

Patent-Based Recommendation Method for Potential Enterprise R&D Partners: Postprint

Authors: Zhai Dongsheng, Guo Cheng, Zhang Jie, Xia Jun

Date: 2017-11-08T00:00:00+00:00

Abstract

[Objective] To address the challenge of accurately identifying potential R&D partners, this study proposes a patent-based recommendation method for potential enterprise R&D partners. **[Method]** Based on TRIZ theory, semantic features including function, scientific effect, and efficacy are extracted from relevant patents to construct a domain-specific patent technology tree. Key information from enterprise technology requirements is extracted and matched onto the technology tree. Potential R&D partners are identified based on patent holders and evaluated using the Analytic Hierarchy Process (AHP). **[Results]** Derwent patent data in the water heater anti-scale technology domain is collected to identify and evaluate potential R&D partners, demonstrating the feasibility of the proposed method. **[Limitations]** Regarding semantic feature extraction, due to the considerable flexibility of Chinese grammatical structures, the extraction accuracy remains to be improved. **[Conclusion]** This method can discover and evaluate potential R&D partners, recommending to enterprises R&D partners capable of addressing their technological needs.

Full Text

Recommending Potential R&D Partners Based on Patents

Authors: Zhai Dongsheng, Guo Cheng, Zhang Jie, Xia Jun

Affiliation: School of Economics and Management, Beijing University of Technology, Beijing 100124, China

Abstract

[Objective] This study proposes a recommendation method for identifying potential R&D partners based on patent data. **[Methods]** Grounded in TRIZ theory, we extract semantic features—including functions, scientific effects, and efficacy—from relevant patents to construct a domain-specific patent technology

tree. Key information from enterprise technology requirements is then extracted and matched within this tree, enabling the identification of potential R&D partners through patent holders and their subsequent evaluation using the Analytic Hierarchy Process. **[Results]** We demonstrate the feasibility of this method by retrieving and evaluating potential R&D partners in the water heater anti-scale technology domain using Derwent patent data. **[Limitations]** The accuracy of semantic feature extraction requires improvement due to the inherent flexibility of Chinese grammatical structures. **[Conclusions]** This method can effectively discover and evaluate potential R&D partners, providing enterprises with recommendations for partners capable of addressing their specific technical needs.

Keywords: Patent; Technology Tree; TRIZ; R&D Partners

Introduction

In today's competitive markets, it is increasingly untenable for enterprises to deliver products and services relying solely on their own capabilities. Consequently, collaborative R&D has become critically important for business success, and the selection of R&D partners has emerged as a key strategic issue that directly impacts the outcome of joint development efforts. Patent documents, as humanity's primary vehicle for technological knowledge, represent highly innovative outputs that embody collective wisdom and contain cutting-edge technical information. Moreover, patents themselves constitute valuable intellectual property rights—a company's patent portfolio serves as a comprehensive indicator of its technological capabilities, making patents a crucial benchmark for assessing corporate innovation and R&D capacity [?]. Patent documents also provide detailed technical descriptions that reveal the fundamental nature of inventions and offer high practical operability [?].

Building upon this research context, this paper proposes a patent-based recommendation method that uses patent data as a technical information carrier to identify potential R&D partners for enterprises. The approach involves: (1) extracting semantic features from domain patents to construct a patent technology tree; (2) analyzing enterprise technology requirement documents to extract key information about technical domains, problems, solutions, and effects, then using this information to precisely locate relevant patents within the technology tree; and (3) developing an evaluation index system for assessing potential R&D partners based on patent holdings, ultimately identifying partners that can address specific enterprise needs.

2.1 Technology Tree Construction Methods

A technology tree is a hierarchical diagram representing relationships among product components, technologies, and technical functions within a specific do-

main. Two primary approaches exist for constructing domain technology trees.

The first approach centers on TRIZ' s (Theory of Inventive Problem Solving) technology evolution theory. This method analyzes the current state of domain technology development, selects the most relevant evolution path as a template, and adds core technical features of existing technologies as branches to the tree structure. By mapping domain-specific characteristics onto the main evolution path, this approach enables the understanding and forecasting of technology development directions.

