Technology Opportunity Discovery through Integrated Review Topic Identification and Multi-dimensional Technical Attribute Analysis: Postprint

Authors: Wu Yiping, Bai Rujiang, Liu Mingyue, Wang Xiaoyue

Date: 2023-04-01T16:02:49+00:00

Abstract

[Purpose/Significance] This study proposes a technical opportunity discovery method that integrates review topic identification with multi-dimensional analysis of technical attributes. From a technology demand-driven perspective, it identifies technical opportunities and provides decision-making support for enterprises to proactively plan R&D directions and scientific research management. [Method/Process] Using product online reviews as the research data source, the method first employs the LDA topic model to identify technical topics from reviews, and proposes two indicators—technical review topic intensity and topic novelty—to screen emerging key technical review topics. Subsequently, technical attribute terms are manually selected from academic papers and technical patents, high-frequency review terms are obtained through TF-IDF calculations, and combined with expert knowledge to further screen technical feature terms, thereby constructing a product technical attribute term-technical feature term table. Through correlation calculations, technical attributes related to reviews and those related to emerging key technical review topics are respectively identified. Finally, an identification indicator model for product important technical attributes is proposed and a multi-dimensional analysis method is designed to analyze the characteristics of these attributes, ultimately identifying emerging technical opportunities embedded in review texts. [Results/Conclusion] Experimental results demonstrate that the method can effectively identify technical opportunities, providing valuable reference for enterprise product technology R&D management.

Full Text

Preamble

Volume 65, Issue 10, May 2021
ChinaXiv Collaborative Journal

Research on Technology Opportunity Discovery Integrating Comment Topic Identification and Multi-Dimensional Analysis of Technical Attributes

Wu Yiping, Bai Rujiang, Liu Mingyue, Wang Xiaoyue
Institute of Information Management Research, Shandong University of Technology, Zibo 255049

Abstract:
[Purpose/Significance] This paper proposes a technology opportunity discovery method that integrates comment topic identification with multi-dimensional analysis of technical attributes. From a technology demand-driven perspective, this approach identifies technology opportunities to provide decision-making support for enterprises' forward-looking R&D direction planning and scientific research management. [Method/Process] Using online product reviews as the data source, we first employ the LDA topic model to identify technical topics in reviews, proposing two indicators—technical comment topic intensity and topic novelty—to screen for emerging key technical comment topics. Next, technical attribute words are manually selected from academic papers and patents, while high-frequency comment words are obtained through TF-IDF value calculation. Combined with expert knowledge, technical feature words are further screened to construct a product technical attribute word–technical feature word table. Through correlation calculation, we obtain technical attributes related to comments and those related to emerging key technical comment topics, respectively. Finally, we propose an identification index model for important product technical attributes and design a multi-dimensional analysis method to analyze the characteristics of these attributes, ultimately identifying emerging technology opportunities embedded in comment text. [Result/Conclusion] Experimental results demonstrate that this method can effectively identify technology opportunities, providing a reference for enterprise product technology R&D management.

Keywords: technology opportunity discovery; technical attribute analysis; topic identification; comment mining
Classification Number: G251
DOI: 10.13266/j.issn.0252-3116.2021.10.007

Currently, Chinese enterprises face severe challenges from global technological revolutions and scientific competition. Advancing forward-looking technology opportunity discovery research helps enterprises seize first-mover advantages in future market competition and provides crucial intelligence for supporting major scientific and technological innovation decisions and formulating core R&D

strategies. Researchers typically use academic papers and technology patents to identify and monitor emerging technology trends from a technology-driven perspective, yet rarely utilize social media data related to emerging technologies for technology opportunity discovery research []. With the vigorous development of e-commerce and the widespread adoption of online shopping, large volumes of easily accessible and content-rich online product review data have emerged on e-commerce platforms, which hold significant research value for obtaining technical demand feedback information. This study proposes a forward-looking technology opportunity discovery method that integrates comment topic identification with multi-dimensional analysis of technical attributes, shifting the focus of technology opportunity discovery research to the technology demand side. By effectively extracting and analyzing technical demand information in review data, this approach directly and accurately identifies forward-looking technology opportunities.

2. Related Research Progress

Technology opportunities are key influencing factors and important decision-making references for promoting technological innovation activities, and the ability to identify them constitutes one of the most important core R&D competencies for nations and enterprises []. In 1995, Professor A. Porter from the Georgia Institute of Technology formally introduced the concept of "technology opportunities" in *Technology Opportunities Analysis*, referring to the inference of likely emerging technological forms or development points in a field through mining existing technologies' horizontal comparisons among competitive entities and vertical development trends over time [], pioneering the field of technology opportunity discovery research. Li Baoming explained technology opportunities from both technical and economic perspectives, viewing them as opportunities for technological progress and the possibility of successfully applying new technologies provided by enterprises (or society) to production []. As academic research on technology opportunity discovery continues to deepen, its concepts, data sources, and research methods are constantly enriched.

