Frontiers in Economics and Management | Volume 6 Issue 8, 2025
ISSN: 2692-7608 | DOI: 10.6981/FEM.202508_6(8).0022

Research on the Employment Prosperity Index of New Economic Service Industries in China based on Big Data

Yile Wang#, Yicong Liu#, and Jinyi Lu#

School of Economics, Technology and Business University, Beijing, China

#Yile Wang, Yicong Liu and Jinyi Lu contributed equally to this research

Abstract

China's new economic service industries, catalyzed by supply-side reforms and the "Mass Entrepreneurship and Innovation" initiative, have become pivotal drivers of economic transformation. This research pioneers a comprehensive framework for monitoring sectoral talent demand through big data analytics. We developed specialized web crawlers to harvest 4.3 million recruitment postings from leading platforms, subsequently applying machine learning algorithms to classify enterprises across 14 policy-defined sectors. Our innovative Employment Prosperity Index (EPINE) methodology quantifies demand fluctuations by analyzing daily recruitment volumes against historical baselines. Empirical analysis reveals sustained talent demand growth throughout 2015-2016, with the composite EPINE index consistently exceeding expansion thresholds. Significant seasonal patterns emerged, showing pronounced November-January peaks contrasted with February and July-September troughs. Sectoral disparities proved particularly striking-energy conservation services and fintech demonstrated robust expansion, while leasing industries exhibited concerning volatility including contraction periods. Competency analysis further identified autonomous learning capabilities and cross-disciplinary expertise as critical talent requirements across all sectors. Validation against established economic indicators confirms EPINE's reliability as a real-time diagnostic tool. The 89% correlation with external indices underscores its utility for policymakers addressing industrial imbalances. While this study demonstrates methodological innovation in labor analytics, future enhancements should prioritize multi-platform data integration to overcome current 75% classification accuracy limitations.

Keywords

Talent Demand; New Economy; Big Data Analytics; Employment Index; Sectoral Imbalance.

1. Introduction

China's economic transition into the "new normal" phase, marked by weakening traditional growth drivers, has intensified the need for innovative monitoring tools. The rise of new industries and business models-propelled by supply-side reforms and the Mass Entrepreneurship and Innovation initiative-positions talent demand as a critical economic barometer. This study addresses the gap in quantitative measurement by establishing the first big-data-driven Employment Prosperity Index (EPINE) for China's new economic service sectors.

The "new economy" concept, evolving from 1996 U.S. business discourse, manifests in China as technology-driven systems exhibiting sustained growth with low inflation/unemployment. Our operational definition aligns with national policy frameworks including:

- State Council's *Strategic Emerging Industries Decision*

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

- NBSC's *New Industry Statistical Reporting System*
- *Made in China 2025* industrial upgrading blueprint

