Cluster Analysis Based on Indicator System on the Development of Digital Economy in Guangdong

Xue Deng, Keyao Zheng, Ye Xiong

Abstract—**The digital economy is regarded as a new economic form that achieves high-quality economic development, which is a new engine for social and economic development. Research on the social and economic development in the context of the digital economy plays a significant part in optimizing the social economic system in the digital age. We divide the digital economy into four major types, "basic type, technology type, integration type and service type". In this essay, we will select 5 indicators to evaluate these types and set up a digital economy index evaluation system with 5 indicators. Under the indicator system, systematic clustering and K-means clustering were performed on 31 provinces and cities across the country. Compared with each province or cities, the advantages and disadvantages of the overall digital economy and the four digital economy types in different provinces or cities are analyzed based on the comprehensive factor scores, which provides a strong basis for the future development of the digital economy in Guangdong Province. Thus, this essay will give some suggestions for the growth of Guangdong Province in digital economy.**

Index Terms—**Digital economy, Indicator evaluation system, Systematic clustering, K-means clustering, Factor scores**

I. Introduction

As an important role of the global economy, digital economy had become the focus of attention of countries around the world. Developments in the digital economy had become an irresistible trend in present world. Li et al. [1] studied the Digital Economy in Asia. It was found that the digital economy of Asian countries transforms business processes through technological innovation, government promotion policies and digital entrepreneurship. The digital economy had played a huge role in promoting Asian economic development. Rita et al. [2] combined comprehensive indices into a new framework to study a country's social economic system over time. According to the data of Eurostat from 2001 to 2016, Monica et al. [3] carried out research into digital economic growth and the progress of digital in the EU by analyzing and selecting a certain amount of specific indices, and finally determined the policies and measures which can foster the digital economic growth in EU states. The rapid increase was explained by Schweighofer et al. [4] in the importance of technology intensive learning. He pointed out the known relationship between technology intensive learning and digital economy, but the relationship between digital economy and technology promoted learning was rarely studied. Talar [5] assessed the consistency of the concept and new challenges of the European digital economy in 2020. The model of the evolution of digital control was built by Landini [6].

Based on the above research, this paper believes that digital economy is worth studying as a new area in the 20th century. Del et al. [7] explained the understanding of today's digital economy and determined its scale and impact on enterprises, thus demonstrating the need to develop its theoretical basis. In addition, an empirical study was conducted in the context of Spain. Berdykulova et al. [8] revealed the essence of digital economy and studied the shortcomings of Kazakhstan's digital economy through examples. Protopopova et al. [9] studied the rationalization of the relevance of digital economy development in Russia, the comparative analysis between Russia and other major countries in digital economy development, and the reasons and solutions of Russia's inability to establish digital economy in recent ten years. Modern Russia was taken by Skiter et al. [10] as an example to study the development trend and Prospect of national food security under the condition of digital economy. Dong et al. [11] summarized the characteristics and measurement methods of digital economy. And the measuring method of digital economy was studied. Nudurupati et al. [12] discussed the improvement and practice of PMM model to make it more flexible and reflect the progress of digital economy.

However, it is known that the research on digital economy is almost qualitative. The development path of digital economy is lack of data support through qualitative research. Therefore, we can carry out quantitative research on the use of cluster analysis, factor analysis and other statistical methods of digital economy in Guangdong. Economic development path should be optimized. Wei et al. [13] used cluster analysis to make development strategy. Kim et al. [14]

Manuscript received January 22, 2021; revised July 2, 2021. This research was supported by the "Humanities and Social Sciences Research and Planning Fund of the Ministry of Education of China, No. 18YJAZH014-x2lxY9180090", "Natural Science Foundation of Guangdong Province, No. 2019A1515011038", "Guangdong Province Characteristic Innovation Project of Colleges and Universities, No. 2019GKTSCX023", "Soft Science of Guangdong Province, No. 2018A070712006, 2019A101002118".

Xue Deng is a Professor of the school of Mathematics, South China University of Technology, Guangzhou 510640, China, (e-mail: dxue@scut.edu.cn).

Keyao Zheng is a Student of the school of Mathematics, South China University of Technology, Guangzhou 510640, China, (corresponding author to provide e-mail: 364122167@qq.com).

Ye Xiong is a Student of the School of Mathematics, South China University of Technology, Guangzhou 510640, China, (e-mail: workyep@163.com).

showed division of related fields so that we can divide the digital economy into 4 types. An memetic algorithm for clustering tasks was proposed by Jiang et al. [15] . The data processing was used for reference in this paper. For the sake of understanding the status of the 4 digital economic types in China and the overall development level of digital economy in Guangdong. This paper conducts cluster analysis on the data of 31 provinces and cities. A comprehensive factor score model is established through factor analysis to cross verifies the results of cluster analysis, so as to explore the shortcomings of digital economy development in Guangdong compared with other provinces (cities).

In Section II, the basic knowledge of cluster analysis is described. Building a new indicator system is in Section III. In Section IV, cluster analysis is carried out with examples. The Section V summarizes comparative analysis between cluster analysis and factor analysis, and put forward advice.

