# Science-Driven Technology Opportunity Discovery Methods: A Postprint

**Authors:** Han Xiaotong, Zhu Donghua, Wang Xuefeng

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

## Abstract

[Purpose/Significance] The intimate relationship between science and technology renders the analysis of technology opportunities through the integration of academic papers and patents more rational and efficient than utilizing single-source data. This study automates the generation of science-technology relationships, reduces dependence on subjective judgment, achieves finer granularity of technology, and concurrently provides research and development recommendations for technology R&D personnel, thereby applying concepts derived from scientific research to corresponding technological innovations.

[Method/Process] The abstracts of papers and patents are represented as Doc2vec vectors, correlated into a network via text similarity, and subsequently science-technology topic clusters are generated based on the Louvain algorithm to identify technology opportunities propelled by scientific research. Finally, an empirical study is conducted with 3D printing technology as a case study.

[Results/Conclusion] Several technology opportunities driven by scientific research are identified, and the identified opportunities are verified to possess certain technological potential, thereby demonstrating the feasibility and effectiveness of the proposed method.

## Full Text

## Research on Technology Opportunity Discovery Methods Promoted by Science

**Han Xiaotong, Zhu Donghua, Wang Xuefeng**
School of Management and Economics, Beijing Institute of Technology, Beijing 100081

**Abstract:**
[Purpose/Significance] The close relationship between science and technology

makes it more reasonable and efficient to analyze technology opportunities by combining papers and patents rather than using single-source data. This paper aims to automate the generation of science-technology relationships, reduce dependence on subjective judgment, refine technology granularity, and provide R&D suggestions for technology developers to apply concepts from scientific research to corresponding technological innovations. [Method/Process] The abstracts of papers and patents were represented as vectors using Doc2vec, which were then associated into a network through text similarity. Technology opportunity clusters promoted by scientific research were identified based on the Louvain algorithm. Finally, an empirical study was conducted using 3D printing technology as a case. [Result/Conclusion] Several technology opportunities promoted by scientific research were identified and verified to have technological potential, proving the feasibility and effectiveness of the method.

**Keywords:** technology opportunity analysis; scientific impetus; patent analysis; Doc2vec; Louvain algorithm

**Classification Number:** G250 F204
**DOI:** 10.13266/j.issn.0252-3116.2022.10.002

Innovation is one of the most important sources of technological progress and economic development [1-2]. To rationally utilize limited resources and enhance innovation capacity and market competitiveness, researchers and corporate R&D personnel need to identify various technology opportunities to achieve innovation. Technology Opportunity Analysis (TOA), proposed by Professor Alan Porter of the Georgia Institute of Technology's Technology Policy and Assessment Center in the 1990s, is defined as inferring potential technological forms or development hotspots that may emerge in a field by mining the development trends and interrelationships of existing technologies in that domain. Professor A.L. Porter combined technology monitoring with bibliometric analysis to conduct technology opportunity identification research [3-4]. Over the 20+ years since TOA was proposed, scholars worldwide have continuously researched technology opportunity analysis, providing rich methodologies for discovering technology opportunities.

Most existing research on technology opportunity analysis is based on patent data. Patents, as indicators for measuring industrial technological innovation, demonstrate the achievements of technological development [5]. In knowledge-intensive technologies, scientific research forms the foundation of technological development and is considered the seed of technology and innovation [6]. Scientific literature, as the primary means of recording, disseminating, and exchanging scientific knowledge, reflects the process and achievements of scientific development [7]. Therefore, many scholars seek technology opportunities from scientific papers. The relationship between science and technology is inseparable; scientific research provides seeds for technological innovation and supplies nutrients for technological development. Conducting technology opportunity analysis from a science-push perspective facilitates the application of scientific research achievements to technological development. Research combining scientific research and technological development typically uses both papers and patents as data sources. Previous scholars have conducted technology opportunity analysis by either mixing these two data types or using them separately, but neither approach adequately utilizes the relationship between science and technology to mine technology opportunities.

Some scholars have formed two major clusters—science themes and technology themes—from papers and patents respectively, and conducted comparative analysis to overcome the shortcomings of previous studies that failed to fully utilize the science-technology relationship. Such research associates themes formed from papers and patents, identifying technology opportunities as themes that exist in papers but not in patents. One approach forms science and technology clusters based on paper and patent citation networks respectively [6,8-9]. However, since immature technologies have not yet formed complete citation networks, this leads to sparse networks. Another approach builds science and technology clusters separately based on text semantics, then compares the thematic differences between them [10-14]. Due to the different data sources, the two types of clusters are difficult to correspond, requiring heavy reliance on experts.

