Analysis of the Evolutionary Dynamics of the Technology Opportunity Discovery Field Based on Thematic Linkages

Duanwu Yan[1], Yuqian Ma[2*], Chenhe Yang[3]

[1] yanwu123@sina.com, [2*] 2568418253@qq.com, [3] 1198435983@qq.com

[123] School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract. As market competition intensifies, it is essential to effectively identify technology opportunities, but existing research lacks an understanding of the evolutionary dynamics of the field of technology opportunity discovery. To reveal the development of this field, this study uses keyword characterization of research themes to explore the evolutionary capacity and path of the highly cited literature themes in the field of technology opportunity discovery during 2011-2022 by using keyword co-word analysis and cluster analysis methods, based on SciMAT visualization tool, expecting to provide references for future research. The results reveal a steady growth of research results in this field in recent years and a strong evolutionary capacity of some themes. However, thematic differentiation and integration are common, and research centers are scattered with immature research overall, presenting five evolutionary paths in four directions: innovation, genetics, energy, and Internet research.

Keywords: Technology Opportunity Discovery; Theme Evolution; Co-word Analysis; Visualization Theme Analysis; SciMAT

1 Introduction

In the increasingly competitive market, innovation has become an important source of technological progress. As an essential basis for innovation decisions, technological opportunities need to be identified by searching and mining[1]. Therefore, research to help R&D subjects discover potential technological opportunities has started to gain importance, a field known as technology opportunity discovery (TOD). Scholars have conducted related studies. Hall analyzes the relationship between patent applications and R&D activities at the enterprise level[2]. Abbas summarizes the features and weaknesses of various types of patent analysis techniques[3]. Ren analyzes the development trend of research from four aspects[4]. Yi comments on the development focus by sorting out the characteristics of the TOD analysis unit[5]. However, existing studies analyze the frontier themes in a certain period, which cannot dynamically demonstrate the evolutionary process of the theme.

The co-word analysis method calculates the closeness of words by counting the number of simultaneous occurrences in the literature. Further, closely related words can be clustered into independent theme groups by clustering algorithms[6]. Density and centrality metrics[7] can describe the intra-class concentration and extra-class correlation of topic clustering and quanti-

ICEKIM 2023, May 26-28, Nanjing, People's Republic of China

DOI 10.4108/eai.26-5-2023.2337266

fy the development of themes[8]. SciMAT can construct keyword co-occurrence matrix and perform clustering, as well as plot density and centrality-based theme distribution and time-series-based themes evolution paths to reveal research hotspots and time-varying evolutionary paths[9].

Therefore, this study employs SciMAT software to create multiple visual maps for highly cited literature of TOD based on co-word analysis. By analyzing the research hotspots in each period, the thematic evolution paths and patterns of TOD are identified to provide references for subsequent research.

2 Data acquisition and processing

In the core collection of Web of Science, " (Technology opportunity) AND (discovery OR identification OR analysis) " is used as the search term 2011-2022. 2961 valid highly cited papers are obtained according to Price Law (equation 1), where N and n_{max} are the minimum and maximum values of citations per year for a single article. The literature with more than N citations is called highly cited literature.

$$\mathrm{N} = 0.749 \times \sqrt{n_{max}} \tag{1}$$

The singular and plural forms of the same keywords are merged, as well as synonymous keywords with different lexical properties. Semantically broad words are discontinued. To avoid data smoothing, the study interval is divided into three time periods: 2011-2014, 2015-2018, and 2019-2022. The analysis unit is set as a word co-occurrence matrix. The similarity index of network normalization is E index, and the clustering method is a simple central algorithm. The average citations of the literature are selected as assessment indicators, representing the influence of the theme.

3 Analysis of experimental results

3.1 Analysis of the development of the TOD field

As TOD develops, keywords from adjacent periods overlap due to the continuity of research. Table 1 represents the keyword changes in the research in the TOD field.

Table 1. Statistical table of keyword changes in the TOD field during 2011-2022.

