An Approach for Identifying Complementary Patents Based on Deep Learning

Jinzhu Zhang [1], Jialu Shi [1]

[1] *Nanjing University of Science and Technology, No.200 Xiaolingwei Street, Nanjing, 210094, China*

Abstract

As technological complexity continues to increase, it is imperative to analyze and identify complementary relationships among patents in order to facilitate the recombination of multi-domain technical knowledge and foster innovation. Hence, this research proposes to identify patent relationships from the perspective of complementarity according to IPC classification numbers. Specifically, representation learning is performed on the external structure and textual content of patents, and on this basis, a convolutional neural network with attention mechanism is utilized to mine the complementary relationships between patents. It is mentioned that a complementary patent dataset is generated based on the IPC classification numbers of the patents for model training. Empirical analysis in the field of new energy vehicles demonstrates that this approach can effectively identify potential patent complementarities, which can facilitate breakthroughs in key core technologies for enterprises and countries.

Keywords

Complementary patent identification; Patent structural feature representation; Patent textual feature representation

1. Introduction

The complementarity of patents in technology refers to the degree to which two patent subjects with the same broad field of technology focus on different narrow domain technologies [1]. Complementary patent identification exposes the impact and integration of technologies across multiple domains, thereby eradicating knowledge barriers between them and promoting innovative synergy. Additionally, it can stimulate crucial technological advancements that cover various technologies, and enables enterprises to develop new competitive advantages [2].

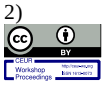
Existing research related to technology mining and analysis is usually based on the similarity between patents, ignoring patent complementarity. It is essential to further utilize the distinctions among various technology components, then form a standardized dataset comprising complementary patents. In addition, the present research relies primarily on patent classification or citation information, which may not accurately reflect the patent complementarity in regards to technical content. Therefore, it is necessary to incorporate their textual content in order to delve deeper into their complementarity. Moreover, certain studies assess patent complementarity according to the co-occurrence of patent classifications, which may not exactly indicate the degree of complementarity between patents. Hence, there is a need to integrate deep learning techniques to establish a quantitative approach to identify complementary patents.

To address the above issues, this paper first constructs a complementary patent dataset for subsequent model training based on the IPC classification numbers arranged in a hierarchy. Secondly, drawing on state-of-the-art representation learning techniques, we characterize the external structure and textual content of patents, and comprehensively generate

Joint Workshop of the 4th Extraction and Evaluation of Knowledge Entities from Scientific Documents and the 3rd AI + Informetrics (EEKE-AII2023), June 26, 2023, Santa Fe, New Mexico, USA and Online

EMAIL: zhangjinzhu@njust.edu.cn

ORCID: 0000-0001-7581-1850 (A. 1); 0000-0002-4447-9952 (A. 2)



the semantic representations. In the end, based on the above two steps, we propose a complementary patent identification approach, which employs deep learning-based model training to evaluate the degree of complementarity between distinct patents.

2. Data and Method

In order to fully exploit the complementary relationships between patents, we firstly gather the complete patent records in the field of new energy vehicles. Then we perform patent semantic representation using external structures and textual contents with a deep representation learning method. Finally, we propose a complementary patent identification method based on deep learning techniques. In this method, a dataset of complementary relationships annotated via IPC classification numbers is constructed for model training. The proposed approach is comprised of three main components, as illustrated in Figure 1.

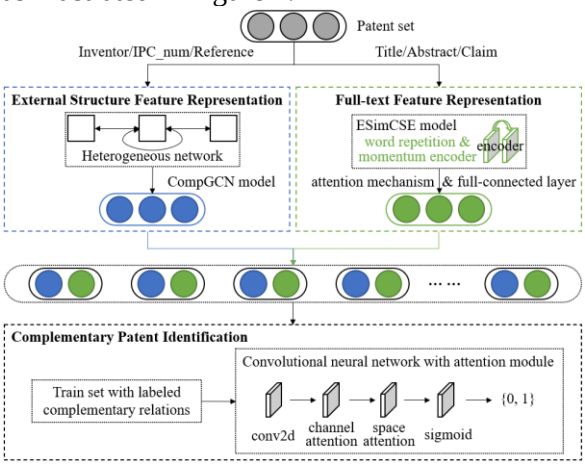


Figure 1: Overall framework of the proposed complementary patent identification method

2.1. Data Description

To demonstrate the application of the proposed approach, we have selected the field of new energy vehicles for our study. This innovative area encompasses a wide range of industry chains, multiple actors and numerous links. Our study collected full-text patent data from USPTO for new energy vehicles published in 2022. We narrowed our search to the title, abstract, and claim of patents, utilizing keywords new energy vehicle (automobile), hybrid vehicle, electric vehicle, plug-in electric vehicle and fuel vehicle. It is worth noting that all of these terms were established by the Ministry of Industry and Information Technology of the People's Republic of China in 2009. Furthermore, we limited our search to invention patents and ultimately retrieved 8267 patents.

