Emerging Technology Prediction from an Outlier Patent Perspective: BERT Model and Deep Neural Network Post-Processing

Authors: Kong Dejing, Dong Fang, Chen Zijing, Liu Yuhan, Zhou Yuan

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] Due to the forward-looking nature of emerging technologies, they often fail to attract considerable attention when first emerging. Current research predominantly identifies emerging technologies by following technological path dependence, thereby overlooking certain R&D efforts that disrupt existing technological trajectories. Analyzing outlier patents with low similarity to mainstream technologies within a field can more effectively identify such R&D activities and predict emerging technologies. [Method/Process] This study proposes a method for outlier patent identification and emerging technology prediction based on deep learning. First, a BERT pre-trained model is employed to construct a similarity network based on patent texts to identify outlier patents. Then, a DNN model is utilized to establish the relationship between outlier patent indicators and technological influence, enabling rapid and accurate prediction of emerging technologies from massive outlier patents. Finally, an empirical analysis is conducted using the CNC system field as an example, retrieving all patents in this field from the Derwent patent database over the past decade. [Results/Conclusion] The empirical analysis results in the CNC system field validate the effectiveness of the proposed model, while also providing important guidance for national technology development policy formulation and enterprise technology layout in related fields.

Full Text

Forecasting Emerging Technologies from the Perspective of Outlier Patents: Based on BERT Model and Deep Neural Networks

Kong Dejing[1], **Dong Fang**[2], **Chen Zijing**[3], **Liu Yuhan**[3], **Zhou Yuan**[2]
[1]School of Modern Post, Beijing University of Posts and Telecommunications,

Beijing 100876 [2]School of Public Policy and Management, Tsinghua University, Beijing 100084 [3]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074

Abstract

[**Purpose/Significance**] Due to their advanced nature, emerging technologies often receive limited attention in their initial stages. Current research predominantly identifies emerging technologies by following technological development path dependencies, which may overlook disruptive R&D efforts that deviate from existing technological trajectories. Analyzing outlier patents—those with low similarity to mainstream technologies within a field—can more effectively identify such R&D activities and forecast emerging technologies.

[**Method/Process**] This paper proposes a deep learning-based method for outlier patent identification and emerging technology forecasting. First, the BERT pre-trained model is employed to construct a similarity network based on patent texts to identify outlier patents. Then, a Deep Neural Network (DNN) model establishes the relationship between outlier patent indicators and technological influence, enabling rapid and accurate prediction of emerging technologies from massive outlier patent datasets. Finally, the method is empirically validated using all patents from the numerical control (NC) system field over the past decade obtained from the Derwent Innovation database.

[**Result/Conclusion**] The empirical results in the NC system domain confirm the model' s effectiveness, offering significant guidance for national technology policy formulation and corporate technology layout in related fields.

Keywords: emerging technologies; deep learning; outlier patents; numerical control system

---

1. Introduction

Emerging technologies represent a new category of "breakthrough" technologies that serve as critical foundations for strategic emerging industries. These technologies may disrupt existing technological systems and paradigms, endowing products, processes, or services with unprecedented performance or substantially improved performance at reduced costs. Emerging technologies exert decisive influence on market rules, competitive landscapes, and industrial boundaries, potentially triggering industrial reshuffling. Consequently, forecasting emerging technologies holds significant importance for technology planning and strategic formulation at national and corporate levels.

Emerging technologies develop rapidly in the short term, exhibit high uncertainty, and possess substantial potential to drive technological progress and societal impact. Most existing research identifies emerging technologies based on technological path dependency, focusing on hotspots and frontiers within mainstream technological trajectories. However, due to their advanced nature, emerging technologies often diverge from established mainstream paths and require extended periods for commercialization, despite their future substantial contributions to industry development. For instance, a decoupled combustion technology patent awarded to the Institute of Process Engineering, Chinese Academy of Sciences in 1998—which reduced exhaust emissions—was not widely recognized at the time due to limited awareness of nitrogen oxide hazards, resulting in large-scale application only in 2017. Similarly, Canon' s 1982 patent application for liquid jet recording heads for printers preceded the mainstream inkjet printer market, with significant citations emerging only after 1990 from major corporations like HP, Xerox, and Google. Path-dependent methods struggle to identify such trajectory-disrupting innovations in a timely manner.

