Research on the Identification of breakthrough technologies driven by science

Dan Wang[1,†], Xiao Zhou[1,*,†], Pengwei Zhao[1], Juan Pang[1] and Qiaoyang Ren[1]

[1] Xidian University, Xifeng 266 710126 Xi'an, Shaanxi Province, China

Abstract

The identification of breakthrough technologies plays a crucial role in driving technological innovation forward. The science-driven technology innovation pattern has emerged as a significant approach for identifying breakthrough technologies. This paper presents a novel framework for identifying breakthrough technologies based on a science-driven technological breakthrough pattern. Firstly, the acquisition of new science is defined as scientific topics that are both novel and impactful, yet have not been integrated into existing technological systems. Subsequently, the introduction of new science into the existing technological system is achieved through the construction of an S-T network. Link prediction is employed to uncover deep semantic links between new science and technology, followed by the application of community detection algorithms to filter subnetworks containing newly formed science-technology links. Finally, the impact of these subnetworks is evaluated using structural entropy to identify breakthrough technologies. This method reveals the intricate relationship between new science and technology, capturing the diffusion pathways and impact scope of new science within existing technological systems. The effectiveness of this model is validated using the field of artificial intelligence as an illustrative example. This method not only assists researchers in accurately identifying the sources and development paths of technological breakthroughs but also provides important information for the formulation of future research and development policies.

Keywords

Innovation, breakthrough technology, knowledge networks, link prediction, structural entropy

1. Introduction

Breakthrough innovation, characterized by its highly revolutionary nature, plays a pivotal role in enabling enterprises to overhaul industry chains, enhance competitiveness, and seize prime opportunities in the increasingly competitive global landscape [1]. Recent research has highlighted the significance of the interplay between science (S) and technology (T) in fostering potential breakthrough technologies [2]. Notably, the S-T model signifies instances where technological advancements stem from scientific discoveries, serving as a key driver of technological innovation. The incorporation of scientific insights into technological progress plays a pivotal role in enhancing national innovation capabilities and competitiveness [3].

The 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2024) and the 4th AI + Informetrics (AII2024), April 22 - 26, 2024, Changchun, China and Online

* Corresponding author.
† These authors contributed equally.
wdan2017@126.com (Dan Wang); belinda1214@126.com (Xiao Zhou); pwzhao@mail.xidian.edu.cn(Pengwei Zhao); pangjuan1006@163.com(Juan Pang); medicine98@foxmail.com(Qiaoyang Ren)
0009-0007-7011-5347 (Dan Wang); 0000-0001-5289-4876 (Xiao Zhou)

Research has demonstrated that emerging scientific disciplines characterized by novelty and impact play a crucial role in driving revolutionary technological breakthroughs [4]. This influence is primarily evident in three key areas. Firstly, new scientific breakthroughs serve as a primary catalyst for propelling technological innovations and enhancing industrial progress [5]. Secondly, within the organizational context, new sciences offer valuable theories, data, and problem-solving capabilities that furnish substantial evidence of novelty and creativity in corporate research and development endeavors [3]. Thirdly, within the field of invention, where invention is seen as a search process for combinations of technology, new sciences have changed the search process for inventors, guiding them to find useful combinations in a more direct way, eliminating ineffective research paths and favouring the generation of breakthrough inventions [6][7]. The higher the quality of the scientific papers cited in a patent, the higher the value of that patent [8]. Patents that cite scientific literature are more likely to be traded, and the greater the reliance on science, the more likely the invention is to be traded [9]. Therefore, the objective of this study is to investigate the progressive evolutionary development of emerging scientific disciplines that lead to significant advancements in technology, with the goal of identifying innovative breakthrough technologies.

Existing methods for identifying breakthrough technologies mainly adopt two different forms to represent technological knowledge. The first form utilizes coarse-grained IPC classification codes or individual patent documents to represent technological knowledge [10], which cannot monitor changes in technological details at a micro level. The second form involves mining and measuring fine-grained technological knowledge based on textual content, using keywords and keyword phrases as the most basic units of representation for knowledge elements [11]. This paper adopts the fine-grained representation approach, considering breakthrough technologies as composed of several closely related scientific and technological knowledge elements. The aim of this paper is to delve into and analyze the connections and interactions between these two types of knowledge elements, exploring the dynamic evolutionary process by which new science triggers technological breakthroughs, thereby identifying breakthrough technologies. To do so, this paper constructs a breakthrough technology identification framework. The core idea of the study is to use new science as a signal of innovation, to deeply explore the mechanisms and evolutionary paths through which new science leads to technological breakthroughs, and on this basis, to identify disruptive technologies.

