1st Workshop on AI + Informetrics - AII2021

Emerging Technology Opportunity Identification Based on Community Detection and Burst Detection: A Case study of Intelligent Robots

Jingwen Luo[1,2], Zou Fang[3], Yujie Peng[4], Ying Huang[1,5*]

1. School of Information Management, Wuhan University, China
2. Information School, The University of Sheffield, United Kingdom
3. School of Public Administration, Hunan University, China)
4. Department of Management and Economics, North China University of Water Resources and Electric Power, China
5. Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Belgium

Abstract: Due to a new round of technological revolution and industrial transformation driven by artificial intelligence, emerging technologies have triggered a new round of unprecedented developments. Understanding the technology development trend and identifying potential technology opportunities has become an essential proposition for academia and industry. Based on patent document data, we apply text analysis to extract technical terms, introduce the community detection model to technology topics, then construct a framework for identifying opportunities for emerging technologies from three perspectives: technological R&D trends, competitive environment, and technology layout. The intelligent robot technology is illustrated as an empirical study to clarify the proposed framework. We conclude three main findings: First, intelligent robot technology has undergone three evolution stages and formed 16 key topics, among which IoT robot, sensor fibre resistance, and pneumatic muscle have become hotspots in recent years; Second, China has an absolute advantage in the scale of intelligent robot technology, but it has a noticeable strength gap with other countries and regions in core technologies; Third, compared to the other leading countries in intelligent robots field, the layout of China's overseas technology market is relatively limited, especially in remote control, robot voice and intelligent detection. The proposed opportunity identification framework can clarify emerging technologies' overall context and provide a helpful reference for emerging technologies' market layout and technology layout.

Keywords: Intelligent Robot; Emerging Technology; Patent Analysis; Topic Identification; Technology Opportunity Analysis

1 Introduction

The technology's iteration is accelerating rapidly, with old technologies always being eliminated and new ones emerging. Only by timely grasping the trend of emerging technology development and taking the lead in occupying the advantageous layout of emerging technology can the nations/enterprise win the competitive advantage in the fierce technological and industrial revolution. An accurate grasp of emerging technologies' development trends relies on the effective identification and dynamic monitoring of technology topics.

The pioneering researcher, Alan L. Porter from Georgia Institute of Technology, who proposed the framework of Technology Opportunity Analysis (TOA), indicated that analyzing the changes of related technology or the socio-economic environment in a specific technology field through technology detection and bibliometrics can play a particular role in predicting the technology development in this field (Porter et al., 1995). TOA is a process of identifying emerging technologies and exploring promising technologies in specific technology fields. Later, scholars research and elaborate TOA based on Porter's systematic explanation, which can be mainly divided into three aspects:

- Connotation and functions. Zhu et al. (1998) established an analysis system based on data analysis and expert evaluation to provide the basis for monitoring and early warning and government decision-making in specific technology fields. Yoon and Park (2005) regarded TOA as one of the research methods of technology forecasting, which is in line with many technology management researchers (Shane, 2000; Zhu et al., 2002; Yang & Peiyang, 2013). Ma et al. (2014) defined technology opportunity as new technology forms that emerged from the existing technologies in a specific field during technology development.
- Methodology and indicators. Yoon et al. (2008) proposed a morphological analysis based on keywords, using patent data to identify technology gaps as technology opportunities. Lee et al. (2009) used a combination of Principal Component Analysis and keyword-based patent maps to identify technology opportunities in the map gaps. Wang et al. (2015) applied text mining and a high-dimensional data object clustering to papers and patents, exploring technology opportunities by mining the gaps between science and technology. Rodriguez et al. (2016) discovered technology opportunity by ranking patent outliers, identifying patent anomalies and patent attributes and citations. Wang et al. (2017) proposed a morphological analysis based on the Subject-Action-Object structure and morphological matrix. Lee et al. (2020) used the word2vec method to navigate product landscape analysis to identify potential technology opportunities.
- Empirical studies. Ho et al. (2014) summarized the main technological barriers and the corresponding potential solutions for Proton Exchange Membrane Fuel Cells and Direct Methanol Fuel Cells technologies. Ma et al. (2014) proposed a TOA framework to analyze the technology opportunities in the dye-sensitized solar cells field. Chan and Miyazaki (2015) explored the knowledge convergence between cloud computing and big data and applied it to analyze Malaysia's emerging technology

opportunities. Zhang and Yu (2020) conducted analogies between source and target domains to predict 5G technology opportunities.