2.1 The Conceptual Connotation of Technology Opportunities Continuously Enriched in Information Science and Enterprise Management Research

Li Baoming [] divides technology opportunities into intensional and extensional categories. Intensional technology opportunities refer to the possibility of improving existing technical specifications or performance, while extensional technology opportunities refer to the possibility of transferring a specific technology to many other technical systems, where it may prove more effective in many functions than currently applied systems after transfer. Chen Zhenhong et al. [] consider technology opportunities as entrepreneurial opportunities brought about by technological changes. Kang Yuhang [] views technology opportunities as the discovery of the latest technological trends and inference of possible technological forms or development points in a technical field through mining existing technology development trends and interrelationships. G. Cecere et al. [] regard technology opportunities as core elements in enterprises' continuous innovation processes. Technology Opportunity Discovery (TOD) can be understood as a research activity that, guided by relevant theories, methods, and technologies, actively mines data sources containing technology opportunity information to discover potential opportunities for technological innovation development. Technology opportunity discovery research helps grasp the latest technological trends and provides intelligence references for national macro-level science and technology decision-making and enterprise technology R&D innovation management.

2.2 Data Sources for Technology Opportunity Discovery Research Primarily Consist of Papers and Patents

A. Porter proposed that technology opportunity discovery is closely related to existing technologies and has complex interactive mechanisms []. F. Malerba et al. [] suggested that patent data, scientific journals, and technical reports reflect the latest technical information and dynamic scientific and technological resources, providing possibilities for mining potential technological innovation opportunities and offering reference data sources for technology opportunity discovery research. Li Xin and Huang Lucheng et al. [] noted that researchers typically use academic papers and patent data to identify and monitor emerging technology trends from a technical perspective. Scientific papers are important carriers of basic science, while technology patents are important carriers of technical information. Therefore, current technology opportunity discovery research still primarily relies on scientific papers and technology patents as data sources.

2.3 Research Methods for Technology Opportunity Discovery Continuously Enriched with Era Development

Traditional technology opportunity discovery research mainly relied on expert knowledge-based methods, which could ensure relatively high efficiency and accuracy within limited scopes or segmented technical fields. The vigorous development of machine learning technologies in the big data era has provided ample methodological support for technology opportunity discovery research. As market environments change dramatically and technology innovation cycles shorten, technology opportunity discovery research methods have also been continuously enriched in empirical studies, including but not limited to comprehensive applications of expert knowledge-based methods [], bibliometrics-based methods [], text mining-based methods [], and social network analysis-based methods []. For example, M.Y. Wang et al. [] proposed that differences between scientific and technological knowledge hold potential for mining technology opportunities, thus using patent text mining combined with clustering algorithms to analyze gaps between scientific and technological knowledge for discovering potential opportunities, conducting empirical research in the microalgae biofuel field. Li Xin, Huang Lucheng et al. [] used bibliometric methods to statistically analyze technological hotspots, frontiers, opportunities, and development trends in the dye-sensitized solar photovoltaic technology industry, constructing an analytical framework for emerging technology industry future development based on bibliometrics, patent analysis, and technology roadmapping methods to objectively identify emerging technology opportunities through statistical data. Wang Jing'an [] identified technology opportunities by comparing keyword clustering network graphs, author clustering network graphs, institution clustering network graphs, and keyword clustering timeline network graphs generated from scientific papers and patent literature in the IoT technology research field, revealing future development trends in the IoT industry.

Although current technology opportunity discovery research has achieved rich results, several issues remain. Regarding data sources, on the one hand, current research primarily uses papers and patents as information carriers of existing technologies. The results often suffer from certain time lags, potentially lagging behind the latest trends in technical fields and failing to meet research needs for obtaining the most forward-looking technology opportunity discovery results. Product review data is easily accessible, updated in real-time, and directly and objectively reflects users' technical demands and perceptions of products, representing valuable scientific and technological data for technology opportunity discovery research. Mining technical demand feedback from user product reviews can more directly and prospectively discover future technology opportunities. However, current research has not paid sufficient attention to online product reviews, leading to certain limitations in combining technology opportunity discovery results with users' direct technical demands. The mechanism of technology opportunity discovery research from a user technology demand-driven perspective requires further exploration.