The most fundamental characteristic of emerging technologies in their early stages is radical innovation, which typically manifests as low similarity and weak association with mainstream technological paradigms, creating an "outlier" status. Analyzing emerging technologies from an outlier perspective more accurately captures their formative characteristics.

Current research on emerging technology identification from an outlier perspective remains limited, with existing studies primarily using patent data to identify outlier patents for technology forecasting. While valuable, these approaches suffer from limitations: (1) they calculate patent similarity using citation coupling methods, lacking semantic understanding of patent texts and resulting in inaccurate similarity measures; (2) prediction methods based on patent indicators and expert judgment are costly and time-consuming. Data-driven deep learning methods can ensure prediction effectiveness while substantially improving efficiency and reducing costs, enabling rapid and accurate identification of emerging technologies from massive patent datasets.

This study constructs an emerging technology forecasting model from the outlier perspective using word vectors and Deep Neural Networks (DNN). The BERT pre-trained model first vectorizes patent texts to construct a semantic similarity network, identifying outlier patents as candidate emerging technologies. The DNN model then learns the relationship between outlier patent indicators and future technological influence. Finally, this relationship model predicts the future influence of current outlier patents, uncovering overlooked patents with potential for significant future impact. The NC system field serves as an empirical case to validate the method' s effectiveness.

---

2. Related Research

2.1 Emerging Technology Forecasting Methods Traditional emerging technology forecasting relies primarily on expert knowledge, such as the Delphi method and Analytic Hierarchy Process (AHP), which have generated numerous patent indicators for characterizing emerging technologies. Some indicators remain stable over time, including IPC count, inventor count, and non-patent literature citations, while others evolve, such as forward citations and patent modification frequency. However, these methods alone cannot predict complex technological growth and application expansion.

Recent advances in computing power have enabled data-driven approaches, with machine learning and deep learning methods for emerging technology forecasting gaining widespread attention. Patent-based emerging technology identification research defines "technology" in two ways: (1) from a technological perspective, defining a technology as all patents belonging to the same IPC class or clustered together; (2) from a patent perspective, treating each patent as a theoretical focal point to identify high-impact patents for emerging technology discovery. This study adopts the second definition, representing each patent as a technological research focus, extracting technical features through bibliometric methods, and using deep learning to predict its likelihood of becoming an emerging technology.

2.2 Outlier Patent Identification A crucial characteristic of emerging technologies is radical innovation, implying strong heterogeneity with existing technologies. Scholars have emphasized the importance of focusing on "outlier patents" during identification, as these may trigger paradigm shifts. Outlier patents are more likely to develop into emerging technologies, and their exclusion from patent analysis causes significant information loss.

Current outlier patent identification relies on two methods: citation coupling and text similarity. Citation coupling assumes patents sharing more co-citation relationships are more similar, while text similarity methods calculate semantic vectors based on SAO (Subject-Action-Object) structures. However, SAO suffers from low computational efficiency and semantic confusion, limiting large-scale application. Recent studies demonstrate that deep learning methods like BERT significantly improve semantic representation performance. This study employs the BERT model to extract patent text information and identify outlier patents based on text similarity.

In summary, forecasting emerging technologies from an outlier perspective effectively identifies overlooked technological points early, while deep learning demonstrates superior performance in text information extraction and prediction tasks. This study builds upon these foundations.

---

3. Research Method

Figure 1 [Figure 1: see original paper] illustrates the overall methodology, comprising five main steps: (1) patent data acquisition; (2) text vectorization using word embedding models to construct patent similarity networks and screen outlier patents as candidate emerging technologies; (3) extraction of early-stage characteristic indicators from patents using bibliometric methods and assessment of future technological influence; (4) fitting the relationship between patent indicators and future influence using deep learning models; (5) model performance evaluation.

