Machine Learning-Based Evaluation of Mass Entrepreneurship and Innovation in Chinese Cities: Postprint

Authors: Shi Mengyi, Shen Yuncong, Li Jiaojiao, Zhang Siwei, Xu Mengyu

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] Based on open data related to scientific and technological innovation, an evaluation system and index are established for the "mass entrepreneurship and innovation" level of cities nationwide. [Method / Process] Employing machine learning algorithms, a "mass entrepreneurship and innovation" ranking for domestic cities is constructed, and the strengths and weaknesses of urban "mass entrepreneurship and innovation" capabilities are analyzed through various indicators and urban agglomeration characteristics. [Result / Conclusion] The "mass entrepreneurship and innovation" level of cities in China is closely related to economic resources, talent reserves, and policy environment. Beijing continues to lead the development of "mass entrepreneurship and innovation" in China, with various regional central cities following closely behind, each with its own advantages and small development gaps. The deficiencies and shortcomings of each city are also very clear; how to formulate reasonable countermeasures for their respective deficiencies and leverage strengths to compensate for weaknesses is the fundamental approach to continuously developing urban innovation capacity.

Full Text

Evaluation System on Public Entrepreneurship and Innovation Based on Machine Learning

Shi Mengyi[1], Shen Yuncong[2], Li Jiaojiao[1], Zhang Siwei[1], Xu Mengyu[1]

[1]Document and Information Center, Fudan University, Shanghai 200433
[2]Department of Mathematics, Fudan University, Shanghai 200433

Abstract

[**Purpose/Significance**] This study establishes an evaluation system and index for the "public entrepreneurship and innovation" (shuangchuang) level of Chinese cities based on open data related to scientific and technological innovation. [**Method/Process**] Machine learning algorithms are employed to develop a national city ranking for shuangchuang performance, analyzing the strengths and weaknesses of urban shuangchuang capabilities through various indicators and agglomeration characteristics. [**Result/Conclusion**] The findings reveal that China's urban shuangchuang level is closely associated with economic resources, talent reserves, and policy environments. Beijing continues to lead China's shuangchuang development, with various regional central cities following closely behind—each possessing distinct advantages and exhibiting small development gaps. The deficiencies and shortcomings of each city are also clearly identifiable; addressing these limitations through targeted measures to leverage strengths and compensate for weaknesses constitutes the fundamental basis for sustaining urban innovation capacity.

Keywords: shuangchuang level; public entrepreneurship and innovation capability; evaluation system; shuangchuang index
Classification Number: F272.2

"Mass entrepreneurship and innovation" has been regarded as one of the "dual engines" driving China's economic development under the new normal. This initiative aims to transform economic development patterns by promoting innovation through entrepreneurship and vice versa, thereby facilitating economic transition and upgrading. With the further development of information technology and diversification of information acquisition methods, entrepreneurship has become an accessible development possibility for ordinary citizens, offering potential solutions to employment challenges. Against this strategic backdrop of innovation-driven development, various regions have undertaken extensive practices to enhance their shuangchuang competitiveness. However, development varies significantly across regions, with distinct problems emerging in different areas. Consequently, establishing a rational and effective shuangchuang evaluation system is particularly crucial.

Through data analysis and interpretation of the shuangchuang environment across Chinese cities, this study proposes and constructs a "public entrepreneurship and innovation" evaluation index system. The objective is to guide Chinese cities in conducting shuangchuang activities, evaluate the rationality and scientific nature of these activities, build upon strengths, address weaknesses, promote sustainable innovation development, and enhance activity quality and effectiveness.

1 Development of Domestic and International Entrepreneurship and Innovation Evaluation Systems

Extensive research has been conducted both domestically and internationally on entrepreneurship and innovation evaluation systems, yielding relatively mature frameworks. Currently, such evaluation systems primarily focus on specific aspects of urban capabilities, with entrepreneurship and innovation evaluations developing as two independent dimensions. Entrepreneurship evaluation systems and innovation evaluation systems have evolved separately, each establishing complete and rigorous assessment frameworks.