Most of the above research on TOA focuses on the patent analysis of emerging technologies, which have narrower meanings and limited development time. In this paper, we proposed a technology opportunity identification framework that can be applied to a broader technology field by identifying the key technology topics and exploring the evolution of the technology. We detected technology opportunities from multiple dimensions to examine the technology opportunities holistically.

In recent years, Artificial Intelligence (AI) has successfully sparked great interest from all sectors of society. It is regarded as a current strategic high technology and a stalwart in leading disruptive industry changes in the future. The combination of AI and robots has made social productivity unprecedentedly magnified. Intelligent robots have huge development prospects and are used in diversified scenarios.

Based on patent data of intelligent robots, we extracted terms and construct co-occurrence networks through text mining, apply community detection for technology clustering, then introduce an emerging topic measurement model and an emergent topic monitoring model to build a framework for identifying technology opportunities from three perspectives: R&D trends, competitive environment, and technology layout. We conducted an empirical study with the intelligent robots patents. We located intelligent robots' key technologies and reveal the overall pulse of emerging technology development using the framework, providing a useful reference for future emerging technologies' market layout and technology layout.

2 Framework and Methodology

2.1 Framework

Fig. 1 illustrated the framework of the research. First, we formulated the patent retrieval query and downloaded the patent documents from the Derwent World Patents Index (DWPI) database. Second, we used text analysis techniques to mine the patent data and extract terms from the patent titles that can characterize the terms' full meaning, including tokenization, POS tagging, stemming, lemmatization and merging terms with similar semantics. Third, after constructing the term co-occurrence network, we used the Leiden algorithm in the community detection to identify intelligent robots' key technology topics; we used the burst detection algorithm to perform topic evolution and hotspot identification of technical terms. Finally, based on the identified key technology topics, we used the TOA framework to identify emerging technology opportunities from three perspectives: technology R&D trends, competitive environment, and technology layout.

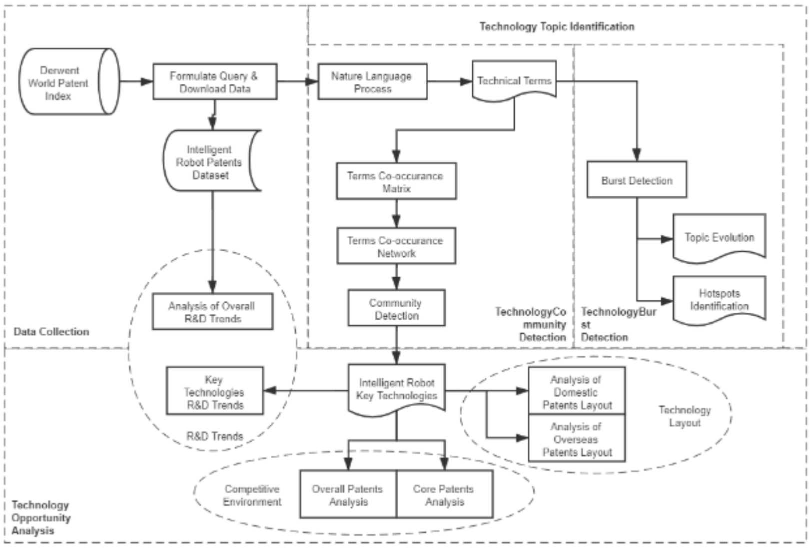


Fig. 1. Research Framework

2.2 Methodology

Natural Language Processing.
Natural language processing (NLP), an important research subject in computer science and artificial intelligence, mainly focuses on making computers understand natural human language. The Natural Language Toolkit (NLTK) package is a set of libraries and programs for NLP developed by Bird et al. (2009). We used the NLTK to process the patent titles, including tokenization, stopwords filtering, POS tagging, stemming, lemmatization. We extracted n-gram terms from each patent title and performed fuzzy semantic merging of terms with similar semantics.

