Identification of High-Value Patents from a Mutation Perspective: A Case Study of Patents in the Field of Nanodrugs Postprint

Authors: Feng Guohe, Li Lin, Liu Renhua, Deng Weiwei

Date: 2025-07-09T14:32:37+00:00

Abstract

The 14th Five-Year Plan proposes the strategic goals and tasks for building a strong intellectual property nation, and for the first time incorporates the protection and incentivization of high-value patents into the main objectives of China' s economic and social development, making the identification of high-value patents of significant importance. Based on BERTopic model' s identification of novelty topics, this study constructs a corresponding catastrophe progression model grounded in an indicator evaluation system covering four dimensions—legal, technical, economic, and inventor—to conduct comprehensive assessment of patents. The identified high-value patents are verified through patent text analysis, emerging industry comparison, and patent market value. The results demonstrate that the high-value patents screened by this method are consistent with relevant data, constituting an effective approach for high-value patent identification.

Full Text

Identification of High-Value Patents from a Mutation Perspective: A Case Study of Patents in the "Nanomedicine" Field

Abstract

The 14th Five-Year Plan outlines the objectives for building a strong intellectual property nation in China, marking the first time that protecting and incentivizing high-value patents has been included as a primary goal in the nation' s economic and social development, which highlights the critical importance of identifying high-value patents. This study develops a mutation series model

based on the BERTopic model to identify novel topics, constructing a comprehensive evaluation system from four dimensions—legal, technical, economic, and inventor-related indicators—to conduct a holistic assessment of patents. The identified high-value patents are validated through patent text analysis, comparison with emerging industries, and examination of patent market value. The results demonstrate that the high-value patents selected through this method are consistent with findings from relevant sources, confirming the effectiveness of this approach for high-value patent identification.

Keywords: high-value patent; BERTopic; mutation series method; index system; content analysis

1. Related Research

1.1 Definition of High-Value Patents

The concept of high-value patents can be traced back to a survey study conducted by European scholar Sanders in 1958. Research has found that only a small percentage of patents achieve licensing transfer and industrial application, while also revealing significant differences in commercialization value among patents. High-value patents refer to effective invention patents that align with national strategic industrial development directions and possess high quality and value. Such patents typically exhibit characteristics of high technical value, high economic value, and high market value. During the 14th Five-Year Plan period, China has included the number of high-value invention patents per 10,000 population as a key indicator, with the current number reaching 12.9, indicating that high-value patents generate substantial value for society. The core task of current high-value patent work involves establishing a scientific evaluation methodology system to efficiently identify high-value patents.

1.2 High-Value Patent Evaluation Indicator Systems

A scientifically sound evaluation indicator system forms the core of patent value assessment models. Existing research predominantly considers legal, technological, and economic dimensions comprehensively, with few studies treating inventor characteristics as a separate dimension. However, patent inventors serve as the core force in patent creation, and their professional levels and collaboration patterns exert non-negligible effects on patent value. Studies have shown that the number of inventors, whether they collaborate with enterprises, and other factors impact patent value, though some scholars argue these relationships are not significant. Despite differing views on the direction and magnitude of impact, these indicators remain important for patent value assessment.

From a legal perspective, the number of claims, independent claims, and patent family size positively correlate with patent value. The number of claims reflects the richness of technical features, while patent families indicate international

protection scope. From a technical viewpoint, citation counts, cited frequency, and specification pages serve as value indicators, with high-value patents typically receiving more subsequent citations. From an economic perspective, transfer frequency, litigation frequency, remaining validity period, and emerging industry classification significantly affect patent value. Some scholars argue that transfer frequency correlates with value, while others suggest high-value patents are often state-protected with restricted transfers. The remaining validity period reflects market lifespan, and alignment with emerging industries indicates technological impact.

1.3 High-Value Patent Evaluation Methods

Traditional patent value assessment methods primarily include Analytic Hierarchy Process and fuzzy evaluation methods, which rely heavily on subjective judgment. Machine learning methods help identify high-value patent characteristics and construct value models, avoiding limitations of subjective evaluation. Researchers have employed various models including self-organizing maps, support vector machines, random forest, convolutional neural networks, and graph neural networks for patent value detection, all demonstrating high accuracy. Studies have found positive relationships between patent novelty and value, though some argue that excessive novelty may increase costs, requiring further investigation. Current mainstream approaches combine multi-dimensional measurement with machine learning methods.