1.1 Entrepreneurship Evaluation Systems

1.1.1 International Entrepreneurship Evaluation Systems Internationally, three major entrepreneurship evaluation systems exist: the Global Entrepreneurship Monitor (GEM), the Global Entrepreneurship and Development Index (GEDI), and the Kauffman Index of Entrepreneurial Activity (KIEA) [1-4].

GEM is an international entrepreneurship research initiative launched by London Business School and Babson College in the United States. Through surveys and analysis of entrepreneurial environments, projects, and outcomes worldwide, it publishes the Global Entrepreneurship Monitor Report, which enjoys a prestigious global reputation. GEM's evaluation system examines how entrepreneurial environments influence entrepreneurial activities by decomposing environmental components and conducting in-depth analyses from different perspectives. It proposes a three-tier environmental factor theory, dividing the entrepreneurial environment into three major elements: basic requirements, efficiency enhancement, and innovation-driven entrepreneurship, encompassing nine aspects including financial support, government policies, government project support, education and training, R&D transfer, business and professional infrastructure, entry barriers, physical infrastructure, and cultural/social norms. This framework possesses considerable breadth and depth, making the GEM Report highly influential [1,5].

The Global Entrepreneurship and Development Index (GEDI) was developed to provide more comprehensive and multidimensional understanding of typical environmental characteristics that promote enterprise establishment and growth during regional development, thereby assessing the degree of efficient entrepreneurial support in a region. The GEDI indicator system comprises three sub-dimensions: entrepreneurial attitudes, entrepreneurial activities, and entrepreneurial aspirations. It includes 34 major institutional and individual factors that promote these three sub-dimensions within regional development systems, enabling prediction of regional entrepreneurial potential [2,5].

The Kauffman Index of Entrepreneurial Activity (KIEA) represents an early and important reference for entrepreneurial activity indices in the United States. It calculates the proportion of new enterprises among the adult non-business-owner population to determine entrepreneurial activity indices, with close attention paid to the first-month business activities of new ventures [3].

1.1.2 Domestic Entrepreneurship Evaluation Systems Domestically, more comprehensive entrepreneurship evaluation systems include the Zhongguancun Index, the Urban Entrepreneurship Environment Evaluation Index System, and the China Urban Entrepreneurship Index [6-8].

The Zhongguancun Index, compiled by the Beijing Municipal Bureau of Statistics, is a comprehensive index for observing the development of Beijing's high-tech industry. It includes five major categories: economic growth, economic efficiency, technological innovation, human capital, and enterprise development indices, representing a macro-level comprehensive evaluation system. The Urban Entrepreneurship Environment Evaluation Index System focuses on the quality of the entrepreneurial environment, assessing urban entrepreneurship environments from both hard and soft environmental indicators—specifically energy, postal and telecommunications, municipal infrastructure, institutional frameworks, management, and humanities. The China Urban Entrepreneurship Index similarly evaluates urban entrepreneurship environments from multiple perspectives.

1.2 Innovation Evaluation Systems

1.2.1 International Innovation Evaluation Systems Internationally, four major innovation evaluation systems currently exist: the Global Innovation Index (GII), the Summary Innovation Index (SII), the Joint Venture Silicon Valley Index (JVSV), and the World Knowledge Competition Index (WKCI) [5].

The Global Innovation Index evaluates institutions and policies, innovation drivers, knowledge creation, enterprise innovation, technology application, and intellectual property rights, providing enterprise leaders and government decision-makers with directions for improving comprehensive competitiveness by identifying gaps and efforts needed. It measures the broad economic innovation capacity of an economy.

After years of indicator adjustments, the EU Innovation Index quantitatively represents regional innovation status and analyzes inter-regional differences from several perspectives: human resources, financial support, enterprise investment, linkages and entrepreneurship, productivity, innovation, and economic effects. Its innovation index system primarily includes the European Innovation Scoreboard, Regional Innovation Scoreboard, and Innovation Barometer Survey.

The Silicon Valley Index focuses on measuring basic conditions for innovation from a macro perspective, including population, economic, social, spatial, and management indicators.

The World Knowledge Competition Index specifically targets the knowledge economy, with primary indicators including human capital, knowledge capital, financial capital, economic output, and knowledge sustainability, though it does not establish indicators for innovation output.