Community Detection.
Community detection is used to understand the structure of large and complex networks. Many scholars have employed community detection in the identification of technology topics. Louvain algorithm by Blondel et al. (2008) as the representative hierarchical clustering algorithm is a commonly used community detection method. Louvain algorithm is recognized as one of the fastest non-overlapping community detection algorithms, but it may produce poorly connected communities. Traag et al. (2019) improved the Louvain algorithm and introduced the Leiden algorithm, ensuring a good connection between communities. We used the Leiden algorithm to conduct community detection on the co-occurrence network. We obtained clearer clusters based on word frequency and word contribution by removing some terms with little meaning from the co-occurrence matrix.

Burst Detection

In 2002, Kleinberg (2003) proposed the Burst Detection algorithm, which can identify the sudden increase or "burst" of terms' use frequency over time. Kleinberg used text classification technology to classify documents that need to be detected. The classified documents are defined as a time-sensitive sequence according to their arrival time. When an event occurs in the real world, the event's documents increase, causing time interval to become shorter. The state at this time is the Burst State, and the opposite is the Normal State. The change of time interval can detect the transition from the Normal State to the Burst State. In this way, the burst is defined as the transition between states, and detect the burst is to detect the time interval between the arrival of two documents. We used burst detection to identify bursts of term frequencies and obtain the burst hot technology topics in this paper.

2.3 Data collection.

We collected patent data from DWPI, which contains patent data from patent licensing agencies in over 100 countries and regions. It provides access to published patents and scientific literature worldwide.

We initially formulated the retrieval query by reviewing relevant literature and consulting domain experts; after repeated verification and justification, the final retrieval query was set as follows: ABD=(((Autonomous OR Bio* OR Humanoid OR Smart OR Intelligent) ADJ2 Robot*) OR Bio-robot* OR Biorobot* OR (Robot* Cognit*) OR (Robot* Percept*) OR (Robot* Sens*) OR (Robot* Act*)). Considering the emergence of intelligent robot patents and the time lag between patent application and publication, we set the time interval from 1956 to 2018. A total of 10,433 intelligent robot-related patents were retrieved (the retrieval was conducted on 20 February 2020).

3 Key Technologies Identification of Intelligent Robots

3.1 Clustering key technological topics

We extracted patent titles from these intelligent robot patent documents, employed text mining, co-occurrence analysis, and community detection, and obtained 34 clusters, 33 of which are valid. We merged clusters sharing similar topics and finally got 16 topic clusters, as shown in Fig. 2.

After manual analysis and interpretation, we obtained 16 key technologies: Humanoid Robot, Robot Joint, Mechanical Arm, Driving Mechanism, Sensor, Wireless Communication, Servo Motor, System Controller, Remote Control, Path Planning, Robot Voice, Robot Vision, Detection Mechanism, Robot Interaction, Robotic Fish, and Intelligent Robots for Different Usage (including Internet of Things Robots, Underwater Robots, Sweeping Robots, Welding robots, Painting robots, etc.).