To overcome these limitations, this paper takes the relationship between science and technology as its theoretical foundation. It first associates paper and patent data through the text similarity of their abstracts to form an association network, then generates science-technology theme clusters to identify technology opportunities promoted by scientific research, thereby enhancing the correspondence between science and technology themes while reducing dependence on experts.

## 1 Literature Review

### 1.1 The Interaction Between Scientific Research and Technological Development

The close relationship between scientific research and technological development forms the theoretical basis of this paper. Since M. Price's 1965 study [15], researchers have found that as innovation cycles shorten, the connection between technology and scientific research becomes increasingly close [16-17]. S. Breschi revealed the science-technology connection through the close ties between scientists and inventors [18]. Abundant scientific research can stimulate innovation and technological development [19]. In the long run, the most important source of new technological opportunities is the accumulation of scientific knowledge. Science can promote industrial technology in two ways: first, by providing theories, data, and problem-solving capabilities; second, by directly developing new technological possibilities and proposing new solutions to old problems [20]. Without recent academic research, there would be 10% fewer new products and processes [21]. Conversely, technologies with less scientific exploration may also inspire important scientific breakthroughs. Most applied research work begins with needs or goals before returning to science to achieve those goals [20]. Therefore, science and technology are interdependent [22-23].

In empirical research, Lai Yuangen explored the relationship between scientific research and technological development by mapping CLC classification numbers to IPC categories based on term correspondence to connect papers and patents [24]. Many empirical studies have shown that large amounts of scientific information are organized and compiled in patents, indicating that increasingly more technological development depends on science [25-27]. The mutual citation phenomenon between scientific papers and patents also demonstrates the interconnection, mutual influence, and mutual promotion between science and technology, as well as between basic research and technological innovation [28-30]. L. Fleming confirmed through patent data analysis that science changes inventors' search processes, helping them eliminate ineffective research paths, find useful combinations more directly, and motivate them to continue despite negative feedback [31]. Since scientists rarely know what industrial applications their discoveries might have, and companies often don't know which scientific discoveries might help their needs, T. Hellmann constructed a search and matching model between scientists and enterprises to bridge the communication barriers. The study found that patenting scientific discoveries helps the scientific community push their findings to industry [32].

### 1.2 Opportunity Analysis Research Based on Scientific Papers and Patents

Research on technology opportunity analysis based on papers and patents can be roughly divided into three categories: (1) Studies that do not distinguish between paper and patent data sources and mix them for use. For example, Ren Zhijun mined paper and patent texts to implement an emerging technology discovery method, predicting technology trends and emerging technologies based on feature selection results [33]. (2) Studies that distinguish between papers and patents to identify technology opportunities in each separately. For instance, Zhang Fujun used co-word analysis and predicate tree analysis to conduct comparative analysis between papers and patents and among patents to mine technology opportunities in marine science [34]. Wang Xingwang combined three different types of information—scientific papers, patents, and technology public opinion—to identify high-intensity keywords for frontier technology prediction [35]. (3) Studies that form science and technology themes separately from papers and patents for comparative analysis, overcoming the defect of insufficient utilization of the science-technology relationship in the first two categories. This research associates themes formed from papers and patents, identifying technology opportunities as themes existing in papers but not in patents.

One sub-approach forms science and technology clusters based on paper and patent citation networks respectively, such as N. Shibata et al., who built paper and patent citation networks and measured semantic similarity between academic paper sets and patent sets through natural language processing to mine frontier fields existing only in academic research without patents, supporting R&D investment decisions for enterprises and governments [6,8]. Y. Takan and Y. Kajikawa built paper and patent citation networks separately, calculated cosine similarity between paper clusters and patent clusters, and selected emerging clusters with low similarity, more papers, and fewer patents for technology opportunity analysis [9].

Another sub-approach builds science and technology clusters based on text semantics separately, then compares thematic differences. For example, M.Y. Wang mined papers and patents to form science and technology domain clusters, identifying areas with scientific activity but no technological application, providing potential for new technology opportunities [11]. Huang Lucheng and Wang Jingjing extracted SAO structures from paper and patent texts, calculated paper similarity and patent similarity, summarized cluster themes through multidimensional scaling analysis, and identified themes existing in papers but not in patents to determine technology opportunities [13]. Wang Kun et al. extracted keywords from paper and patent texts, built a dissimilarity co-occurrence matrix, used multidimensional clustering to identify research hotspot themes, and recognized themes existing in papers but not in patents as technology opportunities [14]. Wang Jing'an compared keyword clusters and technology research hotspots in scientific papers and patents, finding that the two major technology opportunities for IoT development were GIS technology at the application layer and fusion computing technology under multi-source big data at the platform layer [12]. Han Yan used edge-betweenness centrality indicators and knowledge unit integration methods to extract technology themes, analyzed technology themes considering both scientific research and technological achievements as foundations for technology formation, and finally found technology opportunities manifested as "knowledge association" [36]. X. Li compared the time difference between when technology themes first appeared in scientific papers and patents, combined text mining with expert judgment to predict short-term technology development trends and identify technology opportunities [10]. Table 1 summarizes these third-category studies.