1.2.2 Domestic Innovation Evaluation Systems Domestic innovation evaluation systems concentrate primarily on regional scientific and technological development, drawing upon considerable international experience to establish a complete evaluation mechanism with "scientific and technological innovation capability factors" as the main evaluation criteria. These factors are typically assessed through regional S&T input and output indicators. Regarding S&T input and output, evaluation systems are generally established based on indicators such as basic environment, S&T input, S&T output, S&T achievement transformation, economic development, and innovation benefits.

The China Science and Technology Development Strategy Research Group [9] proposed evaluating regional innovation capacity from five aspects: knowledge creation, knowledge flow, enterprise technological innovation capability, innovation environment, and innovation economic performance, forming a relatively authoritative regional innovation capability evaluation index system.

1.3 Literature Review

International research on urban capability evaluation systems is more in-depth, with more mature indicator selection and establishment. Domestic innovation and entrepreneurship evaluation systems exhibit certain limitations, with content biased towards specific domains. Comprehensive analysis of various innovation and entrepreneurship evaluation systems reveals that indicator establishment primarily concentrates on macro-level dimensions such as policy environment, economic strength, input-output ratios, and talent development, with relatively comprehensive coverage. Therefore, this study synthesizes domestic and international innovation and entrepreneurship evaluation systems while incorporating the implementation guidelines of the State Council's "Mass Entrepreneurship and Innovation" initiative regarding entrepreneurship environment, vitality, and policies. The resulting shuangchuang evaluation system includes detailed indicators such as government investment and infrastructure construction for policy environment; GDP growth and corporate profit growth for economic strength; R&D expenditure ratios and patent output for input-output ratios; and numbers of higher education institutions and research organizations for talent development. Based on these considerations, this study explores shuangchuang capability evaluation methods to provide valuable references for regional shuangchuang assessment in China.

2 Data Sources and Processing

2.1 Data Sources

This study's data primarily originate from the National Bureau of Statistics, the China Economic and Social Development Statistical Database, the Wind Database, the Ministry of Education of the People's Republic of China, the People's Bank of China, and the Guoxin Hongshu Open Database published on Peking University's Open Research Data Platform [10-13].

The National Bureau of Statistics Database and the China Economic and Social Development Statistical Database represent China's most authoritative official statistical databases. From these sources, we collected macro-level measurements such as population and GDP, along with various data reflecting urban infrastructure development levels, supplemented by lists of universities and "Double First-Class" institutions published by the Ministry of Education and financial data released by the People's Bank of China.

The Wind Database is a first-class large-scale financial engineering and economic data warehouse centered on financial securities data. We selected indicators from listed companies' financial reports, including R&D expenditures, employee education levels, technician proportions, industry distribution, and profit and revenue levels, as data support reflecting China's shuangchuang resources in government, economic, and talent aspects.

The Guoxin Hongshu Open Database, published on Peking University's Open Research Data Platform, is a macroeconomic big data analysis database covering extensive data from development and reform, industry and commerce, taxation, and finance sectors. We utilized multiple datasets from the official data space of the First National University Data-Driven Innovation Research Competition, including "Point of Interest (POI) data," "Internet recruitment website data," and "invention patent data," to develop the shuangchuang level evaluation system.

2.2 Data Processing

Data processing includes data cleaning and preprocessing. Data cleaning involves deleting or correcting erroneous, incomplete, improperly formatted, or redundant data in datasets, primarily addressing missing and anomalous data. For missing data, we employed weighted average methods and interpolation fitting to fill null values. During subsequent statistical analysis of cleaned data, extreme values and discontinuities were frequently identified as anomalous data, which were then corrected or weighted according to actual conditions to eliminate their impact on results. Finally, data were normalized as indicator values for shuangchuang level assessment.

Using tools such as Excel and MATLAB, we processed and analyzed data from various sources and types. The "Map POI data" were collected from Gaode Map POI data in summer 2017. Internet recruitment website data serve as an important reference for urban talent levels, with recorded entries, position numbers, and education-level proportions representing key indicators. However, due to the non-authoritative nature of internet recruitment data and urban development disparities, extreme values frequently occur and require special attention. In contrast, invention patent data demonstrate relatively high reliability and simpler statistical processing, requiring only consideration of quality differences among invention, utility model, and design patents.