BERTopic enables automated identification of potential topics through topic modeling and mapping. In practice, enterprises should establish comprehensive patent management systems, continuously improve patent quality and value through ongoing innovation and technological advancement, thereby maintaining competitive advantages.

---

2. Research Design

2.1 Framework and Approach

This study constructs an innovative patent value assessment model that integrates BERTopic for novelty identification and mutation series method for comprehensive evaluation. The research framework involves dividing the collected patent dataset into a control window (2011-2020) and an experimental window (2021-2023). Patent titles and abstracts undergo topic extraction, with domain-specific keyword dictionaries and technology keywords used for tokenization and stop-word removal. The model iteratively determines optimal topic numbers, converts topic words to vectors using Word2vec, calculates similarity between topics across time windows, and identifies novel topics based on similarity thresholds.

2.2 Indicator System Construction

Drawing from the Patent Evaluation Guidelines (GB/T 42748-2023) issued by the China National Intellectual Property Administration and existing research, this study introduces inventor characteristics alongside legal, technical, and economic dimensions to construct a comprehensive evaluation system comprising 4 primary dimensions and 16 specific indicators.

Legal Dimension: Includes quantifiable indicators such as claim stability, protection scope, claim reasonableness, infringement determinability, claim types, and technical feature attributes. Patent family data reflects international protection scope, while independent claims, claim structure, and technical feature count indicate protection reasonableness.

Technical Dimension: Encompasses technical advancement (citations, cited frequency), technical scope (IPC classification count, technical field quantity), technical independence (background technology description, standalone applicability), and specification pages.

Economic Dimension: Covers market control (patentee' s patent portfolio, remaining economic lifespan), policy adaptability (policy support level, industry approval requirements), and patent operation status (transfer/licensing frequency).

Inventor Dimension: Includes inventor quantity, professional titles, administrative positions, patent authorization counts, and enterprise collaboration experience. First inventers with senior professional titles or administrative positions, as well as collaboration with enterprises, positively correlate with patent value.

2.3 Mutation Series Method for Patent Value Assessment

The mutation series method conducts comprehensive evaluation through bottom-up hierarchical ranking. This approach combines catastrophe theory with fuzzy functions, using normalized formulas to characterize indicator relationships while reducing subjectivity compared to multivariate statistical methods. The method involves identifying mutation system types (fold, cusp, swallowtail, butterfly, Indian hut) based on control variable counts, with corresponding potential functions. During normalization, the non-complementary principle applies when control variables lack mutual compensation, using normalized formulas for evaluation. This study employs random forest algorithms for indicator weight analysis to determine mutation system types at each hierarchy.

2.4 Validation Approach

For identified high-value patents, this study employs multi-level validation: (1) Text analysis of high-value patents to examine emerging technology themes; (2)

Cross-analysis with emerging industry classifications to verify alignment with national policy directions; (3) Market value verification through transfer frequency analysis, as transferred patents typically possess higher recognized value.

---

3. Empirical Analysis

3.1 Data Collection and Preprocessing

Using the Innojoy patent database as the data source, this study employs the following search strategies: TA=('nanomedicine' *OR* ' *nanomaterial drug*' OR 'nanodrug' *OR* ' *nano-drug*' OR ' nano drug*'), with a timeframe from 2011-2023 and authorized patents only. After removing duplicates and invalid patents, the data is divided into a control group (2011-2020) and experimental group (2021-2023). Missing value treatment, forward/backward indicator processing, and min-max normalization are performed following established methods.

3.2 Novelty Patent Screening

Patent titles and abstracts from both time windows undergo topic extraction using BERTopic. The model generates 11 topics for the control window and 12 for the experimental window. Topics with high c-TF-IDF values are selected as effective topics. Similarity between topic word vectors is calculated using Word2vec trained on Chinese Wikipedia corpus. Topics with similarity below the mean are identified as novel, with Topic0, Topic6, and Topic8 from the experimental window showing high novelty.

3.3 Patent Value Assessment Based on Mutation Series Method

Secondary indicators under each dimension are aggregated to calculate primary indicators. Random forest algorithms determine indicator weights and identify mutation system types. The legal and technical dimensions use the Indian hut mutation system, the economic dimension uses the butterfly mutation system, and the inventor dimension also employs the Indian hut mutation system. After standardizing the data, normalized formulas are applied following non-complementary principles. Patents ranking in the top 10% of comprehensive scores are designated high-value patents.

