Technical Opportunity Analysis Based on Patent Novelty R&D Portfolio Mapping: A Case Study in the Field of Thrombolytic Drugs (Postprint)

Authors: Tao Chenglin, Wang Wei, Zhang Shiyu

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

Abstract

[Purpose / Significance] To identify novel patents instead of patent gaps represented by a set of keywords, thereby addressing the drawback of overly subjective technology opportunity identification. [Method / Process] This study employs a quantitative method based on systematic processes to identify the novelty degree of patents. Novel patents are identified through the Density-Based Local Outlier Factor (DLOF) algorithm, and technology scope indicators together with indicators of the number of similar patents are utilized to construct an R&D portfolio identification map. [Results / Conclusion] The research results demonstrate that the R&D portfolio identification map based on patent novelty can accurately identify novel patents, providing reference for technology research and development.

Full Text

Preamble

Technology Opportunity Analysis Based on Patent Novelty R&D Portfolio Identification Map: A Case Study of Thrombolytic Drugs

Tao Chenglin, Wang Wei, Zhang Shiyu
Department of Medical Informatics, School of Public Health, Jilin University, Changchun 130021

Abstract

[**Purpose/Significance**] This study identifies novel patents to replace patent gaps typically represented by keyword sets, thereby addressing the excessive subjectivity in technology opportunity identification. [**Methods/Process**] We employ a systematic, quantitative process-based method to assess patent novelty. The Density-based Local Outlier Factor (DLOF) algorithm identifies novel patents, which are then positioned using technology scope and similar patent count indicators to construct an R&D portfolio identification map. [**Results/Conclusion**] The findings demonstrate that the novelty-based patent R&D portfolio map accurately identifies novel patents and provides actionable references for technology R&D.

Keywords: technology opportunity analysis; novel patents; patent novelty R&D portfolio identification map; text mining; local outlier factor
Classification Number: G202
Citation Format: Tao Chenglin, Wang Wei, Zhang Shiyu. Technology opportunity analysis based on patent novelty R&D portfolio identification map: A case study of thrombolytic drugs [J/OL]. Knowledge Management Forum, 2016, 1(4): 276-282 [citation date]. http://www.kmf.ac.cn/p/1/44/.

1 Introduction

As new technologies continue to emerge, enterprises increasingly recognize the inherent risks of technological innovation. In this context, the strategic importance of technology opportunity analysis has grown substantially. More companies are forming expert panels to grasp current technological development status in their industries and identify breakthrough opportunities. However, as technology drives greater value and labor-intensive industries gradually yield to technology-driven ones, expert discussions that consume substantial manpower and time can no longer meet the demands of shortening innovation cycles [1]. Consequently, business leaders seek to discover potential technological opportunities through technology opportunity analysis to reduce production costs and time while creating greater value. Patent analysis based on bibliometric analysis and text mining can be used to analyze technology opportunities and objectively identify enterprise technology development points [2].

Technology opportunity analysis based on patents refers to using a series of techniques and large-scale datasets to mine patent information, discover new technologies, and predict their market prospects, with an emphasis on technology forecasting and foresight. Since patent information contains critical technologies in technical fields, analyzing patent information can provide support for technology opportunities. Technology opportunity analysis results can be represented through patent maps in the form of graphics, tables, and curves to make the results more concise and intuitive [3]. Numerous studies have combined text mining with data dimensionality reduction methods for patent-based technology opportunity analysis. Methods such as patent blank maps [4], patent vacancy maps [5], patent maps based on generative topographic mapping, and semantic patent maps [6] have all been used for technology opportunity identification. Although scholars have researched technology opportunity identification from various perspectives, the precision remains insufficient.

Current research mostly reduces high-dimensional patent data based on technical similarity to display it in a two-dimensional patent map. Blank spaces in the patent map are interpreted as potential technology opportunities, with lower blank point density indicating greater potential opportunities. While such research saves manual operation time, drawbacks persist when analyzing large amounts of unstructured data in practice. Current patent map research emphasizes visualization methods [7], primarily Principal Component Analysis (PCA) and Self-Organizing Maps (SOM). PCA [8] achieves dimensionality reduction through linear rotation transformation by finding directions with maximum variance as coordinate axes and discarding dimensions with smaller variance, but valuable information contained in discarded variables is not considered. The SOM method [9] enables non-linear mapping and clustering of vectors, effectively predicting patent data on a two-dimensional plane. However, it assumes adjacent input vectors are independent, failing to consider that patent information is a time series where adjacent data points have substantial correlation, resulting in information loss when using SOM for patent analysis. Recent research on technology opportunity analysis applies generative topographic mapping [10] to identify patent blank points, using inverse mapping to maintain original topological relationships and generate patent maps that identify keyword strings. Whether these keyword strings are meaningful and the thresholds for judging patent blanks both rely on expert panel opinions, making results overly subjective [11]. Due to this subjectivity in patent blank evaluation, assessment of potential technology opportunities also suffers from subjectivity, and objective indicators for measuring patent novelty are currently lacking.