3.1 Outlier Patent Acquisition After obtaining all patents in the target field from patent databases, a strategy is needed to identify outlier patents. As shown in Figure 2 [Figure 2: see original paper], in a patent similarity network, each patent represents a node, with connections between nodes determined by similarity scores. Outlier patents are "outlier points" in the network—nodes without any connections to other patents.

The acquisition process involves four steps: (1) converting each patent into an n-dimensional vector representation using the BERT pre-trained model; (2) calculating pairwise patent similarities; (3) constructing a patent similarity network by connecting patents with similarities above threshold ; (4) selecting nodes without connections as outlier patents.

The vectorization process, or encoding, uses the BERT pre-trained model as an effective deep learning-based encoder. Unlike traditional frequency-based methods, BERT considers word associations within texts, enabling more comprehensive information extraction. The model' s input is patent text, which is decomposed into three components as shown in Equation (1):

$$X_{emb} = T_{emb} + S_{emb} + P_{emb}$$

The first part is token embedding, carrying word meanings; the second is segment embedding, representing contextual relationships between sentences in long texts; the third is position embedding, representing word order. These three components combine as BERT input.

BERT introduces a multi-head self-attention mechanism to weight keywords during encoding. The input X_{emb} undergoes linear transformation to produce three matrices: Q, K, and V, representing query, key, and value vectors for each word, obtained through:

$$Q = X_{emb}W_Q, \quad K = X_{emb}W_K, \quad V = X_{emb}W_V$$

The self-attention mechanism is implemented as:

$$\text{attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where QK^T represents the correlation between the current word and other sentence parts, and d_k is the vector dimension. The softmax output provides word weights, which multiply with value matrix V to produce weighted vector representations.

Multi-head attention implements multiple attention mechanisms simultaneously, merging results to complete encoding:

$$Y_{emb} = \text{feed}(W_z \cdot \text{multihead\_attention}(Q, K, V) + X_{emb}) + X_{emb}$$

where Y_{emb} is the encoding result, $\text{feed}(x)$ represents forward propagation, and W_z are fully connected layer weights. This encoding process repeats multiple times, mapping each word in patent text to an n-dimensional vector and summing these to obtain the patent text vector.

Patent similarity is calculated using cosine similarity. For two n-dimensional patent vectors $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$:

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i)^2}}$$

3.2 Outlier Patent Indicator Extraction and Technology Impact Assessment 3.2.1 Outlier Patent Indicator Extraction

Bibliometric research offers numerous patent indicators for describing emerging technologies. As shown in Table 1 , this study employs 11 indicators across five categories to measure outlier patent characteristics:

1. **Novelty**: Measured by technological originality (TO) and prior knowledge (PK). TO describes the diversity of referenced technical fields—broader integration increases invention value. PK describes reference frequency to other patents—more citations decrease novelty and commercial value.
2. **Development Speed**: Represented by technology cycle time (TCT), indicating the recency of prior knowledge and development pace.
3. **Knowledge Density**: Represented by scientific knowledge (SK) in patents—higher density suggests greater innovative potential.
4. **Application Scope**: Encompassed by technological scope (TS), commercial scope (CS), independent claims (PCID), and dependent claims (PCD). Larger patent families indicate higher commercial value, while claim types reflect protection scope.
5. **Development Capability**: Represented by collaboration level (COL), inventor count (INV), and total know-how (TKH). Collaboration and multi-inventor patents correlate with higher value, while assignee capability affects future development.

3.2.2 Technology Impact Classification Labels

Forward citation count is the most widely used impact assessment method, reflecting a patent' s contribution to subsequent technological development. Higher citation frequency indicates greater technological influence. This study classifies impact into high and low levels based on citation distribution, using a threshold where patents with \$ $16 forwardcitationsrepresenthigh - impactsamples(TL\{1\})andthosewith < 16representlow - impactsamples(TL\{0\}$\$).

3.3 Emerging Technology Prediction Based on Outlier Patents The key to predicting emerging technologies from outlier patents is constructing a relationship model between patent indicators and future impact. Given the large volume of outlier patents and numerous indicators, a DNN model fits the complex nonlinear relationship between indicators and impact.