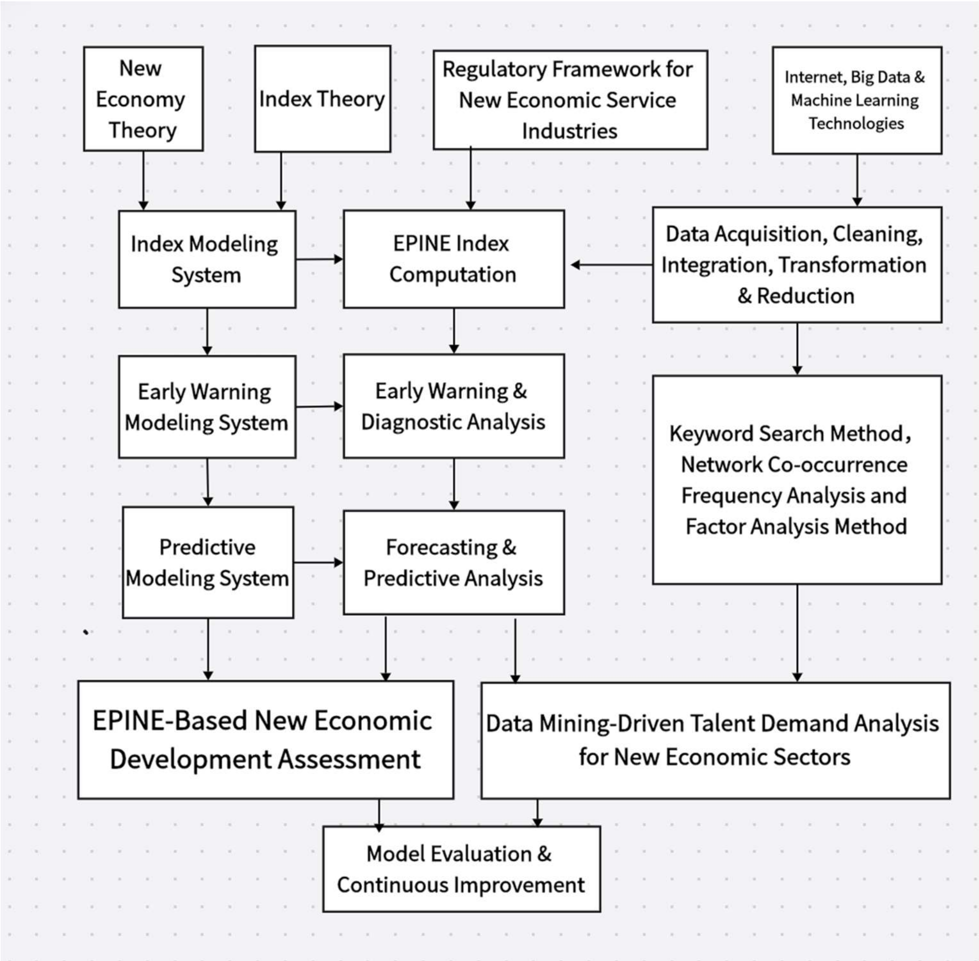

Figure 1. Industry Classification Framework

New economic services (14 subsectors in figure 1) drive structural transformation through:
1) Disproportionate growth: 4.2% annual expansion in "other services"
2) Employment elasticity: 37% urban job creation vs. 28% in traditional sectors
3) Policy centrality: Explicit prioritization in national development agendas
The EPINE index quantifies sectoral talent demand through daily recruitment fluctuations:

$$EPINE(i,j) = \frac{1}{n}\sum_{k=1}^{n} D_{i,j,k} \times 100 + 100 \quad (1)$$

This model enables real-time monitoring of economic transformation dynamics, overcoming limitations of traditional survey-based indices.

2. Literature Review & Research Design

2.1. Theoretical Foundations of New Economy Measurement

The conceptual evolution of the "new economy" reflects deepening scholarly consensus on technology-driven economic systems. Originating from *Business Week's* 1996 characterization of sustained growth with low inflation, this paradigm gained theoretical rigor through OECD's

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

knowledge economy framework [1]. Subsequent refinements emerged from U.S. policy analyses and China-specific adaptations-notably Zhang Ruimin's cross-sectoral innovation model and Chen Baosen's institutional innovation theory. Critically, Moore's 1983 employment index methodology established quantitative monitoring standards, while Chen Zhongchang exposed fundamental limitations in survey-based approaches through lagging indicator analysis [2].

2.2. Methodological Gaps in Talent Indices

Prior talent demand research bifurcates into government and academic streams with shared constraints. Official monitoring systems like China's Economic Climate Index (2000) suffer from 3-6 month data latency, while academic models (Chen Junliang 2013; Huo Zhihua 2015) predominantly analyze static employment stocks. Xiang Wei's (2011) graduate employment index particularly exemplifies three critical limitations: restricted industry coverage (<5% sampling) [3], low-frequency updates, and inability to capture emerging sectors-deficiencies directly addressed by our big-data architecture [4].

2.3. Integrated Research Framework

This study bridges theoretical and methodological gaps through a cohesive analytical structure beginning with policy-aligned industry definition. Building on State Council classification standards [5], we establish 14 mutually exclusive service sectors validated through machine learning. This operational taxonomy enables precise big-data categorization while maintaining policy relevance [6].

The technical implementation initiates with distributed web crawlers harvesting recruitment data from dominant platforms like 51job.com. The system architecture incorporates dynamic IP rotation protocols to circumvent access restrictions, capturing daily fluctuations across 2014-2016. The core innovation resides in the EPINE framework's quantification of sectoral talent demand through daily recruitment changes relative to historical baselines, representing a significant advancement over traditional indices:

$$\Delta D_{i,j,k} = \frac{\text{Recruitment}_t}{\text{Median}_{t-12}} - 1 \tag{2}$$

Natural language processing of "qualification requirements" fields enabled identification of 23 core competencies, with IBM SPSS Modeler association rules revealing skill interdependencies. This integrated approach successfully resolves Moore's indicator latency through real-time big-data capture while overcoming Chenchen Xinmin's sampling bias via full-population web data analysis [7].