Regarding research methods, the technology opportunity discovery method system based on expert knowledge, bibliometrics, text mining, and social network analysis is relatively well-developed. However, the existing implementation technologies and research methods still require sufficient empirical research in terms of identification efficiency, result accuracy, and algorithm applicability for technology opportunity discovery based on product online review data. Current technology opportunity discovery research processes have not fully integrated temporal, brand, model, and other factors for in-depth analysis, thus results often fail to comprehensively reflect characteristics such as time sensitivity of technology opportunities and differences among brand models. Furthermore, theoretical method models and empirical research results for product review-driven technology opportunity discovery remain relatively limited overall, with identification methods being relatively rough and general. More scientifically rigorous research methods and technical means are needed to combine quantitative and content analysis for in-depth technology opportunity discovery from a user technology demand-driven perspective.

3. Research Approach

Based on online smartphone product reviews from e-commerce platforms, academic papers, technical reports, and expert knowledge, this study identifies forward-looking technology opportunities by analyzing user technical demand feedback in review data. The research framework is illustrated in Figure 1 [Figure 1: see original paper]. The approach consists of five main components: (1) using topic models to identify emerging key technical topics in review text; (2) constructing technical attribute word–technical feature word tables; (3) identifying technical attributes related to emerging key technical comment topics based on the first two components; (4) identifying technical attributes related to specific products under key technical topics; and (5) proposing a multi-dimensional analysis model for important product technical attributes to ultimately identify technology opportunities embedded in comment text.

3.1 Identification of Emerging Key Technical Comment Topics

3.1.1 Technical Comment Topic Identification Based on the PLDA Topic Model This paper designs a product online review topic identification method based on the Parallel Latent Dirichlet Allocation (PLDA) model to segment review data temporally and identify topics, aiming to study the technical content in reviews across different periods. The PLDA topic model uses Gibbs sampling for parameter estimation, improving algorithm efficiency and parallel speedup ratio, enabling efficient and accurate identification of topics and topic words in review text [].

3.1.2 Identification of Emerging Key Technical Comment Topics Considering that technology opportunities to be identified should have certain timeliness and user demand characteristics, this paper designs an emerging key technical comment topic identification model comprising two indicators: technical comment topic intensity and topic novelty. Through calculation using this indicator model, topics with relatively high intensity and high novelty are precisely identified as emerging key technical comment topics.

(1) Technical Comment Topic Intensity Indicator. This indicator intuitively reflects the attention and participation level of topics. It is defined as the weight of the number of comments within each topic relative to the total number of comments in that period, calculated as:

$$TI = \frac{X_i}{\sum_{j=1}^{n} X_j}$$

where TI represents the technical comment topic intensity for a certain period; X_i represents the number of comments under a certain technical comment topic identified in that period; and $\sum_{j=1}^{n} X_j$ represents the total number of comments under all technical comment topics in that period.

To determine which topic a comment belongs to and the number of comments under each topic, this study uses cosine similarity calculation to obtain the correlation between comments and topics.

First, a Vector Space Model (VSM) is constructed to describe topics and comments as vectors. In VSM, $Comment$ represents a comment, $Topic$ represents a topic, L represents comment information words or topic words, and w represents the weight of topic words or attribute feature words. A topic vector can be represented by topic words as $Topic_i = \{L_1, L_2, L_3, \dots, L_m\}$; a comment vector can be represented by comment information words as $Comment_j = \{L_1, L_2, L_3, \dots, L_n\}$; the weight vector of topic words is $TopicVector = \{W_1, W_2, W_3, \dots, W_n\}$; and the weight vector of comment information words is $CommentVector = \{W_1, W_2, W_3, \dots, W_n\}$, where each topic word or technical feature word has a weight.

Second, the similarity between comments and topics is calculated, with results ranging between [0,1]. Referring to the cosine similarity calculation method [], the similarity calculation formula between comments and topics is designed as:

$$Sim(Topic_i, Comment_j) = \cos\theta = \frac{\sum_{k=1}^{n} w_k(Topic_i) \times w_k(Comment_j)}{\sqrt{\sum_{k=1}^{n} w_k^2(Topic_i)} \times \sqrt{\sum_{k=1}^{n} w_k^2(Comment_j)}}$$

where the numerator $\sum_{k=1}^{n} w_k(Topic_i) \times w_k(Comment_j)$ represents the dot product of the comment topic vector and comment information word vector; the denominator $\sqrt{\sum_{k=1}^{n} w_k^2(Topic_i)} \times \sqrt{\sum_{k=1}^{n} w_k^2(Comment_j)}$ represents the product of the magnitudes of the comment topic vector and comment information word vector; and $Sim(Topic_i, Comment_j)$ is the similarity between the comment topic vector and comment information word vector.

Set A as the collection of comments under a certain topic, TC_Sim as the similarity between a comment and a topic, and α as the threshold. If the similarity is greater than α, the comment belongs to that topic; otherwise, it does not. This is expressed as $A = \{TC\_Sim \mid TC\_Sim > \alpha, \alpha \in [0,1]\}$.

