators, the weights of social indicators are calculated.

As shown in Figure 6, the population growth rate with the highest weight is 0.48 in the first principal component, 0.72 in the second principal component, and 0.39 in the third principal component. This shows that the population

TABLE 3: Calculation results of principal components of natural indexes.

Principal component	Characteristic value	Variance contribution rate (%)	Cumulative variance contribution rate (%)
1	3.547	35.472	35.472
2	2.381	23.809	59.281
3	1.796	17.963	77.224
4	1.056	10.563	87.807
5	0.404	4.041	91.848
6	0.311	3.112	94.960
7	0.257	2.569	97.529
8	0.114	1.142	98.671
9	0.087	0.871	99.542
10	0.046	0.458	100.000

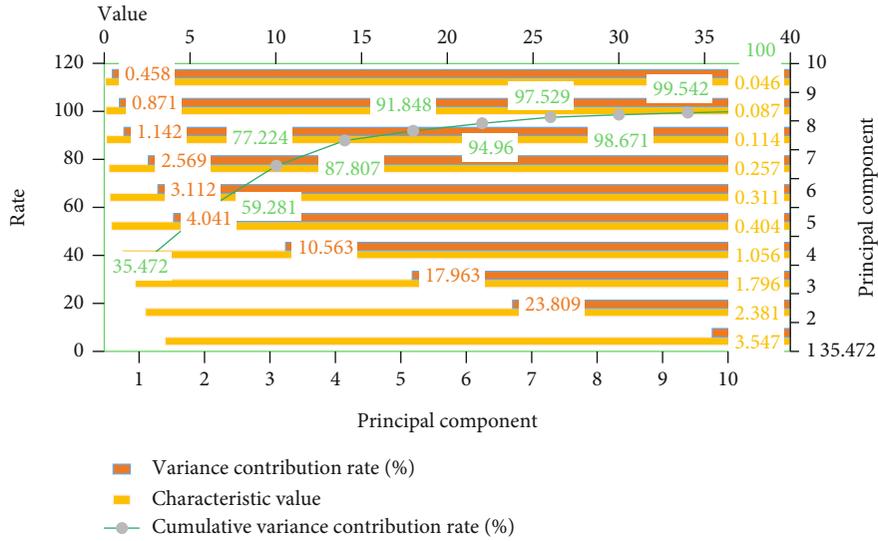


FIGURE 7: Contribution rate of principal component variance of natural indicators.

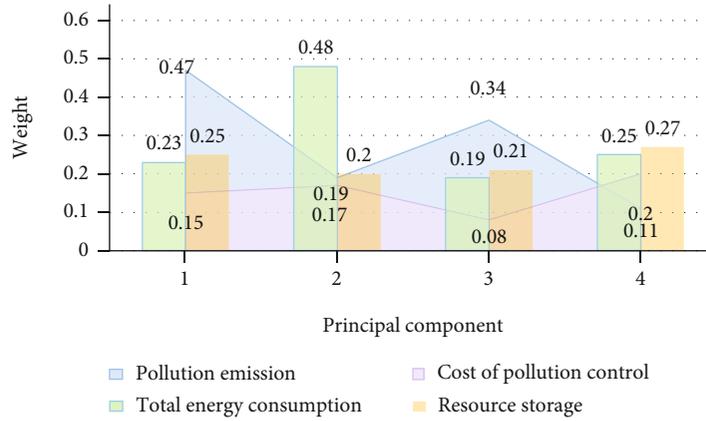


FIGURE 8: Weight values of main components of natural indexes.

growth rate and the average consumption of urban households play a greater role in the social indicators affecting the regional economic development.

4.3. Results of Natural Index Cluster Analysis. The data of pollution emissions and treatment costs, energy consumption, and resource storage in the region from 2015 to 2019 are input into the data mining model based on mobile sensor network positioning technology, and cluster analysis is conducted. Because there are many subindexes in natural indexes, it is necessary to reduce the dimension. The principal component analysis model is used to analyze the data and calculate the eigenvalue, variance contribution rate, and cumulative variance contribution rate.

As shown in Table 3, the eigenvalue of the first principal component is 3.547, and its variance contribution rate is 35.472%. The eigenvalue of the second principal component is 2.381, and its variance contribution rate is 23.809%. The eigenvalue of the third principal component is 1.796, and its variance contribution rate is 17.963%. The eigenvalue of the fourth principal component was 1.056, and its variance contribution rate was 10.563%.