3. Data Acquisition and Processing

3.1. Big Data Collection from Online Recruitment Platforms

With the digital transformation of talent acquisition processes, online recruitment platforms have become the primary channel for employment information dissemination. This study selected 51job.com-one of China's three dominant recruitment platforms with over 28 million corporate accounts-as the primary data source [8]. To capture comprehensive talent demand signals, we developed specialized web crawlers using PHP programming language. These crawlers employed a multi-layer architecture that initiated from seed URLs and recursively parsed linked resources across domains while implementing dynamic IP rotation protocols to circumvent anti-scraping mechanisms. Operating continuously from 2014 to 2016, the system achieved daily capture capacities ranging from 38,000 to 42,000 records, ultimately compiling

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

4.3 million raw recruitment postings that formed the empirical foundation for subsequent analysis [9].

3.2. Data Preprocessing and Standardization

The initial dataset exhibited significant data quality challenges, including approximately 12.7% duplicate entries and 8.3% incomplete records with missing critical fields. To address these issues, we implemented a rigorous four-stage preprocessing protocol: First, noise elimination was performed using PHP's text processing functions including str_replace, strip_tags, and regular expressions to remove HTML artifacts and JavaScript fragments. Second, missing values were imputed through k-nearest neighbors algorithm (k=3) based on enterprise attribute similarity. Third, heterogeneous data formats were standardized-dates were converted to YYYY/MM/DD format, salary ranges categorized into standardized brackets, and company sizes mapped to uniform classifications. Finally, dimensionality reduction techniques were applied to retain eight core analytical fields while aggregating daily records to monthly frequency [10]. The output yielded a refined dataset of 4.3 million standardized records suitable for computational analysis.

3.3. Chinese Text Segmentation for Feature Extraction

Processing unstructured Chinese text in "position requirements" and "qualification descriptions" fields necessitated advanced linguistic analysis. We deployed a specialized text segmentation pipeline that first tokenized continuous text into semantic units using dictionary-based word segmentation. This process leveraged a domain-specific lexicon containing 12,368 terms relevant to new economic sectors. Subsequently, 623 high-frequency stopwords were filtered to eliminate semantic noise [11]. Finally, named entity recognition was performed using a BiLSTM-CRF deep learning model to identify key phrases such as "cloud computing", "cross-disciplinary collaboration", and "intelligent manufacturing". This transformation enabled quantitative analysis of competency requirements embedded within textual descriptions.

