Analysis of Regional Innovation Capability Based on Information Security Data Mining Algorithm

Meng Li

Vocational Technical Education Institute, Xinxiang Vocational and Technical College, Xinxiang, Henan, 453000, China

Abstract.

In order to explore the correlation of population structure and regional innovation capabilities, this paper uses big data technology as the basis to mine massive data and find out the main factors that affect regional innovation capabilities. Moreover, this article studies the influence of population structure on regional innovation capabilities and explores the internal correlation between population structure and regional innovation capabilities, which expands the research field of factors affecting regional innovation capabilities. In addition, this article initially establishes a relatively complete theoretical research framework on the influence of population structure on regional innovation capabilities and constructs a basic model of the correlation between population structure and regional innovation capabilities by combining theory with reality. Finally, this article uses actual data analysis and big data technology to carry out experimental analysis. From the research results, it can be seen that the system constructed in this paper has a certain effect, which demonstrates the correlation of population structure and regional innovation capabilities.

Keywords: Big data technology; population structure; regional innovation; correlation; information security

1. INTRODUCTION

With the continuous development and progress of social economy, a series of problems such as the rapid increase of population, the imbalance of population sex ratio, the aging of population, the unreasonable distribution of population between urban and rural areas, and the increasing difference of population culture and education have become more and more widespread concerns all over the world, and China is facing the same kind of problems.

Historically, population structure is one of the most important and fundamental issues in human economy and society. Whether it is the gender, age, ethnicity, quality structure or urban-rural and employment structure of the population, it has always been a hot topic of concern, research and discussion among scholars from all walks of life. Such issues are essential to our lives. For example, gender imbalance will have an impact on employment and marriage, reducing the happiness index and social stability. In addition, the imbalance in age composition will have a serious impact on social and economic development and welfare, and the aging of the population will increase the pressure on social security and public service systems, which will affect intergenerational relations in society. If this demographic imbalance is not effectively addressed and tends to increase, it will seriously affect the sustainable development of the regional economy, thus affecting the quality of life and income levels of the population and even the stability of the region [1].

After a certain level of economic development, demographic factors will become the main factor affecting innovation [2]. The competition between countries in the future is mainly in the field of innovation, and China, as a country with a large population, is more likely to have a large-scale talent pool and market to gain an innovation advantage. Compared with the 1970s, the demographic structure of China has changed significantly in the new period. As urbanization continues to advance, more people are leaving the countryside to move to live in the cities and towns. With the industrial transformation and upgrading, the number of workers employed in the secondary industry has been decreasing. At the same time, the population popularized by higher education has been increasing as institutions of higher education continue to expand their enrollment quota. As the labor force demographic dividend gradually wanes and disappears, China is gradually moving into an aging society. The changes in these important demographic elements may have an impact on building China into an innovation-driven country [3]. Currently, there are still large gaps in the level of innovation among regions in China, and the innovation capabilities of eastern, central and western regions are extremely unbalanced. Focusing on the actual situation, how to provide impetus for innovation and achieve innovation equalization is a problem that needs to be studied [4]. Therefore, this paper, in the context of China's urbanization, industrial transformation, higher education popularization and population aging, will study the impact of China's demographic changes on the innovation capacity of each province and its neighboring provinces in terms of their demographic structure through empirical tests, and identify the proximity effects arising from each region's innovation development to provide suggestions for the government to formulate demographic and innovation policies.

Based on this, this paper analyzes the correlation between population structure and regional innovation capacity through big data technology.

2.RELATED WORK

The discussion of the factors that can influence technological innovation capabilities is richly supported by a considerable literature both domestically and internationally. The literature [5] empirically examines the specific factors that influence patenting in different countries and shows that while resource inputs can certainly have a significant impact on innovation, policy factors that influence R&D productivity, such as open trade, the degree of technological differentiation, and the different stock of knowledge in each country are more significant variables. The literature [6] demonstrates, using a generational overlap model, that the efficiency of capital markets directly affects the market's choice of technology. Excessive transaction costs in the capital market will reduce the frequency of transactions and when the capital market is active, the capital market will tend to focus on technologies that have longer maturities but higher returns, thus bringing a significant growth impact on the economy. The literature [7] concludes that the allocation of R&D among and within technological innovation agents and the internal and external institutional factors have an important impact on the performance of technological innovation. All these literatures provide valuable references for the study of this paper.