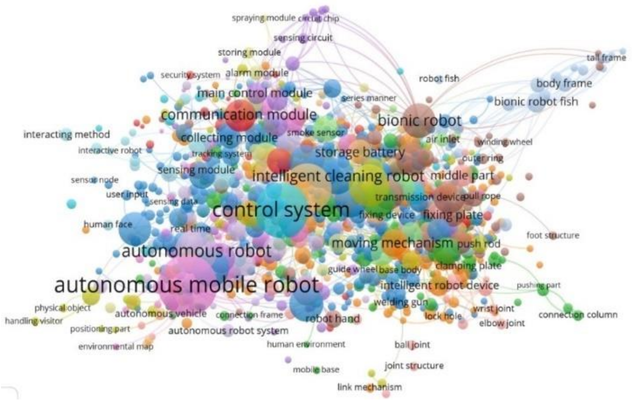


Fig. 2. Clustering Visualization of Intelligent Robot Technologies

3.2 Tracing the topic evolution

We used burst detection algorithms to analyze the rough topic evolutionary lineage of intelligent robots over a more extended period of development. As shown in Fig. 3, we broadly divided the evolution into three time periods: 1979-1993, 1994-2006, and 2006-2018. Between 1979 and 1993, robots were centred on 'Industrial Robots,' which were mainly used in industrial manufacturing and were not yet characterized by intelligence; from 1993 onwards, 'Autonomous Mobile Robots' emerged, which marked the beginning of the transition from powerful stationary machines to highly autonomous and complex mobile platforms. In 2002, the emergence of 'Pet Robots' signalled that robots were no longer only used in industrial scenarios but were beginning to participate in and change people's lives. It also meant that robots were becoming interactive, followed by the "Humanoid Robot" in 2004 and the "Intelligent Mobile Robot" in 2006 when robots became associated with "intelligence" for the first time.

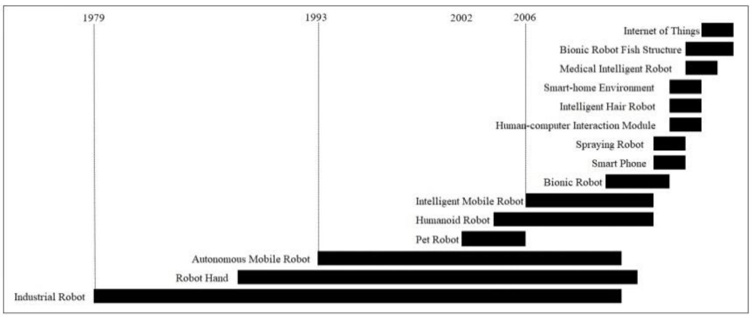


Fig. 3. The Evolution of Intelligent Robot Technologies Topics

The burst weights were calculated and ranked in descending order, and we selected the highest-ranked topics that had not yet reached a steady-state as being related to burst hotspots of intelligent robots.

Table 1. Burst Hotspots of Intelligent Robots

Burst Hotspots	Weight	Duration	Start
Internet-of Thing	5.982	2	2017
Bionic Robot Fish Structure	5.431	3	2016
Charging Pile	5.156	1	2018
Multi-Modal Output	4.858	3	2016
Resistance Layer	4.448	2	2017
Intelligent Robot Human-Computer Interaction	4.399	2	2017
Intelligent Security Robot	4.342	2	2017
Unmanned Aerial Vehicle	4.164	1	2018
Pneumatic Muscle	3.746	1	2018
Grabbing Robot	3.673	1	2018

As illustrated in Table 1, internet-of thing, bionic robot fish structure, robot charging pile, multi-modal output, resistance layer, intelligent robot human-computer interaction, intelligent security robot, unmanned aerial vehicle, pneumatic muscles, and grabbing robot have become hotspots in intelligent robots in recent years.

4 Key Technologies Opportunities Analysis of Intelligent Robot

4.1 R&D Trend Analysis

Although the earliest patent retrieved was published in 1977, the number of published patents from 1975 to 1999 is deficient. For better presentation, Fig. 4 only shows the growth trends for the 16 key technology topics for intelligent robots from 1999 to 2018. As illustrated, Intelligent Robot for Different Usage has always been the main emphasis of technology development, and the majority of key technologies have accelerated in growth since 2014. Driving Mechanism, Humanoid Robots, Mechanical Arm, Robot Vision, Path Planning, and Remote Control are rapidly developing. Still, the growth trend shows that Humanoid Robot and Robot Vision have more momentum for future development. Robot Voice, Sensor, Robot Interaction, System Controller, Wireless Communication, Servo Motor, and Robot Joint are developing relatively slowly. In terms of growth, System Controller and Robot Joint may step into a period of an upward trend. Robotic Fish has a slower growth and fluctuating as a new technology, probably because it has not yet stabilized. There is more potential for future growth.