As shown in Figure 3 [Figure 3: see original paper], the DNN comprises input, hidden, and output layers. Inputs are outlier patent indicator vectors (12-dimensional: 11 patent indicators plus impact label). Hidden layers use ReLU activation. The model training process initializes hidden layer parameters with normally distributed random values, compares predictions with actual labels, and updates parameters using cross-entropy loss until convergence. Testing evaluates performance on unseen data using accuracy, precision, recall, and F1-score:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where P = positive samples, N = negative samples, TP/TN = correctly classified positives/negatives, and FP/FN = misclassified positives/negatives.

---

4. Research Results

4.1 Data Collection and Outlier Patent Identification The NC system technology field validates the method' s effectiveness. As a critical technology for digital transformation in manufacturing and national intelligent manufacturing strategies, forecasting emerging NC technologies holds significant strategic importance.

The experiment retrieved 58,021 patents from Derwent Innovation with priority dates between January 1, 2011, and October 30, 2020. The 22,418 patents from 2011–2015 built the indicator-impact relationship model, while the 35,603 patents from 2016–2020 predicted 2025 emerging technologies.

Outlier identification used Python' s transformers library BERT module with two encoder layers, 8 attention heads, mapping each patent to a 128-dimensional vector. Cosine similarity calculated pairwise relationships. Existing studies select similarity thresholds experimentally. A threshold of 0.5 produced too few outliers (risking omission of disruptive technologies), while 0.7 produced too many (introducing noise). A 0.6 threshold balanced these concerns, yielding 2,747 outliers for 2011–2015 and 15,385 for 2016–2020 as candidate emerging technologies for 2020 and 2025, respectively.

4.2 Indicator Extraction and Model Training 4.2.1 Indicator Extraction

The 2020 candidate emerging technology dataset comprised 2,747 samples, each a 12-dimensional vector (11 indicators plus impact label). Table 2 presents descriptive statistics. Given large value ranges, all indicators were log-transformed (negative values set to zero). Following Section 3.2.2, patents with $ $16 forward citations were classified as high-impact.

4.2.2 Model Training

Although numerous patents emerge annually, few achieve high future impact. In the outlier dataset, the high-to-low impact ratio was only 1:12, creating severe class imbalance. After splitting data 7:3 for training/testing, positive samples were oversampled in the training set to achieve balance.

Models were implemented in Python using scikit-learn: DNN, logistic regression (LR), random forest (RF), and support vector machine (SVM). DNN hyperparameters were optimized through multiple experiments: 3 hidden layers with 512 neurons each, ReLU activation, Adam optimizer, and L2 regularization (0.0001). LR, RF, and SVM parameters were similarly tuned.

Table 3 compares model performance. The DNN significantly outperformed LR, RF, and SVM across all metrics, demonstrating superior ability to identify potential emerging technologies while minimizing false negatives (avoiding missed opportunities) and false positives (preventing resource waste). The DNN better captures complex nonlinear relationships between indicators and impact.

4.3 Emerging Technology Prediction and Analysis 4.3.1 Emerging Technology Prediction

The trained DNN predicted 2025 impact for 15,385 outlier patents from 2016–2020. Among these, 348 (2.26%) were projected as high-impact, representing predicted emerging technologies.

An LDA model analyzed these patents' topics. Perplexity analysis indicated 5 optimal topics with good discrimination. Table 4 presents keywords and themes:

1. **Autonomous Sensing and Connectivity**: Key to intelligent machine tools, based on "instruction domain oscilloscopes" and connectivity protocols. Cloud platforms show application potential.
2. **Process Parameter Optimization**: Critical for machining quality and efficiency. Big data modeling with neural networks optimizes feed rates and spindle power beyond traditional stability models.
3. **External Sensors**: "Internet + sensors" enhances machine tool state perception, enabling adaptive control through temperature/pressure data analysis.
4. **Error Compensation**: Important for quality assurance, including thermal and geometric error compensation using deep learning models with sensor feedback for closed-loop control.
5. **Special Material Processing**: Growing demand for hard, brittle, heat-sensitive materials in aerospace, automotive, and medical fields drives development of novel processing technologies like laser heating and electrochemical methods, with additive manufacturing showing biomimetic material synthesis potential.