3.4 High-Value Patent Text Theme Validation

A corpus of high-value patent texts is constructed and preprocessed. TF-IDF calculations identify key terms with high discriminative power. Analysis of the top 6 high-value patents reveals five core themes: (1) Carrier functionalization and surface modification, (2) Peptide modification technology, (3) Composite material preparation technology, (4) Nanomaterial biological barrier mechanisms, and (5) Natural biological nanomaterials. These themes align with frontier

research and national policies, including the National Nanomedicine Quality Control Research Technical Guidelines and the National Key R&D Program.

3.5 Strategic Emerging Industry Classification Cross-Analysis

High-value patents are analyzed against the Strategic Emerging Industry Classification (2018). The identified patents align with categories such as biomedical materials, drug controlled-release systems, and medical polymer materials, confirming their relevance to national strategic directions.

3.6 High-Value Patent Market Value Verification

Analysis of transfer records shows that 60% of identified high-value patents have transfer frequencies 1, indicating significant market recognition and economic value. This validates the effectiveness of the assessment model.

[Figure 3: see original paper] Topic distance maps for (a) 2011-2020 and (b) 2021-2023, showing topic similarity and novelty identification.

---

4. Conclusion

This study constructs a systematic identification and evaluation framework for high-value patents, integrating BERTopic and mutation series methods to provide valuable references for emerging technology patent research. Methodologically, BERTopic extracts patent topics and their keywords, identifying novelty from a technical content perspective and providing new pathways for high-value patent screening. The indicator system expands traditional evaluation dimensions by incorporating inventor characteristics, enhancing scientific rigor and comprehensiveness. The selected indicators accurately reflect patent value, improving assessment precision and reliability.

The study identifies high-value patents in the nanomedicine field and validates them through in-depth analysis of policy documents and frontier literature. The identified themes demonstrate consistency with cutting-edge research and national policies, confirming the effectiveness and accuracy of the evaluation methodology. Future research can further explore multi-dimensional innovations, expand application domains, and refine the evaluation framework.

References

[1] State Council. Notice on Issuing the National Intellectual Property Protection and Utilization Plan for the 14th Five-Year Plan [EB/OL]. [2025-01-14]. https://www.gov.cn/zhengce/zhengceku/2021-10/28/content_5647274.htm.

[2] National Intellectual Property Administration. Report on Implementation of Key Tasks for the First Half of 2024 [EB/OL]. [2025-01-14]. https://www.cnipa.gov.cn/art/2024/7/31/art_53_194030.html.

[3] Scherer FM. The Economic Impact of Patents [J]. Patent, Trademark & Copyright Journal of Research and Education, 1958: 340-362.

[4] Sanders BS. Some Problems in Measuring the Inventive Output of Patents [J]. Journal of the Patent Office Society, 1958, 40: 340-362.

[5] Lanjouw O, Schankerman M. Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators [J]. Economic Journal, 2004, 114(495): 441-465.

[6] De I, Orsi L, Belussi F. Collaborative Networks and Innovation Performances in Lagging-Behind European Regions [J]. Research Policy, 2018, 47(1): 1-13.

[7] Arts S, Hou J, Gomez J. Natural Language Processing to Identify Creation and Impact of Technologies in Patent Text: Code, Data, and New Measures [J]. Research Policy, 2021, 50(2): 104144.

[8] Devlin J, Chang W, Lee K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.

[9] State Administration for Market Regulation, Standardization Administration. Patent Evaluation Guidelines (GB/T 42748-2023) [EB/OL]. [2025-01-14]. https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=21AB9B32C8B01C3849AF785EBEFC357B.

[10] Ran C, Huang W. Identification of High-Value Patent Technology Opportunities in Universities [J]. Information Source Management Journal, 2024(4): 103-116.

[11] Song K, et al. Research on High-Value Patent Cultivation Based on Mutation Series Method and Patent Portfolio Strategy [J]. China Science and Technology Forum, 2023(11): 108-119.

[12] Hou J, Lin M. Early Detection of Valuable Patents Using Learning Models: Case Study of the Semiconductor Industry [J]. Technological Forecasting and Social Change, 2020, 158: 120146.

[13] Choi J, Jeong B, Yoon B. A Novel Approach to Evaluating Business Potential of Intellectual Properties: A Machine Learning-Based Predictive Analysis [J]. Computers & Industrial Engineering, 2020: 106544.

[14] Chung P, Sohn S. Patent Analysis Using Deep Learning to Identify Technological Trends [J]. World Patent Information, 2020: 101943.