II. Basic Principles of Cluster Analysis

A. System Clustering Algorithm

A.1 Basic content

According to Gao [16], six clustering algorithms are as follows:

(1) The shortest distance method

The shortest distance between all sample pairs in two categories is as follow: $D_{KL}=\min\limits_{i\in G_k, j\in G_L} d_{ij}$. This is the shortest distance method. The distance between sample i and sample j is represented by d_{ij}, $G_1, G_2, \cdots$ is a class, and the distance between G_K and G_L is represented by D_{KL}. The recurrence formula is

$$\begin{aligned} D_{MJ} &= \min_{i\in G_M, j\in G_J} d_{ij} \\ &= \min\{\min_{i\in G_K, j\in G_J} d_{ij}, \min_{i\in G_L, j\in G_J} d_{ij}\} \\ &= \min\{D_{KL}, D_{LJ}\}. \end{aligned} \quad (1)$$

(2) Longest distance method

The distance between the two categories is the farthest distance among the samples in the two categories is as follow: $D_{KL}=\max\limits_{i\in G_k, j\in G_L} d_{ij}$. When class G_K and class G_L synthesize a new class G_M, then the distance between G_M and other classes G_J is $D_{MJ}=\max\{D_{KL}, D_{LJ}\}$.

(3) Intermediate distance method

Taking the middle distance for the shortest distance method and the longest distance method mentioned above. When G_K and G_L are combined into a new class G_M, for a certain class G_J, D_{KL}, D_{LJ}, D_{KJ} are taken as the side lengths to form a triangle, and the center line D_{MJ} of D_{KL} side is made. At this time, we can know from geometric knowledge that:

$$D_{MJ}^2 = \frac{1}{2}D_{KJ}^2 + \frac{1}{2}D_{LJ}^2 - \frac{1}{4}D_{KL}^2. \quad (2)$$

(4) Class average method

The average value of the square distance between every two samples in the two categories is the class average method, and the formula is

$$D_{KL}^2 = \frac{1}{n_K n_L}\sum_{i\in G_K, j\in G_L} d_{ij}^2. \quad (3)$$

The recurrence formula is as follows:

$$D_{MJ}^2 = \frac{n_K}{n_M}D_{KJ}^2 + \frac{n_L}{n_M}D_{LJ}^2. \quad (4)$$

(5) Barycenter method

The distance between the two categories is defined as the Euclidean distance of the center of gravity. Suppose $\bar{x}_K$ and $\bar{x}_L$ are the centers of gravity of G_K and G_L, then the distance between G_K and G_L is

$$D_{KL} = d_{\bar{x}_K\bar{x}_L}^2 = (\bar{x}_K - \bar{x}_L)^T(\bar{x}_K - \bar{x}_L). \quad (5)$$

The recurrence formula is as follows:

$$D_{MJ}^2 = \frac{n_K}{n_M}D_{KJ}^2 + \frac{n_L}{n_M}D_{LJ}^2 - \frac{n_K n_L}{n_M^2}D_{KL}^2. \quad (6)$$

(6) Method of sum of squares of deviation

Let G_M be the combined category of G_K and G_L, then the sums of deviation squares for G_K, G_L and G_M are

$$W_K = \sum_{i\in G_K}(x_{(i)} - \bar{x}_K)^T(x_{(i)} - \bar{x}_K), \quad (7)$$

$$W_L = \sum_{i\in G_L}(x_{(i)} - \bar{x}_L)^T(x_{(i)} - \bar{x}_L), \quad (8)$$

$$W_M = \sum_{i\in G_M}(x_{(i)} - \bar{x}_M)^T(x_{(i)} - \bar{x}_M). \quad (9)$$

Where $\bar{x}_K$, $\bar{x}_L$ and $\bar{x}_M$ are the focus of G_K, G_L and G_M respectively, and the dispersion degree of samples in each class is represented by W_K, W_L and W_M. If G_K and G_L are two classes which are close to each other, then the value of the sum of squares of deviations $W_M - W_K - W_L$ will be smaller after merging into a new class. Therefore, the square distance between G_K and G_L can be defined as: $D_{KL}^2 = W_M - W_K - W_L$. This is the method of sum of squares of deviation in system clustering. The recurrence formula is as follows:

$$D_{MJ}^2 = \frac{n_J + n_K}{n_J + n_M}D_{KJ}^2 + \frac{n_J + n_L}{n_J + n_M}D_{LJ}^2 - \frac{n_J}{n_J + n_M}D_{KL}^2. \quad (10)$$

To sum up, the basic steps of systematic clustering can be summarized. Suppose there are n samples.

A.2 Basic steps

The basic steps of using hierarchical clustering in R software are as follows:

Step 1: Calculate the distance between every two samples;
Step 2: Construct n classes and each sample is a class;
Step 3: Merge the two nearest classes into a new class;
Step 4: Calculate the distance between the new class and the existing class;
Step 5: Repeat Step 3 and Step 4 until there is only one category left;
Step 6: Draw a genealogy, and determine the number of classes and the samples in the class according to the genealogy.

B. K-means clustering

B.1 Basic content

Let $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})^T$, $y_i = (y_{i1}, y_{i2}, \cdots, y_{in})^T$ be the data of two samples. The Euclidean distance is defined as

$$d\left(x_i, y_i\right) = \sqrt{\sum_{m=1}^{n}\left(x_{im} - y_{im}\right)^2}. \quad (11)$$

If C_i is the i -th category and T_i is the number of samples in Category C_i , then the centroid of data in the same category is

$$\mu_i = \frac{1}{\left|T_i\right|}\sum_{x_i \in C_i} x_i. \quad (12)$$

B.2 Basic steps

The basic steps of K-means algorithm are as follows:

Step 1: Determine the number of clusters k ;

Step 2: Determine k initial clustering centers;

Step 3: Calculate the distance from each sample to each cluster center according to Formula (11), and allocate the samples to the nearest cluster;

Step 4: Recalculate the centroids of the existing classes as k new cluster centers by Formula (12);

Step 5: Repeat Step 3 and Step 4. When the centroid remains unchanged, the class and the sample in the category are obtained.