The second approach extracts technical and functional information to build the tree based on the logical structure of the original data. Yoon et al. [?] proposed a systematic technology roadmapping method using quantitative data, employing text mining to extract core information from product designs and patents to identify existing products and technologies, thereby generating detailed product and technology configurations to support innovation. Lee et al. [?] developed a technology roadmap construction method based on literature and qualitative analysis, building a technology tree for broadband network technology from a technical perspective. Cascini et al. [?] utilized a computer-aided patent portfolio analysis system to extract function-related semantic information from patents, calculated patent similarity by comparing invention function trees, and constructed a technology tree for circuit interruption devices from a functional perspective. Fantoni et al. [?] adopted a Function-Behavior-Structure (F-B-S) framework, using text mining combined with keywords, concepts, relationships, and regular expressions to extract functions, behaviors, and structures from patents for technology tree construction. Choi et al. [?] proposed an SAO (Subject-Action-Object) extraction-based method for building technology trees, using K-means clustering to group SAO structures and classify them into products, technologies, materials, and technical attributes, then combining three dimensions of the technology tree to construct the domain-specific tree. Russo et al. [?] introduced a functional search method based on TRIZ' s Function-Behavior-Scientific Effect-Structure ontology, defining functions as the purpose of technical systems and building a local knowledge base and domain technology tree by extracting key information from patents.

Wang Chaoxia et al. [?] conceptualized a patent solution (PS) as comprising Solution Components (SC) and Solution Relationships (SR), further subdividing components into Technical Systems (TS), System Flows (FL), and Technical Attributes (TA), and relationships into Composition Relationships (RS), Attribute Relationships (RA), and Functional Relationships (RF), using pattern matching to extract information and form hierarchical relationships among elements.

In summary, the first construction method relies heavily on the builder' s knowledge and experience, making it difficult to expand and imposing significant limitations. Existing research on the second method primarily focuses on product, technology, and functional perspectives, while neglecting the functional effects dimension. Therefore, this paper adopts the second approach, using Derwent patent data as the source to extract semantic information about functions, sci-

entific effects, and functional effects from patent titles and abstracts, and constructs the domain technology tree using patent indices combined with syntactic structures.

2.2 Research on R&D Partner Selection

Enterprises must establish appropriate criteria and indicator systems when selecting R&D partners, as different criteria reflect different corporate needs. International scholars have established a foundation for understanding factors influencing R&D partner selection. Cantner et al. [?] reviewed relevant research on partner selection factors, drawing on the resource-based view of the firm and using German patent data to verify the positive effects of technological overlap, potential knowledge flow, future collaboration value, and prior cooperation on partner selection. Lhuillery et al. [?] used French Innovation Survey (CIS) data to explore factors contributing to failure in collaborative R&D projects from four perspectives: firm type, prior cooperation experience, knowledge spillovers, and intellectual property rights. Chun et al. [?] applied a probit model to explore factors influencing R&D cooperation among small and medium-sized enterprises (SMEs), finding that internal knowledge spillovers positively and significantly affect SME participation in R&D cooperation.

Domestic scholars have also contributed to this research area. Wang Jinfu et al. [?] applied 3C theory (Capability, Compatibility, Commitment) to evaluate R&D partner selection factors, subdividing them into 9 aspects and 35 influencing factors, which were validated through questionnaire-based statistical analysis. Ji Huisheng et al. [?] argued that knowledge characteristics are primary factors in partner selection, with knowledge mismatches and inefficient knowledge sharing being major causes of collaborative R&D failure, and constructed a knowledge characteristic index system using the Analytic Hierarchy Process. Yuan Xiaodong et al. [?] proposed a patent information analysis method for selecting R&D partners, dividing the analysis into technical and entity perspectives, examining technology development trends, maturity, and core patents from the technical angle, and comparing R&D qualifications, relative patent positions, and advantages from the entity perspective. Su Huishuang et al. [?] identified organizational size, R&D intensity, complementary resources, organizational openness to external environments, mutual trust, cultural compatibility, and cooperation experience as important factors influencing partner selection.