### 1.3 Literature Review Summary

Most technology opportunity analysis studies use single-source data (scientific papers or patents), missing much useful technical information during analysis and yielding incomplete results. There is a close relationship between scientific research and technological development, with scientific research providing impetus for technological R&D. Combining scientific papers with technology patents facilitates more comprehensive technology opportunity analysis than using either alone. However, many studies that jointly analyze papers and patents still treat the two data sources separately, lacking comparison and connection between them and failing to consider their relationship.

In recent years, some scholars (see Table 1) have overcome the limitation of separating the two data types by clustering papers and patents separately to obtain science and technology themes for comparative analysis. However, these studies still have limitations: (1) In studies forming science and technology themes based on citation networks, emerging technologies have not yet formed complete citation networks and may be excluded during network construction, resulting in sparse networks and potential loss of important information. (2) Papers and their references do not necessarily address the same topic. In studies forming theme clusters based on text, the low correspondence between the two clusters requires expert identification of cluster correspondences, resulting in coarse-grained technology granularity and typically only a few matched science-technology theme pairs, which is not conducive to mining more technology opportunities.

## 2 Technology Opportunity Analysis Method Based on Scientific Impetus

The research framework of this paper is shown in Figure 1 [Figure 1: see original paper] and consists of three steps: (1) Data collection and document vector representation: selecting appropriate technology fields, collecting scientific paper and patent data, extracting abstract texts, and representing papers and patents as "document-feature vectors" based on the Doc2vec model; (2) Science-technology theme network construction: calculating cosine similarity between document vectors to build a network linking scientific papers and patents, forming science-technology theme clusters through the Louvain community detection algorithm to divide the paper-patent network into sub-networks; (3) Technology opportunity discovery: constructing "relative scientific impetus" and "relative technology achievement quantity" indicators, establishing a two-dimensional coordinate system to identify potential technology opportunities.

### 2.1 Data Collection and Doc2vec Document Vector Representation

Scientific paper data were obtained from SCI (Science Citation Index) and EI (Engineering Index). SCI is an internationally recognized tool for scientific statistics and evaluation, while EI is a famous comprehensive retrieval tool for engineering technology. The papers indexed by the two have differences and overlaps but both represent documents needed for this study to characterize scientific processes and achievements. Patent data were obtained from the Derwent Innovation Index (DII) platform. Title, abstract text, and year information were extracted from paper and patent data for analysis.

To associate and match paper and patent data, a document vector representation method is needed to map both abstract texts into the same vector space. Traditional document vector representation methods include: (1) Bag-of-Words; (2) Average Word Vectors, which simply average all word vectors in a sentence; and (3) TF-IDF Weighted Word Vectors, which weight all word vectors in a sentence by TF-IDF. These three methods share the common drawback of ignoring word order and semantic information. Doc2vec [37] is currently a more effective document vector representation method. It is an unsupervised algorithm that learns fixed-length feature representations from documents based on word context order, overcoming the disadvantage of models like Bag-of-Words that ignore word order.

The Doc2vec model is inspired by the Word2Vec word vector model. Similar to Word2vec, Doc2vec has two training modes: PV-DM (Distributed Memory Model of Paragraph Vectors) and PV-DBOW (Distributed Bag of Words version of Paragraph Vectors). This paper uses Python's gensim library for document vector training. Following the model authors' recommendations [37], both models were trained with the same parameters and the resulting vectors were combined for use. To ensure paper and patent vectors exist in the same space, the two data sources were marked and trained together. Before training, to achieve better results, two adjacent words with high co-occurrence frequency in the abstract text were combined into a bigram and treated as a single word for model training.

After repeated experimentation and combining with the default parameters set by the model authors [37], the Doc2vec model parameters were set as follows: document vector dimension of 100, sliding window length of 10, minimum word frequency of 4, and adjacent words with co-occurrence frequency greater than or equal to 20 were combined into bigrams. Ultimately, each paper obtained a document vector $PA_i$, and each patent obtained a document vector $PT_j$, with all vectors having equal length.

