Technology Convergence Prediction From a Timeliness Perspective: An Improved Contribution Index in a Dynamic Network*

Jinzhu Zhang*, Bing Yan

Department of Information Management, School of Economics and Management, Nanjing University of Science and Technology, Nanjing China

Abstract

Technology convergence prediction can identify potential trends and directions in technological development, as well as providing valuable guidance for innovation strategies, research investment, and industrial development. Current methods often construct a technological co-occurrence network to explore the potential associations between technologies for technology convergence prediction. However, its calculation of node importance is often based on quantity statistics of frequencies, failing to break down and distinguish the technological features in each co-occurrence, and assuming equal importance for each technology in every convergence. In addition, the current approach for assessing technological timeliness is too broad, making it difficult to accurately capture technological change. The perspective needs to shift from the life cycle to more specific points in time. Therefore, this paper introduces a contribution index designed to measure changes in the importance of technology convergence from a timeliness perspective. Firstly, we extract and filter valid technical topics to represent technology categories. Secondly, we use dynamic time weights to calculate the semantic similarity between technical topics and patent texts, to indicate the contribution of the technology in each convergence. Thirdly, this paper labels the contributions in technology co-occurrence network to build a dynamic technology network that records changes in technology importance. Finally, we utilize a graph neural network to generate node embeddings for link prediction. In experiments within the field of new energy vehicles, the dynamic network prediction model based on contribution features improved the AUC by 8.92%, 3.52%, and 1.11%, compared to the frequency feature network. It proves that the proposed technological contribution index can effectively enhance the accuracy and effectiveness of technology convergence prediction.

Keywords

technology convergence, timeliness, semantic similarity, graph neural network

1. Introduction

The convergence of technologies from different disciplines can solve increasingly complex technical problems and social needs. At the same time, it is a key factor to ensure technological timeliness for increasing competitiveness in research and investment. Therefore, how to efficiently and accurately predict the potential direction of technology convergence from the mass of existing technologies has become a significant task.

Research often explores the co-occurrence of technologies to analyze the current state of technological convergence. And the technology networks are constructed to explore the potential correlation between technologies. Current methods for technological convergence using patent data include approaches based on patent co-classification

Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2024) and the 4th AI + Informetrics (AII2024)

* Corresponding author.
zhangjinzhu@njust.edu.cn (J. Zhang)
0000-0001-7581-1850 (J. Zhang); 0009-0008-0524-9450 (B. Yan)

[1, 2], patent cross-referencing [3], and text mining methods [4]. In these methods, technology categories are typically identified by patent classification numbers [5, 6] and technical topics [7, 8]. In addition, scholars have expanded the research on technology convergence to a broader perspective, such as the construction of market characteristics [9, 10], social impacts [11, 12] and time characteristics [13, 14], etc., to further improve the prediction index system of technology convergence.

However, these methods do not consider differences in the importance of technologies in each convergence and changes in timeliness. They assume that the importance of each technology is the same in each case of technological convergence. In addition, the technology timeliness is often distinguished by technology lifecycle segmentation [15] and linear weight assignment [16], which are too broad and difficult to capture small differences between different technologies. In this paper, a technological combination co-occurring within a same patent is considered a convergence event. As shown in Figure 1(a), if technologies T_1 T_2 T_3 co-occur with the same frequency, their importance is considered equal, and there is no distinction made among the timeliness of their co-occurrence at different points in time. Actually, different technologies contribute differently to the overall technological combination and have different timeliness with each co-occurrence. As a result, their impact within the technological network differs in scope and extent. For example, as shown in Figure 1(b), although technology T_1 is present in each convergence, its contribution declines over time, indicating declining importance and possibly gradual obsolescence. On the other hand, technology T_2 maintains a stable contribution, suggesting that it may be a foundational technology or in a phase of steady development. Meanwhile, technology T_3 shows a higher contribution, indicating a greater impact or greater timeliness within the technology combination, making it more likely to combine with other technologies.

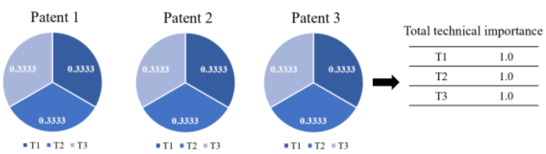

Figure 1(a).