III. BUILDING THE INDICATOR SYSTEM

According to the white paper of the release of Chinese digital economy indicator in 2017 issued by CCID consultants, digital economy can be divided into five types: basic, resource-based, technological, integrated and service-oriented. According to the availability and effectiveness of the indicator data, the digital economy indicator system in line with Guangdong Province is established by combined with the index system in Ning [17], as shown in Table I:

In Table I, the basic digital economy indicators start with the two respects of basic telecommunications and network. The evaluation system select indicators such as optical cable line length, telephone penetration rate, Internet broadband access users and so on to measure the popularization and growth in telecommunications and Internet.

Technological digital economy indicators should start with the view of digital technology, including block-chain, big data, AI and other emerging industries into the indicator system. However, due to the difficulty of data acquisition, its industry is primarily used to measure the development of digital technology. Thus, some representative data are used as the evaluation indicators of technological digital economy, as follows: the information technology service income, the total amount of telecommunication business, the income of software business and so on.

The indicator of integrated digital economy is analyzed from the perspective of industry, agriculture and digital economy. These selected indicator are to measure the application of enterprises, industry and agriculture informatization. Hence the 5 indicators are selected as above.

From the integration of digital economy and service industry, service-oriented digital economy indicator shows the significant element of digital technology in life, study and entertainment. Five indicators are adopted, including amount of public information on government websites, e-commerce transaction volume, amount of electronic reading room terminals in public libraries, amount of digital TV users, scientific research, technical services and social fixed asset investment in geological exploration industry.

TABLE I
EVALUATION INDICATOR SYSTEM OF DIGITAL ECONOMY IN GUANGDONG PROVINCE

Indicator category	Indicator expression	Indicator name	Unit
Basic digital economy	X_1	Length of optical cable line	km
	X_2	Telephone penetration rate	Department / person
	X_3	Internet broadband access users	Ten thousand households
	X_4	Number of websites	Ten thousand
	X_5	Number of domain names	Ten thousand
Technological digital economy	X_6	IT service revenue	Ten thousand yuan
	X_7	Software revenue of embedded system	Ten thousand yuan
	X_8	Total telecom services	100 million yuan
	X_9	Software business income	Ten thousand yuan
	X_{10}	Information transmission, computer service and fixed asset investment in software industry	100 million yuan
Integrated digital economy	X_{11}	Increase of rural e-commerce comprehensive demonstration counties	Unit / year
	X_{12}	Number of P & D projects of Industrial Enterprises above Designated Size	term
	X_{13}	Proportion of enterprises with e-commerce transaction activities	%
	X_{14}	Quantity of enterprise informatization	unit
	X_{15}	Number of enterprises integrated with industrialization and industrialization	unit
Service oriented digital economy	X_{16}	E-commerce transaction volume	100 million yuan
	X_{17}	Amount of public information on government websites	strip
	X_{18}	Amount of terminals in electronic reading room of Public Library	Number
	X_{19}	Amount of digital TV users	Ten thousand households
	X_{20}	Scientific research, technical services and social fixed asset investment in geological exploration industry	100 million yuan