### 2.2 Science-Technology Theme Network Based on Louvain Community Detection Algorithm

After representing papers and patents as document vectors, the cosine similarity between each paper and patent document was calculated:

$$\cos_{sim}(i,j) = (PA_i \cdot PT_j) / (|PA_i|\ |PT_j|) \quad (1)$$

Cosine similarity values range between 0 and 1. A $\cos_{sim}(i,j)$ value closer to 1 indicates greater content similarity between paper i and patent j, while a value closer to 0 indicates less relevance. To extract associations between papers and patents, a threshold was set to exclude pairs with cosine similarity below the threshold, while paper-patent pairs with similarity greater than or equal to the threshold were included as edges in the paper-patent association network, with edge weights reset to 1.

The paper-patent association network is complex and difficult to interpret, requiring clustering into interpretable themes. Since the network contains both paper nodes representing science and patent nodes representing technology, the resulting clusters are called science-technology themes. The Louvain community detection algorithm [38] is a commonly used clustering method in social networks. It is essentially a hierarchical graph clustering method comprising two alternating iterative phases. The first phase continuously traverses network nodes, assuming each node is a community. With N nodes initialized as N communities, each node is attempted to be added to the community that maximizes modularity improvement until all nodes no longer change. Modularity is an indicator measuring community partition quality; higher values indicate more internal edges within communities and fewer external edges, meaning better clustering.

In weighted networks, the modularity calculation formula is:

$$Q = (1/2m) \Sigma_{ij} [A_{ij} - (k_i k_j)/2m] \delta(c_i, c_j) \quad (2)$$

where $A_{ij}$ represents the weight of edges between i and j, $k_i = \Sigma_j A_{ij}$ is the sum of weights of edges connected to node i, $c_i$ is the community assigned to vertex i, $\delta(c_i, c_j)$ is a custom function that equals 1 if $c_i = c_j$ and 0 otherwise, and $m = \Sigma_{i,j} A_{ij}$.

In the second phase, small communities are merged into new nodes to reconstruct the network. These two steps are iterated until the algorithm stabilizes, meaning the sum of modularity across all sub-communities remains unchanged.

In practice, after one Louvain algorithm iteration, a few extremely large clusters exist, covering most network nodes. To achieve smaller and more precise theme granularity, large clusters underwent secondary computation using the Louvain algorithm again, while smaller clusters retained their first-round results, ultimately forming an appropriate number of science-technology theme clusters. Each science-technology theme exists as a network containing both paper and patent nodes, representing a sub-network of the paper-patent association network.

### 2.3 Technology Opportunity Discovery Based on Scientific Impetus

In this paper, scientific impetus (SiPush) is defined as: in a science-technology theme cluster network, if there exists a connected paper-patent pair where the patent publication year is at least n years later than the paper, this pair is called one scientific impetus, and the number of such pairs in the theme is SiPush.

M.P. Carpenter et al. analyzed the time lag of patents citing academic papers, finding that patents typically cite papers 3-5 years after publication (referred to as citation lag), similar to citations among scientific papers [39]. X. Li extracted technology themes and science themes from patents and papers respectively, finding that the same theme appears in patents 1-2 years later than in papers (referred to as theme lag) [10]. In existing research, theme lag and citation lag differ significantly because theme lag is based on content similarity while citation lag is based on citation relationships. The former represents the time difference when science and technology develop relatively independently, while the latter directly reflects the lag of science's impetus on technology—that is, scientific research achievements require at least 3 years to manifest in technological development. Therefore, in paper-patent associations, only when the patent application year is 3 or more years later than the paper publication time can we conclude that the technological invention's impetus comes from scientific research rather than relatively independent development. This paper studies technology opportunities promoted by science, thus limiting the source of scientific impetus for technology opportunities to scientific papers published at least 3 years before the patent application (i.e., n = 3).

Assuming a theme contains SiNum papers, two indicators are constructed:

(1) Relative Scientific Impetus ($R_{SiPush}$), representing the relative magnitude of scientific impetus in a science-technology theme, calculated as:

$$R_{SiPush} = SiPush / SiNum \quad (3)$$

where SiPush is the number of scientific impetus pairs in the theme.

(2) Relative Technology Achievement Quantity ($R_{TeNum}$), representing the relative number of existing technology achievements in a science-technology cluster, calculated as:

$$R_{TeNum} = TeNum / SiNum \quad (4)$$

where TeNum is the number of patents in the theme.

As shown in Figure 2 [Figure 2: see original paper], the network represents a science-technology theme cluster, with hollow nodes representing papers (quantity 7) and solid nodes representing patents (quantity 6). The number of scientific impetus pairs (where patent publication year is at least 3 years later than paper) is 12, giving $R_{SiPush} = 12/7 = 1.71$ and $R_{TeNum} = 6/7 = 0.86$.