Additionally, we employed weighted average methods based on past observations of the same variable arranged chronologically, using time sequence numbers as weights to calculate weighted arithmetic means for predicting future variable trends. For example, based on provincial GDP data, we first calculated the proportion of each prefecture-level city's GDP within its province, then used this proportion as weights to calculate indicator values for other prefecture-level cities in each province. After organizing collected data by indicators, we obtained individual shuangchuang indicator values for each city. However, due to inconsistent data dimensions, uneven data quality, and occasional severe polarization, subsequent operations were hindered. Therefore, using MATLAB and referencing each column's maximum, median, and quartile values, we fitted power formula parameters a, b, c according to the formula: $d_i' = \frac{(d_i-b)^a}{(c-b)^a}$ to normalize data and improve smoothness and uniformity. Here, d_i represents original data, d_i' represents normalized data, and a, b, c are power formula model parameters.

Ultimately, we obtained shuangchuang indicator scores for Chinese cities, as shown in Table 2 .

3 Construction of the Shuangchuang Indicator System

3.1 Indicator System Establishment

This study examines cities and regions with GDP and population data recorded in the 2018 National Bureau of Statistics Database, including municipalities and autonomous prefectures. Synthesizing existing literature [14-16] and systematically considering policy, economic, talent, and infrastructure factors, we adopt four dimensions—subject, resources, environment, and performance—along with twelve factors including shuangchuang capability, shuangchuang confidence, government resources, economic resources, talent resources, infrastructure, economic level, environmental atmosphere, and patent output to establish the shuangchuang level evaluation indicator system, as shown in Table 1 .

Patent output represents shuangchuang output. We ultimately determined the indicator system comprising four levels (subject, resources, environment, performance) and twelve factors, generating 19 indicator scores. However, due to inconsistent data dimensions and uneven quality, we normalized the data as described previously.

3.2 Indicator Weight Consideration

We first used rankings and scores of key national cities published by various organizations as machine learning training sets, including Beijing (99.422), Shenzhen (93.296), Shanghai (89.416), Guangzhou (78.565), Suzhou (78.518), Wuhan (75.353), Hangzhou (74.991), Nanjing (74.893), Tianjin (74.319), Chengdu (70.966), Xi'an (67.097), Changsha (63.500), Wuxi (63.500), Chongqing (61.869), Qingdao (61.673), Hefei (60.241), Ningbo (58.683), Jinan (57.953), Xiamen (55.567), Dalian (54.477), Harbin (49.722), Shenyang (48.395), and Changchun (46.792). Using the indicator system constructed in the previous section as factors, we performed regression in MATLAB using both Support Vector Machine (SVM) and Least Squares methods. The SVM employed a linear kernel, where trained coefficients would represent weights for calculating scores and rankings for all cities. However, results deviated significantly from ideal conditions, with extremely large weight disparities among factors—in most cases, only one or two indicators contributed to the scoring system.

From the weight coefficients, social financing increments, R&D expenditure, and listed company R&D expenditure ratios, as manifestations of innovation and entrepreneurship economic resource levels, hold significant importance for enhancing urban shuangchuang levels. Listed company technician proportions and internet recruitment talent levels also provide talent support for urban innovation and entrepreneurship. National research institution quantities and map POI research institution proportions reflect urban shuangchuang capabilities and atmospheres, occupying important positions in the evaluation system. In contrast, environmental factors such as infrastructure and economic levels have relatively smaller impacts on urban shuangchuang capabilities. Among infrastructure indicators, urban road area, as a transportation representative, more positively influences shuangchuang capabilities compared to water supply and electricity consumption.

However, coefficients obtained through these methods are not necessarily positive. Since training samples consist of high-scoring cities, negative weights could cause abnormal ranking increases for out-of-sample cities with low scores on those indicators. Therefore, negative weights are unreasonable for this data.

Thus, we further refined the machine learning optimization objective. First, we constrained coefficients to be non-negative and established upper and lower bounds based on evaluation requirements and actual conditions. Second, we adjusted weights for each data point in the objective function, assigning greater weights to higher scores. This approach yielded significant improvements. We ultimately selected the following optimization model:

$$\min_{w} \sum_{i \in T} (x_i w - y_i)^2$$

where T represents the training sample set and y_i represents sample scores. This

quadratic optimization problem yields weight coefficient w, as shown in Table 3
.