Literature [8] defines technological innovation as the act of transforming scientific and technological achievements into productivity and commodities. It is a concept produced by the combination of science and technology and economy. It has the essential operability of scientific and technological development and the practicality of macroeconomic measures. In space, technological innovation is manifested as a system, and in time, it is manifested as a dynamic continuum. Literature [9] proposes that technological innovation is a process that integrates technology and economy, so that technology and economy are organically combined and transformed into each other. Specifically speaking, market-oriented, research potential demand, to develop new methods, new processes, and new products to obtain commercial profits and good economic benefits. It is a commercial activity jointly organized by enterprises, entrepreneurs and inventors for the purpose of profiting, starting from market needs and achieving the value of commodities as the end point. Literature [10] pointed out that technological innovation refers to the whole process from the conception of new products and new processes to production and market applications. It is the construction process of the value chain of material wealth with the participation of R&D research and invention. Literature [11] believes that technological innovation is the totality of all technological and economic activities including the production of new products and new process ideas, research and development, the realization of commercial benefits, and the entire process of new technology diffusion. Due to changes in the social and economic environment, the definition of the concept of technological innovation is constantly being given new connotations. Throughout the research and discussion at home and abroad, technological innovation can be regarded as the commercialization process of pursuing economic benefits and commercial profits and transforming scientific and technological achievements [12]. On this premise, the optimal integration of resources through the form of the enterprise, the transformation and upgrading of traditional technologies and models, and even fundamental subversion to promote the establishment of high barriers to entry, achieve market share and squeeze out competitors Purpose, so as to obtain monopoly benefits and market share [13]. Although a small number of companies may monopolize the industry, the entire society may also benefit from it, making major technological breakthroughs and even breaking out a technological revolution, thereby achieving leapfrog progress.

Literature [14] shows that with the intensification of competition, R&D investment will increase at the beginning, but later it may cause a decline in R&D activities to the extent that it is lower than the optimal level of R&D. Literature [15] used detailed data and a larger market sample to study the relationship between innovation speed and market concentration, and finally came to an empirical conclusion that there is a significant positive coefficient relationship between market concentration and innovation speed . Literature [16] believes that the most suitable for innovation is the market structure of monopolistic competition. In this case, it can not only ensure the capital and material conditions required for innovation, but also maintain the enterprise's sense of crisis and expand the driving force of occupying the market.

Literature [17] In a country's leapfrog economy, a market economy with free competition is not enough, and emphasizes the importance of accelerating the establishment of a national innovation system under state intervention. Literature [18] analyzes the composition of the national innovation system from a micro perspective, and explores the internal reasons for the international innovation system. Literature [19] proposes that innovation is a learning process in which producers and users interact, and the core of the national innovation system is learning activities. Literature [20] applies the domestic competitive advantage theory to the international field and proposes the famous "diamond" model.

Literature [21] studied the relationship between innovation performance and the spillover of international technology exchanges, and proposed a policy proposal for more open cooperation and information sharing. A centralized and mutual assistance platform can greatly promote innovation. The literature [22] verifies that there is a very significant causal relationship between the national innovation system and economic growth. At the same time, the government's support for innovation, the political structure and the degree of openness of the economy, and other variables also play an important role in development. Literature [23] believes that there are many theories for reference, but the more important thing to promote innovation is how to transform the theoretical views of the national system innovation school into practical policies that are practical in reality, and how to find out in the process of implementation The best strategy can achieve the best efficiency, which requires the government to do further detailed research and work.

3.ANALYSIS OF BALANCE OF POPULATION STRUCTURE

The balanced population development refers to the state of mutual matching, interdependence and synergistic development of the components of the population system, which means that the supply of population and the demand for population are in a relatively balanced and dynamically coordinated situation in terms of quantity or structure. The balanced development of population structure means that the gender structure, age structure, industrial structure, quality structure, urban-rural structure and ethnic structure of the population are in a relatively balanced or dynamically coordinated state, and the influencing factors gradually change from unequal to equal and from uncoordinated to coordinated.

Firstly, balanced development of population structure means that each structure within the population gradually achieves balance, and these structures interact with each other, promote each other, and constrain each other, and jointly promote the overall population structure to achieve balanced development. Secondly, the population structure balance is not static, but constantly changing, and the structures interact with each other, and from unbalanced to balanced, and from balanced to unbalanced, so that the population structure balance constantly develops to a high level balance, that is, from low level balance to high level balance succession.