A two-dimensional quadrant is established using these two indicators, as shown in Figure 3 [Figure 3: see original paper]. The horizontal axis represents relative technology achievement quantity—the further right a theme is positioned, the more patents exist relative to papers, indicating higher technological realization. The vertical axis represents relative scientific impetus—the higher a theme is positioned, the more paper-patent pairs with a 3+ year time difference exist relative to paper quantity, indicating greater relative impetus from scientific papers.

Quadrant I (upper right): Both scientific impetus and existing technology achievements are relatively high. Themes in this area depend heavily on scientific research for technological realization and have high technological maturity, suggesting fewer future technology opportunities and making it unsuitable for technology opportunity mining.

Quadrant II (upper left): High scientific impetus but few existing technology achievements. Themes in this area depend heavily on scientific research but have not yet produced many technological achievements, indicating high future development potential and easy access to theoretical foundations from scientific research. This quadrant is selected for discovering technology opportunities promoted by science.

Quadrant III (lower left): Both scientific impetus and existing technology achievements are relatively low, with low dependence on scientific research and low technological realization. These themes may be in early or declining stages, with relatively few theories and concepts to draw from scientific research. This quadrant is not discussed in this paper.

Quadrant IV (lower right): Low scientific impetus but many existing technology achievements, indicating mature technology with fewer future opportunities. Low scientific impetus suggests these themes depend less on scientific research or have matured without requiring extensive scientific support. This quadrant is also excluded from discussion.

Based on the research objectives, Quadrant II is selected as the area for discovering technology opportunities.

## 3 Empirical Study: Technology Opportunities in 3D Printing

3D printing (three-dimensional printing), also known as additive manufacturing, is a rapid prototyping technology and one of today's rapidly developing hot technology fields. After decades of development, 3D printing has established a certain foundation and continues to develop rapidly, with considerable and steadily increasing numbers of scientific papers and technology patents. Its wide commercial application makes it a suitable case for this study.

### 3.1 Data Retrieval and Document Vector Model

Using 2010-2019 as the retrieval period and referencing previous scholars' relevant work [40], the search query "TI = ((((3D OR 3-D OR (3 ADJ2 dimension*) OR (three ADJ2 dimension*) OR additive) ADJ (print* OR fabricat* OR manufact*)))" was formulated to retrieve paper and patent data from SCI, EI, and Derwent indexes on February 9, 2021. After merging, deduplicating, and removing invalid data, 23,018 papers from SCI and EI and 22,250 patents from Derwent were obtained. According to the research design, patents applied for at least 3 years after paper publication are considered to have scientific impetus. Additionally, 2018-2019 data were reserved for result validation. Therefore, paper data in this study covers 2010-2014, while patent data covers 2013-2017.

After repeated experimentation and combining with the default parameters set by the model authors [37], Doc2vec model parameters were set as follows: document vector dimension of 100, sliding window length of 10, minimum word frequency of 4, and adjacent words with co-occurrence frequency \$ \$20 were combined into bigrams.

### 3.2 Science-Technology Theme Network Construction

After representing papers and patents as document vectors, pairwise cosine similarity was calculated between all papers and patents. A threshold of 0.5 was applied, with paper-patent pairs having similarity $ $0.5 considered content-related and recorded as an edge, thus constructing a paper-patent association network.

Based on this network, the Louvain community detection algorithm was applied for clustering. After the first clustering, 5 large communities were identified, containing 11,395, 4,635, 7,634, 10,689, and 5,769 nodes respectively, with remaining communities having fewer than 60 nodes. These 5 large clusters underwent secondary clustering using the Louvain algorithm again, excluding communities without scientific impetus, ultimately yielding 107 valid science-technology themes.

### 3.3 Technology Opportunity Discovery

The relative scientific impetus and relative technology achievement quantity were calculated for the 107 science-technology theme clusters. The median values median(R_{SiPush}) = 0.410 and median(R_{TeNum}) = 0.960 were used as quadrant division thresholds. Due to the extreme outlier "1_{14}" affecting visualization, this point was excluded, and the two-dimensional scatter plot shown in Figure 4 [Figure 4: see original paper] was drawn, where each point represents a science-technology theme labeled with its cluster ID (first number = first clustering result, second number = second clustering result).

Twenty-three themes fell in Quadrant II, with relatively high scientific impetus but relatively low existing technology achievements—the target area for technology opportunity identification in this paper. Table 2 shows statistics for these 23 themes, including scientific impetus (SiPush), paper count (SiNum), patent count (TeNum), relative scientific impetus (R_{SiPush}), and relative technology achievement quantity (R_{TeNum}).

