Scholar Publishing Group
Distributed Processing System
https://doi.org/10.38007/DPS.2021.020103
ISSN 2790-0916 Vol. 2, Issue 1: 24-41

Gravity Index Algorithm of Industrial Spatial Agglomeration Center Based on Distributed Collaboration

Laura Hoekstra*

Vrije University Amsterdam, Netherlands

**corresponding author*

Keywords: Distributed Collaboration, Collaborative Industry, Spatial Agglomeration, Central Gravity Index

Abstract: In order to measure the degree of industrial agglomeration as well as the spatial location of industrial agglomeration, this paper proposes the central gravity index method by integrating two kinds of industrial spatial agglomeration measurement methods based on output value and distance. The purpose of this research is to design a central gravity index algorithm for distributed collaborative industrial space integration. In the present study, the data of the industrial data of North - West, Tianjin, Hebei, Yangtze River Delta and Zhu Jiang Delta were surveyed, and the data were verified and verified using the industrial accumulation measurement and the central gravity index algorithm. The results show that SVM has the strongest learning ability, and has good performance in precision (84%), recall (84%) and f1measure (84%); the learning ability of KNN is second only to SVM, and precision (82%), recall (82%) and f1measure (82%) have good performance. It is concluded that the central gravity index algorithm of spatial agglomeration in this study improves the existing industrial spatial agglomeration measurement methods, and makes up for the defects that the output value measurement method ignores the spatial distribution, while the distance based method does not use enough output weight and cannot determine the geographical location; through the geographical location of virtual center enterprises, it provides a new method for tracking the geographical path of industrial transfer, and contribute to the research of regional integration.

1. Introduction

With the deepening of financial liberalization and globalization, countries all over the world have established modern financial systems, and as the strategic goal of economic development, they have built international financial centers. Typical international financial centers include London, New York, Tokyo, etc. These three cities are considered as international financial clusters. Most of the world's banks, securities companies, insurance companies, investment companies have headquarters in these three cities. The financial industry is constantly forming. Because the differences of financial development in the four major economic regions are more and more obvious, we can study the financial integration phenomenon by analyzing the differences. In order to promote regional development, the integration of financial industry has become one of the important means to adjust regional development.

The evolution mechanism of the spatial distribution of producer services in the Yangtze River Delta is investigated by analyzing the integration and industrial characteristics of the producer service. The general evolution rules of the producer service industry can be obtained from existing literature. Based on the actual circumstances of the Delta in the US youth, the Changjiang Delta producers can analyze the evolution mechanisms of specificity, differentiation and specialization. This paper proposes a reasonable and effective policy proposal for optimizing the industrial spatial layout of the Yangtze Delta region by studying the spatial integration of the producer service industry and the mechanism of the evolution of regional distribution. Each region has its own special circumstances. The Yangtze River Delta region provides reasonable and effective policy guidance from the national level in accordance with China's national situation, leading to the theoretical basis of the industrial spatial distribution in the Yangtze River Delta region.

In the research on the gravity index design of industrial agglomeration center, Wang X S constructed the theoretical model and econometric model of the impact of the internal agglomeration index and external agglomeration degree on the market scale effect, and conducted an in-depth empirical study on the impact of spatial agglomeration on the intensity of industrial pollution emissions. He thinks that spatial agglomeration reduces the emission intensity of industrial COD and industrial dust pollution per unit GDP; internal agglomeration has a negative impact on the industrial COD emission gap, and thus narrowing the gap; external agglomeration has a positive impact, which shows that with the enhancement of agglomeration externality, the importance of location is more reflected in the correlation between different regions, that is, the market scale effect is stronger; and the external agglomeration has a negative impact on the gap of industrial COD emission The concentration degree has a significant impact on the threshold effect of industrial pollution emission intensity. The accuracy of his method is low [1]. Fuxiang W constructs the model framework of quadratic utility quasi linear preference utility function, and conducts model derivation and numerical simulation on the causes of this spatial imbalance from two dimensions of individual and region. His research found that, in the long run, industrial spatial layout has certain threshold restrictions on the proportion of differentiated labor portfolio. The dilemma of China's industrial spatial layout is mainly due to the deviation of market optimal agglomeration from social optimal agglomeration, and the lack of intermediary role between global value and domestic value in East China. His method is not stable [2]. In the spatial econometric analysis, Ye C takes into account the two factors that affect the autocorrelation between economy and geographical space, and constructs the economic weight matrix to analyze the panel data of 77 cities in the Yangtze River economic belt. His research shows that the Yangtze River Delta has the characteristics of financial agglomeration, which has a significant positive impact on population urbanization and economic urbanization. Foreign direct investment, fixed assets investment, industrial development, government expenditure and technological progress have promoted the urbanization of population and economy in our region. From the regional perspective, the impact of financial agglomeration on urbanization is different and unbalanced. The financial agglomeration in the western region has little positive impact on population urbanization and economic urbanization, while the financial agglomeration in the eastern and central regions has a greater positive impact on population urbanization and economic urbanization. His method is not practical [3].

In this study, firstly, the definition of industrial spatial distribution and the significance of industrial agglomeration are introduced, and then various theories of industrial agglomeration theory are explained in detail. The main algorithms in this study are industrial agglutination measurement method and central gravity index algorithm. In this study, the calculation and verification of the central gravity index algorithm will select the industrial integration data of three special economic circles in China as the research object. Combined with the results, this paper analyzes the aggregation of decentralized common industry, the design and analysis of spatial aggregation center gravity index algorithm, the analysis of classification model results, the evaluation analysis of results and the expectation of central gravity algorithm. The central gravity index algorithm of this study draws the conclusion that it is designed to improve the spatial agglomeration of decentralized common industries, better realize regional integration and promote economic development.

2. Distributed Collaborative Industrial Agglomeration and Central Gravity Index

2.1. Definition of Industrial Spatial Distribution

The spatial distribution of industry refers to the overall layout of industrial development by adjusting the industrial structure. Industrial zones are areas where large-scale industries are concentrated in the production process. Companies in the same place depend on the economic innovation and companies in the same place. Sporadic corporate locations were innovated. Companies will determine Merritt and the cost in the same area with a comprehensive consideration. This paper analyzes industrial integration from the perspective of polarization and diffusion. As the poles of growth, advanced areas have the capacity to promote economic development. The new industrial park is an industrial estate with a certain social culture, and its essence is flexible specialization. It is inevitable for companies to choose integration to reduce transaction costs [4-5].