Set the threshold for technical comment topic intensity S as β. If a topic's intensity exceeds β, it indicates high topic intensity. Higher topic intensity suggests greater user attention and participation, making it a potential emerging key technical comment topic.

(2) Technical Comment Topic Novelty Indicator. This indicator reveals whether a technical comment topic's development trend over time is emerging, developing, or declining. The indicator is calculated as:

$$TN = \frac{\sum_{i=1}^{n} date_i}{total\_num}$$

where TN represents technical comment topic novelty; $total\_num$ is the total number of comments under that technical comment topic; $date_i$ is the publication date of each comment under that technical comment topic; and $\frac{\sum_{i=1}^{n} date_i}{total\_num}$ yields the average publication date of all comments under that topic, i.e., the technical comment topic novelty. A larger calculated value indicates stronger topic novelty, suggesting newer content in that comment topic.

Set the topic novelty threshold as γ. If topic novelty exceeds γ, it indicates newer layout years, making it a potential emerging key technical comment topic.

Set B as the collection of emerging key technical comment topics. Topics with technical topic intensity S above threshold β and topic novelty N above threshold γ are identified as emerging key technical comment topics belonging to collection B, expressed as $B = \{S, N \mid S > \beta, N > \gamma, \beta \in \mathbb{R}, \gamma \in \mathbb{R}\}$.

3.2 Construction of Product Technical Attribute Word–Technical Feature Word Table and Identification of Comment-Related Technical Attributes

3.2.1 Construction of Product Technical Attribute Word–Technical Feature Word Table To subsequently identify which technical attributes are contained in emerging key topics, this paper proposes a method for constructing a product technical attribute word–technical feature word table.

Step 1: Conduct text preprocessing and TF-IDF calculation on product online review data to identify high-frequency words in comment content, and further screen technical feature words combined with expert knowledge.

Step 2: Integrate academic papers, patent information, and expert reports to select authoritative technical attribute words.

Step 3: Based on the authoritative technical attribute words screened in Step 2, match them with the technical feature words selected in Step 1. For example, for the technical attribute word "sound," prioritize screening words containing "sound," "volume," "tone," "noise," "ring," etc., that can characterize "sound" from high TF-IDF words, and further select appropriate technical feature words to classify them under the "sound" technical attribute. After construction, conduct expert analysis and review to finalize the product technical attribute word–technical feature word table, with partial results shown in Table 1 .

3.2.2 Identification of Comment-Related Technical Attributes To determine which technical attribute a comment belongs to and the number of comments under each attribute, this study uses the cosine similarity calculation method (see Formula (2)) to obtain the correlation between product comments and the product technical attribute word–technical feature word table. Set C as the collection of comments under a certain attribute, AC_Sim as the similarity between an attribute and a comment, and δ as the threshold. If the similarity exceeds δ, the comment belongs to that technical attribute;

otherwise, it does not belong to any technical attribute. This is expressed as $C = \{AC\_Sim \mid AC\_Sim > \delta, \delta \in [0, 1]\}$.

3.3 Analysis of Technical Attributes Related to Emerging Key Technical Comment Topics

To determine which technical attributes emerging key technical comment topics belong to and clarify the scope of technology opportunity discovery, this study designs a cosine similarity-based calculation method for topic-technical attribute correlation to identify technical attributes related to comment topics (see Formula (2)). Set D as the collection of topics under a certain attribute, AT_Sim as the similarity between an attribute and a topic, and ε as the threshold. If the similarity exceeds ε, the topic belongs to that technical attribute; otherwise, it does not belong to any technical attribute. This is expressed as $D = \{AT\_Sim \mid AT\_Sim > \varepsilon, \varepsilon \in [0, 1]\}$.

3.4 Identification of Important Product Technical Attributes

To improve the accuracy of important product technical attribute identification results, this paper designs a cosine similarity-based calculation method for comment-attribute similarity and constructs an identification index model for important product technical attribute characteristics comprising two indicators: technical attribute intensity and technical attribute novelty. The model evaluates technical attributes and identifies important product technical attributes (including current important technical attributes and potential important technical attributes) by confirming that they contain emerging key technical comment topics.

3.4.1 Comment Technical Attribute Intensity Indicator This indicator intuitively reflects user attention and comment participation regarding technical attributes. It is defined as the weight of the number of comments within each attribute relative to the total number of comments in that period, calculated as:

$$AI = \frac{C_i}{\sum_{j=1}^{n} C_j}$$

where AI represents comment technical attribute intensity for a certain period; C_i represents the number of comments under a certain identified comment technical attribute in that period; and $\sum_{j=1}^{n} C_j$ represents the total number of comments under all comment technical attributes in that period.

Higher attribute intensity indicates greater user attention and comment participation, suggesting that the attribute may be an important product technical attribute.