Ten larger themes were selected as examples for result interpretation and validation. Themes "5_9", "3_9", and "5_{10}" were excluded due to ambiguous meanings. Theme content was inferred from keywords in paper and patent titles. Table 3 shows high-frequency terms (including scientific terms from paper titles and technical terms from patent titles) for each theme, limited to the top 5 terms for space considerations. Weighted frequency was used, where title term frequency was calculated based on occurrence counts in paper-patent association pairs. Documents appearing in multiple associations have their title terms counted repeatedly, as they theoretically play more important roles in the cluster. The inferred theme contents are shown in Table 3, representing the identified technology opportunities. Figure 5 [Figure 5: see original paper] visualizes high-frequency scientific and technical terms, with scientific terms shown as light gray nodes labeled "[si]" and technical terms as dark gray nodes labeled "[te]".

Theme "1_1" is 3D printing display devices. Based on high-frequency paper title terms like "holographic fabrication," "three-dimensional photonic crystal" and patent terms like "liquid crystal," "display device," technology developers can mine theoretical foundations of holography from scientific research to develop 3D printing holographic displays. Theme "3_1" is 3D printing electrodes. Based on paper terms like "supercapacitor performance," "high performance," "electrochemical performance" and patent terms like "organic solvent," "electrode material," opportunities exist in various organic solvents and electrode materials, with scientific impetus expected to drive development toward high performance.

Theme "4_1" is ceramic 3D printing. Technology developers can draw on scientific research to innovate or improve ceramics and sintering processes. Theme "3_4" is 3D inkjet printing, where developers can conduct technological development in printing ink directions and acquire relevant knowledge from scientific research. Theme "2_1" is bone tissue 3D printing, where developers can track and absorb knowledge and concepts from scientific research to develop or improve bone tissue 3D printing technologies for bone regeneration and repair. Theme "4_7" is powder bed 3D printing, where technical opportunities can be pursued in powder bed and laser 3D printing technologies. Theme "5_9" is food 3D printing, where scientific development drives technological progress, bringing opportunities in food printing raw material extraction and food/feed 3D printing technologies. Theme "4_6" is laser metal 3D printing, with opportunities in metal powder materials, laser 3D printing technology, and memory materials, while also potentially borrowing relevant concepts from scientific research achievements. Theme "5_{13}" is 3D printing equipment quality control and cloud service systems, where fault diagnosis concepts and security knowledge can be extracted from scientific research and applied to improve 3D printing equipment quality and perfect cloud service systems. Theme "3_5" is 3D printing composite materials, where various chemical experiments and material studies from scientific research can be referenced for innovation and improvement.

### 3.4 Result Validation

To verify whether the identified opportunities have significant technological potential and whether this potential is largely driven by scientific research, patent counts for each technology theme from 2013-2019 and the number of patents with scientific impetus were statistically analyzed, as shown in Figure 6 [Figure 6: see original paper]. The bar charts represent total patent counts in each theme, while line charts represent patents with scientific impetus.

All themes show rapidly increasing patent trends (bar charts), with scientific impetus trends also growing rapidly (line charts). The apparent decline in 2019 may be due to the time lag between patent application and publication, resulting in incomplete 2019 patent data in the Derwent database. Moreover, the proportion of patents with scientific impetus (line charts) to total patents (bar charts) exceeds the majority and shows an upward trend. Overall, these statistics demonstrate that the technology opportunities identified by this method indeed have technological potential and are largely driven by scientific research, validating the method's effectiveness.

To compare patent growth trends across quadrants, Figure 7 [Figure 7: see original paper] shows patent counts and patents with scientific impetus for themes in Quadrants I, III, and IV from 2013-2019. In Quadrant I, although the proportion of patents with scientific impetus (line charts) is high, total patents (bar charts) show no clear upward trend, suggesting these themes may be in mature development stages with limited technology opportunities. In Quadrants III and IV, the proportion of patents with scientific impetus remains consistently low, indicating that technology development in these quadrants is less influenced by scientific research. Quadrant IV also shows no clear upward trend in total patents, suggesting limited development potential for these technologies under scientific impetus.

In other scholars' technology opportunity analyses of 3D printing, Wang Jinfeng [41] identified optimal technology opportunities including switchable multi-color printer heads, laser sintering, and composite material combinations through patent text mining and morphological analysis. Wang Jing'an [14] identified opportunities for scientific theory patenting and industrialization from the perspective of combining papers and patents: changing material properties through direct laser metal sintering and deposition technology, and testing metal material properties through non-linear modeling and simulation. These results partially overlap with or contain technical elements from the laser metal 3D printing, 3D printing composite materials, and ceramic 3D printing themes identified in this paper. However, this paper identifies a broader range and greater number of technology opportunities, further validating the method's effectiveness and advantages from another perspective.