TABLE 4: Calculation results of principal components of educational science and technology indicators.

Principal component	Characteristic value	Variance contribution rate (%)	Cumulative variance contribution rate (%)
1	5.045	50.453	50.453
2	2.089	20.891	71.344
3	1.800	18.003	89.347
4	0.692	6.924	96.271
5	0.373	3.729	100.000

As shown in Figure 7, the cumulative variance contribution rate of the first four principal components reaches 87.807%, which indicates that only the numerical changes of the first four principal components need to be analyzed.

As shown in Figure 8, the pollution emission with the highest weight in the first principal component is 0.47; the highest weight in the second principal component is the

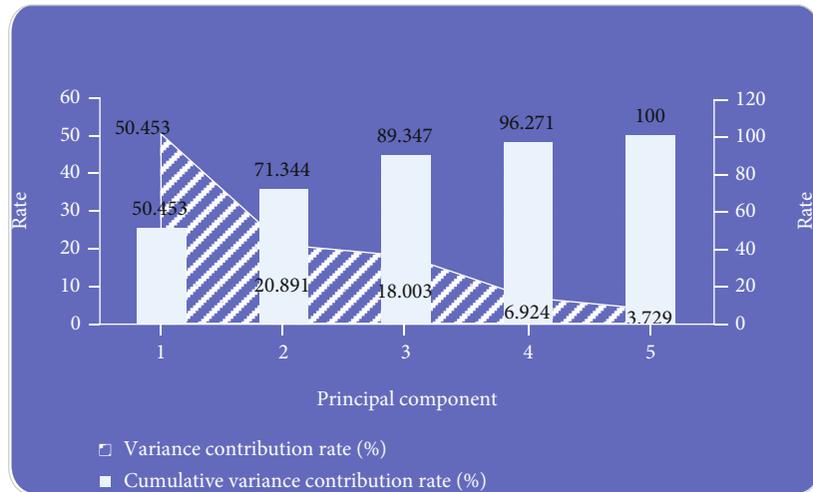


FIGURE 9: Contribution rate of principal component variance of educational science and technology indicators.

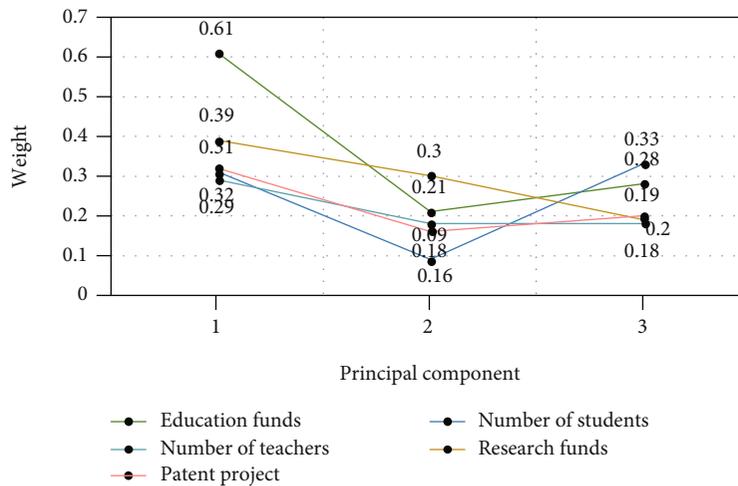


FIGURE 10: Main component weight values of educational science and technology indicators.

total energy consumption, which is 0.48; and the pollution emission with the highest weight in the third principal component is 0.34. In the fourth principal component, the highest weight is energy storage, which is 0.27. This shows that the two factors of pollution emission and total energy consumption play an important role in the natural indicators of regional economic development.

4.4. Cluster Analysis Results of Educational Science and Technology Indicators. The data of education funds, the number of students and teachers, scientific research funds, and patent applications in this region from 2015 to 2019 are input into the data mining model based on mobile sensor network positioning technology, and cluster analysis is conducted.

As shown in Table 4, the eigenvalue of the first principal component is 5.045, and its variance contribution rate is 50.453%. The eigenvalue of the second principal component is 2.089, and its variance contribution rate was 20.891%. The eigenvalue of the third principal component is 1.8, and its variance contribution rate is 18.003%.

As shown in Figure 9, the cumulative variance contribution rate of the first three principal components reaches 89.347%, which indicates that only the numerical changes of the first three principal components need to be analyzed.