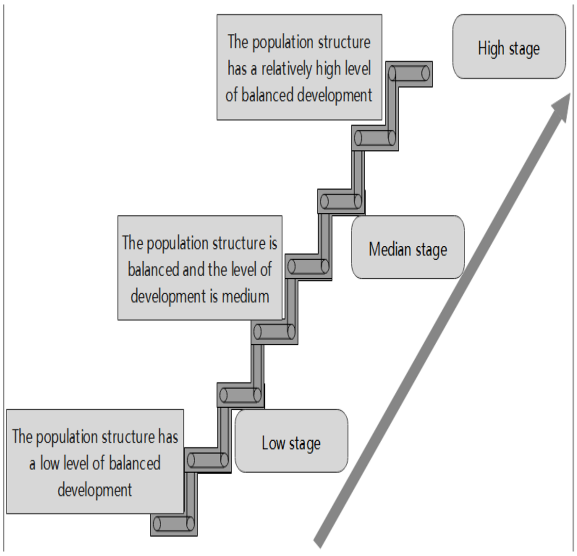

Figure 1 Gradient diagram of balanced development of population structure

From the definition of balanced development of population structure, it includes not only the coordinated development of sub-structures within the population structure, but also the evolution of balanced population structure, i.e., the development from lower to higher levels, both of which are indispensable. Therefore, this paper evaluates the demographic structure from two perspectives: the level of balanced development of the demographic structure and the degree of balanced demographic structure.

3.1. Construction of evaluation index system and data processing

(1) Construction of evaluation index system

In order to better reflect the specificity of the population structure, this paper follows the principles of scientificity, systematicity, independence and operability to construct the index system, as shown in Table 1.

Table 1 The balanced development index system of population structure and its weight

Target layer	Constraint layer	Index layer	Unit	Indicator attributes	Index Weight
Balanced population structure development	Gender structure	Sex ratio X1	%	Moderate index	0.001
	Age structure	Proportion of working-age labor force X2	%	Positive index	0.143
		Total dependency ratio X3	%	Reverse indicator	0.036
	Industrial structure	Proportion of employed population in the secondary and tertiary industries X4	%	Positive index	0.151
	Quality structure	Average years of education per capita X5	year	Positive index	0.235
	Urban-rural structure	Proportion of urban population X6	%	Positive index	0.202
	National structure	National ratio X7	%	Moderate index	0.233

(2) Data processing and index weight determination

Due to the different dimension and magnitude of the original index data, in order to facilitate the convenience and accuracy of data processing in the article, the original index data is standardized first. The calculation formulas are:

If the index value is larger, the better:

$$X_{ti}^{'} = \frac{X_{ti} - X_{t\min}}{X_{t\max} - X_{t\min}} \quad (1)$$

If the index value is smaller, the better:

$$X_{ti}^{'} = \frac{X_{t\max} - X_{ti}}{X_{t\max} - X_{t\min}} \quad (2)$$

In the formula, $X_{ti}^{'}$ represents the standardized value of the i-th index of the population t substructure, X_{ti} represents the i-th variable of the population t substructure, and $X_{t\max}$ and $X_{t\min}$ represent the maximum and minimum values of the i-th variable in the t substructure, respectively.

In order to determine the importance of each index in the system, it is necessary to calculate the index weight of each index. Here, the coefficient of variation method is used to calculate the index weight. The calculation formula is:

$$\alpha_{ti} = \frac{\delta_{ti}}{\sum_{i=1}^{n} \delta_{ti}} \quad (3)$$

Among them,

$$\delta_{ti} = \frac{\sigma_{ti}}{\overline{Z_{ti}^{'}}} \quad (4)$$

In the formula, δ_{ti} is the standard deviation coefficient of the i-th index in the t substructure, and σ_{ti} and $\overline{Z_{ti}^{'}}$ respectively represent the standard deviation and the mean value of the standardized value of the i-th index in the t substructure.

3.2. Construction of a balanced evaluation method for population structure

(1) Development level construction of population structure

From the above, it is clear that the level of development of population structure involves gender structure, age structure, industrial structure, quality structure, urban-rural structure, and ethnic structure. The method of measuring the level of development of population structure is as follows.

$$E = \sum_{i=1}^{m} \alpha_t P_t \quad (5)$$

$$P_t = \sum_{i=1}^{n} \alpha_{ti} X_{ti}^{'} \quad (6)$$

In the formula, E is the population structure development level, α_t is the weight of the population t substructure, P_t is the development level of the population t substructure, α_{ti} is the weight of the i index of the population t substructure, and $X_{ti}^{'}$ is the standardized value of the i-th index of the population t substructure.

(2) Construction of population structure balance

From the perspective of the concept of balanced development of the population structure, the essence of the population structure and each other is to achieve balance from the previous imbalance and to achieve coordination from the previous no coordination. The specific steps are as follows:

1) The establishment of the set of factors to be judged:

$$X = \left\{ x_1^{'}, x_2^{'}, \cdots, x_n^{'} \right\} \quad (7)$$

2) If

$$X^m \subset R^m \to [0,1] \quad (8)$$

Then, $D(x)$ is the balance degree of X, where

$$D(x) = 1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^{'} - \frac{1}{n}\sum_{j=1}^{n} x_j^{'} \right)^2} \quad (9)$$

In the formula, $X_i^{'}$ represents the standardized value of each indicator in the balanced development indicator system of the population structure. When the value of D is larger, it indicates that the balance of population structure is higher. When the value of D is smaller, the balance of population structure is worse. According to the size of balance, it can be divided into nine states.