2.2. Industrial Agglomeration

Industrial accumulation refers to the phenomenon that the same industry is very concentrated in a particular geographical position. As industrial concentration increases, accumulation of industrial capital will continue to converge to this geographical position. Industrial accumulation is an imbalanced distribution of relevant economic activities. With the general law of industrial development. Industrial unification in this field is completed if a similar company that produces a particular kind of product collects related support companies and service industries to achieve a common goal. In other words, based on the observance of the corresponding economic rules, the relevant firms may form industrial aggregates within a particular geographical range, and the main role is the spatial imbalance of relevant economic activities [6-7].

2.3. Industrial Agglomeration Theory

(1) Neoclassical theory of location agglomeration

The theory of central place and the theory of new industrial zone play a key role in the study of the agglomeration theory of new occupation. Among them, the research hypothesis of the central place theory is: there is no spatial heterogeneity in the homogeneous plane space, that is, the population density and the endowment stock of resources. According to the scale and level of goods and services provided, the central part can be divided [8-9]. It's a central area at different levels. The level of the central area of a region depends on its size, and there is a positive correlation, but the number of central areas of the same size in the region is negatively correlated; the definition of new industrial area is as follows: "the production complex facilities formed by organizations or individuals with the same background in a specific spatial area. In the new industrial zone, SMEs use competition and cooperation to improve the economic development level of the new industrial zone. The main characteristics of this economic activity are related division of labor and externality of interaction between enterprises, emphasizing that industrial integration needs to promote the improvement of group efficiency through the role of technological innovation [10].

(2) Industrial Location Theory

Industrial cluster has two stages of development, one is the simple expansion stage, the other is the formation of scale economy. Transportation costs, personnel costs and integrated economy determine the layout of industrial enterprises [11-12]. An important consideration for enterprises in deciding layout design is the three most appropriate combinations, with the purpose of minimizing the cost of enterprises [13]. Moreover, if multiple enterprises can be gathered to obtain benefits and cut costs, industry integration will be formed. Moreover, there are two reasons for industrial integration. In other words, internal scale economy and external scale economy [14-15].

(3) Growth pole theory

The growth pole is defined as an industrial group that continuously develops in the city and promotes the development of the whole area through independent influence on the surrounding area. Therefore, the growth poles show the characteristics of industrial integration. In addition, the economic space is different from the former geographical space, it does not exist in a specific place, and exists in the economic relationship between the regions through the data. The driving force of economic development is technological progress and innovation. Some companies with scale dominance tend to have more innovative conditions, interact better with other companies, promote industry chain development, promote aggregate aggregation, produce scale effects, and promote the development of surrounding areas [18].

(4) New economic geography theory

Based on the theory of cosmic economy, the influence mechanism of industrial integration is composed of unbalanced spatial distribution. The agglomerated central force and centrifugal force can change the spatial structure characteristics of relevant economic activities under the influence of spatial imbalance. The "center and periphery" model is a typical regional model of spatial economics [19-20]. In the "center and periphery" model, and industrial agglomeration has a further impact on the spatial structure of national or regional economic activities. In the process of industrial agglomeration, the existence of external economy is an important driving force, which helps to promote the integration of specific industries in geographical location [21].

2.4. Measurement Method of Industrial Agglomeration

The integration of producer services is usually represented by the level of industrial integration, which is mainly represented by the degree of industrial integration. Although there are many methods to measure the integration degree of an industry, the viewpoint is different if the measurement methods are different. Therefore, the results may be different when measuring the cohesion of the same industry in the same region. This paper briefly introduces the following main industrial integration measurement methods, and selects cruise crew spatial Gini coefficient to measure producer service industry in the Yangtze River Delta more accurately [22-23].

In order to maintain the consistency of the variables in the following indexes, this paper will use the following basic variables to illustrate: R_{ij} represents a certain indicator (GDP, added value, employment) of J industry in region I, M represents the total number of regions, and N represents the total number of industries.

$$S_{ij} = R_{ij} \Big/ \sum_{j=1}^{n} R_{ij} \tag{1}$$

S_{ij} represents the proportion of a certain indicator (GDP, added value, employment, etc.) of industry J in region I to all industries in the region [24].

$$S_{ji} = R_{ij} \Big/ \sum_{i=1}^{m} R_{ij} \tag{2}$$

S_{ji} indicates the proportion of a certain indicator (GDP, added value, employment, etc.) of J industry in the Yangtze River Delta region.

$$S_{i} = \sum_{j=1}^{n} R_{ij} \Big/ \sum_{i=1}^{m} \sum_{j=1}^{n} R_{ij} \tag{3}$$

S_{i} represents the proportion of a certain indicator (GDP, added value, employment, etc.) in all industries in the Yangtze River Delta region.

$$S_{j} = \sum_{i=1}^{m} R_{ij} \Big/ \sum_{i=1}^{m} \sum_{j=1}^{n} R_{ij} \tag{4}$$

S_{j} represents the proportion of a certain indicator (GDP, added value, employment, etc.) of J industry in all industries in the Yangtze River Delta region.

(1) Industry concentration CRn

Industry concentration refers to the indicators (gross output value, added value, employment, etc.) of the first few industries or regions of the whole industry (gross output value, added value, employment).

$$CR_{n} = \sum_{j=1}^{n} S_{ij} \tag{5}$$

Among them, CRn is the industry concentration, S_{ij} is the proportion of a certain indicator of industry J in region I accounting for all industries in the region, and $\sum_{j=1}^{n} S_{ij}$ is the proportion of specific indexes of all industries in region I. The industry concentration CRn ranges from 0 to 1. The value of n is related to the index of the first n enterprises or regions. The larger the CRn, the higher the industry concentration, and the higher the initial concentration of N enterprises or regions, which tends to monopolize the industry market. On the contrary, the lower the concentration of CRn industry. The lower the concentration of the top n companies or regions, the higher the competitiveness of the industry market.