IV. Empirical Research

A. Experimental Data

TABLE II-(a)
STANDARDIZED DATA

Provinces（cities）	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Beijing	-1.06	3.51	-0.64	2.80	2.74	2.68	-0.35	-0.03	2.25	0.29
Tianjin	-1.24	0.04	-0.90	-0.48	-0.53	-0.04	-0.22	-0.86	-0.15	-0.24
Hebei	0.57	-0.36	0.85	-0.07	-0.27	-0.55	-0.38	0.35	-0.58	0.27
Shanxi	-0.16	-0.45	-0.28	-0.47	-0.44	-0.68	-0.38	-0.46	-0.65	-0.95
Inner Mongolia	-0.51	-0.03	-0.72	-0.68	-0.67	-0.68	-0.38	-0.58	-0.66	-0.52
Liaoning	-0.02	0.26	-0.06	-0.12	-0.37	0.12	-0.21	-0.11	0.17	-0.72
Jilin	-0.79	0.10	-0.70	-0.60	-0.58	-0.47	-0.24	-0.64	-0.45	0.52
Heilongjiang	-0.37	-0.41	-0.51	-0.54	-0.50	-0.63	-0.36	-0.54	-0.61	0.10
Shanghai	-0.75	1.87	-0.49	1.00	0.67	1.30	-0.16	-0.26	0.96	-0.56
Jiangsu	2.69	0.44	2.24	0.69	0.34	2.20	3.57	1.75	2.61	3.29
Zhejiang	1.98	1.53	1.53	1.21	0.75	1.24	0.09	1.36	0.94	0.82
Anhui	0.65	-1.19	0.23	-0.36	-0.27	-0.58	-0.32	0.01	-0.54	0.56
Fujian	0.10	0.52	0.30	0.79	3.10	0.26	-0.01	0.04	0.26	0.83
Jiangxi	0.07	-1.22	-0.12	-0.54	-0.43	-0.64	-0.38	-0.33	-0.62	-0.27
Shandong	1.09	-0.34	1.73	0.81	0.32	0.69	1.02	0.94	1.04	0.75
Henan	0.72	-0.81	1.13	0.42	0.17	-0.54	-0.36	0.96	-0.54	0.26
Hubei	0.27	-0.80	0.20	-0.17	-0.04	-0.21	-0.30	0.00	-0.10	-0.43
Hunan	0.64	-1.06	0.21	-0.33	-0.03	-0.54	-0.27	0.16	-0.50	0.86
Guangdong	1.50	1.56	2.50	3.29	2.39	2.63	3.65	3.81	2.87	2.32
Guangxi	-0.14	-0.90	-0.16	-0.51	-0.41	-0.63	-0.38	-0.15	-0.63	0.03
Hainan	-1.22	0.27	-1.05	-0.65	-0.56	-0.62	-0.38	-0.97	-0.62	-0.84
Chongqing	-0.32	0.13	-0.31	-0.48	-0.45	-0.16	-0.15	-0.39	-0.22	-0.83
Sichuan	1.52	-0.20	1.19	0.39	0.10	0.44	-0.26	0.67	0.38	0.60
Guizhou	-0.35	-0.59	-0.65	-0.67	-0.58	-0.61	-0.38	-0.13	-0.61	-0.90
Yunnan	-0.03	-0.91	-0.37	-0.63	-0.55	-0.65	-0.38	0.19	-0.64	-0.45
Tibet	-1.30	-0.65	-1.24	-0.76	-0.73	-0.69	-0.38	-1.25	-0.67	-1.46
Shanxi	-0.17	0.24	-0.24	-0.42	-0.46	0.03	-0.20	-0.01	-0.07	0.08
Gansu	-0.61	-0.52	-0.69	-0.70	-0.63	-0.66	-0.38	-0.63	-0.65	-0.95
Qinghai	-1.23	-0.02	-1.17	-0.75	-0.72	-0.69	-0.38	-1.09	-0.66	-0.92
Ningxia	-1.27	0.15	-1.13	-0.73	-0.71	-0.68	-0.38	-1.05	-0.66	-1.05
Xinjiang	-0.25	-0.16	-0.66	-0.71	-0.67	-0.65	-0.38	-0.75	-0.64	-0.48

TABLE II-(b)
STANDARDIZED DATA

Provinces（cities）	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
Beijing	-1.45	-0.30	2.89	0.03	0.50	2.30	-0.48	-1.13	-0.22	-0.48
Tianjin	-1.45	-0.04	-0.62	-0.46	-0.49	-0.16	-0.88	-0.95	-0.66	0.55
Hebei	0.66	-0.17	-0.78	-0.04	0.67	-0.27	-0.64	0.38	0.16	0.71
Shanxi	0.40	-0.55	-0.91	-0.59	-0.62	-0.52	-0.52	0.09	-0.62	-0.50
Inner Mongolia	0.57	-0.59	-1.01	-0.72	-0.46	-0.42	0.54	0.08	-0.81	-0.37
Liaoning	-0.07	-0.30	-1.43	-0.15	-0.03	-0.24	-0.37	0.41	0.19	-0.33
Jilin	-0.41	-0.59	-1.59	-0.57	-0.53	-0.65	-0.23	-0.77	-0.34	-0.19
Heilongjiang	-0.41	-0.53	-1.56	-0.70	-0.85	-0.65	-0.50	-0.23	-0.10	0.06
Shanghai	-1.45	-0.06	0.69	0.07	0.17	2.14	-0.71	-0.78	-0.25	-0.61
Jiangsu	-1.45	2.75	-0.07	2.52	2.46	0.67	1.18	0.92	2.14	2.37
Zhejiang	-1.45	2.85	1.14	1.88	1.56	0.39	0.93	1.43	1.75	-0.24
Anhui	-0.33	0.21	0.79	0.29	0.21	-0.09	0.63	0.60	-0.18	0.58
Fujian	-0.24	0.10	0.45	0.44	0.13	-0.30	-0.86	0.04	0.15	-0.34
Jiangxi	0.10	-0.27	-0.46	-0.25	-0.53	-0.35	1.78	0.19	-0.13	-0.22
Shandong	-0.50	1.46	0.04	1.82	1.99	1.52	0.77	1.55	2.23	4.04
Henan	-0.03	0.08	-1.04	0.89	0.01	0.00	0.12	0.99	0.22	0.33
Hubei	-0.20	-0.08	-0.02	0.27	0.54	-0.14	0.26	0.22	0.83	0.00
Hunan	0.66	-0.15	0.00	0.14	-0.53	-0.26	-0.95	0.15	0.78	0.92
Guangdong	-0.76	2.58	0.33	2.71	2.91	3.42	0.78	2.69	2.22	0.30
Guangxi	0.57	-0.56	-0.10	-0.52	-0.58	-0.54	0.84	-0.03	-0.23	-0.18
Hainan	-1.05	-0.68	1.88	-0.98	-1.06	-0.64	-0.83	-1.61	-1.12	-0.71
Chongqing	0.01	-0.19	0.31	-0.28	-0.22	-0.18	-0.58	-0.57	-0.51	-0.67
Sichuan	2.12	-0.16	0.71	0.23	-0.05	-0.17	3.67	1.53	0.89	-0.30
Guizhou	1.56	-0.58	0.23	-0.56	-0.68	-0.50	-0.03	-0.46	-0.22	-0.68
Yunnan	2.03	-0.50	0.45	-0.54	-0.73	-0.46	-0.09	0.40	-0.51	-0.73
Tibet	-0.29	-0.71	1.22	-1.06	-1.06	-0.74	-1.00	-1.61	-1.33	-0.81
Shanxi	1.60	-0.46	0.32	-0.38	-0.15	-0.51	-0.06	-0.14	0.00	0.14
Gansu	1.13	-0.62	-0.45	-0.79	-0.50	-0.62	-0.40	-0.49	-1.02	-0.56
Qinghai	-0.20	-0.69	-0.11	-1.01	-0.89	-0.68	-0.85	-1.43	-1.24	-0.74
Ningxia	-0.24	-0.63	-0.04	-0.96	-0.29	-0.71	-0.84	-1.31	-1.19	-0.75
Xinjiang	0.53	-0.65	-1.27	-0.73	-0.89	-0.62	-0.68	-0.15	-0.90	-0.59