As shown in Figure 10, in the first principal component, the highest weight is education funding, which is 0.61; in the second principal component, the highest weight is scientific research funding, which is 0.3; in the third principal component, the number of students is the highest, which is 0.33. This shows that the education funds, scientific research funds, and the number of students play an important role in the development of the regional economy.

5. Conclusions

The C-means clustering method has the advantages of less calculation, stronger robustness, and faster processing speed, which is suitable for clustering analysis of economic samples. Therefore, this paper combines the mobile sensor network localization technology based on sequential Monte Carlo

and C-means clustering method to establish the data mining model. Following the principles of comprehensiveness, objectivity, periodicity, and independence, the indicators of economy, society, nature, and education science and technology which can reflect the comprehensive content of regional economic sustainable development are selected.

Through cluster analysis and principal component analysis of the relevant index data of a certain region, we find that among many factors, financial income, average consumption level of urban households, pollution emissions, total energy consumption, and education funds are particularly important. This paper accomplishes the following: (1) a comprehensive analysis of the regional economy based on emerging mobile sensor network technology and data mining techniques; (2) production function analysis is used to measure economies of scale and agglomeration economies. The relationship between total output and the agglomeration index of factor inputs is derived as an effective measure of agglomeration economy to measure the state of unity of industry distribution sectors.

Due to the limited time and knowledge, there are still some deficiencies in this study. In the selection of sample data, only the data of nearly five years are used, but the growth of some development indicators shows a long periodicity, and the data of five years may be difficult to reflect its change law. The C-means clustering method is suitable for the analysis of economic samples, but its own shortcomings are ignored in this study. Therefore, in the future research work, we need to increase the selection range of sample data and improve the C-means clustering method to make its performance better.

Data Availability

No data were used to support this study.

Conflicts of Interest

There are no potential competing interests in this paper.

Authors' Contributions

The author has seen the manuscript and approved its submission.

Acknowledgments

This work was supported by the 2020 Guangdong Province Ordinary University Characteristic Innovation Project "Research on the Path of Guangdong Foreign Trade Enterprises' Response to the Epidemic from the Perspective of Digital Economy" (Project No. 2020WTSCX114); the 2020 Guangdong Provincial Science and Technology Projects "Guangdong Province Cold Chain Standardization," the Joint Co-construction of Engineering Technology Research Center" (Project No. 2020440121000082); and the 2020 Disciplinary Co-construction Project of the "13th Five-Year Plan" of Philosophy and Social Sciences of Guangdong Province in 2020 "Research on the Digital High-quality Develop-

ment Path of Guangdong Foreign Trade Enterprises Under the Visual Threshold of Post-epidemic" (Project No. GD20XYJ23) funding.

References

- [1] G. Dartmann, H. B. Song, and A. Schmeink, *Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things*, Elsevier, 2019.
- [2] Q. Lu, Q. L. Han, B. Zhang, D. Liu, and S. Liu, "Cooperative control of mobile sensor networks for environmental monitoring: an event-triggered finite-time control scheme," *IEEE Transactions on Systems, Man, and Cybernetics Part B, Cybernetics*, vol. 47, no. 12, pp. 4134–4147, 2017.
- [3] V. Chaurasia and S. Pal, "A novel approach for breast cancer detection using data mining techniques," *Socientia Electronic Publishing*, vol. 3297, no. 1, pp. 2320–9801, 2017.
- [4] L. Zhu, Z. Yu, and H. Zhan, "Impact of industrial agglomeration on regional economy in a simulated intelligent environment based on machine learning," *Access*, vol. PP (99), pp. 1–1, 2020.
- [5] Z. Min and Q. Rui, "Data mining and economic forecasting in DW-based economical decision support system," *International journal of reasoning-based intelligent systems*, vol. 11, no. 4, pp. 300–307, 2019.
- [6] M. Arends-Kuenning, K. Baylis, and R. Garduno-Rivera, "The effect of NAFTA on internal migration in Mexico: a regional economic analysis," *Applied Economics*, vol. 51, no. 10, pp. 1052–1068, 2019.
- [7] S. K. Singh and P. Kumar, "A load balancing virtual level routing (LBVLR) using mobile mule for large sensor networks," *Journal of Supercomputing*, vol. 75, no. 11, pp. 7426–7459, 2019.
- [8] I. Butun, P. Österberg, and H. Song, "Security of the Internet of Things: vulnerabilities, attacks, and countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 616–644, 2020.
- [9] S. Wan, "Topology hiding routing based on learning with errors," *Concurrency and Computation: Practice and Experience*, no. 6, p. e5740, 2020.
- [10] N. Ganganath, C. T. Cheng, and C. K. Tse, "Distributed anti-flooding algorithms for dynamic coverage of mobile sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1795–1805, 2016.
- [11] S. Sharma, D. Puthal, S. K. Jena, A. Y. Zomaya, and R. Ranjan, "Rendezvous based routing protocol for wireless sensor networks with mobile sink," *The Journal of Supercomputing*, vol. 73, no. 3, pp. 1168–1188, 2017.
- [12] S. Sun, M. Kadoch, L. Gong, and B. Rong, "Integrating network function virtualization with SDR and SDN for 4G/5G networks," *IEEE Network*, vol. 29, no. 3, pp. 54–59, 2015.
- [13] S. H. Semnani and O. A. Basir, "Multi-target engagement in complex mobile surveillance sensor networks," *Unmanned Systems*, vol. 5, no. 1, pp. 31–43, 2017.
- [14] G. Yogarajan Revathi et al., "Nature inspired discrete firefly algorithm for optimal mobile data gathering in wireless sensor networks," *Wireless Networks*, vol. 24, no. 8, pp. 2993–3007, 2018.
- [15] K. Karunanithy and B. Velusamy, "CSDGP: cluster switched data gathering protocol for mobile wireless sensor networks," *IET Communications*, vol. 13, no. 18, pp. 2973–2985, 2019.