(2) Herfindahl Hirschmann index HHI

HHI refers to the sum of squares of the proportion of certain indicators (GDP, added value, employment) of an industry in all regions [25].

$$HHI = \sum_{j=1}^{n} S_{ij}^{2} \tag{6}$$

$\sum_{j=1}^{n} S_{ij}^2$ is the sum of the specific index of all industries in region I and the square of all industries in region I. The range of HHI is 0-1, and the value of HHI is related to the number of regions n. The higher the HHI, the higher the industry concentration and the more uneven spatial distribution. The smaller the HHI, the lower the industry concentration and the more balanced the spatial distribution.

(3) Krugman concentration index

The Krugman concentration index shows the ratio of the number of employees in a particular industry to that of all industries in the region.

$$Krugman = \sum_{i=1}^{m} \left| S_{ji} - S_i \right| \tag{7}$$

Among them, Si represents the proportion of employment in region I to all industries in all regions, and $\sum_{i=1}^{m} \left| S_{ji} - S_i \right|$ represents the absolute sum of the proportion of employment in industry J in all regions and industries in all regions. The Krugman concentration index takes values from 0 to 2. The larger the Krugman concentration index, the higher the concentration of an industry in the region; the smaller the Krugman concentration index, the lower the concentration of an industry in the region. When Krugman concentration index is 2, an industry is completely concentrated in the same region; when Krugman concentration index is 0, an industry is evenly distributed in different regions.

2.5. Central Gravity Index Algorithm

(1) Gravity model between two points

It is divided into 1 point and 2 points, each mass (which can be expressed by population, production and GDP in economic analysis) is M1 and M2, and the distance between two points is M1 and M2, which is $d_{1,2}$ (in economic analysis, it can be geographical distance or other spatial variables). Then the gravity between the two points is $g_{1,2}$. The formula is as follows:

$$g_{1,2} = \varphi \frac{m_1 m_2}{d_{1,2}^2} \tag{8}$$

Where φ is the gravitational coefficient.

(2) One factor gravity model between two points

The general form is that the space point is set with P, and the total number of points is n. the gravity between any two points I and j is as follows:

$$g_{i,j} = \varphi \frac{m_i^{\alpha} m_j^{\beta}}{d_{i,j}^{\gamma}} \tag{9}$$

Where: α, β and γ are constants greater than 0.

(3) Multi factor gravity model between two points

Economic analysis is mostly composed of multiple points of space, while investigating the interaction of multiple elements, so the two-point multi-element gravity model will become more complex.

$$G_{i,j} = f(X_i, X_j, S_{i,j}) \tag{10}$$

Where x is a series of elements of the survey object, s is a vector with spatial attributes between points. Considering the K coefficients, the total gravity between the two points is as follows

$$G_{i,j} = \sum_{1}^{k} \varphi \frac{m_{i,k}^{\alpha} m_{j,k}^{\beta}}{d_{i,j}^{\gamma}} \tag{11}$$

(4) Estimation of central gravity index

In Z industry with n enterprises, n gravity point pairs are formed between enterprises and virtual center enterprises. Because of the gravity of multiple factors, the point-to-point relationship is formed between the enterprise and the virtual center enterprise. The distance between the center point and the virtual center enterprise is $d_{O,i}$. The central gravity of all factors is summed up to get the total value of $G_{O,i}$.

$$G_{O,j} = \sum_{1}^{k} \varphi \frac{m_{O,k}^{\alpha} m_{j,k}^{\beta}}{d_{O,j}^{\gamma}} \tag{12}$$

The central attraction of all enterprises in the industry is G_Z.

$$G_Z = \sum_{j=1}^{N} G_{O,j} \tag{13}$$

The location of the central enterprise is set as point O, and the distance r is the radius to form a circular area, $r \in [\min d_{O,i}, \max d_{O,i}]$. The radius r gradually extends from the center o point to the farthest enterprise distance $\max d_{O,i}$. The enterprises in the circular area form the enterprise cluster in the core area of economic agglomeration, which is recorded as Z_O, and the central gravity value of all enterprises is recorded as G_O.

$$G_O = \sum_{i=1}^{n} G_{O,i} \tag{14}$$

The degree of industrial agglomeration of industry Z is called CRZ, which is expressed by the ratio of G_O of the total gravity of enterprises in Z_O and G_z of all enterprises in industry Z.

$$CR_Z = \frac{G_O}{G_Z} = \frac{\sum_{i=1}^{n} G_{O,i}}{\sum_{j=1}^{N} G_{O,j}} \tag{15}$$

3. Design of Center Gravity Index Algorithm for Distributed Industrial Agglomeration

3.1. Data Set Selection

The test data set used in the experiment is ucikdarchiveftp FTP: / / FTP. ICs. UCI. Edu / . At the same time, some changes are made to the original data in consideration of the variance conditions. Set. Of course, the six datasets increase according to the amount of data, the maintenance is 27, and the data amount is about 15000 to 55000. The amount of data is 957 and 7 according to the amount of maintenance. Input parameters $\xi = 6$, density threshold τ = 0.2, data are evenly distributed among nodes. Input parameters and density threshold are the most ideal parameters obtained by preprocessing and evaluating sample data.

3.2. Experimental Environment

The following experimental environment is constructed to simulate the environment of the DPA clique algorithm. A total of four VMware virtual servers are implemented on the intelxon e56202.4ghz * 8 CPU configuration and 48g storage blade servers. Each virtual machine's CPU consists of four cores: 8 g storage, and the operating system is windows 2008 R2. Four servers are in the same virtual network switch, one of which is the master node and service bus server with the MSMQ and IIS services installed. The other three are slave node servers with slave agents installed.

3.3. Selection of Driving Factors

Seven indicators of economic development environment, industrial development levels, regional human capital, policy supply capacity, infrastructure construction, information technology level, openness affect the comprehensive level of China's stifks industry. The economic environment is an important condition for the development of the logistics industry. This study reflects the economic environment using state and urban regional GDP. The development of logistics industry is closely related to manufacturing. This study is inseparable and shows the level of industrial development using the added value of industry. Securing rich people is an essential element for the development of logistics industry. The industry is labor intensive industry. The development of logistics industry is inseparable from technology and management ability. This survey represents the talent of using the number of practitioner practitioners.