2. Research framework and methodology

The framework for identifying breakthrough technology is shown in Figure 1, encompassing a total of five stages.

The first phase involves data collection and preprocessing, where papers and patents are utilized as carriers of science and technology, respectively. We use Web of Science (WOS) and Incopat patent databases as data resources to collect data, and use search queries related to the research topics to download relevant scientific papers and patents.

The second phase focuses on acquiring new science. We consider new science to meet two criteria: novelty and impact, and absence within the existing technological system. Firstly, We adopted Sentence-BERT (SBERT) [12] and Local Outlier Factor (LOF) [13] to quantify the novelty of papers, while utilizing citation counts as a metric for assessing paper impact. Building upon this foundation, we proposed a yearly cumulative iterative strategy for the recognition of innovative papers, as illustrated in Fig.2. Secondly, the KeyBERT [14] algorithm was employed to extract topics from innovative papers and patent texts. We categorize the topics into three types: 1) New science topics, which represent the topics that only appear in scientific innovation papers but not yet covered in patents; 2) Shared topics, which refer to the topics that appear in both scientific innovation papers and patents; and 3) Technological topics, which represent the topics that only exist in patents but not in the scientific knowledge network.

The third phase entails constructing a science-technology network(S-T network) by integrating new science topics into the existing technological system, thereby establishing an S-T network to reveal the connectivity between new science and existing technologies. The specific integration process is shown in Figure 3.

The fourth phase involves the identification of "new science-technology" association subnetworks based on link prediction and community detection. We adopt an attribute feature-based graph convolutional network (GCN) [15] for link prediction in the S-T network to discover potential linkages between new science topics and technological topics. Subsequently, adhering to our definition that breakthrough technologies are composed of several closely related new science topics and technological topics, we further identify subnetworks covering "new science-technology" associations through community detection algorithms as candidate breakthrough technologies.

The objective of the fifth phase is to delve into the profound impact that the integration of new science and technology may have, by assessing the influence of subnetworks containing new links. We employ the structural entropy measure proposed by Xu et al. [16] to calculate the structural entropy influence of each subnetwork. This process enables us to filter out subnetworks that significantly affect the overall structural entropy, thereby identifying potential breakthrough technologies.

3. Empirical analysis

To assess the efficacy of the suggested approach, the domain of artificial intelligence (AI) is selected as a representative case study. Following a methodology similar to that outlined by Tsay et al. [17] and subsequent removal of duplicate records, a total of 236,333 publications and 29,468 patents related to AI, published between 2014 and 2018, were identified.

First, the process of identifying new science topics begins with the conversion of scientific paper texts into 384-dimensional vector representations using the SBERT model. Subsequently, outlier detection is conducted through the LOF algorithm, with the citation count threshold set to the top 10% of papers. Employing a cumulative annual approach, a total of 4667 scientific innovation papers were identified between 2014 and 2018. The K-Means algorithm was then utilized to eliminate papers irrelevant to the domain of study, resulting in 4021 domain-relevant scientific innovation papers being retained. The KeyBERT algorithm was applied to extract keywords from innovative papers and patents. Ultimately, 201 new science topics, 478 shared topics, and 407 technological topics were identified.

Second, S-T network building. S-Net and T-Net are built separately. Shared topics serve as bridges, facilitating the introduction of new scientific topics into T-Net. Consequently, this culminates in the formation of an S-T network under the auspices of new science.

Third, the study employs a GCN for link prediction in the S-T network. This paper introduces comparative experiments, selecting the graph sample and aggregated(GraphSAGE) and graph attention network(GAT) models as benchmarks. The results indicate that the GCN model outperforms both in terms of AUC and AP evaluation metrics. Next, the trained GCN model is applied to the complete dataset for prediction. From the prediction results, the top 3000 new links related to science and technological topics, ranked by link probability, are selected, and added to the original S-T network to create the revised S-T network. After link prediction, Liu et al.'s method [18] is used to partition the S-T revised network into 13 communities. Two communities that do not contain new science topics are excluded, leaving 11 communities for further investigation.