4 Analysis and Results Discussion

4.1 Shuangchuang Index Score Analysis

From perspectives encompassing economy, politics, talent education, and social environment, and based on the data model established in the previous section, we comprehensively assessed the innovation and entrepreneurship capabilities of major Chinese cities. The final rankings and scores are presented in Table 4
.

From a regional distribution perspective, significant development disparities exist between eastern and western regions, with coastal areas generally outperforming inland regions. The Yangtze River Delta, Beijing-Tianjin-Hebei, and Pearl River Delta regions demonstrate clear regional agglomeration trends, with shuangchuang development exhibiting pronounced regionalization.

The comprehensive scores reveal that Beijing and Shanghai possess obvious advantages in shuangchuang capabilities, ranking in the first tier of shuangchuang cities. Shenzhen, Nanjing, and Chengdu follow closely behind with minimal score differences, constituting the second tier. The remaining cities form a third tier with small score gaps from the second tier, remaining highly competitive in shuangchuang strength.

From the radar chart of top 5 cities' indicator distributions (Figure 1 [Figure 1: see original paper]), government investment rankings align with overall rankings, indicating that government investment represents a crucial influencing factor. Scientific research innovation output also occupies an important position in shuangchuang capabilities.

Beijing exhibits outstanding advantages across all dimensions, with only minor shortcomings in "number of higher education institutions per 10,000 people" and "R&D proportion." Resource allocation emerges as Beijing's primary challenge. Shanghai, Shenzhen, Nanjing, and Chengdu show minimal gaps across indicators, each with distinct strengths. Shanghai's overall development resembles Beijing's, with balanced indicator performance but similar challenges in rational education resource allocation. Shenzhen shows clear disadvantages in higher education and research institution quantities but excels in R&D expenditure and patent output, demonstrating relatively complete innovation input and achievement transformation. Nanjing boasts abundant education resources, ranking first in higher education institutions per 10,000 people, but suffers from insufficient national research institutions and imperfect talent retention systems. Chengdu possesses numerous national research institutions, showing advantages in research output and proportions, with superior infrastructure compared to Shanghai, Shenzhen, and Nanjing, though its strengths are not prominent enough and overall performance remains relatively balanced.

(1) TOP 1: Beijing. As China's capital, Beijing's innovation and entrepreneurship capabilities are particularly prominent among Chinese cities, inseparable from its powerful political, economic, and cultural resources. Abundant high-level talent reserves and substantial shuangchuang human and financial investments constitute the primary factors behind Beijing's leadership. Beijing's government investment ranks first among Chinese cities at 46.3 billion RMB, providing a foundation for innovation and entrepreneurship activities. Simultaneously, Beijing's rich talent reserves, with numerous universities and research institutes, ensure a continuous talent supply for innovation and entrepreneurship. Based on abundant resources, excellent integration platforms have rapidly emerged, with 125 national-level maker spaces ranking first in China. Innovative incubators represented by Innovation Works, 36Kr, and Legend Star constitute the most dynamic components.

However, Beijing's innovation and entrepreneurship challenges are also evident. Conflicts between population and resources/environment are prominent, with resource allocation becoming a major constraint. Rising costs for innovation and entrepreneurship activities and severe environmental pollution have become obstacles. Government resource allocation tensions also hinder activity development.

(2) TOP 2: Shanghai. Shanghai ranks second in innovation and entrepreneurship capabilities among Chinese cities. As China's most important financial and commercial center, Shanghai has gradually established its economic advantages by leveraging the Yangtze River Delta urban agglomeration economic hinterland. In terms of policy, besides ranking second in government investment, Shanghai has introduced a series of supporting policies, including overseas talent introduction, household registration policies, international talent pilot zones, and professional title system reforms, continuously optimizing the innovation and entrepreneurship talent environment and establishing multiple shuangchuang demonstration bases.