4. Center Gravity Index Algorithm for Spatial Agglomeration of Distributed Collaborative Industries

4.1. Distributed Collaborative Industrial Agglomeration

(1) Industry comprehensive location quotient of tertiary industry in three economic zones

The economic development level of the three economic circles is the highest, and the service industry is the highest in China, accounting for the majority of the national economy, and a part of the service industry is mainly distributed in the three economic circles. These industries are mainly capital intensive industries, which are not only beneficial to China in terms of operating profit, but also the number of employees in China. Therefore, the distribution of industrial agglomeration is relatively concentrated. As shown in Figure 1, the comprehensive location quotient of subdivided industries in the three economic zones.

It can be seen from the data in Figure 1 that the comprehensive distribution index of the three industries in Beijing Tianjin Hebei region is relatively large. The three industrial zones have three common advantages, such as computer industry and industrial leasing. In these industries, information computer software industry and leasing industry are capital intensive industries, which not only need sufficient capital, but also need higher level of human capital and corresponding market demand. The three major economic zones are all economic development areas, which not only have developed business environment and market demand, but also have a sound education system to cultivate talents, and attract talents from other regions. Moreover, this advantage is also a manifestation of the economic development of the three economic circles. In the last comparison, the per capita income and per capita income of the three economic circles were compared. If the per capita income of China is the same as that of China's three circles, it can be said that the per capita income of China is the same. Stand at the forefront and promote the development of resident service industry.

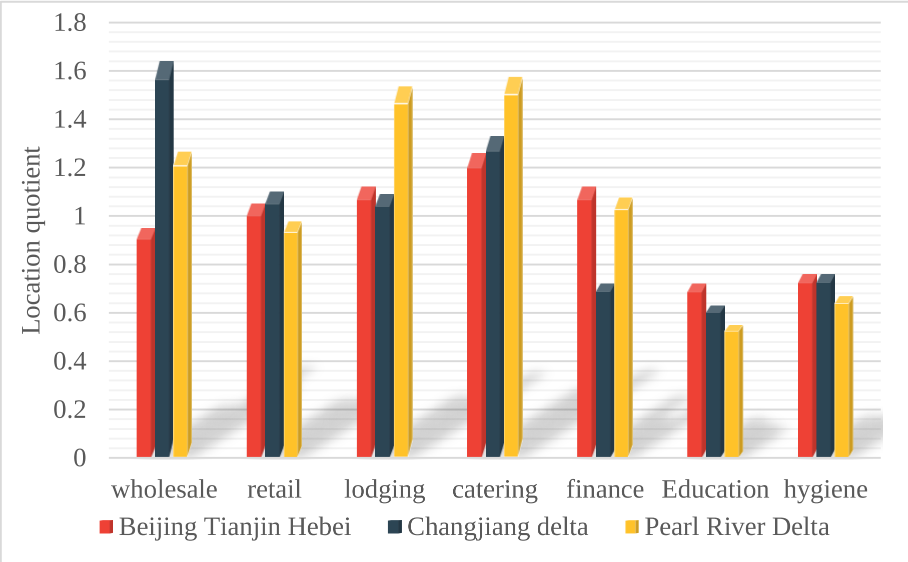

Figure 1. The comprehensive location quotient of sub industries in the three economic zones

(2) The effect of industrial spatial agglomeration on economic growth

As shown in Table 1, the spatial agglomeration results of economic growth of wood processing industry. The figures in the table are the statistical index values, and the expectation is in brackets.

Table 1. Spatial agglomeration results of economic growth of wood processing industry

Particular year	Moran'S I	Girley Index C	GETIS & ord's index G
2010	0.499(-0.043)	0.416(1.000)	0.252(0.178)
2011	0.514(-0.043)	0.341(1.000)	0.260(0.178)
2012	0.504(-0.043)	0.342(1.000)	0.258(0.178)
2013	0.492(-0.043)	0.371(1.000)	0.266(0.178)
2014	0.447(-0.043)	0.441(1.000)	0.279(0.178)
2015	0.435(-0.043)	0.453(1.000)	0.285(0.178)
2016	0.418(-0.043)	0.463(1.000)	0.288(0.178)
2017	0.412(-0.043)	0.462(1.000)	0.282(0.178)
2018	0.420(-0.043)	0.464(1.000)	0.281(0.178)
2019	0.448(-0.043)	0.470(1.000)	0.287(0.178)

From Table 1 to 2019, Moran'si of economic growth of China's wood processing industry is statistically 1%, which is obviously correct, and Moran'si values are all [0,1]. Therefore, Moran's index Moran's global spatial cluster test results confirm that there is a spatial autocorrelation with the economic growth of China's wood processing industry. The results were examined and further observed. These indices have a meaningful C-level in a year, not a C-level. This is the same as the Moran's index Moran'si. GETIS & ord's index g is significantly positive, with a meaningful level of 5%. Compared with the g value and its expected value from 2010 to 2019, the economic growth of China's wood processing industry is in the hot zone every year. Wood processing industry tends to have high total output and high concentration.

4.2. Gravity Index Algorithm for Spatial Agglomeration Center

(1) Performance bottleneck of spatial agglomeration index

The biggest advantage of spatial aggregation index algorithm is that it does not need to set the cluster center in advance, and only provides the similarity between data to calculate the cluster representative (cluster center) of each cluster. However, the spatial aggregation index algorithm O (N^2) is very complex in time and space. If the sample data increases, the spatial aggregation index algorithm implemented on a single machine will take a lot of time. Reorder matrix and time. N increases in proportion to the square. As the amount of data increases, the clustering time under different dimensions is shown in Figure 2.