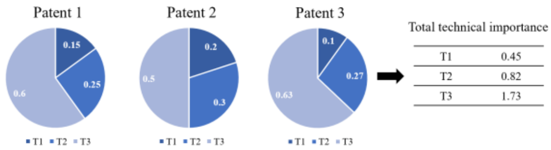

Figure 1(b).

Figure 1: Changes in the contribution of technology within patents

Therefore, this paper proposes an index designed from a timeliness perspective to measure changes in the importance of technology. We obtain the contribution of the technology in each co-occurrence by calculating the semantic similarity between the technical topic and the patent, and then combining the dynamic time weights to obtain the final value. To capture changes in the importance of technology in each convergence, we improve the timeliness of technology by refining it from the lifecycle and dates to more precise convergence time points, constructing a dynamic technological co-occurrence network. Finally, we use link prediction to explore the prediction of technology convergence, aiming to better evaluate the timeliness of technology and its impact on convergence.

2. Data and Method

The method for predicting technology convergence from a timeliness perspective includes three parts, as shown in Figure 2. Firstly, this paper extracts and filter out the valid technical topics characterizing the technical categories in patent texts. Then, cosine similarity is used for semantic similarity computation on the patent texts and technical topics to measure the contribution of different technical topics in each co-occurrence. To obtain the total contribution score for technical topics, this study introduces dynamic time weights and follows the principle of time decay to sum the contributions from each co-occurrence. Then we extract co-occurrence relationships to construct a technological network. Label the contribution of each technical topic on the matching nodes to build a dynamic technical topic co-occurrence network. Finally, graph neural networks are used to learn the node representations of technical topics, and quantitative evaluation is performed by link prediction.

2.1. Data collection

In this paper, the full text data of patent applications were batch downloaded from the USPTO (United States Patent and Trademark Office) patent search platform in December 2023, parsed and stored in a

PostgreSQL database. We use SQL queries to search for relevant patents in the field of new energy vehicles, as shown in Figure 3. A total of 23,792 relevant patents were retrieved and the titles, abstracts and application time of the patents were extracted as the data source for the study. A total of 16,975 patents were used as training data from 2012-2021 and 6,817 patents were used as test data from 2022-2023. The training set contains 192,602 co-occurring relationships. Relationships that were not present in the training set were filtered out to create the actual test set. An equal number of negative samples were generated, resulting in a final test set of 51,562 relationships.

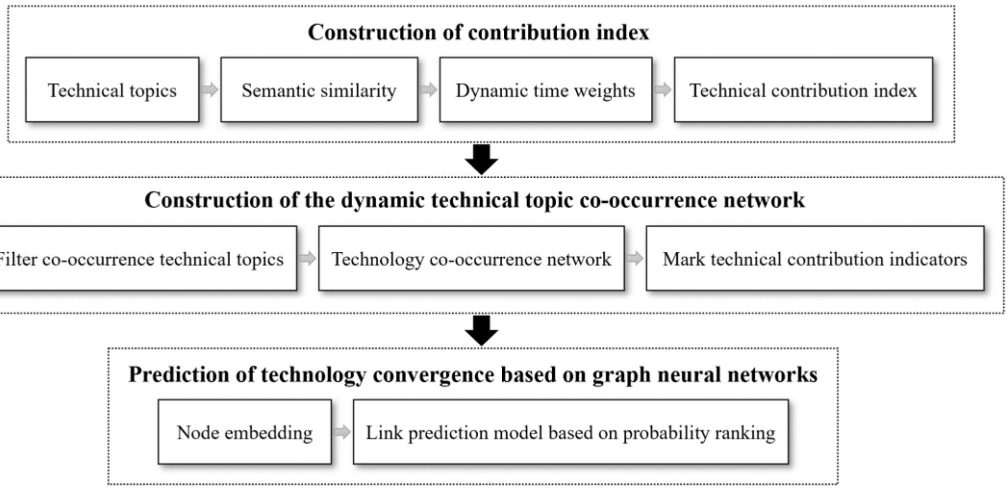

Figure 2: Framework of the method

```
SELECT *
FROM uspto_patents
WHERE
    (publication_date >= '2012-01-01' AND publication_date <= '2023-12-31')
    AND (
        LOWER(publication_title) LIKE '%new energy vehicles%'
        OR LOWER(publication_title) LIKE '%nevs%'
        OR LOWER(publication_title) LIKE '%new energy automobile%'
        OR LOWER(abstract) LIKE '%new energy vehicles%'
        OR LOWER(abstract) LIKE '%nevs%'
        OR LOWER(abstract) LIKE '%new energy automobile%'
        OR LOWER(section_class_subclass_groups) LIKE '%B60L%'
    );
```