Shanghai's maker spaces have flourished. By the first half of 2017, Shanghai had over 500 maker spaces, including 100 startup nurseries, 159 incubators, 14 accelerators, 250+ new-type incubators like maker spaces, and over 12,000 technology-based SMEs. In 2016, Shanghai held 9,800+ entrepreneurship guidance activities covering 380,000+ entrepreneurs, creating a low-cost, convenient, comprehensive, and open entrepreneurship network for entrepreneurs and enterprises [14].

Shanghai's shuangchuang development challenges resemble Beijing's: high resource costs, uneven distribution, and local exclusivity all pose challenges.

(3) TOP 3: Shenzhen. As China's first special economic zone established after reform and opening up, Shenzhen has demonstrated rapid development momentum in recent years, attracting worldwide attention. Shenzhen is a young city, and its shuangchuang capabilities reflect this youthful vitality. In terms of infrastructure construction and talent reserves, Shenzhen lacks advantages and appears constrained, with education and human resources representing major shortcomings. However, Shenzhen's patent output rivals Beijing's, ranking second and performing remarkably—primarily attributable to the rapid development of high-tech enterprises. Besides gathering technology-dominated high-tech companies, numerous leading domestic venture capital institutions have taken root and grown in Shenzhen. Simultaneously, Shenzhen has constructed a policy chain around shuangchuang needs, formulated measures to promote maker development, allocated multiple shuangchuang funds, and built a policy system covering different innovation entities and the entire innovation and entrepreneurship process.

In the long term, Shenzhen's scarce educational infrastructure resources require careful consideration. Insufficient talent reserves may lead to unsustainable urban shuangchuang development. Establishing more comprehensive talent introduction systems represents a crucial step for Shenzhen's sustainable shuangchuang environment development.

(4) TOP 4: Nanjing. Nanjing is a city with profound cultural heritage and economic strength. Leveraging its talent reserve advantages and Jiangsu Province's shuangchuang policies, Nanjing has established its own distinctive shuangchuang development path while forming a shuangchuang linkage with Wuxi and Suzhou to jointly promote shuangchuang environment construction and improvement.

Nanjing possesses considerable advantages in education resource allocation with abundant talent reserves. However, recruitment data reveal serious talent loss, indicating deficiencies in talent retention measures. Therefore, talent retention represents a key focus for Nanjing to achieve further breakthroughs in shuangchuang capabilities.

(5) TOP 5: Chengdu. Located in western China, Chengdu is a pearl among central and western urban development. As a "new first-tier city," Chengdu outperformed Tianjin, Hangzhou, Guangzhou, and others to secure fifth place in shuangchuang capability rankings, benefiting from local government support, talent and capital reserves, investment, and well-developed infrastructure.

Chengdu's shuangchuang development challenges are also clear. Its central-western geographical location requires greater efforts in talent attraction. Although Chengdu possesses considerable scientific research institution resources, achievement transformation effects are not obvious, and the scarcity of high-tech enterprises represents a major challenge for Chengdu's sustainable shuangchuang development.

Beyond these five cities, Tianjin, Zhengzhou, Wuhan, Hangzhou, Wuxi, Suzhou, and Guangzhou all possess strong shuangchuang capabilities. Tianjin boasts deep talent education reserves and substantial government investment support, with overall shuangchuang strength not inferior to others. Zhengzhou possesses good scientific research resources with guaranteed innovative research strength, but faces inevitable shortcomings in talent reserves and government investment. Wuhan, as another key shuangchuang development city in central and western China besides Chengdu, ranks at the forefront in talent reserves but shows fatigue in scientific research development. Hangzhou, leveraging rapid e-commerce development, has injected strong momentum into innovation and entrepreneurship activities, rapidly becoming an "entrepreneurship new city" with substantial government support and development vitality. Wuxi and Suzhou are Jiangsu's other two shuangchuang development leaders besides Nanjing. Suzhou outperforms Wuxi slightly in talent reserves, ranking first in Jiangsu for shuangchuang youth numbers (Wuxi ranks third). Suzhou demonstrates strong strength in patent innovation output. Wuxi shows clear advantages in scientific research reserves and emphasizes talent introduction, ranking second nationally in talent recruitment numbers. Wuxi demonstrates more stable performance across indicators with balanced development, while Suzhou excels in talent reserves and scientific innovation output but performs relatively weaker in other aspects. Guangzhou, as a Chinese mega-city, ranks at the national forefront in education resources and government investment, but underperforms in talent introduction, research expenditure, and GDP profit growth over the past three years. While talent reserves are abundant, talent loss has become a constraint on Guangzhou's sustainable shuangchuang capability development.