TABLE III
CLUSTERING RESULTS OF THE OVERALL DEVELOPMENT OF DIGITAL ECONOMY

Methods	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Longest distance	Guangdong	Jiangsu, Zhejiang Shandong	Beijing, Shanghai	Sichuan	others
Class average	Guangdong	Jiangsu, Zhejiang, Shandong	Beijing, Shanghai	Sichuan	others
Barycenter	Beijing	others	Guangdong	Zhejiang	Jiangsu, Shandong
Sum of deviation squares	Zhejiang, Shandong, Guangdong	Hainan, Tibet, Qinghai, Ningxia	others	Beijing, Shanghai	Hebei, Anhui, Fujian, Henan Hubei, Hunan, Sichuan
K-means	Tianjin, Hainan, Tibet, Qinghai, Ningxia	Jiangsu, Zhejiang, Shandong, Guangdong	Beijing, Shanghai	Hebei, Anhui, Fujian, Henan, Hubei, Hunan, Sichuan	others

Above the indicator system established in Section II, combined with the data published on the websites of the National Bureau of Statistics, the Ministry of Commerce and the provincial people's governments, the existing data of 31 provinces and cities from 2015 to 2018 are averaged. On this basis, all the data are standardized, and the final data are as shown in Tables II-A and II-B.

B. Clustering results

According to the standardized data, under the indicator system of digital economy established in Table I, the data of 31 provinces and cities are clustered by four clustering methods (the longest distance method, the class average method, the center of gravity method, the sums of deviation squares) and K-means clustering method, and are divided into five categories. On this basis, this paper studies the categories of the holistic digital economic development in Guangdong. The pedigree map can be obtained by cluster analysis in R software.

According to the pedigree map, the distribution of 31 provinces and cities under each category is summarized, as shown in Table III.

Similarly, the clustering results about the 4 kinds of digital economy are shown in Appendix respectively.

V. RESULT ANALYSIS

A. Analysis

It is impossible to judge the advantages and disadvantages of the clustering results of the digital economic population and the 4 types of digital economy. Hence, the factor analysis is carried out on the digital economic population and the four digital economic types respectively. In the result of factor analysis, k public factor with cumulative contribution rate is greater than 80%. Based on each digital economy type which are as indicators, the corresponding score coefficient matrix $Q=\left(q_{ij}\right)_{m\times k}$ are selected. If m indicators are $X_1, X_2, \ldots, X_m$, according to the score coefficient matrix and variance contribution rate, the score of each common factor can be calculated as well as the comprehensive factor score by Lei et al. [18], and their expressions are shown in Formulae (13) and (14).

$$F_j=\sum_{i=1}^{m} q_{ij} X_i \tag{13}$$

$$F=F_j\times a_j \tag{14}$$

Where a_j is the proportion of the variance contribution rate of the j-th common factor in total variance contribution rate.

TABLE IV
SCORE TABLE OF DIGITAL ECONOMY OF PROVINCES AND CITIES

Provinces（cities）	Overall	Basic	Tech	Integrated	Service oriented
Beijing	-0.19	2.01	-1.10	0.45	-0.57
Tianjin	-0.47	-0.43	-0.53	-0.07	-0.74
Hebei	0.20	-0.03	0.39	-0.04	-0.16
Shanxi	-0.37	-0.37	-0.50	-0.54	-0.25
Inner Mongolia	-0.37	-0.46	-0.35	-0.57	0.15
Liaoning	-0.07	-0.07	-0.54	-0.20	0.01
Jilin	-0.38	-0.40	0.09	-0.45	-0.33
Heilongjiang	-0.31	-0.43	-0.04	-0.54	-0.27
Shanghai	-0.25	0.73	-0.95	0.29	-0.56
Jiangsu	1.94	0.80	2.58	2.05	0.93
Zhejiang	1.10	1.15	0.52	1.79	1.03
Anhui	0.20	-0.35	0.41	0.27	0.37
Fujian	0.10	1.10	0.30	0.22	-0.30
Jiangxi	-0.05	-0.53	-0.13	-0.30	0.78
Shandong	1.37	0.48	0.77	1.33	0.85
Henan	0.41	0.14	0.65	0.17	0.34
Hubei	0.13	-0.20	-0.24	0.20	0.29
Hunan	0.15	-0.26	0.62	-0.22	-0.27
Guangdong	2.09	2.14	2.87	2.09	1.28
Guangxi	-0.17	-0.46	0.09	-0.48	0.33
Hainan	-0.90	-0.44	-0.69	-0.37	-0.90
Chongqing	-0.37	-0.25	-0.54	-0.15	-0.43
Sichuan	0.69	0.32	0.31	-0.25	2.07
Guizhou	-0.40	-0.54	-0.36	-0.64	-0.12
Yunnan	-0.26	-0.54	0.01	-0.68	0.04
Tibet	-0.98	-0.75	-1.09	-0.55	-0.99
Shanxi	-0.13	-0.20	-0.03	-0.44	-0.06
Gansu	-0.54	-0.57	-0.58	-0.65	-0.41
Qinghai	-0.86	-0.58	-0.76	-0.60	-0.87
Ningxia	-0.82	-0.54	-0.81	-0.42	-0.83
Xinjiang	-0.49	-0.47	-0.41	-0.70	-0.41