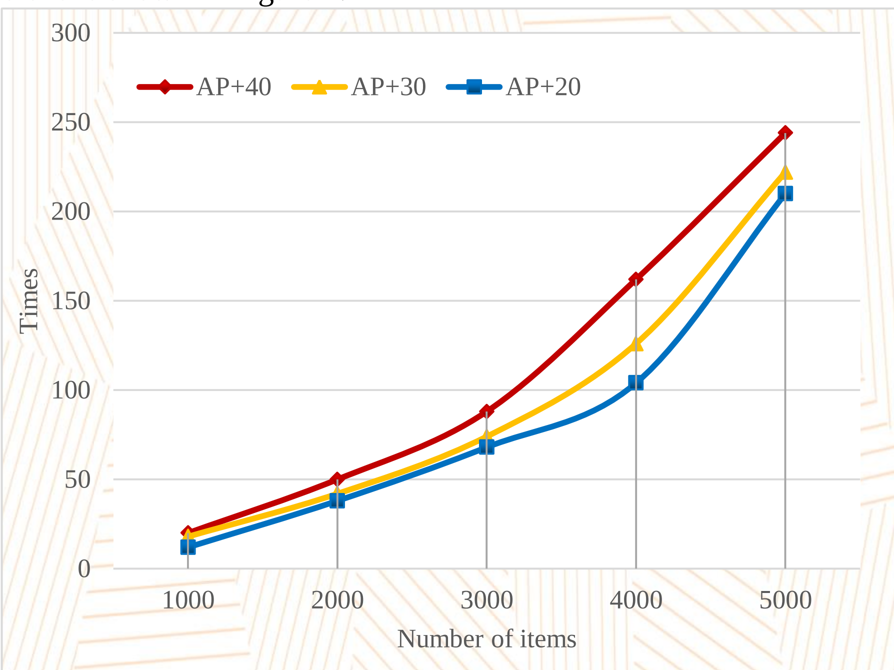

Figure 2. Algorithm time of spatial agglomeration index

As can be seen from Figure 2, the performance of spatial aggregation index algorithm decreases significantly with the increase of data volume. This is the raw data. The size of the dataset will increase to the size of a computer, especially the size of a single memory set. Therefore, the spatial aggregation index algorithm can not adapt to a large number of data processing.

In recent years, with the development of cloud computing technology and the prosperity of distributed computing platform, in order to improve the processing efficiency and scalability of large-scale data, more and more data mining algorithms and distributed platform parallel environment are being transplanted. adapt.

(2) Comparison of gravity index algorithm models for spatial agglomeration centers

The selection of slmpanel model and sempanel model needs to be further judged by LM check. The benchmarks are as follows. If LM (lag) is more important than LM (error), R.LM (lag) is important, and R.LM (error) is not, slmpanel model is selected. At the same time, fixed effect models can be divided into four types: no fixed space, fixed time, fixed time and spatial effects. This paper will discuss these four models respectively. As shown in Table 2, the fixed effect LM Test.

Table 2. Fixed effect LM test

	Test	value	P-value
On-fixed	LM(1ag)	3.359	0.068
	R-LM(1ag)	12.143	0.000
	LM(error)	0.851	0.357
	R-LM(error)	9.635	0.003
Patial fixed	LM(1ag)	0.014	0.909
	R-LM(1ag)	7.044	0.009
	LM(error)	1.042	0.308
	R-LM(error)	8.072	0.005

As can be seen from Table 2, in the unfixed model and the spatial fixed model, r.lm (lag) and IIR.LM (error) makes sense at a meaningful level of 1%, and the rest of the statistics fail in the LM Test. In the timefixed model, r.lm (lag) and r.lm (error) are meaningful at 1% meaningful level, LM (error) is meaningful at 5% meaningful level, and LM (lag) is not meaningful. In the spatial and time fixed model, LM (lag), r.lm (lag), r.lm (error), LM (error) are not important. Therefore, it is not possible to determine whether to choose SLM or SEM.

(3) Design and Simulation of gravity index algorithm for spatial agglomeration Center

As shown in Figure 3, the state trajectory of the closed-loop system is simulated by DMPC method. As shown in Figure 4, the state trajectory of the closed-loop system is simulated by CMPC method.

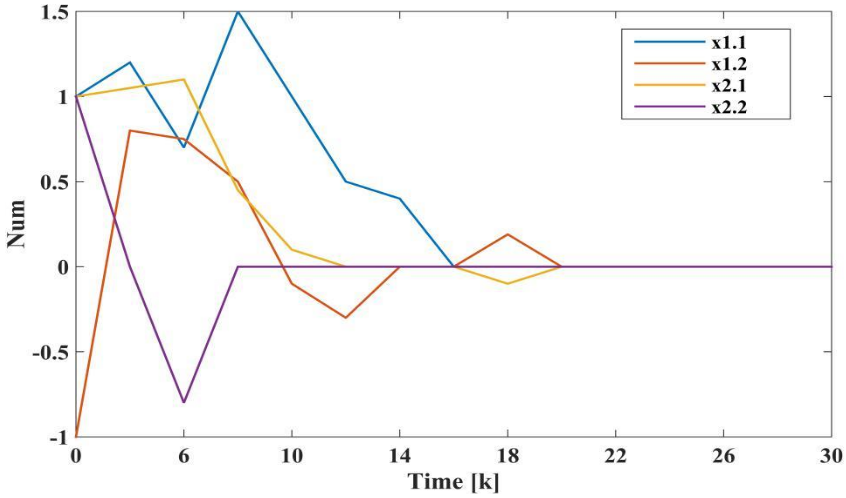

Figure 3. DMPC simulation of state trajectory of closed-loop system

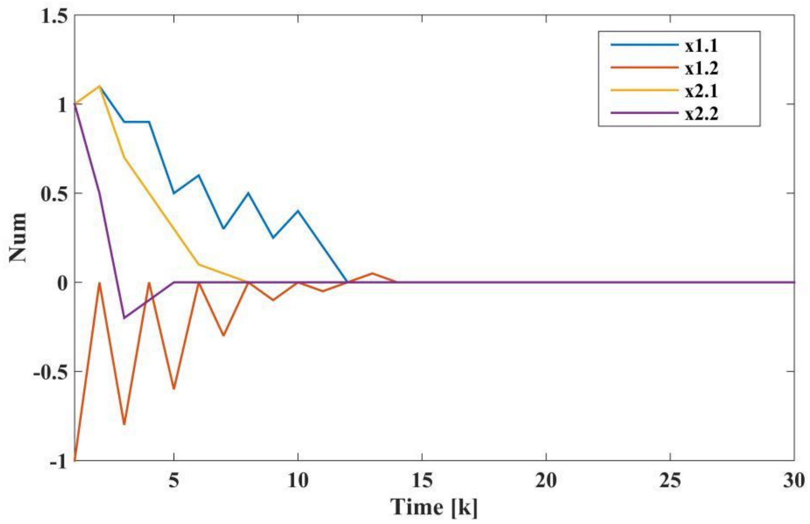

Figure 4. CMPC simulation of closed loop system state trajectory

It can be seen from Figure 3 and Figure 4 that the DMPC controller normally stabilizes the state of all subsystems, and the state and control of the system are limited within the limits. After about 15 steps, the state of the system is very close to equilibrium. However, since the starting point of the system lies at the boundary of the switch interval clone, the system state jumps at different handover intervals.