4.3.2 Prediction Result Analysis

Predicted emerging technologies differ significantly from non-emerging technologies across six indicators: PK (prior knowledge), TCT (technology cycle time), TS (technological scope), CS (commercial scope), INV (inventor count), and TKH (assignee capability). Technology convergence is a key emergence mechanism, making patents spanning multiple IPC classes more promising. Collaboration among high-capability assignees and multiple institutions enhances innovation quality. Higher commercial value increases application likelihood and technology development. Emerging technologies reference more prior knowledge and older literature, indicating that breakthroughs require deep domain investigation rather than spontaneous ideas. Emerging technologies also show slightly higher PCID and PCD values, reflecting stronger protection demands for innovative technologies.

This empirical case predicts promising emerging technology directions in NC systems, analyzes key early characteristics, validates the method, and provides strategic guidance for corporate and government planning.

5. Research Conclusions

This study identifies and forecasts emerging technologies from an outlier perspective using deep learning. Key conclusions include: (1) Identifying emerging technologies through outlier patents based on early characteristics is timely and effective, validated in the NC system case; (2) A BERT-based outlier identification method effectively captures patents diverging from mainstream trajectories through text similarity networks; (3) The DNN model outperforms LR, RF, and SVM in fitting complex relationships between indicators and impact, enabling more effective emerging technology prediction.

This approach complements existing forecasting methods by leveraging patent text information for similarity calculation and modeling high-dimensional indicator-impact relationships. It offers low time costs, broad applicability, and rapid identification of potential emerging technologies in any field. The NC system case provides actionable intelligence for strategic development.

Limitations include: (1) Sole reliance on patent data, focusing on technology-driven forecasting—future work should incorporate social impact and commercial benefit indicators; (2) Deep learning' s complexity prevents deep exploration of indicator impact mechanisms—future research should investigate causal relationships to strengthen theoretical foundations.

References

[1] Xue Lan, Zhou Yuan, Li Yingbo. Research on Innovation Patterns and Industrial Policies for Strategic Emerging Industries[M]. Beijing: Science Press, 2015: 44-56. [2] Valles, Vázquez-Bustelo D. Concurrent engineering performance: incremental versus radical innovation[J]. International Journal of Production Economics, 2009, 119(1): 136-148. [3] Fu Yuxiu, Zhang Hongshi. Breakthrough innovation: conceptual definition and comparison[J]. Journal of Quantitative & Technical Economics, 2004, 21(3): 73-83. [4] Noh H, Song Y-K, Lee S. Identifying emerging core technologies for the future: case study of patents published by leading telecommunication organizations[J]. Telecommunications Policy, 2016, 40(10/11): 956-970. [5] Li He, Yuan Cuimin, Xie Mengfan. Analysis and research on sleeping beauties in patent literature[J]. Library and Information Service, 2019, 63(6): 64-74. [6] Li Jinghai, Xu Guangwen, Yang Lidan, et al. A smokeless coal combustion method and furnace for suppressing nitrogen oxides: CN95102081[P]. 1998-05-20. [7] Sugitani H, Matsuda H, Ikeda M. Liquid jet recording head: US06/394787[P]. 1985-12-10. [8] Rotolo D, Hicks D, Martin B R. What is an emerging technology?[J]. Research Policy, 2015, 44(10): 1827-1843. [9] Zhang Guosheng. Technological change, paradigm shift and strategic emerging industry development: an evolutionary economics perspective[J]. Industrial Economics Research, 2012(6): 26-32. [10] Aharonson B S, Schilling M A. Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution[J]. Research Policy, 2016, 45(1):