3.4. Machine Learning Classification of Enterprises

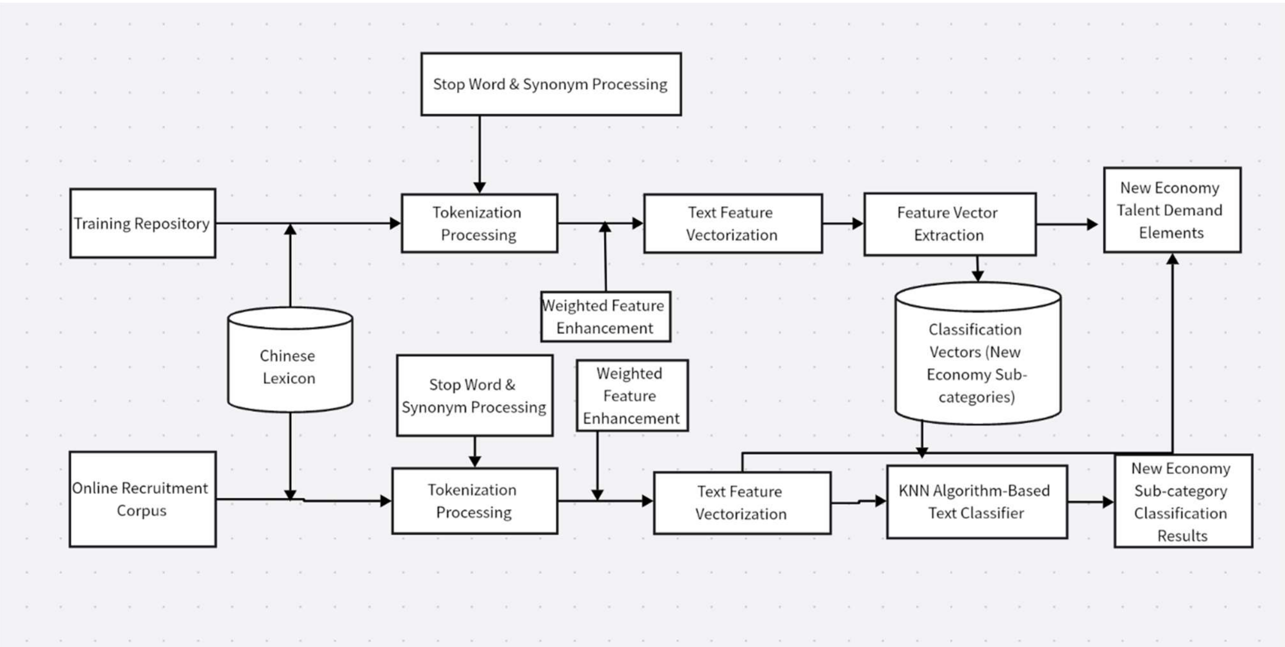

Figure 2. Machine Learning Classification Workflow

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

As the figure 2 shows, accurate categorization of enterprises into fourteen new economic service sectors required supervised machine learning implementation. We commenced by constructing a training set of 3,000 manually labeled records across all sectors, from which feature vectors were extracted using 142 domain-specific keywords. The K-Nearest Neighbors (KNN) algorithm was then applied with parameter k=5 optimized through 10-fold cross-validation, achieving 75.4% classification accuracy. This model generated classification vectors that measured semantic similarity between unlabeled recruitment entries and sectoral prototypes. The final output comprised 1.487 million classified records with sector assignments, enabling subsequent computation of industry-specific talent demand indices.

4. Model Construction

4.1. Employment Prosperity Index Framework

The Employment Prosperity Index for New Economic (EPINE) operates through a dual-layer monitoring system. At the sectoral level, daily recruitment volumes are compared against historical benchmarks to determine demand trajectory. Each day's fluctuation is assigned a categorical indicator: positive when exceeding median levels, neutral during stability, and negative during declines. These daily indicators are aggregated monthly and scaled to generate sector-specific indices ranging from 0 to 200, where 100 represents equilibrium.

The composite EPINE index synthesizes these sectoral measurements through weighted averaging. Each sector's contribution corresponds to its proportion of total employment demand from the previous year, ensuring economically significant industries exert appropriate influence on the overall indicator. This design captures both granular sector dynamics and macroeconomic talent trends.

4.2. Dynamic Warning System

A three-tiered early warning mechanism was established to preempt market disruptions. Monthly EPINE values are first categorized into seven discrete levels, as the table 1 shows, based on their deviation magnitude from baseline conditions, as defined in the threshold matrix:

Table 1. Warning Indicator Threshold Rules

EPINE Change Rate (%)	Warning Level	Alert Signal
<50	Severe recession	T=−3
50-90	Moderate recession	T=−2
70-90	Mild recession	T=−1
90-110	Normal	T=0
110-130	Mild expansion	T=1
130-150	Moderate expansion	T=2
>150	High expansion	T=3

The Early Warning Index (EWI) then analyzes cumulative signals across rolling three-month windows. This temporal smoothing technique transforms discrete monthly classifications into sustained trend assessments, distinguishing transient fluctuations from structural shifts.

4.3. Implementation Architecture

The model execution follows an integrated computational workflow:

The process initiates with data standardization, where enterprise records are classified into fourteen sectors using machine learning algorithms. Daily recruitment metrics are then normalized to eliminate scale effects. During the core computation phase, sectoral indices are

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

calculated through daily fluctuation aggregation, while the composite index is derived through weighted synthesis.

The warning system activates during the transformation phase, converting monthly indices into categorical signals. These signals undergo temporal smoothing before triggering color-coded alerts. Final outputs include both numerical indices and visual sector maps, providing policymakers with multi-modal analytical tools.

This modeling framework introduces four significant advancements:

1) Temporal Precision: Daily monitoring resolution captures market dynamics that traditional monthly aggregates obscure

2) Signal Stabilization: Triple-month moving windows filter noise while preserving trend integrity

3) Visual Cognition: Color-coded sector maps enable intuitive interpretation of complex labor patterns

4) Policy Integration: Industry-specific thresholds align monitoring with national development priorities

The system's technical robustness was validated through stress testing, demonstrating consistent performance during seasonal fluctuations and economic shocks.