4.3. Result Analysis about Classification Model

(1) Algorithm parameters of feature word number change

The average values of the F1 measures with Na ïve Bayes, Rocchio, SVM and KNN increased from 320 to 1120. As shown in Figure 5, the F1 values of different algorithms change with the increase of the number of feature words.

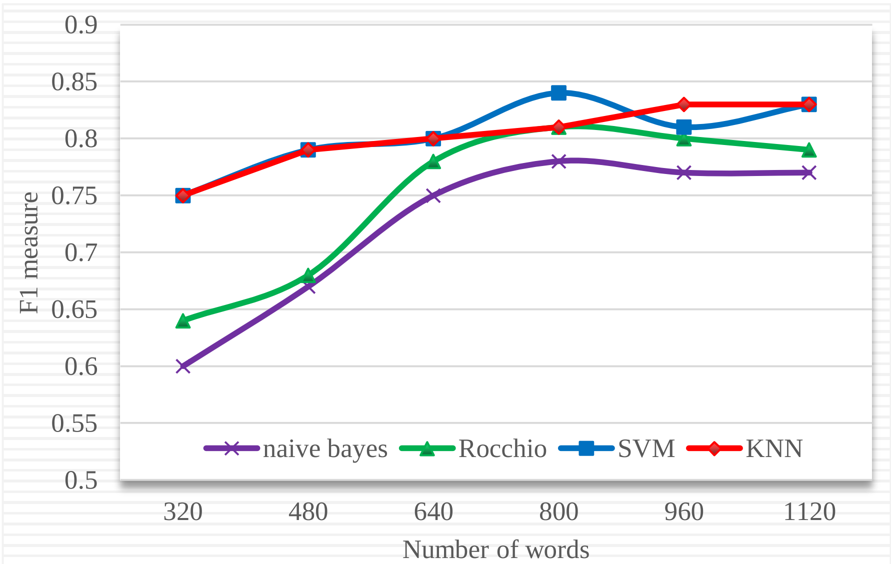

Figure 5. Different algorithms change with the number of feature words

As can be seen from Figure 5, when the number of feature words increases from 320 to 640, the number of F1 main words of Loki, SVM and KNN increases significantly, and the number of feature words increases. It will increase accordingly. When the number of feature words is 640-800, the main increase rate of F1 is slower than that of 320-640. When the number of feature words is 800-1120, the accuracy of all algorithms is basically stable. The performance of KNN and SVM is improved.

(2) Comparison of different algorithms with the same feature number

If the number of feature words is 800, the F1 measurement function of each algorithm is the best. As shown in Figure 6, the comparison of the four algorithms when the number of feature words is 800.

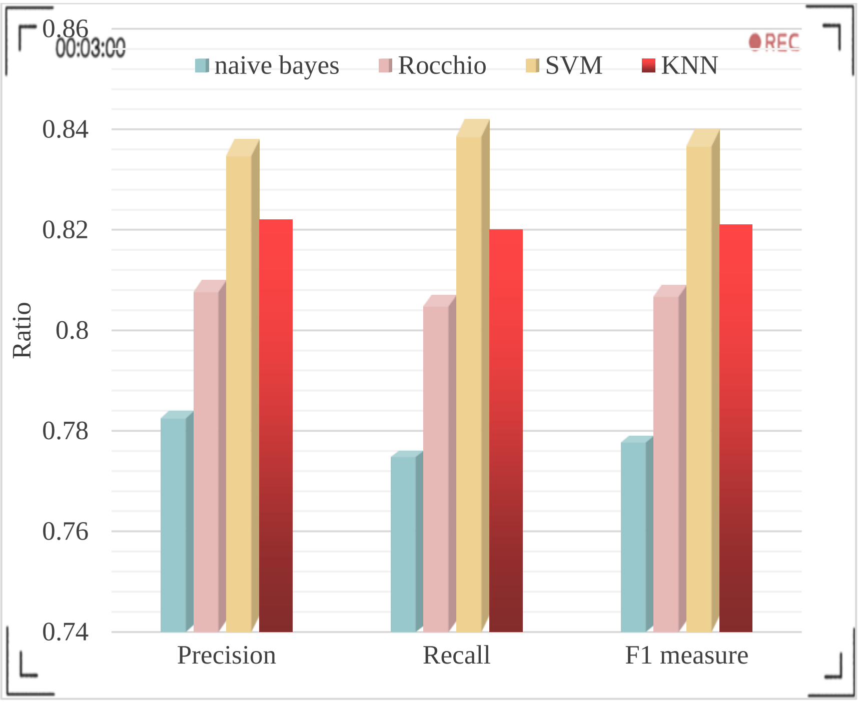

Figure 6. Comparison of four algorithms when the number of feature words is 800

It can be concluded from Figure 6 that in the process of agricultural information classification, when the feature words are 800, the learning ability of SVM is the strongest in terms of accuracy (84%), recall (84%) and F1 main (84%). Both will perform better. The learning ability of KNN is second only to SVM. Precision (82%), recall (82%) and f1measure (82%) can perform well. Rocchio and nathis are not very satisfied. Analysis reason: Rocchio method mainly depends on the calculation of classification center vector, and the captured data may have multiple attributes, so the classification accuracy is reduced.The feature attributes of this experiment are highly correlated, because they are not independent of each other, the classification results become insufficient.

4.4. Evaluation of Results and Expectations

(1) Algorithm execution of different data volume

At the same time, the execution time and precision of CLIQUE algorithm and DPA-CLIQUE algorithm are tested in the data set test. As shown in Table 3, the algorithm execution time.

Table 3. Algorithm execution time

Data volume / Time (s)	15000	25000	35000	45000	55000
CLIQUE algorithm	79	126	142	298	1767
DPA-CLIQUE algorithm	64	76	91	121	566

As shown in Table 4, the algorithm accuracy.