Regarding the establishment of dataset, we utilize the IPC classification numbers to determine the technical fields of patents. If two patents are focused on different technical features but both fall into the same technical category, we categorize them as complementary patents [1]. Specifically, we determine whether a patent pair belongs to the same subclass-level but has distinct group-level according to the IPC classification numbers and assign a binary label of 0 or 1 to indicate their relationship.

2.2. Patent External Structure Feature Representation for Complementary Patent Identification

Initially, the multi-layer association relationships between patents are utilized to construct a patent heterogeneous network denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{O}, \mathcal{R})$, where the nodes in $\mathcal{V}$ contain multiple external features, $\mathcal{E}$ represents the connections between patents and their corresponding features, $\mathcal{O}$ involves various node types in $\mathcal{V}$, while $\mathcal{R}$ denotes the linkage types, such as cite/cited, invent/invented, include/belong, and others. Afterwards, we employ the CompGCN graph neural network model [3] to produce the representation of patents.

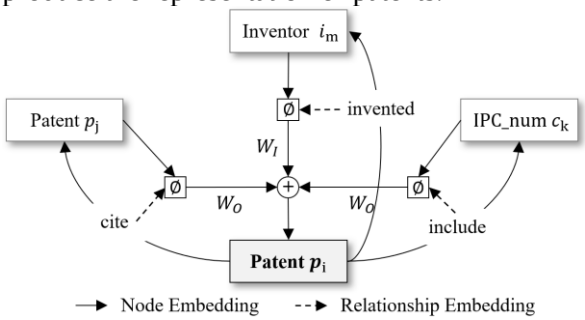


Figure 2: Node and edge aggregation process in the patent heterogeneous information network

Figure 2 shows the nodes and edges aggregation process within the heterogeneous network $\mathcal{G}$. First, a combinatorial operation $\emptyset(\cdot)$ is applied to each edge of the neighborhood of patent p_i based on its initial node and relation embedding. Then the combinatorial embedding is represented using specific weights, W_O and W_I, for convolution of the original and inverse

relations. Finally, the central node's final embedding, patent p_i, is obtained by aggregating messages from all its neighbors.

2.3. Patent Full-text Feature Representation for Complementary Patent Identification

To begin with, sentence vectors are learned through the ESimCSE sentence embedding model [4], which effectively captures the contextual information of the sentences in the text. Afterwards, we utilize attention mechanism to weigh the sentences and calculate the correlation between each dimension in the sentence vectors. Equation 1 demonstrates the specific calculation, where $o_l(s_a, s_b)$ measures the matching degree of the l th pair of text features between the two sentence vectors (s_a, s_b) in the patent. Furthermore, f^l represents the unique heat vector representation of the lth pair of features, with only the lth element being equal to 1 and the rest being 0. The ReLU (Rectified Linear Unit) activation function is represented as $\phi(\cdot)$.

$$o_l(s_a, s_b) = w^T \phi\left(W_{att}\left(l_a \odot l_b \odot f^l\right) + b\right) + c \quad (1)$$

2.4. Complementary Patent Identification Process

By treating complementary patent identification method as a classification problem, we utilize the semantic vector of patents p_i and p_j as input data and employ the CBAM module [5] to attend to channels and spaces that contain crucial information. This involves the assignment of varying weights to different positions in each channel based on the relevance of the information. The structure of CBAM is illustrated in Figure 3.

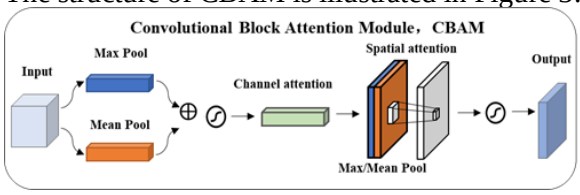


Figure 3: Convolutional block attention module schematic

Afterward, the characteristics are enhanced by a multi-layer neural network. Subsequently, we apply a shifted sigmoid function to the combined features to determine the conclusive term, as Equation 2. Finally, we obtain either a label with 1 or 0 to indicate the existence or absence of a complementary relationship.

$$f_{p_i,p_j} = \sigma\left(\eta_d - c_{p_i,p_j}\right) = \frac{1}{1 + exp\left(c_{p_i,p_j} - \eta_d\right)} \quad (2)$$

2.5. Evaluation metrics

To evaluate the performance of this approach in identifying complementary patents, a quantitative assessment can be conducted by utilizing both predicted outcomes and actual complementary relationships in the test set, through metrics such as precision, recall, and f1 score.