3. The construction of the sustainable development degree of balanced population structure

The realization of the balanced development of the population structure is not achieved overnight, but a gradual and gradual process of realization. The population structure balance sustainable development (SD) index can be used to reflect the sustainable development capacity of the population structure, and its calculation formula is as follows:

$$SD = \sqrt{E \times D} \quad (10)$$

In the formula, SD is the degree of sustainable development of a balanced population structure, E is the balanced development level of the population structure, and D is the balance degree of the population structure. The higher the SD value, the stronger the ability of balanced and sustainable development of the population structure, that is, the better the foundation for further development. Conversely, the smaller the SD value, the worse the ability of the population structure to balance sustainable development, that is, the worse the foundation for further development.

4.ANALYSIS OF THE CORRELATION BETWEEN POPULATION STRUCTURE AND REGIONAL INNOVATION

Based on the principles of scientificity, operability, completeness, dynamism, principal components and independence, this paper starts from the intrinsic linkage between population structure and regional economic development, and then conducts theoretical analysis and preliminary screening of indexes according to the current situation of the study and the specific situation as well as the accessibility and reliability of data. The population structure system was finally determined to include 6 aspects such as age structure, ethnic structure, gender structure, quality structure, industrial structure and urban-rural structure. The regional economic system includes 5 aspects such as total economic volume, economic level, industry above the scale, domestic trade and agricultural economy, as shown in Table 2.

Table 2 Evaluation index system

Coupling system	Indicator type	index
population structure	Age structure	Proportion of population aged 0-14 (X1), proportion of population aged 15-64 (X2), proportion of population aged 65 and above (X3)
	National structure	National ratio (X4)
	Gender structure	Sex ratio (X5)
	Quality structure	The illiterate population (X6), the primary and junior high school population (X7), the senior high school and technical secondary school population (X8), the junior college population and above (X9)
	Employment structure	Proportion of population with primary birth (X10), proportion of population with secondary birth (X11), proportion of population with tertiary birth (X12)
	Urban-rural structure	Proportion of urban population (X13)
Regional economy	Economic aggregate	Gross domestic product (Y1), first gross output value (Y2), second gross output value (Y3), third gross output value (Y4)
	economic level	Per capita GDP (Y5), total social fixed assets (Y6), average wage of employees (Y7), local fiscal revenue (Y8), local fiscal expenditure (Y9), per capita disposable income of urban residents (Y10)
	Above-scale industries	Gross Industrial Output Value (Y11)
	Domestic Trade	Total retail sales of consumer goods (Y12)
	agricultural economy	Agricultural output value (Y13)

(1) Determine the analysis sequence. The two analysis sequences in this article are the population structure sequence group $\left(X_i\right)$ and the regional economic sequence group $\left(Y_j\right)$.

(2) Data processing. Due to the difference in the magnitude and order of magnitude of the original index data, in order to facilitate comparison and analysis, this paper uses the method of extreme difference standardization to dimensionlessly process the data before conducting gray correlation analysis.

$$X_i^{'} = \frac{\left(X_i - X_{\min}\right)}{\left(X_{\max} - X_{\min}\right)} \quad (11)$$

$$Y_j^{'} = \frac{\left(Y_i - Y_{\min}\right)}{\left(Y_{\max} - Y_{\min}\right)} \quad (12)$$

In the formula, $X_i^{'}$ is the dimensionless value of the i-th population structure index, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the i-th population structure index respectively, $Y_j^{'}$ is the dimensionless value of the j-th regional economic index, and $Y_{\max}$ and $Y_{\min}$ are the maximum and minimum values of the j-th regional economic index respectively.

(3) Correlation coefficient. The correlation coefficient is the relative difference between two mutually comparable sequences at the t-th moment (or region), and it is difficult to compare the two sets of sequences as a whole because the information is scattered and has many values. Therefore, the correlation degree is calculated on the basis of the correlation coefficient.

$$\xi_{ij}\left(t\right) = \frac{\min_i \min_j \left|X_i^{'}\left(t\right) - Y_j^{'}\left(t\right)\right| + \rho \max_i \max_j \left|X_i^{'}\left(t\right) - Y_j^{'}\left(t\right)\right|}{\left|X_i^{'}\left(t\right) - Y_j^{'}\left(t\right)\right| + \rho \max_i \max_j \left|X_i^{'}\left(t\right) - Y_j^{'}\left(t\right)\right|} \quad (13)$$

In the formula, $\xi_{ij}\left(t\right)$ is the correlation coefficient between the i-th population structure and the j-th regional economic index at time t in each prefecture; $X_i^{'}$ and $Y_j^{'}$ are the standardized values of the i-th population structure and the j-th regional economic index at time t in each prefecture respectively; ρ is the resolution coefficient. Since population and economy are two equal subsystems, the value is generally 0.5.