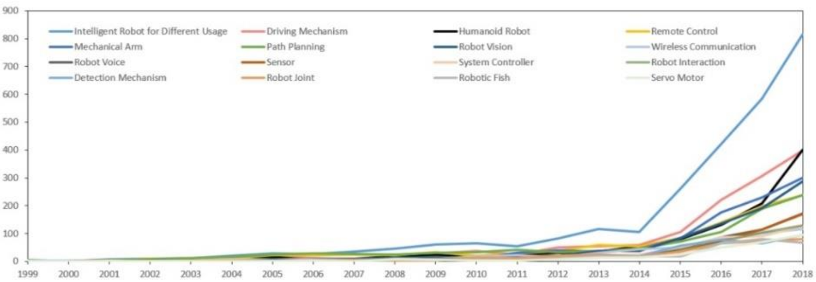
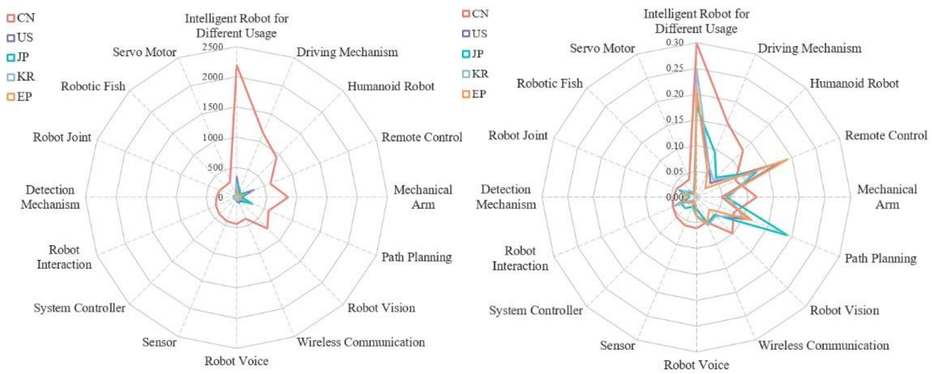


Fig. 4. Analysis of the R&D Trend of Key Technologies of Intelligent Robot

4.2 Competitive Environment Analysis

Fig. 5(a) shows the strength comparison in terms of patent scale in the field of key technologies of intelligent robots in five main competing countries/regions, namely China, Japan, Korea, the US, and Europe. Compared with other countries/regions, China has an absolute advantage in the scale of intelligent robot patents. Also, the layout of key technologies of the five countries/regions is similar to a certain extent.
As the absolute value of the number of patents in the five leading countries/regions varies considerably, to further compare the layout of each key technology for intelligent robots in these countries/regions, we standardized the data to reveal the proportional distribution of each country itself in each technology area, as shown in Fig. 5(b).



(a) Patent Counts (b) Patent Proportion

Fig. 5. Distribution of 16 Key Technologies in Major Countries/Regions

As illustrated, all five countries/regions, especially China and South Korea, emphasize Intelligent Robots for Different Usage. Also, Europe has the most considerable presence in Remote Control; Japan featured prominently in Path Planning; South Korea has the largest share of Wireless Communication; and the US distinguished itself on Remote Control, Path Planning, and Driving Mechanism.

In addition, we explicitly introduced an analysis of the number of core patents to assess the core technological strength of the five countries/regions and form a more scientific understanding of the competitive environment. A core patent has a key position in a particular technology field, making an outstanding contribution to technological development or impacting other patents or technologies (Xu et al., 2014). There are many discussions on the identification of core patents. Still, we mainly use three indicators to measure core patents: the number of citations, the number of patent families, and the number of claims.

4.3 Technologies Layout Analysis

To clarify the five leading countries/regions' technology layout, we constructed a matrix of the distribution of patent priority countries/regions and patent family countries/areas for intelligent robot technologies, as shown in Table 2. Patent priority countries/regions indicate the technology source, while the countries/regions distribution of patent families describe the technology's market layout.