[15] Cheng G, Liu Y. Anti-Parkinsonian Therapy: Strategies for Crossing the

Blood-Brain Barrier and Nano-Biological Effects of Nanomaterials [J]. Nano-Micro Letters, 2022, 14(1): 105.

[16] Huang T, Wang G, Shahbazi M. Surface Decoration of Nanoparticles Enables Efficient Therapy Toward Osteoporosis and Diabetes [J]. Advanced Functional Materials, 2023, 33(2): 2210627.

[17] Li L, Yang H. Natural Nanofibrils as Reinforcements for Preparation of Chitosan-Based Bionanocomposites [J]. Carbohydrate Polymers, 2021: 117214.

[18] National Natural Science Foundation of China. Guidelines for Major Research Plan on Dynamic Modification and Chemical Intervention of Biomacromolecules [EB/OL]. [2024-12-20]. https://www.nsfc.gov.cn/tzgg_229/a791078ac3e6407eb7e29885cb1a5d49.htn

[19] Center for Drug Evaluation, NMPA. Technical Guidelines for Quality Control Research of Nanomedicines [EB/OL]. [2024-12-10]. https://www.cde.org.cn/main/news/viewInfoCommon/

[20] Bernkop-Schnürch A, Dünnhaupt S. Chitosan-Based Delivery Systems [J]. European Journal of Pharmaceutics and Biopharmaceutics, 2012, 81(3): 463-469.

[21] Yang Y. Advances in Silica-Based Nanoparticles for Targeted Cancer Therapy [J]. Nanomedicine: Nanotechnology, Biology and Medicine, 2016, 12(2): 317-332.

[22] National Bureau of Statistics. Strategic Emerging Industry Classification (2018) [EB/OL]. [2024-12-20]. http://www.stats.gov.cn/xw/tjxw/tzgg/202302/t20230202_1893984.html.

[23] Suk Q, Kim J. PEGylation Strategy for Improving Nanoparticle-Based Drug Delivery [J]. Advanced Drug Delivery Reviews, 2016: 28-51.

[24] State Council. National Long-Term Science and Technology Development Plan (2006-2020) [EB/OL]. [2025-01-09]. https://www.gov.cn/gongbao/content/2006/content_240244.htm.

[25] Ministry of Science and Technology. 2021 Annual Project Application Guide for Nanotechnology Frontier Key Special Project [EB/OL]. [2025-01-10]. https://service.most.gov.cn/u/cms/static/202105/B9202197_20210511091057.pdf.

[26] Fang X, et al. Research on Frontier Topic Identification Methods in Medical Field [J]. Library and Information Service, 2018, 62(7): 686-694.

[27] Zhang J, et al. Core Patent Identification: Theoretical and Empirical Research [J]. China Invention & Patent, 2020(6): 100-105.

[28] Liu Q, et al. Theoretical Discussion on High-Value Patent Evaluation in Universities [J]. China University Science & Technology, 2020(S1): 15-18.

[29] Tang Y, et al. High-Value Patent Economic Value Cultivation and Operation Research [J]. Jiangsu Science & Technology Information, 2022(4): 147-152.

[30] Zhou C, et al. High-Value Patent Evaluation Methods, Problems and Countermeasures [J]. Science and Technology Management Research, 2023(7): 133-141.

[31] Liu C, et al. High-Value Patent Influencing Factor Analysis and Topic Mining: A Case Study of Nanomedicine Field [J]. Library Forum, 2024(7): 69-82.

[32] Wu J, et al. Patent Value Re-evaluation: An Empirical Study of University Patent Transfer [J]. Science Research, 2022(9): 1608-1620.

[33] Guo Y, et al. High-Value Patent Identification Method Effectiveness Empirical Study [J]. Science Research, 2019(1): 11-17.

[34] Ren H, et al. Core Patent Identification Comprehensive Value Model Construction and Empirical Research [J]. Information Source Management Journal, 2021(12): 195-202.

[35] Feng G, et al. High-Value Patent Identification Under Domain Knowledge Fusion Impact [J]. Data Analysis and Knowledge Discovery, 2025(3): 69-82.

[36] Feng G, et al. Emerging Technology Identification from Semantic and Evolutionary Perspectives: A Case Study of Nanomedicine Patents [J/OL]. [2025-03-25]. http://kns.cnki.net/kcms/detail/11.1567.G3.20250122.1659.013.html.