(4) Obtain the relational degree. The expression of the relational degree γ between population structure and regional economy is:

$$\gamma_{ij} = \frac{1}{k}\sum_{j=1}^{k}\xi_{ij}\left(t\right) \quad (14)$$
$$k = 1, 2, \cdots, n$$

In the formula, k is the number of samples, that is, the number of regional economic indexes (or population structure indexes) selected in this article. The value range of relational degree is $\left(0{,}1\right]$, and the larger the value, the greater the correlation between systems.

On the basis of the incidence matrix, the relational degree model can be obtained by averaging the rows and columns respectively:

$$d_i = \frac{1}{l}\sum_{j=1}^{l}\gamma_{ij} \quad (15)$$
$$i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, l$$

$$d_j = \frac{1}{m}\sum_{i=1}^{m}\gamma_{ij} \quad (16)$$
$$i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, l$$

In the formula, d_i is the average relational degree of the i-th index in the population structure system and the regional economic system, d_j is the average relational degree of the j-th index in the regional economic system and the population structure system, and m and l are respectively the number of indicators of population structure and regional economic system.

4.ANALYSIS MODEL OF POPULATION STRUCTURE AND REGIONAL INNOVATION CAPABILITY BASED ON BIG DATA TECHNOLOGY

With the informatization, systematization and globalization of national societies, innovation ecosystems in a given country or region cannot be isolated, but rather are closely linked to the outside world. Therefore, in the innovation ecosystem, it is the frequent entry of innovative species from the outside that drives the internal population competition and community change, resulting in the flourishing or decay of the whole system. Therefore, the innovation subjects, innovation resources and innovation environment inside and outside the innovation ecosystem must be interconnected, constrained and promoted to form a synergistic and open network, so that the innovation ecosystem can evolve and develop continuously. Based on the above concepts, characteristics and evolutionary mechanism, this paper summarizes the structure of innovation ecosystem operation mechanism (Figure 2). In the innovation ecosystem, the roles of the main innovation agents are as follows: universities, governments, intermediaries, and financial institutions mainly provide talent, policies, information, and financial flows for the innovation ecosystem, respectively, and research institutions provide basic R&D for enterprises and are rewarded by enterprises, and users, as innovation application groups, provide various feedbacks for enterprises. These different innovation agents intertwine with each other and generate a huge innovation network system with competition and symbiosis, self-organized dynamic evolution and open collaboration through various resources in the system environment.

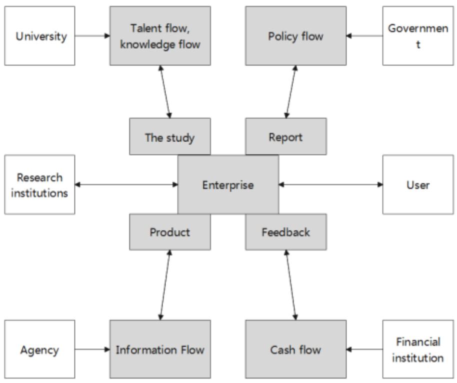

Figure 2 Structure diagram of the operation mechanism of the innovation ecosystem

The radiality of the city's innovation capacity is determined by the characteristics of the city itself. The city itself has the following characteristics: scientific and technological progress drives the formation of the city; modern industry gathers in the city; the city integrates multi-centers of industry, transportation, finance, culture and art, post and telecommunications, and information. Moreover, modern material and technical equipment, convenient transportation and modern means of information transmission, abundant library materials, and the vast majority of society's productive forces are concentrated in cities, which determine the centrality of cities in regional technological innovation. The radiation of the city's innovation capacity is mainly reflected in the ratio and linkage of the city constantly transferring and spreading technological innovation results to the surrounding areas, transporting scientific and technological talents, providing advanced equipment and improving various technological means through technological innovation development, prompting the changes of the regional economic structure to be constantly rationalized and providing strong support for the economic and social development of the surrounding areas. The structure of the city's innovation capacity is shown in Figure 3.