Set the comment technical attribute intensity threshold as Q. If an attribute's intensity exceeds Q, it indicates high attribute intensity and may be an important technical attribute containing technology opportunities.

3.4.2 Comment Technical Attribute Novelty Indicator This indicator reflects the temporal distribution of comments on technical attributes, revealing the temporal characteristics of user attention to the technical attribute. The comment technical attribute novelty is calculated as:

$$AN = \frac{\sum_{i=1}^{n} date_i}{total\_num}$$

where AN represents comment technical attribute novelty; $total\_num$ is the total number of comments under that comment technical attribute; $date_i$ is the publication date of each comment under that technical attribute comment; and $\frac{\sum_{i=1}^{n} date_i}{total\_num}$ yields the average publication date of all comments under that attribute, i.e., the comment technical attribute novelty. Stronger attribute novelty indicates newer content in that comment attribute.

Set the attribute novelty threshold as X. If attribute novelty exceeds X, it indicates newer attribute layout years and may be an important technical attribute containing technology opportunities.

Set E as the collection of current important technical attributes of the product. Technical attributes with comment technical attribute intensity Q above threshold ζ, attribute novelty X above threshold η, and containing emerging key technical comment topics are identified as current important technical attributes of the product, belonging to collection E, expressed as $E = \{Q, X, AT\_Sim \mid Q > \zeta, X > \eta, AT\_Sim > \varepsilon, \zeta \in \mathbb{R}, \eta \in \mathbb{R}, \varepsilon \in [0,1]\}$.

Set F as the collection of potentially important technical comment topics. Technical attributes with comment technical attribute intensity Q below threshold ζ, attribute novelty X above threshold η, and containing emerging key technical comment topics are identified as potentially important technical attributes of the product, belonging to collection F, expressed as $F = \{Q, X, AT\_Sim \mid Q < \zeta, X > \eta, AT\_Sim > \varepsilon, \zeta \in \mathbb{R}, \eta \in \mathbb{R}, \varepsilon \in [0,1]\}$.

Set G as the collection of important product technical attributes, where $G = E \cup F$, $E \subseteq G$, and $F \subseteq G$.

3.5 Multi-Dimensional Analysis of Important Product Technical Attributes and Technology Opportunity Discovery

To improve the comprehensiveness and accuracy of technology opportunity discovery results, this study designs a multi-dimensional analysis method for important product technical attributes. It conducts analysis on the temporal distribution patterns of topic quantities under technical attributes, temporal evolution patterns of topic content under technical attributes, temporal distribution patterns of comment quantities under technical attributes, and temporal distribution patterns of product model comments under technical attributes. Finally, combining comprehensive research analysis results from network surveys, academic papers, technical reports, and expert knowledge, technology opportunities are discovered.

Temporal distribution pattern analysis of topic quantities under technical attributes aims to reveal user feedback and attention breadth regarding the technical attribute across different periods.

Temporal evolution pattern analysis of topic content under technical attributes aims to reveal temporal changes in user feedback content regarding the technical attribute. To calculate similarity between comment topics under attributes (see Formula (2)), with results ranging between [0,1], set topic similarity as TT_Sim and threshold as θ. If similarity exceeds θ, the two topics are similar comment topics; otherwise, they are different comment topics. Set F as the collection of similar comment topics, expressed as $F = \{TT\_Sim \mid TT\_Sim > \theta, \theta \in [0,1]\}$.

Temporal distribution pattern analysis of comment quantities under technical attributes aims to reveal temporal distribution patterns of user comment participation and attention regarding the technical attribute.

Temporal distribution pattern analysis of product model comments under technical attributes aims to reveal the distribution patterns of user technical feedback and demands for different brand models across different periods.

4. Empirical Study on Technology Opportunity Discovery Driven by Product Review Data

This study uses smartphone review data from the top 14 products on JD.com as the data source. The data source adopts the JSON format provided by the JD platform. Taking Apple iPhone 11 as an example, its URL is: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_{comment98}&productId A Python crawler program was used to collect data on April 14, 2020, obtaining 13,870 comment entries. After screening for comment usability and removing duplicates, 12,889 comment entries remained. The specific products crawled, comment quantities, and temporal distribution are shown in Table 2 .

The experimental environment used Python, Baidu Aistudio, data mining software KNIME, Excel, and other platforms and software for data processing and analysis.

4.1 PLDA Topic Identification and Emerging Key Technical Comment Topic Identification

The review data was temporally segmented into five time-series groups based on specific comment quantities in each period. PLDA topic identification was performed on each group's review data, identifying 50 comment topics with 15 keywords per topic. Keywords were automatically generated by the PLDA topic model based on probability of occurrence in topics. For statistical convenience and analysis, and combining specific comment collection conditions in each period, this study uniformly named the topics: Period I (2017-2018), Period II (January-April 2019), Period III (May-August 2019), Period IV (September-December 2019), and Period V (January-April 2020). Partial topic and topic word identification results are presented in matrix form, with partial results shown in Table 3 .