Various methods have been applied to partner selection, including the Analytic Hierarchy Process and goal programming techniques. Song et al. [?] proposed a patent portfolio-based approach for evaluating potential R&D partners, overcoming the limitation of focusing solely on individual firm capabilities. Lee et al. [?] developed a method for large companies to select SMEs as R&D partners, using technology roadmaps as essential elements, semantic analysis to generate suitable candidate SMEs, and Bayesian networks based on patent information

to finalize partner selection.

In summary, while international scholars have begun using patent data to identify and evaluate R&D partners, domestic research in this area remains limited. This paper addresses this gap by employing Derwent patent data to match domain patent technology trees with enterprise requirement documents, thereby facilitating more effective partner selection.

3 Design of Patent-Based Potential R&D Partner Recommendation Method

This study utilizes TRIZ theory and text mining methods to extract semantic features—including functions, scientific effects, technical attributes, and functional effects—from relevant patents. Technical attribute extraction assists in scientific effect extraction, and domain patent technology trees are constructed based on co-occurrence relationships among these semantic features. Enterprise requirement documents are then analyzed to extract technical domains, problems, solutions, and effects. By matching this key information within the domain technology tree, we precisely locate target patents that align with enterprise needs, forming a local requirement patent database. The enterprise requirement patent acquisition scheme is illustrated in [Figure 1: see original paper].

3.1 Patent Semantic Feature Extraction Scheme

We extract three types of semantic features from patent text information (titles and abstracts): functions, scientific effects, and functional effects. Patents in the Derwent database are professionally rewritten, making their titles and abstracts more standardized. The abstracts are formatted into distinct sections: the NOV section describes core innovations, including the most prominent improvements over prior art and the scientific effects employed; the ADV section primarily introduces functional effects achieved (e.g., more environmentally friendly, safer, more convenient) and sometimes includes functional descriptions; the USE section describes potential application domains and related technologies.

Based on these characteristics, we designed a semantic feature extraction scheme that extracts functions primarily from titles and ADV sections, scientific effects from NOV sections, and functional effects from ADV sections. The extraction scheme is shown in [Figure 2: see original paper].

3.2 Semantic Feature Annotation and Extraction

Before extraction, patent texts undergo preprocessing: segmentation into paragraphs, sentences, and words; word frequency statistics; part-of-speech annotation based on word characteristics, context, and text structure; and removal of

irrelevant words using stopword lists. The extraction process comprises three components:

(1) Function Extraction

Function represents the abstract description of input-output management for design intentions in specific environments—in essence, the value and specific role of an object. Function expressions follow three primary patterns: (1) verb+noun combinations (e.g., increase temperature, reduce space); (2) input-output transformations involving energy, material, or information; and (3) transformations between behaviors or states (e.g., converting liquid water to steam). The first pattern is most common in patents. Therefore, we extract functions from patent titles and ADV sections using the first pattern as the primary rule, supplemented by the second pattern through verifying whether nouns following verbs represent energy, material, or information.

(2) Scientific Effect Extraction

Scientific effects describe the transformation process between system inputs and outputs, governed by scientific principles and system attributes with accompanying phenomena. We compiled a local scientific effects knowledge base using 100 common effects, categorizing them into mechanical, thermal, acoustic, electromagnetic, chemical, and biological classes, with sub-classifications based on specific effect descriptions. Using NOV sections as source data, we identify characteristic words from this knowledge base to classify patents by scientific effects. A single patent may correspond to multiple scientific effects.

(3) Functional Effect Extraction

Functional effects characterize features of functions, serving as modifiers. When patents achieve specific functions through scientific effects, they simultaneously produce related functional effects (e.g., improved safety, enhanced reliability, speed). Functional effects appear in two forms: (1) adverbs modifying functional verbs (e.g., quickly sterilize, safely disinfect); and (2) verb+noun combinations (e.g., improve safety, increase efficiency). The second pattern is more common in patents. The key difference from function expressions is that functional effect verbs are followed by nouns like safety, efficiency, or stability rather than energy, material, or information terms. We extract functional effects from ADV sections using both patterns.