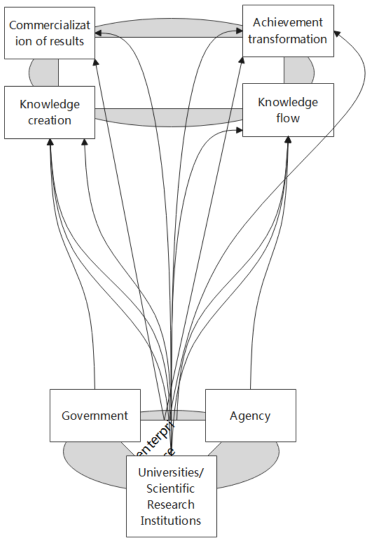

Figure 3 The structure of the city's innovation capability

Although economists generally agree on the scale effect from population agglomeration, they set different conditions on the optimal size of urban population agglomeration and come to different research conclusions. The whole economy can be considered as an aggregate or urban system and is specifically modeled as shown in Figure 4.

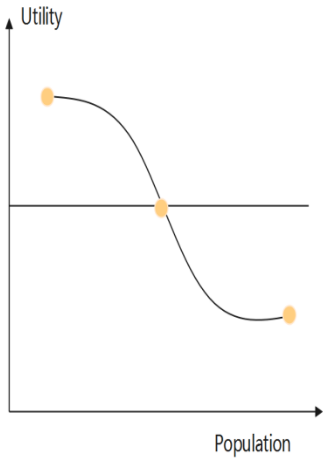

Figure 4 City size and utility

Point D in Figure 4 is the best size of all cities, and any point in the arc om indicates that the city size is too large. If a large number of people move to the new city of the best size, people will be able to make profits. Different cities have different effects. External economies tend to occur in specific industries, but external diseconomies depend on the overall size of the city, and it is this urban size diseconomy that makes it pointless to cluster industries that do not have spillover effects on each other in a city. For example, since the steel industry and the publishing industry do not share a common external economy, then steel mills and publishing houses should be located in different cities. This would create neither congestion nor high land rents. Therefore, each city should specialize in one or a few industries that bring external economies to it. The size of these external economies is likely to vary considerably with industry. As in Figure 5, a textile city may have only a handful of factories, while a financial center should ideally include almost all of a country's financial sectors. Therefore, the optimal size of a city depends on the role of its industry.

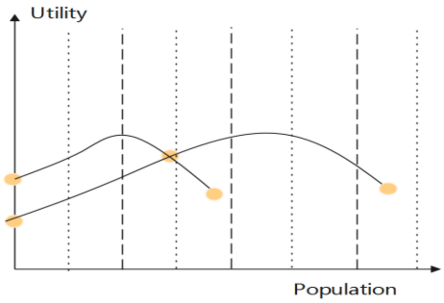

Figure 5 Population size and utility

1. Analysis of the correlation between population structure and regional innovation capabilities based on big data technology

A super-efficient DEA model is used, which requires the number of decision units to be greater than the product of the number of input and output indicators and greater than three times the sum of the number of input and output indicators. In this paper, we choose the innovation data of 30 regions from 2011-2020 as different decision units for efficiency measurement, and the number of decision units is 240, which meets the model requirements.

The innovation input and output data are imported into the model, and the calculation process is completed by using MyDEA software to finally obtain the innovation capacity measurement values for 30 regions for a total of 10 years. The measurement results are shown in Table 3 and Figure 6.

Table 3 Statistical table of the measurement values of regional innovation ability

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	1.0	1.0	1.1	1.0	1.1	1.0	1.1	1.1	1.1	1.1
2	0.8	0.8	0.8	0.8	0.9	0.9	1.0	1.0	0.9	1.0
3	0.4	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.6
4	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.5	0.5	0.5
5	0.4	0.3	0.2	0.2	0.5	0.3	0.3	0.3	0.3	0.4
6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.8	0.8
7	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7
8	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7
9	0.8	0.8	0.9	1.0	1.0	0.8	0.9	0.9	0.9	1.0
10	0.9	0.9	0.9	0.9	0.9	0.9	1.0	0.9	1.0	1.0
11	0.9	0.9	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0
12	0.6	0.6	0.6	0.8	0.7	0.8	0.8	0.7	0.7	0.7
13	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.8
14	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.7	0.7	0.7
15	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
16	0.5	0.4	0.4	0.4	0.4	0.5	0.5	0.6	0.6	0.6
17	0.7	0.6	0.6	0.7	0.7	0.6	0.7	0.7	0.6	0.7
18	0.6	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7
19	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9
20	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.6	0.7	0.7
21	0.6	0.6	0.5	0.4	0.5	0.4	0.5	0.6	0.6	0.6
22	0.6	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.8	0.8
23	0.6	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.8
24	0.5	0.4	0.4	0.4	0.5	0.7	0.7	0.7	0.6	0.7
25	0.5	0.6	0.5	0.4	0.5	0.5	0.6	0.5	0.5	0.6
26	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.7
27	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7