Last, identification of breakthrough technology based on structural entropy. We adopt Haiyun Xu's approach [16], replacing the co-occurrence frequency metric with the probability of edges predicted by the link prediction model. We then use this probability as the basis for calculating structural entropy. We utilized the median as a threshold and identified 5 subnetworks above the median as potential breakthrough technologies. The final results were made in conjunction with expert opinions. Ultimately, the study identified 5 breakthrough technologies. As shown in Figure 4, triangles represent new science topics, while circles represent technological topics. The five breakthrough technologies we identified are: multi-modal natural language processing, humanoid robot, hybrid intelligence, intelligent voice, and drug discovery. Among them, the impact of drug discovery, as the 12th technology, is particularly significant. We conducted a detailed analysis of this breakthrough technology. Deep learning can train models with large-scale biological data to predict the activity, toxicity, and other properties of compounds, thereby rapidly screening out candidate drugs with potential therapeutic effects [19]. Among these technologies, AI-discovered molecules were listed on the Massachusetts Institute of Technology (MIT)'s top ten breakthrough technologies list in 2020. In recent years, drug discovery based on deep learning algorithms has gradually transitioned from research and development to technology development. In 2020, the from-scratch drug design based on deep learning algorithms was recognized by the MIT as a breakthrough in successfully applying artificial intelligence to the drug design process [20].

4. Discussion and Conclusion

This paper proposes a framework for identifying breakthrough technology, starting with new sciences as an innovation signal and tracking the evolution of technological breakthroughs stemming from them. Firstly, building upon the identification of new science, this paper constructs an S-T network incorporating new science. Link prediction is used to mine the latent semantic association between new science and technology, thereby expanding the knowledge structure of the existing technology system. Community discovery algorithms are employed to filter out subnetworks containing links between new science topics and technological topics. Finally, structural entropy is introduced to evaluate the impact of subnetworks, thereby identifying potential breakthrough technologies. The effectiveness of the framework is validated through the example of the field of artificial intelligence.

The key contributions of this study can be listed as follows. First, This study proposes a novel method for identifying breakthrough technologies based on the innovation pattern of science-driven technological breakthroughs, enabling the dynamic tracking and measurement of the innovation process triggered by new science. Second, it provides an in-depth characterization of the essence and core features of new science. Furthermore, employing a topic-based fine-grained approach, the study identifies breakthrough technologies, tracking the dynamic interaction trajectories between new science and technology at the semantic level.

5. Acknowledgements

This work was supported by the General Program of National Natural Science Foundation of China(Grant No. 72374165) .

6. References

[1] D. S. Hain, J. L. Christensen, Capital market penalties to radical and incremental innovation. European Journal of Innovation Management, 23(2): 291-313, 2019.

[2] Z. Ba, K. Meng, Y. Ma, Y. Xia, Discovering technological opportunities by identifying dynamic structure-coupling patterns and lead-lag distance between science and technology. Technological Forecasting and Social Change, 200: 123147, 2024.

[3] A. D. Malva, S. Kelchtermans, B. Leten, R. Veugelers, Basic science as a prescription for breakthrough inventions in the pharmaceutical industry. The Journal of Technology Transfer, 40: 670-695, 2015.

[4] R. Veugelers, J. Wang, Scientific novelty and technological impact. Research Policy, 48(6): 1362-1372, 2019.

[5] J. Dedrick, K. L. Kraemer, Who captures value from science-based innovation? The distribution of benefits from GMR in the hard disk drive industry. Research Policy, 44(8): 1615-1628, 2015.

[6] V. Narayanamurti, J. Y. Tsao, How technoscientific knowledge advances: A Bell-Labs-inspired architecture. Research Policy, 53(4): 104983, 2024.

[7] Q. Ke, An analysis of the evolution of science-technology linkage in biomedicine. Journal of Informetrics, 14(4): 101074, 2020.

[8] D. Popp, From science to technology: The value of knowledge from different energy research institutions. Research Policy, 46(9): 1580-1594, 2017.

[9] A. Arora, S. Belenzon, J. Suh, Science and the Market for Technology. Management Science, 68(10): 7176-7201, 2022.

[10] K. B. Dahlin, D. M. Behrens, When is an invention really radical?: Defining and

measuring technological radicalness. Research policy, 34(5): 717-737, 2005.

[11] B. Sun, S. Kolesnikov, A. Goldstein, G. Chan, A dynamic approach for identifying technological breakthroughs with an application in solar photovoltaics. Technological Forecasting and Social Change, 165: 120534, 2021.

[12] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks. ArXiv preprint arXiv, 1908: 10084, 2019.

[13] D. Jeon, J. Lee, J. M. Ahn, C. Lee, Measuring the novelty of scientific publications: A fastText and local outlier factor approach. Journal of Informetrics, 17(4): 101450, 2023.