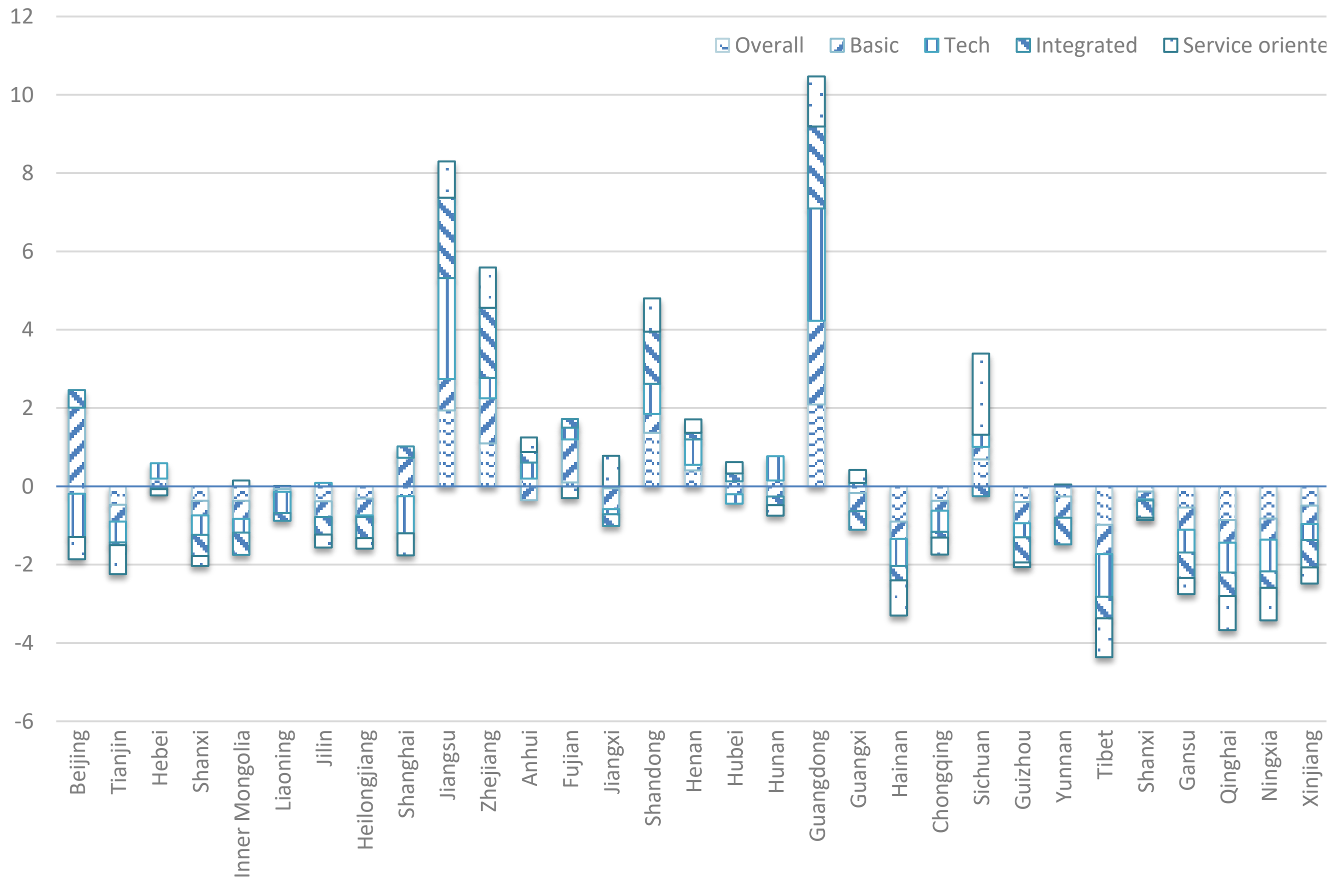

Fig. 1. Score of digital economy of provinces and cities

According to Formulas (13) and (14), we can calculate the comprehensive factor scores of 31 provinces and cities under the overall digital economy and the four major types of digital economy. The results are shown in Table IV.

With comparing the data in Table IV, the clustering results match the comprehensive factor scores obtained by factor analysis based on the works of MacCallum et al. [19]. Provinces (cities) with higher scores can often form a category. In addition, those with lower scores or close scores can also be divided into one category.