Time period	Number of existing keywords	Number of dead keywords	Number of emergence keywords
2011-2014	4286	3003	\
2015-2018	7170	5168	5887
2019-2022	8686	\	6684

The total number of keywords and the number of newborn keywords keep increasing. The number of dead keywords in each phase is less than the number of emergence keywords, which means that the whole field is booming. The number of overlapping keywords accounts for a relatively small number, which proves the strong innovation capacity of the field.

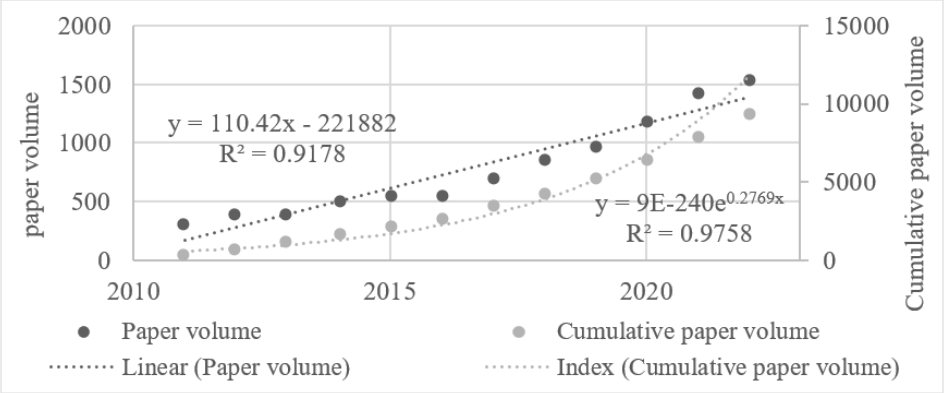

Fig. 1. Trends in the output of papers in the field of TOD during 2011-2022.

Fig.1 represents the distribution of the paper output in the field of TOD, which also proves the above point. The research outputs in this field have shown linear growth in recent years, and the volume of accumulated papers also shows a significant exponential growth trend. This reflects the continuous innovation of research and good research prospect. However, most studies have not received long-term attention and in-depth research from scholars, resulting in their immaturity.

3.2 Theme evolution capability analysis

The nodes indicate the clustering themes and their literature volume. The horizontal and vertical axes are centrality and density, which represent the strength of the association of the theme with external themes and internal keywords[10]. Therefore, this study uses the preprocessed keywords to construct the word co-occurrence matrix and perform clustering, and draw the theme distribution as shown in Fig.2.

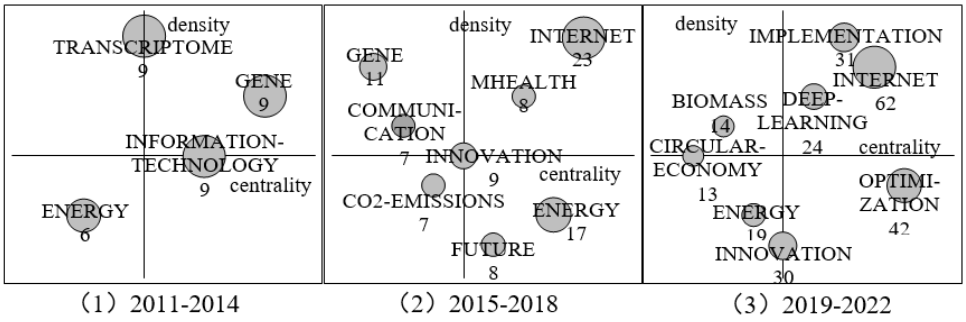

Fig. 2. Theme Distribution Map of TOD Field during 2011-2022.

2011-2014 has four themes. (i) Upper right quadrant: Themes located in this quadrant are hot themes in a growing state that have strong evolution capabilities and have evolved their paths later. (ii) Lower left quadrant: ENERGY is dying, but it has a high average citation rate and may gain continued attention.