## 4 Summary and Discussion

This paper introduces a technology opportunity discovery method promoted by science from the perspective of integrating scientific research and technological development. The method uses scientific papers and technology patents as data sources, employs the Doc2vec model to represent both abstracts as vectors in the same space, constructs a paper-patent association network through cosine similarity calculation, uses the Louvain algorithm to form science-technology theme clusters from the network, and identifies technology opportunities as themes with relatively high scientific impetus but relatively low current technology achievements. Technology developers should focus on these opportunities, drawing theoretical knowledge and relevant concepts from scientific research to apply to corresponding technological innovations.

In the 3D printing case study, the method effectively formed document vector representations of papers and patents, constructed paper-patent association networks and science-technology theme networks, ultimately identifying technology opportunities promoted by science. Validation confirmed that the identified opportunities possess technological potential, proving the method's effectiveness. Unlike previous studies identifying technology opportunities based on paper-patent relationships, this method does not require manual pairing of science and technology themes. It connects papers and patents before theme formation, reducing partial dependence on subjective judgment, making the process more automated and refining technology granularity.

This research has some limitations: (1) The cosine similarity threshold for filtering the paper-patent association network is set manually, potentially excluding some near-threshold but genuinely related connections and causing edge loss in the network. Future research will design more complete methods for constructing association networks. (2) The relationship between science and technology is mutual—science drives technology, and technology also pulls science. This paper only considers the former; future research could consider technology's pull on science to mine research opportunities. (3) Adding citation relationships to the network construction method would increase precision. However, in practice, because patents cite papers infrequently, the network becomes extremely sparse, making the impact of citation relationships negligible on text-based networks. Currently, only text-based methods are used for network construction. Future research will attempt to address this issue by appropriately incorporating citation relationships into text-based networks for optimization. (4) In the innovation chain, besides scientific research and technological development, there is also commercial application, with all three stages interacting. Future technology opportunity analysis could be extended to the entire innovation chain, combining multiple stages to identify technology opportunities more comprehensively.

## References

[1] SCHUMPETER JA. The theory of economic development[J]. Journal of political economy, 1911, 1(2): 170-172.
[2] PENROSE E. The Theory of the Growth of the Firm[M]. New Jersey: Blackwell, 2009.
[3] PORTER AL, MICHAEL J, DETAMPEL. Technology opportunities analysis[J]. Technological forecasting & social change, 1995, 49(3): 237-255.
[4] PORTER AL, JIN XY, GILMOUR JE, et al. Technology opportunities analysis-integrating technology monitoring, forecasting, and assessment with strategic-planning[J]. SRA-Journal of the society of research, 1994, 26(2): 21-31.
[5] ROBINSON DKR, HUANG L, GUO Y, et al. Forecasting Innovation Pathways (FIP) for new and emerging science and technologies[J]. Technological forecasting and social change, 2013, 80(2): 267-285.
[6] SHIBATA N, KAJIKAWA Y, SAKATA I. Extracting the commercialization gap between science and technology-Case study of a solar cell[J]. Technological forecasting and social change, 2010, 77(7): 1147-1155.
[7] DING Y, CHOWDHURY GG, FOO S. Journal as markers of intellectual space: journal co-citation analysis of Information Retrieval area, 1987-1997[J]. Scientometrics, 2000, 47(1): 55-73.
[8] SHIBATA N, KAJIKAWA Y, SAKATA I. Detecting potential technological