Using the designed identification index model comprising topic intensity and topic novelty indicators, emerging key technical comment topics were identified. The topic intensity threshold was set at 0.08, and the topic novelty threshold at 2019. Topics with both topic intensity and topic novelty above thresholds were identified as emerging key technical comment topics, yielding 19 such topics. Partial results are shown in Table 4 . In addition to descriptions of technical attributes and product features, emerging key technical comment topics also reflected concerns about user groups such as the elderly and seniors, as well as attention to product cost-effectiveness.

4.2 Construction of Product Technical Attribute Word–Technical Feature Word Table

High-frequency words were extracted from all comments through TF-IDF calculation. Combined with expert review, words that could characterize product technical features were screened from high-frequency words to form a product technical feature word table. Based on expert knowledge, technical attribute words were abstracted from papers, patents, reports, and other sources. A total of 13 technical attributes were screened, and technical attribute words were matched with technical feature words through expert knowledge. The constructed product technical attribute word–technical feature word table is partially shown in Table 5 .

4.3 Analysis of Technical Attributes Related to Emerging Key Technical Comment Topics

The similarity between emerging key technical comment topics and the product technical attribute word–technical feature word table was calculated (see Formula (2)) to identify technical attribute-related comments. The similarity threshold between comments and technical attributes was set at 3%. Formula (4) was used to calculate technical attribute intensity, and Formula (5) to calculate attribute novelty. The attribute intensity threshold was set at 0.08, and the attribute novelty threshold at 2019.7. Technical attributes with both attribute intensity and attribute novelty above thresholds and related to emerging key technical comment topics were identified as current important technical attributes among important product technical attributes and should receive primary attention. Technical attributes with attribute intensity below threshold, attribute novelty above threshold, and containing emerging key technical comment topics were identified as potentially important technical attributes among important product technical attributes. Identification results for important product technical attributes are shown in Table 6 and Table 7 .

The similarity threshold was set at 1%, and the Echarts visualization platform was used to present the correlation between partial emerging key technical comment topics and technical attributes in a heatmap format, as shown in Figure 2 [Figure 2: see original paper] (for illustration purposes, multiplied by 100). The horizontal axis represents identified emerging key technical comment topics, the vertical axis represents technical attributes constructed in this study, and colored blocks represent correlations between topics and attributes, with darker colors indicating stronger correlations. The same topic may be related to multiple technical attributes. Technical attributes with relatively high similarity to emerging key technical comment topics recently include TECH_8 (battery), TECH_3 (touch/NFC/smart remote control technology), TECH_7 (storage), and TECH_{10} (accessories), reflecting high recent user attention to these attributes. Technical attributes widely distributed in emerging key technical comment topics include TECH_1 (sound), TECH_8 (battery), and TECH_{12} (appearance design), reflecting broad recent user attention to these attributes where technology opportunities may exist.

4.4 Identification of Important Product Technical Attributes Based on Multi-Dimensional Indicator Model

This study uses the constructed identification index model for important product technical attribute characteristics to evaluate technical attributes and identifies two types of important product technical attributes—current important technical attributes and potentially important technical attributes—by confirming that they contain emerging key technical comment topics.

4.5 Multi-Dimensional Analysis Method for Important Product Technical Attributes and Technology Opportunity Discovery

4.5.1 Temporal Distribution Patterns of Topic Quantities Under Important Product Technical Attributes The similarity between topics and the product technical attribute word–technical feature word table was calculated using Formula (2). Since topic word extraction results are relatively concise and the constructed technical attribute–technical feature word table is relatively precise, a 1% similarity threshold was set. Topics with similarity above the threshold were considered attribute-related topics. The temporal distribution of topic quantities under important product technical attributes is shown in Figure 3 [Figure 3: see original paper].

As shown in Figure 3, during Period I, comment topics primarily involved TECH_8 (battery), followed by TECH_0 (processor/network/data transmission technology) and TECH_2 (gaming). During Period II, comment topics primarily involved TECH_3 (touch/NFC/smart remote control technology), followed by TECH_0 (processor/network/data transmission technology), TECH_1 (sound), and TECH_{12} (appearance design). During Period III, comment topics primarily involved TECH_6 (camera function) and TECH_8 (battery), as well as TECH_{12} (appearance design), followed by TECH_0 (processor/network/data transmission technology), TECH_1 (sound), and TECH_{10} (accessories). During Period IV, comment topics primarily involved TECH_1 (processor/network/data transmission technology) and TECH_8 (battery), followed by TECH_6 (camera function), TECH_{10} (accessories), and TECH_{12} (appearance design). During Period V, comment topics primarily involved TECH_2 (gaming), TECH_3 (touch/NFC/smart remote control technology), and TECH_6 (camera function), followed by TECH_0 (processor/network/data transmission technology), TECH_1 (sound), TECH_7 (storage), TECH_8 (battery), and TECH_{12} (appearance design).