Eight themes are clustered for 2015-2018. (i) Upper right quadrant: INTERNET has the highest popularity and is an emerging core theme with great development potential. MHEALTH is a mature theme with low heat. INNOVATION inherits the secondary keywords of IT and still has a strong evolution capability. (ii) Upper left quadrant: COMMUNICATION and GENE with decreasing influence and popularity have inherited the core keywords of GENE and IT respectively, and are professional themes. (iii) Lower left quadrant: CO2-EMISSIONS has great influence and is a new hot theme. (iv) Lower right quadrant: FUTURE is a theme that has received attention but has not yet been developed. ENERGY absorbs new keywords, increasing the heat and relevance to the field, and has a strong evolutionary capacity.

Eight themes are clustered for 2019-2022. (i) Upper right quadrant: After absorbing new content and the FUTURE, INTERNET still maintains the status of the central theme. DEEP-LEARNING and IMPLEMENTATION are hot and influential in general and are growing popular themes. (ii) Upper left quadrant: BIOMASS inherited the secondary keywords of ENERGY, and differentiated to form a new branch. Both it and CIRCULAR-ECONOMY are not very hot and influential, and they are professional themes with stable development. (iii) Lower left quadrant: ENERGY shows thematic fragmentation with a decline in the importance and influence of the field, which may lead to extinction. (iv) Lower right quadrant: INNOVATION and OPTIMIZATION are differentiated from the previous INNOVATION. The former has increasing influence and is a popular theme with high development potential, while the latter is less hot and influential, and is a basic theme to be developed.

3.3 Theme evolution path analysis

The evolution path map shows the flow of data between themes. Nodes are thematic clusters whose size depends on the amount of literature. The solid or dotted lines between nodes in adjacent periods indicate that two topics share core or secondary keywords, and the thickness of the connection line is proportional to the degree of correlation between the themes. Outliers are independent themes. Fig.3 presents the theme evolution path diagram drawn based on the relevance of themes in adjacent periods.

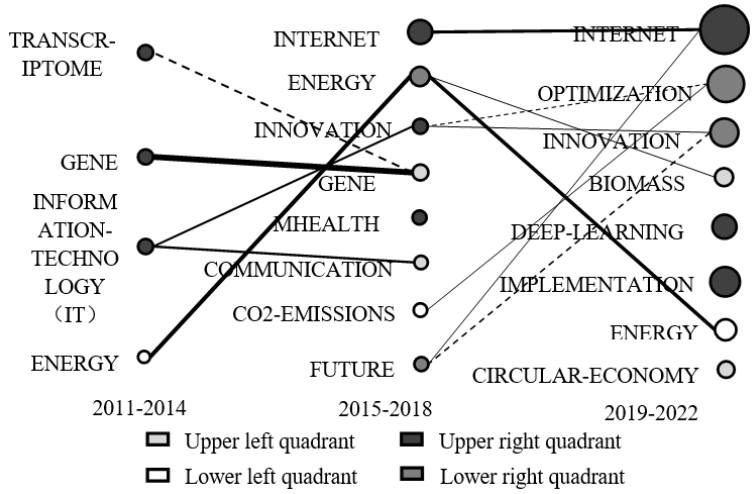

Fig. 3. Theme Evolution Path Map of TOD Field during 2011-2022.

(1) Innovation research: IT→INNOVATION, COMMUNICATION→INN-OVATION. It pays attention to the TOD methods and innovations in the fields of business and manufacturing. Themes shifted from upper right, upper left to lower right, with average popularity and influence. However, after the integration of new content, the amount of literature has increased, and the development in the field is more important, which is a basic path with development potential.

(2) Genetic research: GENE, TRANSCRIPTOME→GENE. It focuses on the opportunity discovery of genetic technologies such as transgenics and genome sequencing. After the theme is merged, it is transferred from the upper right quadrant to the upper left quadrant, which is a more professional research direction in the field. However, the popularity in the development process is not high, and due to the lack of new knowledge injection, the influence continues to decline, which leads to the interruption of the path during 2019-2022.