Thus, the results of clustering and factor score well confirm the growth of digital economy in various regions.

B. Results and Advice

1) From the view of the overall digital economic development, Guangdong's digital economy is divided into a group with Jiangsu, Zhejiang and Shandong under the method of sum of squares of deviation and K-means clustering, but it is divided into a separate category under the method of longest distance, class average and gravity center. Then, the comprehensive factor score is also the highest. It illustrates that although other digital economy provinces are developing better, the fact still be not able to shake the leading position of Guangdong as the largest digital economy province.
2) From the analysis of the results of basic digital economic clustering, although Guangdong is in the first place in most clustering methods, Beijing can also become a separate category under the longest distance method, class average method and barycenter method. The difference of comprehensive factor score between Beijing and Guangdong is only 0.13, indicating that the gap between Beijing and Guangdong in the development of basic digital economy has not been widened.
3) From the aspect of technology-based digital economy, Jiangsu and Guangdong are in the same category under K-means clustering and two kinds of systematic clustering, and the scores of comprehensive factors are also very close, indicating that the scale of technological digital economy is not the unique one in Guangdong, and Jiangsu is likely to catch up with or even surpass Guangdong in the future.
4) According to the clustering results of the integrated digital economy, except for the gravity method, Guangdong and other provinces can be grouped into the same category, and the factor comprehensive score difference is not big, especially in Jiangsu, the score difference is only 0.04. It is thus clear that the integration of digital economy, primary industry and secondary industry in Guangdong Province should be strengthened.
5) Based on the clustering results about the service-oriented digital economy, Guangdong is a separate category under the three clustering methods, but the results of comprehensive factor score show that there is a big difference of 0.79 between Guangdong and Sichuan. In addition, the clustering results obtained by K-means and sum of squares of deviation show that Guangdong and Sichuan do not form a category, showing that the application degree of digital economy in service industry in Guangdong is quite different from that in Sichuan.

The as seen in Fig. 1, we easily draw a clear conclusion that Guangdong is the industry leader of digital economy, and has great potential in service-oriented.

Therefore, it is necessary to strengthen the construction of service-oriented digital economy. Making use of the achievements of digital economy rapid development to benefit the public should become the next goal of Guangdong.

The results obtained from the development of digital technology should not only play a role in some important fields, but also be close to life and bring convenience to people's lives. When the digital products are universal and popular, people can enjoy the benefits brought by the development of information technology.

Consequently, the development of service-oriented digital economy can go up to a higher level.

APPENDIX

CLUSTERING RESULTS OF BASIC DIGITAL ECONOMY

Methods	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Longest distance method	Guangdong	Beijing	Shanghai, Fujian	others	Hebei, Jiangsu Zhejiang, Shandong Henan, Sichuan
Class average method	others	Hebei, Jiangsu Zhejiang, Shandong Henan, Sichuan	Guangdong	Beijing	Shanghai, Fujian
Barycenter method	Fujian	Jiangsu, Zhejiang	Beijing	others	Guangdong
Method of sum of deviation squares	Beijing, Shanghai Fujian	Jiangsu, Zhejiang Guangdong	others	Hebei, Shandong Henan, Sichuan	Anhui, Jiangxi Hubei, Hunan Guangxi, Yunnan
K-means clustering	others	Hebei, Shanxi Liaoning, Anhui Jiangxi, Henan Hunan, Guangxi Hubei, Shanxi	Jiangsu, Zhejiang Shandong, Sichuan	Guangdong	Beijing, Shanghai Fujian

CLUSTERING RESULTS OF TECHNOLOGICAL DIGITAL ECONOMY

Methods	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Longest distance method	Jiangsu, Guangdong	Zhejiang, Shandong	Beijing, Shanghai	Hebei, Jilin Anhui, Fujian Henan, Hunan Sichuan	others
Class average method	Jiangsu	Guangdong	Beijing	others	Shanghai, Zhejiang Fujian, Shandong Sichuan
Barycenter method	Jiangsu	Guangdong	Beijing	Shanghai	
Method of sum of deviation squares	Shanxi, Hainan Guizhou, Tibet Gansu, Qinghai Ningxia, Xinjiang Inner Mongolia	others	Jiangsu, Guangdong	Beijing, Shanghai	Zhejiang, Fujian Shandong, Sichuan
K-means clustering	others	Beijing, Shanghai	Jiangsu, Guangdong	Hebei, Jilin Anhui, Henan Jiangxi, Hubei Hunan, Guangxi Yunnan, Shanxi Heilongjiang	Sichuan, Fujian Shandong

CLUSTERING RESULTS OF INTEGRATED DIGITAL ECONOMY

Methods	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Longest distance method	Jiangsu, Zhejiang Shandong, Guangdong	Beijing, Hainan Tibet	Sichuan, Guizhou Yunnan, Shanxi	Shanghai, Anhui Fujian, Hubei	others
Class average method	Jiangsu, Zhejiang Shandong, Guangdong	Sichuan, Guizhou Yunnan, Shanxi	others	Beijing	Hainan, Tibet
Barycenter method	Zhejiang	Jiangsu, Shandong Guangdong	Beijing	Hainan, Tibet	others
Method of sum of deviation squares	Jiangsu, Zhejiang Shandong, Guangdong	Tianjin, Shanghai Anhui, Fujian Hubei	Beijing, Hainan Tibet	Sichuan, Guizhou Yunnan, Shanxi	others
K-means clustering	Beijing, Shanghai Hainan, Tibet	Sichuan, Yunnan Shanxi, Gansu Guizhou	Zhejiang, Shandong Guangdong, Jiangsu	others	Hebei, Anhui Fujian, Henan Hunan, Hubei Chongqing