4.2 Spatial Agglomeration Characteristics of National Urban Shuangchuang

Beyond differences in cities' own economic, human, and environmental levels, spatial correlation characteristics of urban shuangchuang levels vary considerably. He Shunhui et al., in "Analysis of Spatiotemporal Pattern Evolution and Influencing Factors of Innovation Capacity in Chinese Prefecture-Level and Above Cities," found that adjacent cities demonstrate high consistency in innovation levels, exhibiting agglomeration phenomena in spatial distribution. To further reflect local spatial correlation characteristics, we categorize local agglomeration patterns into four types [15]: H-H agglomeration areas (high-efficiency type)—cities with high innovation capacity levels whose surrounding adjacent cities also exhibit high innovation levels, forming China's "high-efficiency" innovation regions; H-L agglomeration areas (polarization type)—cities with relatively high innovation capacity compared to surrounding cities but weak ability to drive surrounding areas' innovation capacity improvement, forming a "polarization" pattern of high center and low periphery; L-H agglomeration areas (hollow type)—cities with poor innovation levels, minimally influenced by surrounding cities, with innovation capacity significantly lower than surrounding cities, forming a "hollow" pattern of high periphery and low center; L-L agglomeration areas (low-efficiency type)—cities with relatively low innovation capacity nationally and small differences from surrounding cities, representing low-value agglomerations of innovation output.

Using MATLAB and geographic coordinate information to quantitatively analyze cities' shuangchuang indices and their surrounding cities', we can clearly observe these four city types' national distribution, as shown in Tables 5 through 8 .

H-H agglomeration areas (high-efficiency type) are mainly distributed in the Yangtze River Delta, Bohai Rim, and Pearl River Delta regions. At the same scale, Beijing and Tianjin in the Beijing-Tianjin-Hebei region exhibit high shuangchuang levels, yet surrounding cities like Langfang, Chengde, and Zhangjiakou have average shuangchuang capabilities, lowering regional innovation levels. In contrast, Shanghai, Nanjing, and Hangzhou in the Yangtze River Delta region exhibit high shuangchuang levels, while surrounding cities like Shaoxing, Wuxi, and Nantong also demonstrate relatively high shuangchuang levels, fully leveraging urban agglomeration advantages for efficient cooperation and resource exchange.

H-L agglomeration areas (polarization type) are mainly distributed in inland regions and provincial capitals in Northeast China. For example, central and western cities like Chengdu, Wuhan, and Chongqing, and northeastern provincial capitals like Changchun, Shenyang, and Harbin attract talent and capital from their entire provinces, demonstrating remarkable shuangchuang capabilities but weak driving effects on surrounding cities. These regions require strengthened regional exchange and cooperation to pursue mutual benefits.

L-H agglomeration areas (hollow type) are primarily located in eastern and central China, such as Langfang, Hengshui, and Zhoushan. These cities are surrounded by relatively high innovation levels but, constrained by geographical or economic conditions, cannot form effective cooperation with surrounding cities.

L-L agglomeration areas (low-efficiency type) are widely distributed across inland China, such as Leshan in Sichuan, Datong in Shanxi, and Kaifeng in Henan. These cities demonstrate overall low shuangchuang levels and should improve their economic, human, and environmental factors to enhance shuangchuang capabilities.

5 Conclusion

Based on the National Bureau of Statistics, China Economic and Social Development Statistical Database, Wind Database, Ministry of Education, People's Bank of China, and Guoxin Hongshu Open Database published on Peking University's Open Research Data Platform, this study constructed a shuangchuang level evaluation indicator system. Through application of internet recruitment big data and urban map POI big data, the evaluation indicators demonstrate significantly improved authenticity, timeliness, and richness compared to traditional indicators. With continuous database updates, the constructed indicators enable real-time tracking of urban shuangchuang levels, accelerating decision-making feedback speed.