3.3 Technology Tree Construction Based on Patent Semantic Features

The construction process involves three steps: (1) Determining internal hierarchical structures of extracted semantic features and merging synonymous concepts to form independent knowledge blocks; (2) Establishing connections between “function” and “scientific effect” concepts, and between “scientific effect” and “functional effect” concepts based on their co-occurrence relationships within individual patents; (3) Adding bibliographic data (patent numbers, patentees, citation counts) to leaf nodes, enabling comprehensive representation from bibliographic to semantic features. The final technology tree structure is shown in

[Figure 3: see original paper].

3.4 Semantic Information Extraction from Enterprise Requirements

Enterprise requirements are structured as forms with fields including title, requirement description, background, potential technical directions, and excluded technical solutions. We extract information from these fields: titles, descriptions, and backgrounds are matched to technical effects and problems using verb-object structures; potential directions and excluded solutions serve directly as technical means fields. Technical problems represent urgent issues requiring solutions—equivalent to functions that future technologies should achieve. In Chinese, function descriptions typically follow a (adverb)+verb+(adjective)+noun pattern (e.g., quickly eliminate scale, safely provide oxygen). We compile functional verbs from requirement documents as shown in . Technical effects are identified through adverbs modifying functional verbs or effect-expressing verb-object structures.

3.5 Matching Algorithm for Required Patents in Domain Technology Tree

The domain patent technology tree comprehensively displays all functions achieved by scientific effects and their corresponding functional effects, with patent numbers indexing which patents apply which effects and achieve which functions. The matching process proceeds as follows: first, technical problems extracted from requirements are matched with function fields in the technology tree to filter patents meeting basic functional needs; second, technical means and effects from requirements are matched with scientific effects and functional effects in the tree to further refine the patent set. The matching method is illustrated in [Figure 4: see original paper].

4 Evaluation of Potential R&D Partners Based on Patents

Enterprises must consider multiple factors when seeking R&D partners: technical characteristics (technical strength, R&D capability), openness characteristics (R&D openness, cooperation history with the enterprise), and collaborative effectiveness (evaluation of past joint R&D outcomes). This paper proposes a patent-based evaluation framework comprising three main categories: technical strength, R&D openness, and collaborative R&D effectiveness, as shown in [Figure 5: see original paper].

Technical Strength represents a partner's capabilities in the cooperation domain, including technical and operational knowledge and experience. It is assessed through four sub-criteria: technology share, technology leadership, technology impact index, and patent family impact index. **R&D Openness** is measured through organizational openness and joint ownership. **Expected**

Collaborative R&D Effectiveness includes collaborative patent citation index and collaborative patent family index. The calculation methods for each sub-indicator are detailed in .

5 Empirical Study

5.1 Data Collection

This study examines the water heater anti-scale technology domain, involving keywords such as water heater, scale, anti-scale, and scale removal. The search query was formulated as: $TS = (((scale^* \text{ and } (prevention^* \text{ or } prevent^* \text{ or } inhibit^* \text{ or } control^* \text{ or } avoid^* \text{ or } removal^* \text{ or } reduc)) \text{ or } antiscale \text{ or } scaleinhibition \text{ or } anti-scale) \text{ and } water \text{ and } heater^*)$, covering the period from January 1, 1963, to [end date].

5.2 Patent Semantic Feature Extraction

We annotated and extracted three semantic features: patent functions, applied scientific effects, and achieved functional effects. Functions were extracted from titles and ADV sections, scientific effects from NOV sections, and functional effects from ADV sections. Example extraction results are shown in .

The extracted functions primarily focused on preventing or inhibiting scale formation, with a minority addressing scale removal. The scientific effects were consolidated into 11 categories: Magnetic field, Ultrasound, Electric field, Vibration, Gas/steam, Phase change, Circulation flow, Control temperature, Control pressure, Functional ceramics, and Scale inhibitor (chemical). Vibration could be applied to both scale inhibition and removal, while other effects like magnetism, ceramics, temperature control, and chemical inhibitors primarily achieved functional effects such as enhanced safety, speed, convenience, and performance.