28	0.4	0.5	0.4	0.3	0.3	0.3	0.3	0.4	0.4	0.5
29	0.4	0.4	0.4	0.2	0.3	0.4	0.4	0.4	0.5	0.5
30	0.4	0.4	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.6

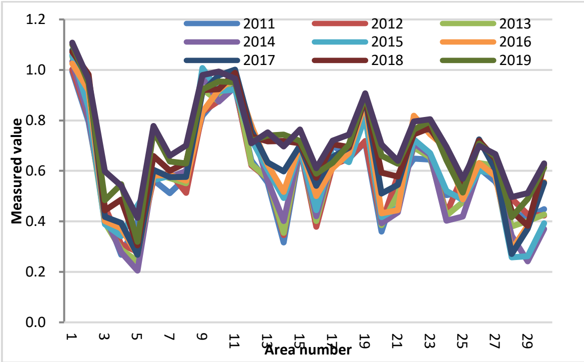

Figure 6 Statistical diagram of the measurement values of regional innovation ability

Based on the spatial adjacency weight matrix of Rook's assignment method, Stata software is applied to calculate the global Moran I index and Geary C index to verify the existence of spatial correlation of regional innovation capacity.

Table 4 Global spatial correlation test of regional innovation capability

Years	Moran's I	Z	P
2011	0.176	1.902	0.057
2012	0.196	2.075	0.038
2013	0.176	1.894	0.058
2014	0.223	2.296	0.022
2015	0.222	2.304	0.021
2016	0.236	2.418	0.016
2017	0.266	2.704	0.007
Years	0.298	3.001	0.003
2008	0.256	2.631	0.009
2009	0.292	2.952	0.003
2010	Geary's C	Z	P
2011	0.707	-1.906	0.057
2012	0.746	-1.683	0.092
2013	0.742	-1.745	0.081
2014	0.669	-2.344	0.019
2015	0.542	-3.063	0.002
2016	0.689	-2.193	0.028
2017	0.644	-2.385	0.017
2018	0.633	-2.388	0.017
2019	0.676	-2.061	0.039
2020	0.576	-2.725	0.006

According to the test results in Table 4, the global MoranI index values of regional innovation capacity in 2011-2020 for all 30 regions are between 0 and 1, and the GreayC index is less than 1, indicating that China's innovation capacity in general shows a positive spatial autocorrelation characteristic. This positive spatial correlation indicates that high innovation regions tend to be adjacent to other high innovation regions, while low innovation regions tend to be adjacent to other low innovation regions, which validates the concentration of regions close in color to each other in the quadratic map of the spatial distribution of regional innovation capabilities. This is due to the fact that regions with comparable innovation capabilities have certain similarities in terms of human resources, financial flows, and policies. As a result, there are frequent interactions between these regions. In terms of temporal trends, the MoranI index values showed a fluctuating upward trend overall and the GreayC index showed a fluctuating downward trend, and both indices passed the 10% significance level test. It indicates that regional innovation capacity has a significant homophilic influence on its neighboring regions, and the trend of regional innovation capacity clustering is gradually strengthening over time.

The above analysis effectively estimates the effect of demographic structure on regional innovation capacity. The empirical results show that the demographic structure of higher education has the most significant contribution to regional innovation capacity, followed by the demographic structure of urban areas, population density, and the demographic structure of secondary industries. At the same time, the interaction effect of innovation capacity between each region and neighboring regions is obvious.

5.CONCLUSION

Since China entered the new century, the regional innovation capacity has been substantially improved and the innovation and entrepreneurship environment has been optimized, which has effectively promoted the optimization and transformation of China's economic structure. At the same time, the demographic structure also changes with the economic situation, and presents a new period different from the previous characteristics. Technological innovation is an important force for industrial upgrading in the future economic development of China, and an important support for long-term sustainable development. In the context of the great changes that are taking place in all aspects of society in the new era, the exploration of the factors influencing technological innovation is of great practical importance. This paper analyzes the correlation of demographic structure on regional innovation capacity by combining big data technology. From the results of the study, it is clear that the demographic structure has a large impact on regional innovation capacity, so innovation capacity enhancement can be targeted for implementation.

ACKNOWLEDGE:

The paper is a funded project of 2023 Henan Soft Science(Henan Provincial Department of Science and Technology),results of “Study on the adaptability of Higher Vocational Education and High-Quality Economic Development in Henan Province”.