4.5.2 Temporal Evolution Patterns of Topic Content Under Important Product Technical Attributes Using the current important technical attribute TECH_{10} (accessories) as an example, Formula (2) was used to calculate similarity between different topic contents under this attribute. Since topic word extraction results are relatively concise, a 7% similarity threshold was set for more precise analysis of topic content temporal evolution patterns. Topics with similarity above this threshold were considered similar topics. The DyData visualization platform was used to present the evolution relationships of similar topics across adjacent periods in a Sankey diagram, analyzing the temporal evolution patterns of topic content under this technical attribute, as shown in Figure 4 [Figure 4: see original paper].

As shown in Figure 4, topics under TECH_{10} (accessories) continuously intersect, differentiate, and merge during the research development process. Topic III_{topic}7, involving Huawei phone accessories such as screen protectors and cases as well as logistics distribution technical services, evolved into topic IV_{topic}2 involving Xiaomi phone logistics distribution technical services, topic IV_{topic}5 involving accessory effects and features, and topic IV_{topic}8 involving Apple phone film, protective cases, and photo/charging-related accessory effects and functions. Topic IV_{topic}8 merged and evolved with other topics into topic V_{topic}4 regarding feedback on accessories for Apple and Huawei phones, topic V_{topic}5 regarding feedback on Apple phone accessories, and topic V_{topic}6 regarding feedback on phone charging and photo accessories.

4.5.3 Temporal Distribution Patterns of Comment Quantities Under Important Product Technical Attributes Formula (2) was used to calculate similarity between comments and the product technical attribute word–technical feature word table. Since preprocessed comment information is relatively concise and the constructed technical attribute–technical feature word table is relatively precise, a 3% similarity threshold was set. Comments with similarity above the threshold were considered technical attribute-related comments. The temporal distribution of comment quantities under important product technical attributes is shown in Figure 5 [Figure 5: see original paper].

As shown in Figure 5, comment quantities under important product technical attributes such as sound, camera function, storage, and battery show a rapid growth trend from Period III to Period V, indicating that these technical attributes contain technology opportunities with future development potential.

4.5.4 Temporal Distribution Patterns of Product Model Comments Under Important Product Technical Attributes The temporal distribution patterns of product model comments under important product technical attributes are shown in Figure 6 [Figure 6: see original paper].

Researchers can conduct targeted investigations and analyses of specific product models' related attributes based on these temporal distribution patterns, combined with diversified information channels such as network surveys, academic papers, technical reports, and expert knowledge. For example, during Period V, the number of comments under TECH_{10} (accessories) for Huawei 9X is relatively prominent. Targeted analysis reveals that comment feedback primarily focused on the phone accessory charging plug, which had low power and did not support fast charging, resulting in long charging times that affected user experience []. R&D personnel can discover technology opportunities based on this technical demand feedback and make targeted improvements to products to meet user technical demands.

Conclusion

This study fully utilizes product online review data to propose a forward-looking technology opportunity discovery method integrating comment topic identification with multi-dimensional analysis of technical attributes. Experimental results demonstrate that: First, technology opportunities discovered from smartphone product reviews are primarily embedded in important product technical attributes such as sound, camera function, storage, battery, accessories, and appearance design, with technology opportunities varying across different brand models. Second, users' technical comment topic content exhibits trends of mutual intersection, penetration, and fusion, requiring R&D personnel to improve technical insight and learn from each other's strengths. Finally, respecting and meeting product technical demands for user groups such as the elderly mentioned in comment topics represents an important technology opportunity;

technical R&D personnel need to demonstrate more humanistic care in product design and technical R&D, continuously improving product technological content and user experience.

On one hand, this study can enrich technology opportunity discovery research methods and mechanisms from a technology demand-driven perspective, discover forward-looking technology opportunities embedded in product reviews, and improve the accuracy of technology opportunity identification results, providing scientific and objective research foundations for technology opportunity discovery. On the other hand, this study will more closely integrate technology opportunity discovery research with user demand, providing timely, forward-looking, and scientific technology opportunity research theories and methods for enterprises to accurately meet market technical demands, grasp scientific R&D opportunity trends, and optimize technological innovation resource allocation.

References

[] Li X, Xie Q, Jiang JJ. Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter mining: the case of perovskite solar cell Technology. Technological Forecasting & Social Change, 2019, 146: 687-705.