Comprehensive analysis reveals that shuangchuang capability is closely related to government investment, talent reserves, innovation output, and policy environments. Government investment tends to favor China's mega-cities, with Beijing, Shanghai, and Guangzhou possessing richer resources and inherent advantages. However, various regional central cities spare no effort in improving policy environments and strengthening talent introduction to find breakthroughs from intellectual and scientific innovation resources.

Shuangchuang capability strength represents a comprehensive evaluation outcome, while sustainable shuangchuang development constitutes a dynamic process. Beijing continues leading China's shuangchuang development, with various regional central cities following closely—each with distinct advantages and small development gaps. Simultaneously, each city's deficiencies and shortcomings are clearly identifiable. Addressing these limitations through targeted measures to leverage strengths and compensate for weaknesses constitutes the fundamental basis for sustaining urban innovation capacity. We anticipate that cities will, while maintaining advantages, correctly recognize deficiencies, adopt timely corresponding measures, comprehensively improve innovation and entrepreneurship competitiveness, and create a more prosperous landscape of flourishing development.

References

[1] HAY M, COX L W, REYNOLDS P D, et al. Global entrepreneurship monitor (Gem)-2002 executive report[J]. Social science electronic publishing, 2002, 3(1):66–70..

[2] Acs, Z.J., Szerb, L., Autio, E. The global entrepreneurship and development index[J]. Regional studies, 2015, 49(12): 1977-1994..

[3] FAIRLIE R W. Kauffman index of entrepreneurial activity by veteran status 1996-2011[J]. Ssrn electronic journal, 2012(21): 28.

[4] CHEN X, LIU Z, MA C. Chinese innovation-driving factors: regional structure, innovation effect, and economic development—empirical research based on panel data[J]. The annals of regional science, 2017, 59(1): 45-65.

[5] Hu Ping. Discussion on the evaluation system and method of "mass entrepreneurship and innovation"[J]. Science and Management, 2017, 37(5): 14-21.

[6] Zhou Li, Wu Yuming. Regional innovation capability in China: factor analysis and cluster research—also on factor analysis as an alternative method for comprehensive evaluation of regional innovation capability[J]. China Soft Science, 2006(8): 96-103.

[7] He Jianwen. Analysis report on Guangdong's regional innovation capability—based on "China Regional Innovation Capability Report 2015"[J]. Guangdong Science and Technology, 2017, 26(2): 45-49.

[8] Bai Jia. Evaluation and comparison of regional technological innovation capability in China[J]. Scientific Management Research, 2012, 30(1): 15-18.

[9] China Science and Technology Development Strategy Research Group. 2003 China Regional Innovation Capability Report[M]. Beijing: Economic Management Press, 2004.

[10] National Bureau of Statistics of the People's Republic of China[EB/OL]. [2018-01-25]. http://www.stats.gov.cn/.

[11] China Economic and Social Development Statistical Database[EB/OL]. [2018-01-27]. http://tongji.cnki.net/kns55/index.aspx.

[12] Ministry of Education of the People's Republic of China[EB/OL]. [2018-01-27]. http://www.moe.gov.cn/.

[13] Guoxin Hongshu[EB/OL]. [2018-01-27]. http://www.ndrcbd.org/.

[14] Zang Ming. Shuangchuang promotes upgrading and expands new kinetic energy | Shanghai already has 7 national shuangchuang demonstration bases[EB/OL]. [2018-01-20]. http://www.thepaper.cn/newsDetail_{{{forward}}_{{1794996}}}.

[15] He Shunhui, Du Debin, Jiao Meiqi, et al. Analysis of spatiotemporal pattern evolution and influencing factors of innovation capacity in Chinese prefecture-level and above cities[J]. Scientia Geographica Sinica, 2017, 37(7): 1014-1022.

Author Contributions

Shi Mengyi: Data processing, results analysis, paper structure development
Shen Yuncong: In-depth data processing, model design
Li Jiaojiao: Data collection and preprocessing, paper writing, revision and improvement
Zhang Siwei: Literature review writing and paper writing
Xu Mengyu: Data processing, model design

Note: Figure translations are in progress. See original paper for figures.