- [16] D. Chen, Q. Lu, D. Peng, K. Yin, C. Zhong, and T. Shi, "Receding horizon control of mobile robots for locating unknown wireless sensor networks," *Assembly Automation*, vol. 39, no. 3, pp. 445–459, 2019.
- [17] S. M. H. Tabatabaie, H. Tahami, and G. S. Murthy, "A regional life cycle assessment and economic analysis of camelina biodiesel production in the Pacific Northwestern US," *Journal of Cleaner Production*, vol. 172, part 3, pp. 2389–2400, 2018.
- [18] H. T. Van and D. Ushakov, "Analysis of economic imbalances under the conditions of regional agrarian markets' integration," *E3S Web of Conferences*, vol. 175, no. 1, pp. 13034–13034, 2020.
- [19] Y. Xiong, H. Wang, and X. Gao, "Evaluation of regional economic efficiency in Hubei province based on cross data envelopment analysis algorithm," *Journal of Interdisciplinary Mathematics*, vol. 20, no. 8, pp. 1715–1729, 2017.
- [20] H. Kasseeah, "Investigating the impact of entrepreneurship on economic development: a regional analysis," *Journal of Small Business and Enterprise Development*, vol. 23, no. 3, pp. 896–916, 2016.
- [21] T. P. Sari and F. Rahmawati, "The analysis of excellent economic sector in regional economic building in Kediri City 2012–2015," *KnE Social Sciences*, vol. 3, no. 3, pp. 91–91, 2018.
- [22] B. Crane, C. Albrecht, M. K. Duffin, and C. Albrecht, "China's special economic zones: an analysis of policy to reduce regional disparities," *Regional Studies, Regional Science*, vol. 5, no. 1, pp. 98–107, 2018.
- [23] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Biotechnology Journal*, vol. 15, no. C, pp. 104–116, 2017.
- [24] S. B. Tsai, Y. C. Lee, and J. J. Guo, "Using modified grey forecasting models to forecast the growth trends of green materials," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 228, no. 6, pp. 931–940, 2014.
- [25] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.
- [26] C. Helma, T. Cramer, S. Kramer, and L. De Raedt, "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 4, pp. 1402–1411, 2004.
- [27] X. Zhao and J. Tang, "Crime in urban areas," *Acm Sigkdd Explorations Newsletter*, vol. 20, no. 1, pp. 1–12, 2018.
- [28] J. H. Kao, T. C. Chan, F. Lai et al., "Spatial analysis and data mining techniques for identifying risk factors of out-of-hospital cardiac arrest," *International Journal of Information Management*, vol. 37, no. 1, pp. 1528–1538, 2017.
- [29] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart disease diagnosis and prediction using machine learning and data mining techniques: a review," *Advances in Computational Ence and Technology*, vol. 10, no. 7, pp. 2137–2159, 2017.