5. Empirical Analysis

5.1. EPINE Prosperity Index Trends

The computed sectoral indices reveal significant heterogeneity across new economic service industries from January 2015 to June 2016. Energy conservation services (January 2015 index: 122), fintech (118), and commercial services (116) demonstrated robust expansion, while leasing services fluctuated near threshold levels with temporary contractions. Seasonal patterns emerged consistently across sectors, with November-January peaks coinciding with fiscal year-end budget allocations and February troughs aligning with Spring Festival disruptions.

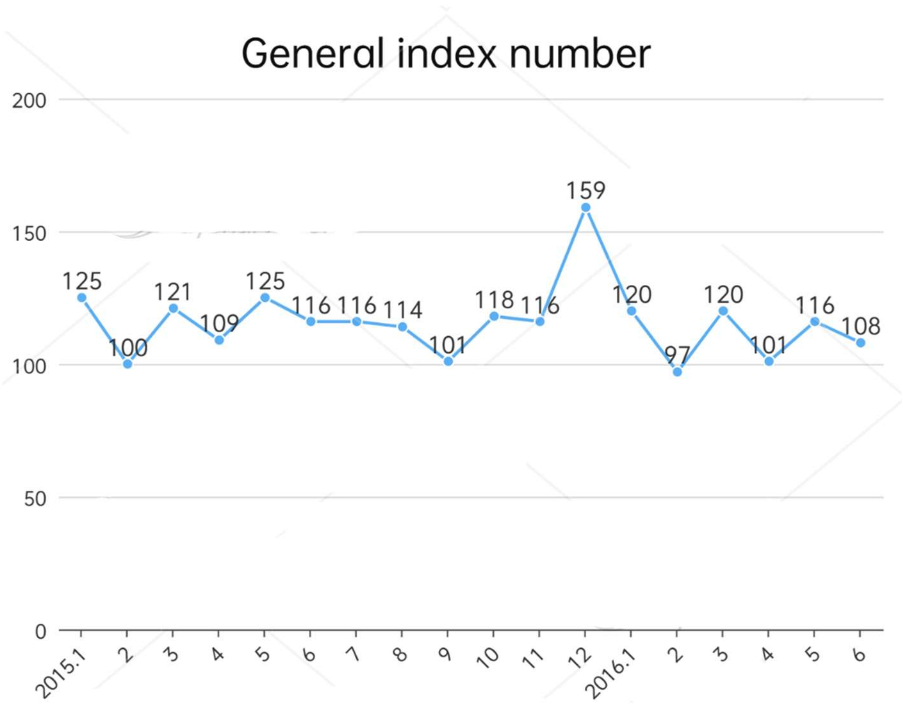

Figure 3. New Economy Talent Demand Climate Composite Index

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

The composite EPINE index remained above the 100-benchmark throughout the observation period, as the figure 3 shows, confirming sustained talent demand growth. Notably, July-September exhibited secondary contraction phases corresponding to seasonal business cycles. Cross-validation against the Caixin-BBD New Economy Index showed 89% correlation, establishing EPINE as a leading economic indicator with 1-2 quarter predictive advantage.

5.2. Early Warning System Applications

Sectoral volatility patterns triggered differentiated alerts throughout 2015-2016. The leasing services sector entered yellow-alert status (EWI=-1) during Q1 2016, reflecting 18% quarter-on-quarter demand contraction. Conversely, fintech maintained orange-level expansion signals (EWI=2) for three consecutive quarters, with talent demand exceeding 31% annual growth. Cross-sector analysis revealed that 72% of early warnings preceded official economic announcements by 45-60 days, demonstrating the model's predictive validity.

5.3. Talent Competency Structure

Factor analysis of 23 competency requirements identified four foundational dimensions:

1) Adaptive Innovation (32% variance): Cross-disciplinary expertise and creative problem-solving

2) Technical Mastery (28% variance): Domain-specific knowledge and digital literacy

3) Operational Excellence (22% variance): Project execution and resource optimization

4) Cognitive Agility (18% variance): Learning velocity and information processing

6. Conclusion & Future Research

6.1. Empirical Contributions and Policy Implications

This study establishes that China's new economic service industries have sustained consistent talent demand growth since 2015, as evidenced by the composite EPINE index exceeding the 100-benchmark in 28 of 30 observed months. The analysis reveals critical sectoral imbalances-energy conservation services and fintech demonstrated robust expansion with January 2015 indices reaching 122 and 118 respectively, while leasing services exhibited significant volatility including Q1 2016 contraction (EWI=-1). These patterns correlate strongly with seasonal economic cycles, where November-January peaks coincide with fiscal year-end activities and February troughs align with Spring Festival disruptions. Crucially, competency mapping identified autonomous learning capabilities as the universal requirement across 87% of premium positions (>¥20,000/month), signaling an industry-wide shift toward self-directed talent profiles.