Table 4. Algorithm accuracy

Data volume / Accuracy	15000	25000	35000	45000	55000
CLIQUE algorithm	0.94	0.97	0.97	0.99	0.96
DPA-CLIQUE algorithm	0.95	0.98	0.99	0.99	0.99

As shown in Figure 7, the algorithm execution time comparison chart.

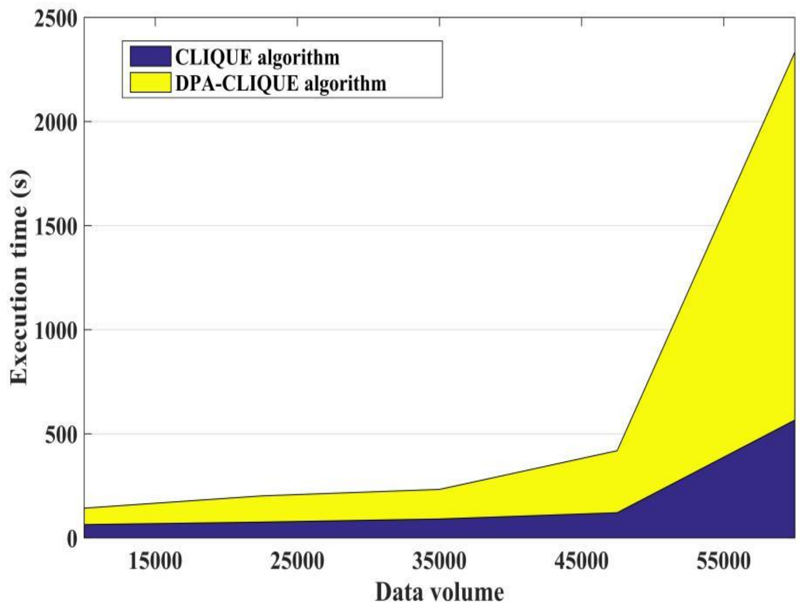

Figure 7. Comparison chart of algorithm execution time

Figure 7 shows that DPA-CLIQUE algorithm is better than CLIQUE algorithm in the case of the same spatial dimension and different data volume. The increase of time, especially when the amount of data exceeds 35000, is basically close to the linear ratio of the increase in the number of records. The efficiency of DPA-CLIQUE algorithm is much higher than that of CLIQUE algorithm.

(2) Implementation of algorithms in different spatial dimensions

The execution time and accuracy of CLIQUE algorithm and DPA-CLIQUE algorithm are tested. As shown in Table 5, the algorithm execution time and accuracy.

Table 5. Algorithm execution time and accuracy

	Dimension	6	12	15	21	27	51	75
Time (s)	CLIQUE algorithm	17	17	48	79	110	1095	4094
	DPA-CLIQUE algorithm	46	44	50	71	79	628	1051
Accuracy	CLIQUE algorithm	0.83	0.94	0.93	0.98	0.97	0.98	0.98
	DPA-CLIQUE algorithm	0.93	0.96	0.98	0.99	0.99	0.99	0.99

As shown in Figure 8, the comparison chart of algorithm execution time of different spatial dimensions is shown.

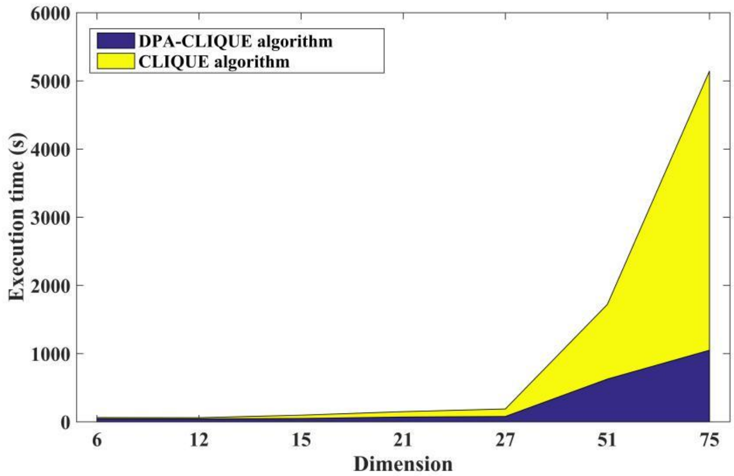

Figure 8. Comparison of algorithm execution time in different spatial dimensions

It can be seen from Figure 8 that, under the same amount of data, the efficiency of dpa-clique algorithm is slightly lower than that of clique algorithm in low dimensional data space for data sets with different spatial dimensions. However, after 15 dimensions, the gap gradually narrowed. When the dimension is larger than 27 dimensions, the execution time of dpa-clique algorithm is much shorter than that of clique algorithm.

Some test data sets can not guarantee the clustering quality of clique algorithm. The accuracy of dpa-clique algorithm has good functions in all test data sets. The higher the dimension of data space, the larger the data set, the higher the clustering quality.

5. Conclusion

The higher the degree of financial intensity, the more obvious the corresponding radiation effect, the higher the regional interests, which may promote economic development. Therefore, in order to improve the competitiveness and integration of the financial industry, it is necessary to promote regional economic development. First of all, according to the situation of each economic region and city and the difference of each economic region and city, different regional industrial policies should be formulated to promote the effective distribution of resources. Second, promote the development of regional integration. Through the construction of a unified market, promote the flow of resources, eliminate the phenomenon of regional protection, and improve the efficiency of economic operation. Third, we should continue to promote the reform and upgrading of the financial industry. Industrial transformation and upgrading can not only improve competitiveness, but also promote investment and promote the formation of financial integration.

By analyzing the advantages and disadvantages of K-means and SVM, the classification method of combining K-means and SVM is realized. The feasibility and performance of hybrid model are proved by experiments. This paper also implements two methods to obtain the most advanced data based on the center vector. According to the idea of keeping the center vector and keeping the adjacent samples of the center vector, two models are designed and implemented. The training time of the two methods is carried out before and without cutting. Measure the number of training samples and F1. The classification model of agricultural information classification based on K-means clustering algorithm and SVM classification algorithm can reduce the classification accuracy and shorten the training time.