Table 2. Patent Priority and Patent Family of Intelligent Robot Technologies in Main Countries/Regions

	China (7342)	US (1790)	Japan (1475)	Korea (890)	Europe (617)
China (6905)	6896	51	29	11	22
U.S. (1091)	168	1056	188	59	230
Japan (1111)	102	250	1086	47	86
Korea (737)	51	240	46	725	63
Europe (73)	21	49	41	17	72

As depicted, China is the most prominent source of technology and the largest technology market. Still, as much as 93.93% (6896/7342 = 0.9393) of China's intelligent robots patents are filed domestically. In other words, the proportion of China's overseas intelligent robots patents is deficient, which may weaken China in future international market competition.

Generally, overseas patents' technology content is higher, and the scale of overseas patent layout can somehow reflect those countries/regions' technological competitiveness. The proportion of overseas patents in China is the lowest among the five countries/regions, implying that China should focus on quality and innovation while developing its patent scale to enhance its technological advancement and competitiveness to gain a dominant global position in intelligent robots. Furthermore, the US, Europe, Korea, and Japan have few patent applications in China, with the highest number being only 168 of the US, which indicates that the domestic market in China is relatively saturated. China may need to explore opportunities overseas in the future.

The US's overseas homologous ratio is 41.01%, and its overseas patent applications are mainly distributed in Japan and South Korea. Simultaneously, the number of patents filed in the US by Japan and South Korea is also the highest except for their own countries, which indicates that the three countries have close patent technology ties. Although the patent scale of Europe is small, the ratio of overseas patent applications is significant, at 88.33%, in contrast to China's conservative inward-looking technology layout. This indicates Europe has pioneering consciousness, attaching importance to the overseas technology market and rapid technology marketization.

5 Conclusion

As an emerging technology that has been developing rapidly in recent years, intelligent robots have gone through three evolutionary stages: "Industrial Robot," "Autonomous

Mobile Robot," and "Intelligent Robot." It has developed sixteen key technologies, namely Humanoid Robot, Robot Joint, Mechanical Arm, Driving Mechanism, Sensor, Wireless Communication, Servo Motor, System controller, Remote Control, Path Planning, Robot Voice, Robot Vision, Detection Mechanism, Robot Interaction, Robotic Fish and Intelligent Robot for Different Usage. IoT Robot, Bionic Robot Fish Structure, Robot Charging Pile, Multi-modal Output Data, Resistance Layer, Intelligent Robot Human-Computer Interaction, Intelligent Security Robot, Unmanned Aerial Vehicle, Pneumatic Muscle, and Grabbing Robot become burst hotspots recently.

Among the key technology topics of intelligent robots, the Intelligent Robot for Different Usage has been the focus, Humanoid Robot will develop rapidly in the coming period, and Robotic Fish has more significant potential for development as a burst hotspot that has not yet stabilized. Globally, China's intelligent robot patents have an absolute advantage in terms of scale. However, its core patented technologies left much to be desired, particularly in Remote Control and Path Planning. China's intelligent robots technology mainly filed domestically, lacking overseas markets, making China underprivileged to compete in the international market in the future. China could invest more in Remote Control, Path Planning, Robot Voice, Detection Mechanism, and Robot Joint in the future to seek a balanced layout.

This paper's main contributions are summarized below: 1) When extracting terms from patent titles, instead of extracting individual words, n-gram term with thematic significance are extracted, providing convenience for topic identification and hotspots analysis. 2) We constructed a framework for analyzing intelligent robots' key technologies' opportunities from three perspectives: technology R&D trends, competitive environment, and technology layout. It considers the overall patent scale and introduces core patents analysis so that the countries/regions' technological strengths can be assessed more scientifically, and the competitive environment can be understood more holistically.

Limitations of this paper: 1) When extracting terms from patent titles, the overly long terms were not accurately segmented, so some of the terms not being counted in the co-word analysis may affect the results of the topic clustering identification. 2) The proposed framework was based solely on the patents' perspective, without considering the impact of market and technology policies on technology opportunities. For a competitive environment, we only assess the competition between countries at a macro level, but the competition between specific individuals, such as commercial enterprises and research institutions, was neglected.