Figure 3: SQL statement for querying patents related to new energy vehicles

2.2. Construction of Contribution Index and Dynamic Network

Firstly, this paper extracts technical topics from patent text, representing specific technical categories. Secondly, we sum the semantic similarity between technical topics and patent texts using dynamic time weights, to represent the contribution of the technology. Then, we construct a dynamic network of technical topics by integrating the technological contribution index. This will help the network to reflect changes in the contribution of technology over time and provide more technical clues.

2.2.1. Extraction of technical topics

Technical topics offer a more flexible and comprehensive expression of technical content, making them more explainable. Therefore, we choose to use technical topics to represent different technical categories.

This paper determines the optimal number of topics based on the topic coherence score. And each technical topic has 20 representative keywords to reduce overlap between topics. As shown in Figure 4, *u_mass* and *c_v* gradually converged when the number of topics was around 500. After comparing extreme values, 507 was identified as the optimal number of topics for this paper. Secondly, the TF-IDF weighting is applied to improve the LDA model's process of generating feature words for technical topic extraction, with the aim of improving the representativeness of the topic words.

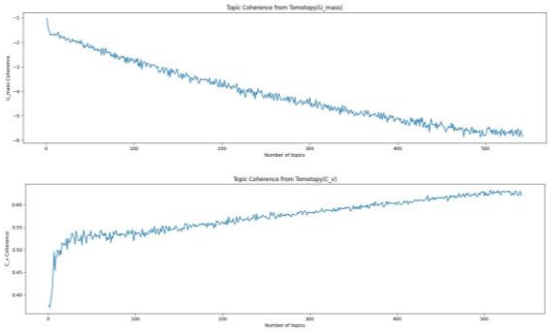

Figure 4: U_mass and C_v variation curves

2.2.2. Calculation of technical contribution index

As the technical topics and patents in this paper are both textual content, and the higher the similarity between technical topics and patent texts, the higher the weight of that technology in the patent. Therefore, we use the semantic similarity between technical topics and patent texts to represent the contribution of technology in each co-occurrence. The study uses Doc2vec [17] to obtain semantic representations of technical topics and patent texts respectively. Then, it applies cosine similarity to calculate the semantic similarity between them, obtaining the contribution values of different technologies in each co-occurrence, as shown in Formula (1). Thirdly, it is important to consider the timeliness of technology. The further away from the current moment, the lower the timeliness tends to be. To address this, dynamic time weights are introduced, based on the retention function of memory capacity [18]. This assigns weighted sums to the contribution of technical topics in each convergence, resulting in the final contribution index score for the given technical topic, as shown in Formula (2).

$$con_{T_{i_n}} = cos(\theta) = \frac{\sum_{i=1}^{m}(P_i \times T_i)}{\sqrt{\sum_{i=1}^{m}(P_i)^2} \times \sqrt{\sum_{i=1}^{m}(T_i)^2}}, \quad (1)$$

Formula (1) defines $con_{T_{i_n}}$ as the contribution of technical topic T_i in the nth co-occurrence. The semantic representation m-dimensional vectors of patent i and technical topic i are denoted as P_i and T_i, respectively.

$$con_{T_i} = \sum_{1}^{n} con_{T_{i_n}} \times Time_{weight}, \quad (2)$$

$$Time_{weight} = \frac{e^{0.42}}{(t_0 + t_i)^{0.0225}}, \quad (3)$$

In Formula (2), con_{T_i} represents the weighted sum of contributions of technical topic T_i in all co-occurrences, $Time_{weight}$ is the dynamic time weight, T_0 is the current year, and T_i is the year when the nth convergence occurs.

2.2.3. Construction of the dynamic technical topic co-occurrence network

A dynamic technical topic co-occurrence network construction primarily involves the following two steps. The first step is to identify the technical topics present in the patent, we set the probability distribution threshold to 0.2 [15]. Technical topics exceeding this threshold are considered to be present in the patent, resulting in the generation of a technology co-occurrence matrix. Then we extract co-occurrence relationships using the networkx package, forming node pairs that represent technical topics. Finally, we mark the obtained technical topic contributions from Section 2.2.2 on the matching nodes, establishing a dynamic co-occurrence network of technical topics.