CLUSTERING RESULTS OF SERVICE-ORIENTED DIGITAL ECONOMY

Methods	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Longest distance method	Hebei, Liaoning Zhejiang, Anhui Fujian, Jiangxi Henan, Hubei Hunan, Guangxi Inner Mongolia	others	Sichuan	Guangdong	Jiangsu, Shandong
Class average method	Zhejiang, Sichuan	Beijing, Shanghai	others	Guangdong	Jiangsu, Shandong
Barycenter method	Guangdong	Sichuan	Jiangsu, Shandong	Zhejiang	others
Method of sum of deviation squares	others	Hainan, Tibet Qinghai, Ningxia	Beijing, Shanghai	Hainan, Tibet Qinghai, Ningxia	Jiangsu, Shandong Guangdong
K-means clustering	Tianjin, Jilin Chongqing, Tibet Gansu, Ningxia Hainan, Qinghai Ningxia, Xinjiang	Zhejiang, Sichuan Jiangxi	others	Jiangsu, Shandong Guangdong	Beijing, Shanghai

REFERENCES

[1] K. Li, D. J. Kim, K. R. Lang, R. J. Kauffman, M. Naldi, "How Should We Understand the Digital Economy in Asia?" *Critical Assessment and Research Agenda*, pp101004-101004, 2020

[2] S. Rita, S. Marlies, V. Simone, "A systemic perspective on socioeconomic transformation in the digital age," *Journal of Industrial and Business Economics*, vol. 46, no.3, pp361-378, 2019

[3] S. B. Mihaela, R. S Monica, "Regional patterns and drivers of the EU digital economy," *Inclusive and sustainable economic growth. Challenges, measures and solutions*, pp1-25, 2019

[4] P. Schweighofer, S. Grunwald, M. Ebner, "Technology Enhanced Learning and the Digital Economy," *International Journal of Innovation in the Digital Economy*, 2015

[5] S. Talar, "The EU Concept of Digital Economy Against the Internet Economy Model," *Proceedings of the 4 th International Conference on European Integration 2018*, 2018

[6] F. Landini, "The Evolution of Control in the Digital Economy," *Economics Department Working Papers*, 2012

[7] A. A. R. Del, A. Padilla, C. Serarols, J. M. Veciana, "Digital economy and management in Spain," *Internet Research*, vol. 13, no.1, pp6-16, 2003

[8] G. M. K. Berdykulova, A. I. U. Sailov, S. Y. K. Kaliazhdarova, E. B. U. Berdykulov, "The Emerging Digital Economy: Case of Kazakhstan," *Procedia - Social and Behavioral Sciences*, vol. 109, pp1287-1291, 2014

[9] N. I. Protopopova, V. D. Grigoriev, S. Y. Perevozchikov, "Information and Digital Economy as an Economic Category," *Perspectives on the use of New Information and Communication Technology (ICT) in the Modern Economy*, Springer, Cham, 2017

[10] N. N. Skiter, O. A. Donskova, L. S. Sagatelova, E. A. Komarova, D. I. Nesterenko, "Food Security in the Conditions of Digital Economy: Tendencies and Perspectives," *Perspectives on the use of New Information and Communication Technology (ICT) in the Modern Economy*, pp238-243, 2017

[11] H. Dong, W. Han, "Analysis of Digital Economy Concepts and Measurement Methods at Home and Abroad," *Proceedings of the 2019 4th International Conference on Financial Innovation and Economic Development*, 2019

[12] S. S. Nudurupati, S. Tebboune, J. Hardman, "Contemporary performance measurement and management (PMM) in digital economies," *Production Planning & Control*, pp1-10, 2015

[13] J. T. Wei, S. Y. Lin, Y. Z. Yang, H. H. Wu, "Using a combination of RFM model and cluster analysis to analyze customers' values of a veterinary hospital," *IAENG International Journal of Computer Science*, vol. 47, no.3, pp442-448, 2020

[14] D. Y. Kim, S. T. Park, M. H. Ko, "A study on the analysis of IT-related occupational cluster using big data," *IAENG International Journal of Computer Science*, vol. 45, no.1, pp7-11, 2018

[15] L. Jiang, D. T. Xie,"An efficient differential memetic algorithm for clustering problem," *IAENG International Journal of Computer Science*, vol. 45, no.1, pp118-129, 2018

[16] H. X. Gao, "Applied multivariate statistical analysis," *Beijing: Peking University Press*, 2005(Chinese)

[17] X. J. Ning, "Research on the Development of Digital Economy in Hubei Province Based on Technical-Economic Paradigm," *Huazhong University of Science and Technology*, 2018 (Chinese)

[18] P. Lei, X. Y. Gu, "Aggregate Analysis on the Economic Development of the Belt and Road Countries --Based on Cluster Analysis and Factor Analysis," *Foreign Economic Relations & Trade*, no.2, pp10-15+9, 2020 (Chinese)

[19] R. C. MacCallum, K. F. Widaman, S. B. Zhang, S. Hong, "Sample size in factor analysis," *Psychological Methods*, vol. 4, no.1, pp84-99, 1999