81-96. [11] Yoon J, Kim K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection[J]. Scientometrics, 2012, 90(2): 445-461. [12] Song K, Kim K, Lee S. Identifying promising technologies using patents: a retrospective feature analysis and prospective needs analysis on outlier patents[J]. Technological Forecasting and Social Change, 2018, 128: 118-132. [13] Luo Suping, Kou Cuicui, Jin Jin, et al. Disruptive technology prediction based on outlier patents: a case study of traditional Chinese medicine patents[J]. Information Studies: Theory & Application, 2019, 42(7): 165-170. [14] Zhou Y, Dong F, Liu Y, et al. Forecasting emerging technologies using data augmentation and deep learning[J]. Scientometrics, 2020, 123(1): 1-29. [15] Choy Y Y, Jeong G H, Kim S H. A Delphi technology forecasting approach using a semi-Markov concept[J]. Technological Forecasting and Social Change, 1991, 40(3): 273-287. [16] Lee S, Kim W, Kim Y M, et al. The prioritization and verification of IT emerging technologies using an analytic hierarchy process[J]. Technological Forecasting and Social Change, 2017, 119: 80-97. [17] Geum Y, Lee S, Yoon B, et al. Identifying and evaluating strategic partners for collaborative R&D: index-based approach using patents and publications[J]. Technovation, 2013, 33(6/7): 211-224. [18] Song B, Seol H, Park Y. A patent portfolio-based approach for assessing potential R&D partners: an application of the Shapley value[J]. Technological Forecasting and Social Change, 2016, 103: 156-165. [19] Lanjouw J O, Schankerman M. Stylized facts of patent litigation: value, scope and ownership[J]. National Bureau of Economic Research, 1999: w6297. [20] Sternitzke C, Bartkowski A, Schramm R. Visualizing patent statistics by means of social network analysis tools[J]. World Patent Information, 2008, 30(2): 115-131. [21] Meyer M. Does science push technology? Patents citing scientific literature[J]. Research Policy, 2000, 29(3): 409-434. [22] Narin F, Noma E, Perry R. Patents as indicators of corporate technological strength[J]. Research Policy, 1987, 16(2): 143-155. [23] Lanjouw J O, Pakes A, Putnam J. How to count patents and value intellectual property: the uses of patent renewal and application data[J]. The Journal of Industrial Economics, 1998, 46(4): 405-432. [24] Aristodemou L, Tietze F. The state-of-the-art on intellectual property analytics (IPA): a literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data[J]. World Patent Information, 2018, 55: 37-51. [25] Kong D, Zhou Y, Liu Y, et al. Using the data mining method to assess the innovation gap: a case of industrial robotics in a catching-up country[J]. Technological Forecasting and Social Change, 2017, 117: 228-237. [26] Zhou Y, Dong F, Liu Y. Research on knowledge fusion process in emerging fields: a case study of bioinformatics[J]. Library and Information Service, 2019, 63(8): 127-224. [27] Hassan S-U, Imran M, Iqbal S, et al. Deep context of citations using machine learning models in scholarly full-text articles[J]. Scientometrics, 2018, 117(3): 1645-1662. [28] Zhou Y, Dong F, Liu Y, et al. Forecasting emerging technologies using data augmentation and deep learning[J]. Scientometrics, 2020, 123(1): 1-29. [29] Geum Y, Kim C, Lee S, et al. Technological convergence of IT and BT: evidence from patent analysis[J]. ETRI Journal, 2012, 34(3): 439-449. [30] Kim G, Bae J. A novel approach to forecast promising technology through patent analysis[J]. Technological Forecasting and Social Change, 2014, 87: 292-304. [31] Zhou Y, Lin H, Liu Y, et al. A novel method to identify emerging technologies using a semi-supervised topic clustering model: a case of 3D printing industry[J]. Scientometrics, 2019, 120(1): 167-185. [32] Hou Jianhua, Zhu Xiaoqing. Research on patent-based technology forecasting evaluation index system and empirical study[J]. Library and Information Service, 2014, 58(18): 77-82. [33] Lee C, Kwon O, Kim M, et al. Early identification of emerging technologies: a machine learning approach using multiple patent