Given these limitations, an improved method for identifying and evaluating technology opportunities is necessary. This paper proposes a systematic, quantitative process-based approach to identify novel patents, replacing previous patent blanks represented by keyword sets to make potential technology opportunity identification more precise. The core of this method combines text mining with local outlier factors. After extracting patent keywords through text mining, local outlier factors measure the novelty degree of a dataset. Unlike current technology forecasting methods based on knowledge flow [12] and knowledge linkage [13], this approach uses keyword usage similarity—that is, patent information novelty degree—as the measurement indicator. Distinguishing itself from existing research, this study employs technology scope indicators to replace citation frequency indicators and constructs a two-dimensional R&D portfolio identification map with similar patent count indicators, making the patent map concise, accurate, and more suitable for analyzing Chinese text patent information.

2.1 Constructing the Patent Matrix

Traditional patent matrices decompose a system into several mutually exclusive and collectively exhaustive two-dimensional matrices, allowing separate processing of each dimension before aggregating into a complete data analysis. This paper mines patent text information to construct a patent matrix with keywords

(S11, …, Snn) as the horizontal axis and patent numbers (P1, …, Pn) as the vertical axis, using patent issue dates, patent numbers, and patent keywords as input for local outlier detection. In the sample patent matrix shown in Table 1 , issue dates and patent numbers are represented in text form, while keyword vectors use binary values, where "1" indicates the patent is related to the listed keyword and "0" means it is not.

In the matrix, P5 was published on YMD5, has keyword S1i in dimension D1, and keyword S21 in dimension D2. The specific structure is shown in Table 1 .

2.2 Identifying Novel Patents

The density-based local outlier mining algorithm no longer treats outliers as a binary attribute (only outlier vs. non-outlier) but uses the Local Outlier Factor (LOF) to represent the degree of anomaly. The LOF of object p reflects its anomaly degree; a larger LOFk(p) indicates greater likelihood of being anomalous data, while a smaller value indicates lower likelihood.

This proposed method uses density-based Local Outlier Factor (LOF) to assess patent anomaly degree, enabling objective interpretation of quantitative results. This method can detect arbitrarily shaped natural clusters while filtering outliers. A point's local outlier value is obtained through the ratio of its average density to surrounding points' densities. The specific calculation involves four steps: (1) The k-distance of object p (k-distance(p)) is the Euclidean distance between p and its k-th nearest neighbor, where k is defined as the minimum distance for parameter clustering; (2) q is defined as the reachable distance of p, expressed as reachDistk(p,q), obtained through max{d(p,q), k-distance(p)}, where d(p,q) is the Euclidean distance between p and q; (3) Nk(p) is defined as the set of p's k-nearest neighbors, and density-based reachable distance is expressed as lrdk(p), as shown in Formula (1); (4) The LOF value of p for surrounding objects k is shown in Formula (2). By determining an appropriate k value, LOF values for all patents can be calculated.

The patent novelty assessment process consists of two main steps: (1) LOF calculation. Assume PSj is defined as a set of patents published in year j, where each patent is represented by a keyword vector (S11, S12, … Snm) in the morphological matrix. For patent Pi, we define the calculated LOFj(Pi) as the local outlier value of patent Pi. In this proposed method, the k value is defined as the number of patents, as is typical for most morphological structure patent data. (2) LOF standardization. While LOF values can analyze the novelty degree of patents in a given year, differences in LOF values across years for the same patent make temporal changes difficult to track. Even if a patent has the same LOF value in two different years, its novelty differs due to different patent sets. To address this, we introduce kernel density estimation to standardize LOF values. Kernel density estimation is a non-parametric method for determining probability distribution functions from discrete samples. The probability distribution function for patents published in year J is defined as fj(LOF), calculated through kernel density estimation as shown in Formula (3).