2.3. Prediction of technology convergence based on graph neural networks

We initially employ a graph neural network model to aggregate the structural and nodal attribute information of the technological co-occurrence network. This helps to address the issue of sparse feature dimensions in technology convergence prediction, resulting in a more accurate representation of node features. Secondly, we transform the research on predicting technology convergence into a link prediction problem. Probability scores are then calculated for the technology combinations formed between technical topic nodes, and the model's performance is evaluated using the AUC metric.

2.3.1. Node embedding based on graph neural network model

Graph neural network model can automatically capture high-level abstract representations of networks by aggregating low-level information, avoiding the need for complex feature engineering. These models combine both topological structure and attribute information for learning, effectively aggregating attribute features and topological structure information from neighboring nodes, to obtain a more accurate feature representation for the target node.

This paper uses technical topics in patents as nodes, with co-occurrence relationships between topics serving as edges in the graph. The technical contribution features are combined and used as node attribute information. Specifically, we first use the co-occurrence relationships in the training set as the graph structure. The contribution index of corresponding nodes is input as node attribute information into the graph neural network for training, thereby obtaining the embedding vectors of known technical topics. Secondly, using a link prediction model, we calculate the probability of fusion between technical nodes, obtaining fusion scores between nodes. The link prediction model is introduced in Section 2.3.2. Additionally, different graph neural network models have their own characteristics. This paper will compare models and choose the one most suitable for prediction of technology convergence.

2.3.2. Link prediction model based on probability ranking

The prediction of technology convergence can be simplified as predicting the emergence of a new link edge. In this context, technologies can be seen as nodes, and the relationships between them as convergence links. Thus, this paper transforms the task of predicting technology convergence opportunities into a link prediction problem for research.

The link prediction method proposed in this paper relies on a co-occurrence graph of technical topics, where the relationships between technical topics serve as edges. The technology contribution is trained as node features on the co-occurrence relationships of technical topics. Once the representations of the technical topic nodes are obtained, the probability score for the technology combination formed by two points is calculated. This probability score can be regarded as the link prediction score. The higher the score, the greater the possibility of a future link between the two nodes, indicating a higher probability of convergence between these two technical topics. Finally, we choose AUC as the evaluation metric to assess the performance of the prediction model based on graph neural networks.

3. Results

This paper generated three co-occurrence networks with different features, to compare and validate the effectiveness of the proposed method. The first network, T-Co1, only considers the frequency of co-occurrence of technical topics. The second network, T-Co2, includes centrality indices as features for technical topic nodes. The third network, T-Co3, integrates technical contribution as features for technical topic nodes. The centrality measure chosen here is degree centrality, which reflects the number of connections a node has. A higher degree centrality indicates a stronger node centrality, signifying greater importance.

For the three types of co-occurrence networks, we use three graph neural network models, namely GCN, GNN and GAT, to learn node representations, using link prediction for quantitative evaluation. The main difference between GCN and traditional GNN lies in the use of convolutional operators for information aggregation, while GAT uses self-attention mechanisms for node weight allocation. The results for different feature networks and methods are shown in Table 1.

Table 1
AUC of Different Methods

	T-Co1	T-Co2	T-Co3
GCN	0.7120	0.7106	**0.7998**
GNN	0.7106	0.7125	**0.7236**
GAT	0.6379	0.6411	**0.6731**

The results show that the performance of T-Co3 is generally superior to T-Co1 and T-Co2 across different model representations, with GCN performing best on T-Co3. In the GCN model, the AUC value of T-Co3 has increased by 8.78% compared to T-Co1 and 8.92% compared to T-Co2. In the GNN and GAT models, the AUC value of T-Co3 has also increased by 1.3% and 3.52%, respectively. Compared to other indicators of importance, the contribution index reflecting technological timeliness provides better, more comprehensive, and accurate clues for predicting technological convergence. And in this experiment, the GCN model performed better and showed better discriminative capabilities for different features. It is more suitable for the technology convergence prediction task in this paper.

4. Conclusion

This paper refines the assessment of technological importance from a timeliness perspective, shifting from traditional distinctions based on lifecycle and dates to a more precise measurement within each convergence event. We replace frequency indicators in the co-occurrence network with the technological contribution index for building dynamic technology networks. The results show that this approach outperforms frequency-based models. As a next step, we aim to improve the technological timeliness index by incorporating additional temporal cues. In addition, the exploration of more efficient embedding models is expected to improve predictive performance.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 72374103, 71974095) and the Postgraduate Research Practice Innovation Program of Jiangsu Province (No. SJCX23_0161).