The central gravity index method can effectively measure spatial aggregation. Through the use of more mature gravity model, the integration of output value and distance, and the use of many mathematical processing methods from the existing measurement methods, the establishment of a virtual center enterprise, the mathematical calculation has been greatly simplified and consolidated theoretical basis. In particular, the confidence interval and the threshold value of absolute index judgment are constructed by relative indexes, and the classification by specific agglutination degree can effectively carry out the scientific structure analysis of the spatial agglutination degree and geographical spatial distribution characteristics of economic activities. The effectiveness of this method is also reflected in the validation of sporadic enterprise data from the New York River Delta.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Wang X S , Yu C Y . Impact of spatial agglomeration on industrial pollution emissions intensity in China. Zhongguo Huanjing Kexue/China Environmental ence, 2017, 37(4):1562-1570.

[2] Fuxiang W , Yue C . A welfare economics analysis of China's industrial layout restructuring. China Political Economy, 2018, 1(2):263-283. https://doi.org/10.1108/CPE-10-2018-018

[3] Ye C , Sun C , Chen L . New evidence for the impact of financial agglomeration on urbanization

from a spatial econometrics analysis. Journal of Cleaner Production, 2018, 200(11):65-73. https://doi.org/10.1016/j.jclepro.2018.07.253

[4] *Anderson R C , Fitton J W . Index theory-based algorithm for the gradiometer inverse problem. GEM - International Journal on Geomathematics, 2016, 7(1):147-161. https://doi.org/10.1007/s13137-015-0072-x*

[5] *Esponda H , Vazquez E , Andrade M A , et al. A Setting-Free Differential Protection for Power Transformers Based on Second Central Moment. IEEE Transactions on Power Delivery, 2019, 34(2):750-759. https://doi.org/10.1109/TPWRD.2018.2889471*

[6] *Khazri D , Gabtni H . Geophysical methods integration for deep aquifer reservoir characterization and modeling (Sidi Bouzid basin, central Tunisia). Journal of African Earth Sciences, 2018, 138(2):289-308.*

[7] *Hiramatsu Y , Sawada A , Kobayashi W , et al. Gravity gradient tensor analysis to an active fault: a case study at the Togi-gawa Nangan fault, Noto Peninsula, central Japan. Earth, Planets and Space, 2019, 71(1):1-8. https://doi.org/10.1186/s40623-019-1088-5*

[8] *Chen L , Xiangmu J . Measurement of Spatial Interaction between Central Towns Based on the Gravity Model. entia Geographica Sinica, 2016, 36(5):724-732.*

[9] *Ayad A , Bakkali S . Interpretation of potential gravity anomalies of Ouled Abdoun phosphate basin (Central Morocco). Journal of Materials and Environmental ence, 2017, 8(9):3391-3397.*

[10] *Ranganai R T , Gwavava O , Ebinger C J , et al. Configuration of Late Archaean Chilimanzi and Razi Suites of Granites, South-Central Zimbabwe Craton, From Gravity Modelling: Geotectonic Implications. Pure and Applied Geophysics, 2020, 177(2):1043-1069.*

[11] *Ndehedehe C E , Ferreira V G , Agutu N O . Hydrological controls on surface vegetation dynamics over West and Central Africa. Ecological Indicators, 2019, 103(8):494-508. https://doi.org/10.1016/j.ecolind.2019.04.032*

[12] *Bannon M . The changing centre of gravity of office establishments within central Dublin, 1940 to 1970. Irish Geography, 2016, 6(4):480-484.*

[13] *He Z , Romanos M . Spatial agglomeration and location determinants: Evidence from the US communications equipment manufacturing industry. Urban Studies, 2016, 53(10):329-330. https://doi.org/10.1177/0042098015586698*

[14] *Wei T , Tian Y . On Spatial Spillover and Industrial Agglomeration of Financial Crises to Real Economy. Journal of Mathematical Finance, 2020, 10(3):464-482. https://doi.org/10.4236/jmf.2020.103028*

[15] *Xie H . Spatial Heterogeneity Strategies for Pollution Agglomeration Control in China: Based on the Coordination Between Industrialization and Urbanization. Arabian Journal of Geoences, 2020, 13(19):1-15.*

[16] *Flores M , Villarreal A , Flores S . Spatial Co-location Patterns of Aerospace Industry Firms in Mexico. Applied Spatial Analysis and Policy, 2017, 10(2):233-251.*

[17] *Nguyen T X T , Diez J R . Multinational enterprises and industrial spatial concentration patterns in the Red River Delta and Southeast Vietnam. Annals of Regional ence, 2017, 59(1):1-38. https://doi.org/10.1007/s00168-017-0820-y*

[18] *Li L , Hong X , Peng K . A spatial panel analysis of carbon emissions, economic growth and high-technology industry in China. Structural change and economic dynamics, 2019, 49(6):83-92.*

[19] *Scott B , Loonam J , Kumar V . Exploring the rise of blockchain technology: Towards distributed collaborative organizations. Strategic Change, 2017, 26(5):423-428. https://doi.org/10.1002/jsc.2142*

[20] Palau A S , Dhada M H , Bakliwal K , et al. An Industrial Multi Agent System for real-time distributed collaborative prognostics. Engineering Applications of Artificial Intelligence, 2019, 85(10):590-606.

[21] Xie Y , Wang Y , He H , et al. A General Collaborative Framework for Modeling and Perceiving Distributed Network Behavior. IEEE/ACM Transactions on Networking, 2016, 24(5):3162-3176.

[22] Roberts D J , Bar C , Cencetti M M , et al. Collaborative virtual reality platform for visualizing space data and mission planning. Multimedia Tools and Applications, 2019, 78(23):33191-33220. https://doi.org/10.1007/s11042-019-7736-8

[23] Elhag T , Eapen S , Ballal T . Moderating claims and disputes through collaborative procurement. Construction innovation, 2020, 20(1):79-95. https://doi.org/10.1108/CI-02-2019-0020

[24] Franz S , Irmler R , Uwe Rüppel. Real-time collaborative reconstruction of digital building models with mobile devices. Advanced Engineering Informatics, 2018, 38(10):569-580. https://doi.org/10.1016/j.aei.2018.08.012

[25] Stone B R , Wald M O , Gorrell S E , et al. Collaboration Task-Technology Fit for Student Distributed Engineering Design Teams. The international journal of engineering education, 2018, 34(5):1687-1700.