(3) Energy research: (i) ENERGY→ENERGY→ENERGY, BIOMASS; (ii) CO2-EMISSIONS→OPTIMIZATION. It identifies technological opportunities for improving energy efficiency and reducing emissions, including methods such as life cycle studies and simulation. The energy pathway has a general influence, a complex evolutionary process, and a clear divergence at a later stage. The biomass branch weakened the extinction trend of this path and became a professional direction with stable development. The optimization path develops from the new theme in the lower left quadrant to the basic theme in the lower right quadrant, and the popularity rises, which is a basic direction with great development potential in the field.

(4) Internet research: INTERNET, FUTURE→INTERNET. It is a path that emerged during 2015-2018, focusing on technological opportunities in the Internet of Things, blockchain, and other directions. This path moved to the upper right quadrant after the theme fusion phenomenon appeared, and the amount of literature and popularity continued to increase. It is a path in a growing state with steady development.

4 Conclusions

With the increasing competition in the market, companies and the market need to effectively identify technological opportunities to improve their competitiveness, and the field of technology opportunity discovery has developed, but the evolution of this field is poorly understood in existing studies. Based on the co-occurrence analysis method, this study uses the keywords of 2961 highly cited documents in this field from 2011 to 2022 as the research object and uses SciMAT to map the theme distribution and theme evolution paths, to analyze the knowledge evolution pattern of the field and provide a reference for future research. The results show that the research content has gradually enriched over time, and some themes have strong evolution capabilities, but the phenomenon of theme differentiation and integration is common. In the later stages, isolated points and scattered research centers indicate that the overall research is immature. Five evolution paths are formed in four directions. Among them, Internet research is a hot direction of continuous development; the optimization sub-path of energy research and innovation research is a basic direction with great development potential, and the energy sub-path is a professional direction with stable development; the genetic research path is broken in the later stage.

However, the selection of keywords still needs to be further explored. Missing keywords in this study are replaced with database keywords that are not strongly referential to the theme, which may lead to new themes not being identified quickly. Follow-up studies can further address the problem and improve the accuracy.

References

[1] Venkatraman, N. (1989) Strategic orientation of business enterprises: The construct, dimensionality, and measurement. Management Science, 35(8): 942-962. https://doi.org/10.1287/mnsc.35.8.942.

[2] Hall, B H., Griliches, Z., Hausman, J A. (1986) Patents and R and D: Is There a Lag. International Economic Review, 27(2): 265-283. 10.3386/w1454.

[3] Abbas, A., Zhang, L., Khan, S U. (2014) A literature review on the state-of-the-art in patent analysis. World Patent Information, 37: 3-13. https://doi.org/10.1016/j.wpi.2013.12.006.

[4] Ren, H Y., Wang, Q. (2020) Technology Opportunity Discovery Methods: A State-of-the-art Review. Journal of Intelligence, 39(4): 51-59. 10.3969/j.issn.1002-1965.2020.04.008.

[5] Yi, H F., Liu, X W., Long, Y X. (2021) A Review of Technology Opportunity Discovery from the Full Perspective of Technology Innovation. Library and Information Service, 65(7): 132. 10.13266/j.issn.0252-3116.2021.07.013.

[6] Feng, L., Leng, F H. (2006) Development of Theoretical Studies of Co-Word Analysis. Journal of Library Science in China, (02): 88-92. 10.3969/j.issn.1001-8867.2006.02.021.

[7] Law, J., Bauin, S., Courtial, J., et al. (1988) Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. scientometrics, 14(3-4): 251-264. https://doi.org/10.1007/bf02020078.

[8] Stegmann, J., Grohmann, G. (2003) Hypothesis generation guided by co-word clustering. scientometrics, 56(1): 111-135. 10.1023/a:1021954808804.

[9] Cobo, M J., López-Herrera, A G., Herrera-Viedma, E., et al. (2012) SciMAT: A new science mapping analysis software tool. Journal of the American Society for Information Science and Technology, 63(8): 1609-1630. 10.1002/asi.22688.

[10] Liu, Y H., Hua, W N., Qian, A B. (2018) Analysis of Themes Evolutions of International Research Papers on Health Information. Journal of Modern Information, 38(07): 135-141. 10.3969/j.issn.1008-0821.2018.07.019.