References

[1] C. S. Curran, J. Leker, Patent Indicators for Monitoring Convergence – Examples from Nff and Ict, Technological Forecasting & Social Change 78 (2011) 256-73. doi:10.1016/j.techfore.2010.06.021

[2] M. Karvonen, T. Kassi, Patent Citation Analysis as a Tool for Analysing Industry Convergence, 2011 Proceedings of PICMET' 11: Technology Management in the Energy Smart World (PICMET), IEEE, Portland, USA, 2011.

[3] K. Sung, H. K. Kong, and T. Kim, Convergence Indicator: The Case of Cloud Computing, Journal of Supercomputing 65 (2013) 27-37. doi:10.1007/s11227-011-0706-1

[4] I. Park and B. Yoon, Technological Opportunity Discovery for Technological Convergence Based on the Prediction of Technology Knowledge Flow in a Citation Network, Journal of Inforindexs 12 (2018) 1199-222. doi:10.1016/j.joi.2018.09.007

[5] Li C, Zhou J, Yang Z, Research on the Identification of Technology Fusion Growth Points from the Perspective of Dynamic Evolution Process, Library And Information Service 66 (2022) 99-109. doi: 10.13266/j.issn.0252-3116.2022.07.010

[6] Zhang J, Li Y, Technology Convergence Prediction by the Semantic Representation of Patent Classification Sequence and Text, Journal of the China Society for Scientific and Technical Information 41 (2022) 609-24. doi:10.3772/j.issn.1000-0135.2022.06.006

[7] Zhang Y, Zhang G, Chen H, Porter AL, Zhu D and Lu J, Topic analysis and forecasting for science, technology and innovation: methodology with a case study focusing on big data research, Technological Forecasting And Social Change 105 (2016) 179-91. doi:10.1016/j.techfore.2016.01.015

[8] Wang L, Liu X, Measuring Diffusion of Technology Topics with Patent Data, Data Analysis and Knowledge Discovery 6 (2022) 1-10. doi: 10.11925/infotech.2096-3467.2021.0915

[9] Daim T, Preschitschek N, Niemann H, Leker J, G. Moehrle M, Anticipating industry convergence: semantic analyses vs IPC co-classification analyses of patents, Foresight 15 (2013) 446-64.

[10] Liu W, Yang Z, Cao Y, Huo J, Discovering the influences of the patent innovations on the stock market, Information Processing and Management 59 (2022) doi: 10.1016/j.ipm.2022.102908

[11] Zhang Y, Wu M, Miao W, Huang L, Lu J, Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies, SSRN Electronic Journal 15 (2021). doi:10.1016/j.joi.2021.101202

[12] Caferoglu H, Elsner D, Moehrle MG, The Interplay Between Technology and Pre-Industry Convergence: An Analysis in the Technology Field of Smart Mobility, IEEE Transactions on Engineering Management 70 (2021) 1504-1517. doi:10.1109/TEM.2021.3092211

[13] Fu Q, Sun Y, Evaluation of Technological Influence: Based on Patent Timeliness and PageRank Algorithm, Journal of Systems & Management 27 (2018) 352-8.

[14] Wenjing Zhu, Bohong Ma, Lele Kang, Technology convergence among various technical fields: improvement of entropy estimation in patent analysis, Scientometrics 127(2022) 7731-50.

[15] J. Ma, C. Wang, L. Yan, and S. Yao, Analysis of Patent Technology Topic Evolution Based on Product Life Cycle, Journal of the China Society for Scientific and Technical Information 41 (2022) 684-691. doi:10.3772/j.issn.1000-0135.2022.07.003

[16] Zhang X, Liu H, Shi J, Mao C, Meng G, LSTM and artificial neural network for urban bus travel time prediction based on spatiotemporal eigenvectors, Journal of Computer Applications 41(2021) 875-880. doi:10.11772/j.issn.1001-9081.2020060467

[17] Le, Q. V. and Mikolov, T, Distributed Representations of Sentences and Documents, JMLR.org, 2014. doi:10.48550/arXiv.1405.4053.

[18] Jiang Z, On the forgetting function - a mathematical discussion on the psychology of memory, Advances in Psychological Science 56(1988) 56-60.