References

1. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc. https://doi.org/10.17509/ijal.v1i1.106
2. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. https://doi.org/10.1088/1742-5468/2008/10/P10008
3. Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2004.08.011
4. Chan, S. K., & Miyazaki, K. (2015). Knowledge convergence between cloud computing and big data and analysis of emerging technological opportunities in Malaysia. *Portland International Conference on Management of Engineering and Technology*. https://doi.org/10.1109/PICMET.2015.7273134
5. Ho, J. C., Saw, E. C., Lu, L. Y. Y., & Liu, J. S. (2014). Technological barriers and research trends in fuel cell technologies: A citation network analysis. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2013.06.004
6. Kleinberg, J. (2003). Bursty and Hierarchical Structure in Streams. *Data Mining and Knowledge Discovery*. https://doi.org/10.1023/A:1024940629314
7. Lee, C., Jeon, D., Ahn, J. M., & Kwon, O. (2020). Navigating a product landscape for technology opportunity analysis: A word2vec approach using an integrated patent-product database. *Technovation*. https://doi.org/10.1016/j.technovation.2020.102140
8. Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation*. https://doi.org/10.1016/j.technovation.2008.10.006
9. Ma, T., Porter, A. L., Guo, Y., Ready, J., Xu, C., & Gao, L. (2014). A technology opportunities analysis model: applied to dye-sensitized solar cells for China. *Technology Analysis and Strategic Management*. https://doi.org/10.1080/09537325.2013.850155
10. Porter, Alan L, Jin, Xiao-Yin, Gilmour, Joseph E, Cunningham, Scott, Xu, Huaidong, Stanard, Christopher, & Lin, Wang. (1994). Technology opportunities analysis: Integrating technology monitoring, forecasting, and assessment with strategic planning. SRA Journal, 26(2), 21.
11. Rodriguez, A., Tosyali, A., Kim, B., Choi, J., Lee, J. M., Coh, B. Y., & Jeong, M. K. (2016). Patent Clustering and Outlier Ranking Methodologies for Attributed Patent Citation Networks for Technology Opportunity Discovery. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2016.2580619
12. Shane, S. (2000). Prior Knowledge and the Discovery of Entrepreneurial Opportunities. *Organization Science*. https://doi.org/10.1287/orsc.11.4.448.14602
13. Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*. https://doi.org/10.1038/s41598-019-41695-z
14. Wang, M. Y., Fang, S. C., & Chang, Y. H. (2015). Exploring technological opportunities by mining the gaps between science and technology: Microalgal biofuels. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2014.07.008
15. Wang, X., Ma, P., Huang, Y., Guo, J., Zhu, D., Porter, A. L., & Wang, Z. (2017). Combining SAO semantic analysis and morphology analysis to identify technology opportunities. *Scientometrics*. https://doi.org/10.1007/s11192-017-2260-y

16. Xu, H., Yue, Z., Lei, B., & Fang, S. (2014). Core Patents Mining based on Cross Co-occurrence Analysis to Patent Technology-Effect Subject Terms and Citations. *Library and Information Service*. https://doi.org/10.13266/j.issn.0252-3116.2014.04.010
17. Yang, L., & Peiyang, X. (2013). Analysis on Research Status Quo of Technology Opportunity at Home and Abroad. *Information Research*. https://doi.org/10.3969/j.issn.1005-8095.2013.01.002
18. Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: Application of text mining. *R and D Management*. https://doi.org/10.1111/j.1467-9310.2007.00493.x
19. Zhang, J., & Yu, W. (2020). Early detection of technology opportunity based on analogy design and phrase semantic representation. *Scientometrics*. https://doi.org/10.1007/s11192-020-03641-z
20. Zhu, D., Porter, A. L., & Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting and Social Change*. https://doi.org/10.1016/S0040-1625(01)00157-3
21. Zhu, D., Wu, C., & Mao, J. (1998). Analysis of Technology Opportunities Based on Data Network Environment. *Industrial Engineering Journal*. 1(04), 14