5.3 Domain Patent Technology Tree Construction

Using patents as indices, we established connections among semantic features based on their co-occurrence and contextual relationships within patents. After merging synonymous concepts from the water heater anti-scale domain, we constructed the technology tree. Since semantic features originate from patent text and each patent carries bibliographic information (patentee, application date, jurisdictions), we integrated these data into the tree's leaf nodes, creating a comprehensive multi-dimensional representation. A partial view of the enhanced tree is shown in [Figure 7: see original paper], which displays both semantic information (e.g., patent JP2015021723-A uses magnetic effects to inhibit scale formation with improved safety) and bibliographic details (Japanese patent, published 2015).

5.5 Identification and Evaluation of Potential R&D Partners

Company A served as the case study enterprise. Its requirements for water heater anti-scale technology are partially listed in . Using the matching algorithm, we aligned extracted semantic features from the requirements with the domain technology tree to identify relevant patents.

The requirement document' s semantic features were extracted as follows: technical domain (water heater, personal health, purification), technical problems (scale formation, dirt deposition, bacterial growth, inability to self-clean), technical means (ultrasound, magnetic methods), and technical effects (convenience, performance, speed, cost), as summarized in .

The matching process first filtered 1,219 patents addressing the identified technical problems by matching them with function fields in the technology tree. Subsequent matching with technical effects (convenience, performance, speed) refined the set to 932 patents. Analyzing the scientific effects in these patents revealed all 11 previously identified effects.

Applying TRIZ' s eighth evolution law (evolution toward micro-level and field applications) and considering the temporal distribution of scientific effects in the domain, we identified advanced technologies including Magnetic field, Ultrasound, Electric field, and Control temperature. The 612 patents applying these effects were designated as advanced patents, and their patentees were identified as potential R&D partners.

We evaluated the top five patent holders using the proposed indicator system. Panasonic' s scores are shown in , and normalized results for all five candidates are presented in . The composite evaluation values, calculated by summing normalized scores, yielded the final ranking shown in .

Conclusion

This study extracts semantic features (functions, scientific effects, functional effects) from patents in specific technology domains to construct patent technology trees. For enterprise requirements, we extract key information about technical domains, problems, means, and effects. By mapping scientific effects to TRIZ evolution laws, we identify advanced technologies and precisely locate relevant patents to find potential partners. We also develop an evaluation index system assessing partners from three dimensions: technical strength, R&D openness, and collaborative effectiveness.

Limitations include the need for improved extraction pattern accuracy due to Chinese language flexibility. Future research should define more accurate patterns based on deeper analysis of Chinese documents to enhance information extraction precision.

References

- [1] Bao Changhuo, Xie Xinzhou. Competitive Intelligence and Enterprise Competitive Power[M]. Beijing: Huaxia Publishing House, 2001: 137-145.
- [2] Xie Wei. Research on China' s Patent Output[D]. Cheng Du: University of Electronic Science and Technology of China, 2005.
- [3] Yoon B, Phaal R, Probert D. Morphology Analysis for Technology Roadmapping: Application of Text Mining[J]. R&d Management, 2008, 38(1): 51-68.
- [4] Lee S, Kang S, Park E, et al. Applying Technology Roadmaps in Project Selection and Planning[J]. International Journal of Quality & Reliability Management, 2008, 25(1): 39-51.
- [5] Cascini G, Zini M. Measuring Patent Similarity by Comparing Inventions Functional Trees[A]//Computer-aided Innovation (CAI)[M]. New York: Springer US, 2008: 31-42.
- [6] Fantoni G, Apreda R, Dell' Orletta F, et al. Automatic Extraction of Function-Behaviour-State Information from Patents[J]. Advanced Engineering Informatics, 2013, 27(3): 317-330.
- [7] Choi S, Park H, Kang D, et al. An SAO-based Text Mining Approach to Building a Technology Tree for Technology Planning[J]. Expert Systems with Applications An International Journal, 2012, 39(13): 11443-11455.
- [8] Russo D, Montecchi T, Liu Y. Functional-based Search for Patent Technology Transfer[C]//Proceedings of the 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Chicago: American Society of Mechanical Engineers, 2012: 295-304.
- [9] Wang Zhaoxia, Qiu Qingying, Feng Peien, et al. Information Extraction Method of Technical Solution from Mechanical Product Patent[J]. Journal of Mechanical Engineering, 2009, 45(10): 198-206.
- [10] Cantner U, Meder A. Technological Proximity and the Choice of Cooperation Partner[J]. Journal of Economic Interaction and Coordination, 2007, 2(1): 45-65.
- [11] Lhuillery S, Pfister E. R&D Cooperation and Failures in Innovation Projects: Empirical Evidence from French CIS Data[J]. Research Policy, 2009, 38(1): 45-57.
- [12] Chun H, Mun S B. Determinants of R&D Cooperation in Small and Medium-sized Enterprises[J]. Small Business Economics, 2012, 39(2): 419-436.
- [13] Wang Jinfu, Wei Zhen, Liu Jiangnan, et al. Research on Influencing Factors of the IUR Strategic Alliance's R&D Partner Selection—Based on the Perspective of 3C Theory[J]. Forecasting. 2014, 32(4): 74-80.