[37] Hou G, et al. Patent Value Assessment and Transfer Efficiency of Stem Cell Technology [J]. China Biotechnology, 2019(1): 99-109.

[38] Kuan C, et al. High-Quality Patent Automatic Identification Method Based on Graph Convolutional Network [J]. Data Analysis and Knowledge Discovery, 2022(1): 88-95, 124.

[39] Liu R, et al. Disruptive Technology Identification Model Based on Improved Mutation Series [J]. Library and Information Service, 2023(3): 61-71.

[40] Feng G, et al. Risk Formation, Evolution and Control Mechanism of Railway Dangerous Goods Transportation Systems [D]. Southwest Jiaotong University, 2019.

[41] Wang Z, et al. Regional Scientific and Technological Innovation Capacity Evaluation Based on Improved Mutation Series [J]. China Soft Science, 2012(6): 90-101.

[42] Lu Z, et al. Technology Opportunity Identification Based on Topic Mining and Patent Evaluation: A Case Study of Smart Agriculture [J]. Library and Information Service, 2021(10): 7-14.

[43] Liu Q, et al. Enterprise Anti-Competitive Intelligence Capability Evaluation Based on Mutation Series Method [J]. Science and Technology Management Research, 2012(10): 12-16.

[44] Song K, et al. Cultural Media Industry Performance Evaluation Based on Mutation Series Algorithm [J]. Science and Technology Management Research, 2011(11): 38-41, 37.

[45] Hou G, et al. Subject Hotspot Research Based on Time-Weighted Keyword Frequency Analysis [J]. Journal of Hebei University of Engineering (Natural Science Edition), 2019(3): 67-71.

[46] Liu R, et al. Rockburst Grade Evaluation Based on Improved Entropy Weight-Mutation Series Method [J]. Journal of Hebei University of Engineering (Natural Science Edition), 2019(3): 67-71.

[47] Feng G, et al. Patent Value Assessment and Influencing Factor Identification from Multi-Method Fusion Perspective [J]. Data Analysis and Knowledge Discovery, 2022(15): 127-139.

[48] Feng G, et al. Patent Value Assessment Research Review and Trend Outlook [J]. Library and Information Service, 2022(15): 127-139.

[49] National Intellectual Property Administration. Patent Evaluation Guidelines [EB/OL]. [2025-01-14]. https://www.cnipa.gov.cn/.

[50] Choi S, et al. Qualitative and Quantitative Analysis of Patents in Nanomedicine: Bridging Research and Practical Applications [J]. World Patent Information, 2020: 101943.

[51] Cheng J, et al. High-Value Patent Identification Method and Its Application in Artificial Intelligence Patents [J]. Information Science, 2020(9): 195-202.

[52] Zhang L, et al. Patent Technology Emerging Degree and Its Technical Impact [J]. Information Science, 2018(5): 21-29.

[53] Li M, et al. Patent Domain Market Dominance Identification Based on Patent Value Assessment [J]. Electronic Intellectual Property, 2018(5): 21-29.

[54] Wang Y, et al. Patent Value Assessment and Classification Based on Self-Organizing Maps [J]. Data Analysis and Knowledge Discovery, 2019(5): 117-124.

[55] Zhang X, et al. Multi-Feature Based Technology Fusion Relationship Prediction and Value Evaluation [J]. Data Analysis and Knowledge Discovery, 2022(Z1): 33-44.

[56] Yang L, et al. Patent Value Assessment Method Based on Multiple Regression Model [J]. Patent Appraisal, 2023(2): 183-192.

[57] Chen Y, et al. Network Platform Patent Value Assessment Method Based on Grey Relational Analysis and Random Forest Regression [J]. Information Theory and Practice, 2019(10): 109-116.

[58] Zhao Z, et al. Patent Intrinsic Value Evaluation Indicator Data Mining Based on Patent Documents [J]. China Invention & Patent, 2020(S1): 15-18.

[59] Wu H, et al. High-Value Patent Identification Method Based on Cloud Matter-Element Model with Combined Weighting [J]. Science and Technology Management Research, 2023(7): 133-141.

[60] Liu Q, et al. High-Value Patent Evaluation and Identification Research Based on Grounded Theory [J]. Information Theory and Practice, 2022(2): 136-144.

[61] Feng G, et al. High-Value Patent Mining and Influencing Factor Identification from Multi-Method Fusion Perspective [J]. China Science and Technology Forum, 2023(11): 108-119.

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