For LOFj(Pi), when i=1,…,n are sample points, n(PSj) is the number of patents, K is the Gaussian kernel function, and h is the smoothing factor. The calculated Rj(Pi) represents the relative novelty rate for each LOFj(Pi), as shown in Formula (4). A higher relative novelty rate Rj(Pi) compared to patent Pi published in year j indicates a lower LOF value, allowing us to treat Pi as an expanded value for novel patents. This enables comparison of LOF values for patents in a given year and dynamic assessment of specific patent novelty.

2.3 Constructing the Novel Patent R&D Portfolio Map

Based on identified novel patents, this paper uses two indicators—patent technology scope and similar patent count—to comprehensively analyze and identify patent technology opportunities. Current research uses patent citation frequency to assess technology opportunities, assuming higher citation frequency indicates greater technical and economic impact. However, a time lag exists between patent examination and publication. Increasing research indicates that citation frequency for earlier-published patents and new patents is constrained by time. Therefore, this study uses technology scope to replace citation frequency. Technology scope refers to the number and range of classification codes for a patent. Beyond its primary classification code, more classification codes indicate broader technology coverage and greater influence. With the development of intellectual property protection, patent infringement frequency and associated economic compensation have increased annually. Constrained by yield rates and patent value, similar patent count can indicate a patent's imitability to some degree—fewer similar patents mean higher imitability. Based on these factors, the novel patent R&D portfolio map uses technology scope and patent family count as evaluation indicators.

3 Empirical Study: Thrombolytic Drug Domain

In recent years, with population aging intensifying, the incidence of peripheral vascular disease has increased annually. In the United States, up to 5% of men and 2.5% of women over 60 suffer from intermittent claudication. Thrombolytic and antiplatelet drugs are effective medications for reducing cardiovascular and cerebrovascular disease incidence [14]. Pharmaceutical companies are focusing on developing new, efficient thrombolytic drugs, making technology opportunity identification in this field essential.

3.1 Data Sources and Methods

This study's data source is the China National Knowledge Infrastructure (CNKI) patent database, retrieving all invention and utility model patents in the thrombolytic drug field up to November 15, 2015. A total of 561 relevant patent families were obtained. After deduplication, a patent database was established using Microsoft Office Access containing entries for title, applicant, address, publication date, publication number, etc., and a patent matrix was constructed based on patent number, publication date, and abstract.

3.2 Constructing the Patent Matrix

Using text mining software TextAnalysis 2.1, 25 important keywords across 8 dimensions characterizing thrombolytic drugs were identified based on TF-IDF index. Salton index was then used to find keyword co-occurrence relationships, making keyword determination more precise and efficient.

The 8 dimensions in the thrombolytic drug field are: preparation methods, medical devices, action sites, traditional Chinese medicine compositions, pharmaceutical compositions, treatment methods, gene expression, and biopharmaceuticals. The specific matrices are shown in Table 2 and Table 3 .

3.3 Identifying Novel Patents

After determining the main patent count through expert discussion, k value determination involves two steps: first calculating the majority patent count, then measuring cosine similarity between keyword vectors in the patent matrix to help experts determine the specific k value. Cosine similarity is considered the most common metric for calculating similarity between two unstructured documents, calculated as shown in Formula (5), where A and B represent keyword vectors in documents. The similarity range is defined as 0 to 1, with higher values indicating greater document similarity. Through calculation, the k value was ultimately determined to be 10. LOF values and standardized LOF values were calculated using Matlab to obtain novelty assessment values for each patent. Partial patent relative novelty degrees are shown in Table 4 .

3.4 Novel Patent R&D Portfolio Map

The top 5% patent R&D portfolio map is shown in Figure 1 [Figure 1: see original paper]. Based on patent matrix information, patents are represented by circles, with circle size indicating patent importance. The classification lines on horizontal and vertical axes represent standardized averages of patent technology scope and similar patent count, respectively. Patent application quantity also determines influence magnitude, while patent value and potential value require investigation. Using this constructed method, patents CN102311396A and CN102210666A were identified as technology opportunities in the thrombolytic drug field. Pyrazine derivatives have antioxidant and thrombolytic effects for preparing new drugs to treat cardiovascular and cerebrovascular diseases and degenerative aging diseases caused by excessive free radicals or thrombosis. Salvianolic acid A can be used to prepare drugs that increase cAMP content in blood or plasma, prepare drugs that inhibit platelet phosphodiesterase activity, and particularly for preparing drugs to prevent or treat myocardial ischemia-reperfusion injury caused by thrombolysis, percutaneous coronary intervention, or coronary artery bypass grafting. Relevant R&D enterprises should focus on developing technologies related to these two patents.