indicators[J]. Technological Forecasting and Social Change, 2018, 127: 291-303. [34] Cao Yiwen, Xu Haiyun, Wu Huawei, et al. Breakthrough prediction of emerging technology topics based on citation curve fitting: a case study of stem cell field[J]. Library and Information Service, 2020, 64(5): 100-113. [35] Zhang Y, Lu J, Liu F, et al. Does deep learning help topic extraction? A kernel k-means clustering method with word embedding[J]. Journal of Informetrics, 2018, 12(4): 1099-1117. [36] Devlin J, Chang M-W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2021-05-31]. https://arxiv.org/pdf/1810.04805.pdf. [37] Yang W, Zhang H, Lin J. Simple applications of BERT for ad hoc document retrieval[EB/OL]. [2021-05-31]. https://arxiv.org/abs/1903.10972.pdf. [38] Bessen J. The value of US patents by owner and patent characteristics[J]. Research Policy, 2008, 37(5): 932-945. [39] Fernández-Ribas A. International patent strategies of small and large firms: an empirical study of nanotechnology[J]. Review of Policy Research, 2010, 27(4): 457-473. [40] Haupt R, Kloyer M, Lange M. Patent indicators for the technology life cycle development[J]. Research Policy, 2007, 36(3): 387-398. [41] Bierly P, Chakrabarti A. Determinants of technology cycle time in the US pharmaceutical industry[J]. R&D Management, 1996, 26(2): 115-126. [42] Kaya A A, Waters R C. An empirical evaluation of the technology cycle time indicator as a measure of the pace of technological progress in superconductor technology[J]. IEEE Transactions on Engineering Management, 1999, 46(2): 127-131. [43] Cozzens S, Gatchair S, Kang J, et al. Emerging technologies: quantitative identification and measurement[J]. Technology Analysis & Strategic Management, 2010, 22(3): 361-376. [44] Day G S, Schoemaker P J. Avoiding the pitfalls of emerging technologies[J]. California Management Review, 2000, 42(2): 8-33. [45] Guellec D, de la Potterie B V P. Applications, grants and the value of patent[J]. Economics Letters, 2000, 69(1): 109-114. [46] Ma Z, Lee Y. Patent application and technological collaboration in inventive activities: 1980-2005[J]. Technovation, 2008, 28(6): 379-390. [47] Meyer M. Are patenting scientists the better scholars? An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology[J]. Research Policy, 2006, 35(10): 1646-1662. [48] Harhoff D, Narin F, Scherer F M, et al. Citation frequency and the value of patented inventions[J]. Review of Economics and Statistics, 1999, 81(3): 511-515. [49] Dong Fang, Liu Yufei, Zhou Yuan. Multi-classification of emerging technologies based on LDA-SVM paper abstracts[J]. Journal of Intelligence, 2017(7): 40-45. [50] Chen J, Yang J, Zhou H, et al. CPS modeling of CNC machine tool processes using an instruction-domain-based approach[J]. Engineering, 2015, 1(2): 247-260. [51] Chen J, Hu P, Zhou H, et al. Toward intelligent machine tool[J]. Engineering, 2019, 5(4): 679-690. [52] Saffar R J, Razfar M R. Simulation of end milling operation for predicting cutting forces to minimize tool deflection by genetic algorithm[J]. Machining Science and Technology, 2010, 14(1): 81-101. [53] Li Z, Wang Y, Wang K. A data-driven method based on deep belief networks for backlash error prediction in machining centers[J]. Journal of Intelligent Manufacturing, 2020, 31(7): 1693-1705. [54] Chen Xiaoli. Review of composite processing technology for hard and brittle materials[J]. Aeroengine, 2010(3): 57-60. [55] Pollini B, Pietroni L, Mascitti J, et al. Toward a new material culture. Bio-inspired design, parametric modeling, material design, digital manufacture[C]//Design in the Digital Age Technology, Nature, Culture. Milano: Politecnica University Press, 2020: 208-212.

Author Contributions

Kong Dejing: Framework design, paper revision, and writing guidance
Dong Fang: Conceptual refinement and experimental guidance
Chen Zijing: Experimental design, implementation, and paper drafting
Liu Yuhan: Experimental implementation
Zhou Yuan: Writing guidance

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