fronts by comparing scientific papers and patents[J]. Foresight, 2011, 13(5): 51-60.
[9] TAKANO Y, KAJIKAWA Y. Extracting commercialization opportunities from text mining of the gaps between science and technology: the case of perovskite solar cell technology[J]. Technological forecasting and social change, 2019, 138(1): 45-68.
[10] LI X, XIE Q, DAI MT, et al. Forecasting technology trends using text mining of the gaps between science and technology: the case of microalgal biofuels[J]. Technological forecasting and social change, 2019, 146(9): 432-449.
[11] WANG MY, FANG SC, CHANG YH. Exploring technological opportunities by mining the gaps between science and technology: the case of perovskite solar cells[J]. Technological forecasting and social change, 2015, 92(3): 182-195.
[12] WANG Jing'an, TANG Yue, WANG Kun. Research on technology opportunity discovery based on Citespace: a case study of IoT technology development[J]. Modern Intelligence, 2018, 38(2): 130-137, 170.
[13] HUANG Lucheng, WANG Jingjing, LI Xin, et al. Technology opportunity analysis of perovskite solar cells based on papers and patents[J]. Journal of Intelligence, 2016, 35(7): 686-695.
[14] WANG Kun, WANG Jing'an, TANG Yue, et al. Research on technology opportunity identification based on patents and scientific papers: a case study of metal 3D printing technology[J]. Science and Technology Management Research, 2018, 38(7): 73-79.
[15] PRICE M. Recent Studies in the Restoration and Eighteenth Century[J]. Studies in english literature, 1965, 5(3): 553-574.
[16] NARIN F, HAMILTON KS, OLIVASTRO D. The increasing linkage between U.S. technology and public science[J]. Research policy, 1997, 26(3): 317-330.
[17] NARIN F, OLIVASTRO D. Status report: Linkage between technology and science[J]. Research policy, 1992, 21(3): 237-249.
[18] BRESCHI S, CATALINI C. Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks[J]. Research policy, 2010, 39(1): 14-26.
[19] ROSENBERG N. Inside the Black Box[M]. Cambridge: Cambridge University Press, 2010.
[20] KLEVORICK AK, LEVIN RC, NELSON RR, et al. On the sources and significance of interindustry differences in technological opportunities[J]. Research policy, 1995, 24(2): 185-205.
[21] MANSFIELD E. Academic research and industrial innovation[J]. Research policy, 1991, 20(1): 1-12.
[22] MEYER M. Tracing knowledge flows in innovation systems-an informetric perspective on future research science-based innovation[J]. Economic systems research, 2002, 14(4): 323-344.
[23] PETRESCU AS. Science and technology for economic growth, new insights from when the data contradicts desktop models[J]. Review of policy research, 2009, 26(6): 839-880.
[24] LAI Yuangen. Research on linking journal papers and patent literature[J].

Library and Information Knowledge, 2011(1): 63-69.
[25] MCMILLAN GS, NARIN F, DEEDS DL. An analysis of the critical role of public science in innovation: the case of biotechnology[J]. Research policy, 2000, 29(1): 1-8.
[26] NARIN F, NOMA E. Is technology becoming science?[J]. Scientometrics, 1985, 7(3): 369-381.
[27] TIJSSEN RJW. Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows[J]. Research policy, 2001, 30(1): 35-54.
[28] ZHAO Liming, GAO Yang, HAN Yu. Application of patent citation analysis in research on knowledge transfer mechanisms[J]. Studies in Science of Science, 2002, 20(3): 297-300.
[29] LIU Li, WANG Yaode. The important role of public science in technological innovation from the perspective of patent citations[J]. Studies in Science of Science, 2003, 21(4): 428-432.
[30] YIN Yuanyuan. Research on science-technology interaction based on paper-patent citation relationships: an empirical analysis of stereoscopic display[J]. Library and Information Service, 2012, 56(16): 65-70, 74.
[31] FLEMING L, SORENSON O. Science as a map in technological search[J]. Strategic management journal, 2004, 25(8-9): 909-928.
[32] HELLMANN T. The role of patents for bridging the science to market gap[J]. Journal of economic behavior & organization, 2007, 63(4): 624-647.
[33] REN Zhijun, QIAO Xiaodong, XU Shuo, et al. Research on technology opportunity discovery model based on data mining[J]. Journal of Intelligence, 2015, 34(6): 174-177, 190.
[34] ZHANG Fujun, YE Quanhui, YU Luyun. Technology opportunity analysis in marine science field based on knowledge graph[J]. Science and Technology Management Research, 2017, 37(24): 165-170.
[35] WANG Xingwang, DONG Jue, YU Tingting, et al. Research on frontier technology prediction method based on multi-type informetric analysis[J]. Journal of Intelligence, 2018, 37(10): 70-75, 89.
[36] HAN Yan, PENG Aidong. Research on technology opportunity identification based on three elements of technology formation: a case study of medical service robot technology[J]. Information Studies: Theory & Application, 2020, 43(5): 156-162.
[37] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]//ICML. Proceedings of the 31st International Conference on Machine Learning. Beijing: JMLR, 2014.
[38] BLONDEL VD, GUILLAUME JL, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008(10): P10008.
[39] CARPENTER MP, COOPER M, NARIN F. Linkage Between Basic Research Literature and Patents[J]. Research management, 1980, 23(2): 30-35.
[40] HUANG Y, ZHU D, QIAN Y, et al. A hybrid method to trace technology evolution pathways: a case study of 3D printing[J]. Scientometrics, 2017, 111(1): 185-204.

[41] WANG Jinfeng, WU Min, YUE Junju, et al. Research on technology opportunity identification path in innovation process: based on patent mining and morphological analysis[J]. Information Studies: Theory & Application, 2017, 40(8): 82-86.

**Author Contributions:**
Han Xiaotong: Methodology design, programming, data processing, paper writing;
Zhu Donghua: Conceptual guidance;
Wang Xuefeng: Conceptual guidance, methodology design, writing guidance, paper revision.

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