The novel patent R&D portfolio map constructed in this paper evaluates a specific cross-section in time, and identified results will continuously change with technology development. If a specific technology opportunity enters a new stage, its information should be removed from the map. While the patent R&D portfolio map structure is not fixed, the initial patent matrix can be reused by simply adding information from newly produced patents.

4 Conclusion

This study proposes a method for technology opportunity identification through novel patent detection based on systematic processes and scientific methods. By identifying anomalous patents rather than keyword combinations representing patent gaps, technology opportunity identification becomes more precise. Additionally, patent novelty degree can be compared through quantitative indicators, making results more objective. Technology opportunity analysis strategy is considered an effective means of generating competitive intelligence for emerging technologies [15]. This paper proposes a technology opportunity identification method based on quantitative data and systematic processes, extending the application of density-based local outlier (LOF) from process control and fault detection to integrating and interpreting novel patent information through text mining. Unlike current research, this study uses technology scope to replace citation frequency and constructs a two-dimensional R&D portfolio map with similar patent count, making patent maps more concise and accurate.

The limitation of this proposed method is that the map does not present a definitive set of new patents but identifies patents with high novelty potential, still requiring some expert evaluation. Future improvements should gradually increase the degree of quantitative measurement and reduce subjective evaluation to further optimize the technology opportunity identification process.

References

[1] LEE C, SONG B, PARK Y. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships[J]. Technology analysis & strategic management, 2013, 25(1): 23-38.

[2] LEE C, JEON J, PARK Y. Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach[J]. Technological forecasting & social change, 2011, 78(4): 2489-2500.

[3] YOON B. Strategic visualisation tools for managing technological information[J]. Technology analysis & strategic management, 2010, 22(3): 377-397.

[4] YOON B, YOON C, PARK Y. On the development and application of a self-organizing feature map-based patent map[J]. R & D management, 2002, 32(4):

291-300.

[5] LEE S, YOON B, PARK Y. An approach to discovering new technology opportunities: keyword-based patent map approach[J]. Technovation, 2009, 29(6/7): 481-497.

[6] LEE C, SON C, YOON B, et al. An instrument for discovering new mobile service opportunities[J]. International journal of mobile communications, 2013, 11(4): 374-392.

[7] LEE C, SON C, YOON B, et al. An instrument for discovering new mobile service opportunities[J]. International journal of mobile communications, 2013, 11(4): 374-392.

[8] 李靖华, 郭耀煌. 主成分分析用于多指标评价的方法研究——主成分评价 [J]. 管理工程学报, 2002, 16(1): 39-43.

[9] 余健, 郭平. 自组织映射 (SOM) 聚类算法的研究 [J]. 现代计算机 (专业版), 2007(3): 7-8.

[10] SON C, SUH Y, JEON J, et al. Development of a GTM-based patent map for identifying patent vacuums[J]. Expert systems with applications, 2012, 39(39): 11850-11859.

[11] LEE C, SON C, YOON B, et al. An instrument for discovering new mobile service opportunities[J]. International journal of mobile communications, 2013, 11(4): 374-392.

[12] REITZIG M. What determines patent value?[J]. Research policy, 2003, 32(1): 13-26.

[13] CUNNINGHAM S W. Analysis for radical design[J]. Technological forecasting & social change, 2009, 76(9): 1155-1164.

[14] 刘健. 口服抗血小板药物在冠心病治疗中的应用 [J]. 中国循环杂志, 2012, 27(4): 103-104.

[15] PORTER A L, JIN X Y, GILMOUR J E. Technology opportunities analysis: integrating technology monitoring, forecasting, and assessment with strategic planning[J]. SRA journal, 1994, 26(2): 21-31.

Author Contributions

Tao Chenglin: Determined research topic, methodology, and research field; processed data; summarized conclusions and wrote the paper.
Wang Wei: Determined research direction and structure; revised and finalized the paper.
Zhang Shiyu: Participated in determining research topic and field; verified data; revised the paper.

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

Source: ChinaXiv — Machine translation. Verify with original.