REFERENCES

[1] Akhmetshin E M, Kolpak E P, Sulimova E A, et al. Clustering as a criterion for the success of modern industrial enterprises[J]. International Journal of Applied Business and Economic Research, 2017, 15(23): 221-231.

[2] Belso-Martinez J A, Diez-Vial I, Lopez-Sanchez M J, et al. The brokerage role of supporting organizations inside clusters: how does it work?[J]. European Planning Studies, 2018, 26(4): 706-725.

[3] Chain C P, Santos A C, Castro L G, et al. Bibliometric analysis of the quantitative methods applied to the measurement of industrial clusters[J]. Journal of Economic Surveys, 2019, 33(1): 60-84.

[4] Chandrashekar D, Bala Subrahmanya M H. Exploring the factors of cluster linkages that influence innovation performance of firms in a cluster[J]. Economics of Innovation and new Technology, 2019, 28(1): 1-22.

[5] EL WAATMANI A. Industrial clusters and promotion of CSR: the case of developing countries[J]. Journal of Academic Finance, 2018, 9(2): 69-80.

[6] Götz M, Jankowska B. Clusters and Industry 4.0–do they fit together?[J]. European Planning Studies, 2017, 25(9): 1633-1653.

[7] Guowu L, Bai G A O. Globalization and Domestic Coping Strategies: The Development of China’s Industrial Clusters[J]. İstanbul University Journal of Sociology, 2020, 40(2): 625-647.

[8] Hou B, Hong J, Wang H, et al. Academia-industry collaboration, government funding and innovation efficiency in Chinese industrial enterprises[J]. Technology Analysis & Strategic Management, 2019, 31(6): 692-706.

[9] Isaksen A, Trippl M. Innovation in space: the mosaic of regional innovation patterns[J]. Oxford Review of Economic Policy, 2017, 33(1): 122-140.

[10] Jones C, Pimdee P. Innovative ideas: Thailand 4.0 and the fourth industrial revolution[J]. Asian International Journal of Social Sciences, 2017, 17(1): 4-35.
[11] Kuksa I, Shtuler I, Orlova-Kurilova O, et al. Innovation cluster as a mechanism for ensuring the enterprises interaction in the innovation sphere[J]. Management Theory and Studies for Rural Business and Infrastructure Development, 2019, 41(4): 487-500.
[12] Lazzeretti L, Capone F, Caloffi A, et al. Rethinking clusters. Towards a new research agenda for cluster research[J]. European Planning Studies, 2019, 27(10): 1879-1903.
[13] Liu Z, Chen X, Chu J, et al. Industrial development environment and innovation efficiency of high-tech industry: Analysis based on the framework of innovation systems[J]. Technology Analysis & Strategic Management, 2018, 30(4): 434-446.
[14] Moradi M, Zandi Paak R. The Effect of Companies' Participation or Non-Participation in the Creation of Industrial Clusters on Their Innovation Performance with an Emphasis on the Mediatory Role of Knowledge Management[J]. Innovation Management Journal, 2023, 5(1): 1-28.
[15] Novotná J, Novotný L. Industrial clusters in a post-socialist country: The case of the wine industry in Slovakia[J]. Moravian Geographical Reports, 2019, 27(2): 62-78.
[16] Schot J, Steinmueller W E. Three frames for innovation policy: R&D, systems of innovation and transformative change[J]. Research Policy, 2018, 47(9): 1554-1567.
[17] Sosnovskikh S. Industrial clusters in Russia: The development of special economic zones and industrial parks[J]. Russian Journal of Economics, 2017, 3(2): 174-199.
[18] Sozinova A A, Okhrimenko O I, Goloshchapova L V, et al. Industrial and innovation clusters: development in Russia[J]. International Journal of Applied Business and Economic Research, 2017, 15(11): 111-118.
[19] Stavroulakis P J, Papadimitriou S. Situation analysis forecasting: the case of European maritime clusters[J]. Maritime Policy & Management, 2017, 44(6): 779-789.
[20] Tödtling F, Trippl M. Regional innovation policies for new path development–beyond neo-liberal and traditional systemic views[J]. European Planning Studies, 2018, 26(9): 1779-1795.
[21] Turkina E, Van Assche A. Global connectedness and local innovation in industrial clusters[J]. Journal of International Business Studies, 2018, 49(6): 706-728.
[22] Ugalde Hernández O. Local industrial agglomerations vis-à-vis global competitive networks: marshallian notions of clusters, innovation and territorial development[J]. Economía y Sociedad, 2020, 25(57): 42-62.
[23] Žižka M, Valentová V, Pelloneová N, et al. The effect of clusters on the innovation performance of enterprises: traditional vs new industries[J]. Entrepreneurship and Sustainability Issues, 2018, 5(4): 780-794.