[] Zhu DH, Porter A. Automated extraction and visualization of information for technological intelligence and forecasting. Technological Forecasting & Social Change, 2002, 69(5): 495-506.

[] Porter A, Michael J. Detampel. Technology opportunities analysis. Technological Forecasting & Social Change, 1995, 49(3): 237-255.

[] Li Baoming. Technical opportunity and decision-making in technological innovation. Scientific Management Research, 1990(5): 61-62.

[] Chen Zhenhong, Dong Junwu. Research on the identification process of entrepreneurial opportunities. Science and Technology Management Research, 2005(2): 133-136.

[] Kang Yuhang, Su Jingqin. Visual identification of technological innovation opportunities—an empirical analysis based on patentometrics. Studies in Science of Science, 2008(4): 695-701.

[] Cecere G, Rexhäuser S, Schulte P. From less promising to green? technological opportunities and their role in (green) ICT innovation. Economics of innovation and new technology, 2019, 28(1), 45-63.

[] Orsenigo L, Malerba F. Technological regimes and sectoral patterns of innovative activities. Industrial and corporate change, 1997, 6(1): 83-117.

[] Li Xin, Xie Qianqian, Hong Zhisheng, et al. Research on emerging technology development trends based on social sensing analysis—taking perovskite solar cell technology as an example. Science & Technology Progress and Policy, 2018, 35(10): 15-24.

[] He Chuanqi. Prediction and analysis of the new scientific and technological revolution. Chinese Science Bulletin, 2017, 62(8): 785-798.

[] Li Xin, Huang Lucheng. Research on technology roadmapping method exploration and practical application—based on bibliometrics and patent analysis perspectives. Science & Technology Progress and Policy, 2016, 33(5): 62-72.

[] Cho TS, Shih HY. Patent citation network analysis of core and emerging technologies in Taiwan: 1997-2008. Scientometrics, 2011, 89(3): 795-811.

[] Lee S, Yoon B, Park Y. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation, 2008, 29(6): 481-497.

[] Zhai Dongsheng, Guo Cheng, Zhang Jie, Li Dengjie. Research on technology opportunity identification method using anomaly detection. New Technology of Library and Information Service, 2016(10): 81-90.

[] Wang Jing'an, Tang Yue, Wang Kun. Research on technology opportunity discovery based on Citespace—taking IoT technology development as an example. Journal of Modern Information, 2018, 38(2): 130-137, 170.

[] Wang MY, Fang SC, Chang YH. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels. Technological Forecasting & Social Change, 2015, 92: 182-195.

[] Song K, Kim KS, Lee S. Discovering new technology opportunities based on on patents: text-mining and F-term analysis. Technovation, 2017(60/61): 1-14.

[] Han Xiaotong, Liu Yanxin, Ren Zhijun, et al. Identification of R&D directions of technology competitors based on patent mining. Science of Science and Management of S.& T., 2018, 39(2): 23-35.

[] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of machine learning research, 2003(3): 993-1022.

[] Zhang Zhenya, Wang Jin, Cheng Hongmei, et al. Research on text spatial indexing method based on cosine similarity. Computer Science, 2005, 032(009): 160-163.

[] Honor 9X charging and battery life test: charging time is troubling but battery life performance is surprising [EB/OL]. [2020-11-14]. https://www.sohu.com/a/331268635_{120156943}.

Author Contributions

Wu Yiping: Data collection, analysis, and paper writing;
Bai Rujiang: Research proposition formulation and paper framework design;
Liu Mingyue: Data analysis and paper detail revision;
Wang Xiaoyue: Paper framework design and detail revision.

Academic Integrity Statement for *Library and Information Service* Submissions

Library and Information Service has always upheld the mission of publishing excellent academic paper results and promoting industry academic exchanges, and is committed to purifying the academic publishing environment and creating a sound academic ecosystem. In 2013, the journal took the lead in formulating, releasing, and implementing the "Joint Statement of Library Science Journals on Abiding by Academic Ethics and Purifying the Academic Environment" (hereinafter referred to as the "Statement") (see: http://www.lis.ac.cn/CN/column/item202.shtml). Subsequently, the journal also took the lead in formulating and releasing the "Joint Action Plan of Chinese Library and Information Science Journals to Resist Academic Misconduct" (hereinafter referred to as the "Joint Action Plan") (see: http://www.lis.ac.cn/CN/column/item247.shtml). To implement and realize this philosophy, this journal hereby solemnly declares that, effective immediately, all submitting authors must commit that: papers submitted to this journal must comply with the above "Statement" and "Joint Action Plan," consciously uphold academic ethics, and resolutely resist academic misconduct. *Library and Information Service* maintains zero tolerance for all forms of academic misconduct, including plagiarism and infringement, and will adopt corresponding disciplinary measures.

Library and Information Service Editorial Office