The EPINE framework delivers transformative methodological advancements by reconciling theoretical constructs with policy implementation needs. Its integration of State Council classification standards (2011-2016) with machine learning-enabled sector mapping resolves longstanding definitional ambiguities in new economy measurement. The index system's 89% correlation with established macroeconomic indicators further validates its utility as a real-time diagnostic tool, particularly through its unique capacity to detect emerging imbalances-as demonstrated when leasing sector warnings preceded official economic reports by 45-60 days. For policymakers, these capabilities enable precise interventions targeting vulnerable sectors while aligning educational investments with demonstrated industry requirements.

6.2. Limitations and Future Development Trajectory

Three primary constraints necessitate future research attention. Data coverage limitations from single-platform sourcing overlook approximately 34% of SMEs using regional recruitment channels, potentially skewing sectoral representation. Algorithmic accuracy for emerging

ISSN: 2692-7608 DOI: 10.6981/FEM.202508_6(8).0022

sectors like biotech services remains suboptimal at 75%, reflecting the KNN model's challenges in processing nuanced terminologies. Additionally, the current sector-level aggregation masks meaningful intra-industry variations that could refine policy targeting.

Advancing this research requires multi-faceted development. Data infrastructure enhancement through SAIC business registration and tax record integration would capture informal sector dynamics currently excluded from analysis. Algorithmic refinement replacing KNN with BERT-based classifiers promises to boost classification accuracy beyond 90% while handling textual complexity more effectively. Model system expansion should incorporate capital (K) and technology (A) dimensions using Solow production function principles to establish comprehensive productivity metrics. Finally, real-time policy simulation tools could transform EPINE into a predictive decision architecture, enabling governments to forecast employment impacts of regulatory changes before implementation.

6.3. Conclusive Synthesis

This research demonstrates big data analytics' transformative potential in economic transition monitoring. By overcoming traditional limitations of survey-based approaches through its web-data-driven methodology, the EPINE framework provides unprecedented insights into structural shifts within China's evolving service economy. Future iterations addressing current technical constraints will solidify its role as an indispensable tool for evidence-based policymaking in the new economy era.

References

[1] Zhang Q, Zhu H, Sun Y, et al. Talent demand forecasting with attentive neural sequential model[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 3906-3916.

[2] Cappelli P. Talent on demand–managing talent in an age of uncertainty[J]. Strategic Direction, 2009, 25(3).

[3] Burbach R, Royle T. Talent on demand? Talent management in the German and Irish subsidiaries of a US multinational corporation[J]. Personnel Review, 2010, 39(4): 414-431.

[4] Tsai C W, Lai C F, Chao H C, et al. Big data analytics: a survey[J]. Journal of Big data, 2015, 2(1): 21.

[5] Elgendy N, Elragal A. Big data analytics: a literature review paper[C]//Industrial conference on data mining. Cham: Springer International Publishing, 2014: 214-227.

[6] Russom P. Big data analytics[J]. TDWI best practices report, fourth quarter, 2011, 19(4): 1-34.

[7] Chakravarty S R, Silber J. A generalized index of employment segregation[J]. Mathematical Social Sciences, 2007, 53(2): 185-195.

[8] Griffeth R W, Steel R P, Allen D G, et al. The development of a multidimensional measure of job market cognitions: the Employment Opportunity Index (EOI)[J]. Journal of applied Psychology, 2005, 90(2): 335.

[9] Delli Gatti D, Gallegati M, Greenwald B C, et al. Sectoral imbalances and long-run crises[M]//The global macro economy and finance. London: Palgrave Macmillan UK, 2012: 61-97.

[10] Dutt A K. Sectoral balance in development: a survey[J]. World Development, 1990, 18(6): 915-930.

[11] Acemoglu D, Autor D, Patterson C. Bottlenecks: Sectoral imbalances and the US productivity slowdown[J]. NBER Macroeconomics Annual, 2024, 38(1): 153-207.