[14] M. Q. Khan, A. Shahid, M. I. Uddin, M. Roman, A. Alharbi, W. Alosaimi, W. Alosaimi, J. Almalki, S. M. Alshahrani, Impact analysis of keyword extraction using contextual word embedding. PeerJ Computer Science, 8: e967, 2022.

[15] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks. In: Proceedings of ICLR 2017 (pp. 1 – 14). 2017.

[16] H. Xu, et al, A methodology for identifying breakthrough topics using structural entropy. Information Processing & Management, 59(2): 102862, 2022.

[17] M. Y. Tsay, Z. W. Liu, Analysis of the patent cooperation network in global artificial intelligence technologies based on the assignees. World Patent Information, 63: 102000, 2020.

[18] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach. ArXiv preprint arXiv, 1907: 11692, 2019

[19] B. Bhinder, C. Gilvary, N. S. Madhukar, O. Elemento, Artificial intelligence in cancer research and precision medicine. Cancer discovery, 11(4): 900-915, 2021.

[20] Singh S, Gupta H, Sharma P, Sahi S, Advances in Artificial Intelligence (AI)-assisted approaches in drug screening. Artificial Intelligence Chemistry, 2(1): 100039, 2024.

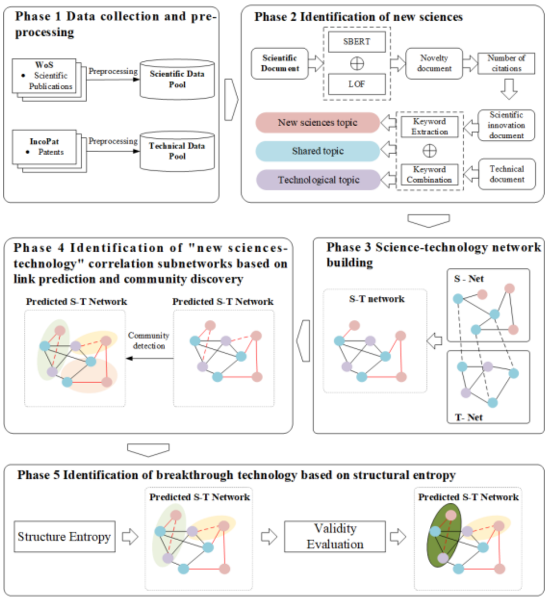

Figure 1: Research framework for identifying breakthrough technology

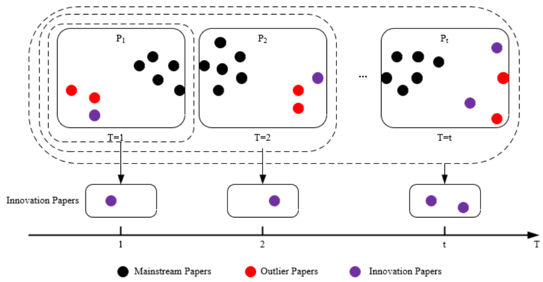

Figure 2: The approach of identifying innovative papers based on yearly cumulative iterative

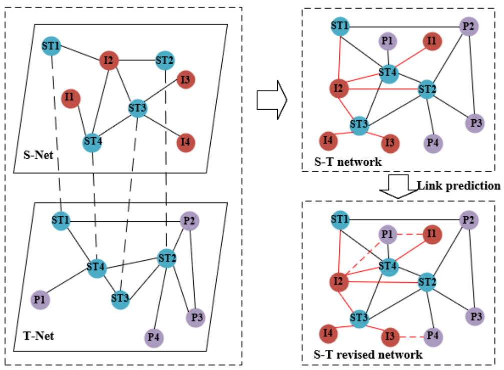

Figure 3: S - T semantic linkage integrative model

7. Appendix A.

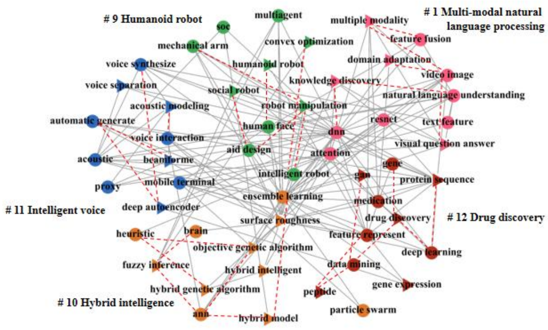

Figure 4： Diagram of breakthrough technology