- [14] Ji Huisheng, Wang Hongwei, Lu Qiang. R&D Partner Selection of Enterprise Based on Knowledge Characteristics[J]. Journal of Shenyang University of Technology: Social Science Edition, 2011, 4(2): 137-140.
- [15] Yuan Xiaodong, Chen Jing. The Application of Patent Information Analysis in the Choice of Cooperative Technological Innovation Partners[J]. Intelligence Journal, 2011, 30(8): 22-27.
- [16] Su Huishuang, Lan Yanfei, Yi Lanwen. Review on Selecting Enterprises R&D Partners: Based on Perspective of Influence Factors[J]. Modern Management Science, 2013 (6): 48-50.
- [17] 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.
- [18] Lee K, Yoon B. A Method for Partner Selection in R&D Collaboration Between Large Companies and SMEs Using Patent Information[C]//Proceedings of the 2013 Technology Management in the IT-Driven Services (PICMET). 2013: 1385-1392.
- [19] Zhai Dongsheng, Xia Jun, Zhang Jie, et al. Research on Construction of Technology Tree Based on Patent Feature Extraction[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(7): 717-724.

Author Contributions: Zhai Dongsheng proposed the research topic and framework, and revised the final manuscript. Zhai Dongsheng, Guo Cheng, and Zhang Jie designed the research methodology. Guo Cheng and Xia Jun processed data, implemented the recommendation method, conducted experiments, and drafted the manuscript.

Conflict of Interest Statement: All authors declare no conflicts of interest.

Supporting Data: See journal website <http://www.infotech.ac.cn>.

[1] Zhai Dongsheng, Guo Cheng. Patent1909.xlsx. Water heater anti-scale technology patent data.

Received: August 29, 2016

Revised: October 27, 2016

News Item: HighWire Partners with Hypothesis to Integrate Annotation in Publications

HighWire Press and Hypothesis recently announced a partnership to add open annotation capabilities to over 3,000 journals, books, reference works, and conference proceedings on the HighWire JCore platform.

Publishers on the JCore platform can implement and control annotation levels and visibility in their publications. Annotation is a fundamental activity for researchers and scholars. To date, journal annotation solutions have been

severely limited, restricting their utility. This is changing with the emergence of standards-based, open-source, interoperable annotation paradigms.

“HighWire, together with its partners, has created an innovative, open platform solution that improves research conditions and delivers more value to readers. We have been supporting the Annotating All Knowledge Coalition since 2015,” said Dan Filby, CEO of HighWire Press. “Open annotation integrated with JCore supports faster research workflows and more efficient communication. We look forward to working with publishers who want to use Hypothesis to improve the interactivity of their content and workflows.”

Hypothesis is a non-profit annotation technology organization founded in 2011 that works with publishers, educators, researchers, and journalists to enable annotation on the internet. Use cases in research environments include pre-publication peer review, post-publication annotation and community review, author annotations on their own works (including updates to previous articles, invited discussions, enhanced footnotes, corrections, and errata). Over 70 major publishers, platforms, and technology organizations have joined forces to support the interoperable vision of the Annotating All Knowledge Coalition.

(Compiled from: <http://home.highwire.org/sites/default/files/HighWire%20Hypothesis%20PR%20AM%20Correspondence>)
(Journal correspondence)

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

Source: ChinaXiv – Machine translation. Verify with original.