3. Result

To affirm the effectiveness of the proposed approach, a comparison with different methods that utilize either the structural or textual dimension was conducted. As depicted in Table 1, all the three methods yielded precision above 85%, demonstrating their ability to accurately identify complementary relationships with a high degree of matching and predict effectively. Out of the three methods, the combined structure and text method produced the greatest precision rate of more than 90%. Notably, the textual dimension method had the highest recall rate, indicating its proficiency in accurately identifying complementary relationships with a limited degree of matching. In particular, our proposed method attained the highest F1 score, indicating its superiority in identifying complementary relationships between patents.

Table 1

Precision, Recall and F1 score on different methods

Method	Precision	Recall	F1_Score
Structural	85.7	73.1	78.9
Textual	86.8	**76.9**	81.6
Structural& Textual	**90.3**	75.8	**82.4**

In addition, we have computed the probabilities of complementary associations among patent pairs in the empirical dataset, and carried out a specialized examination of the ten

patent pairs exhibiting the maximum complementary relationship probabilities. Table 2 presents the specifics, including patent numbers and complementary probabilities of these ten patent pairs.

Table 2

Patent complementarity prediction results based on the proposed method

Patent_nums	Probs	Patent_nums	Probs
US11431046 US11522184	0.907	US11495861 US11322313	0.865
US11289723 US11437666	0.901	US11294551 US11289723	0.864
US11502350 US11243260	0.889	US11237217 US11277011	0.857
US11322313 US11502350	0.881	US11515534 US11335926	0.856
US11225166 US11294551	0.876	US11217793 US11431046	0.856

High probability indicates the likelihood of room for cooperation, so we take the first patent pair as an example to carry out the analysis. The first patent, US11431046, describes an energy storage device that can function as an electrochemical battery with both positive and negative electrodes. Conversely, patent US11522184 presents a technique for preparing positive active material. This technique can be utilized as one of the pathways to enhance the energy storage efficiency of the former patent, and thereby a potential direction for collaboration. Thus, this approach is proved to be valuable in assessing the possibility of a patent complementary relationship as well as determining technological compatibility, which can help enterprises in making informed decisions.

4. Conclusion

Our paper aims to examine the correlation between patents in terms of complementarity and proposes an approach for complementary patent identification utilizing deep learning. This approach solely utilizes fundamental patent data as input with minimal pre-processing and eliminates the need for costly and labor-intensive feature engineering. We confirmed the effectiveness of this method in identifying complementary relations for new energy vehicle patents by conducting ablation experiments. Our forthcoming research will involve adding patent image data to structural and textual features to boost patent retrieval and matching tasks.

5. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 71974095) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. SJCX22_0152).

6. References

[1] Marianna Makri, Michael A. Hitt,Peter J. Lane. (2010), "Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions". Strategic Management Journal, 31 (6): 602-628. doi: https://doi.org/10.1002/smj.829.

[2] Douglas Henrique Milanez, Leandro Innocentini Lopes de Faria, Roniberto Morato do Amaral,José Angelo Rodrigues Gregolin. (2017), "Claim-Based patent indicators: A novel approach to analyze patent content and monitor technological advances". World Patent Information, 50 (9): 64-72. doi: https://doi.org/10.1016/j.wpi.2017.08.008.

[3] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin,Partha Talukdar. (2020), "Composition-based multi-relational graph convolutional networks". Proceedings of the 8th International Conference on Learning Representations (ICLR 2020). Addis Ababa, ETHIOPIA: https://doi.org/10.48550/arXiv.1911.03082.

[4] Xing Wu, Gao Chaochen, Zang Liangjun, Han Jizhong, Wang Zhongyuan,Songlin Hu. (2022), " ESimCSE: Enhanced sample building method for contrastive learning of unsupervised sentence embedding". Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, KOREA: 3898–3907.doi: https://doi.org/10.48550/arXiv.2109.04380.

[5] Sanghyun Woo, Jongchan Park, Joon-Young Lee,In So Kweon. (2018), "CBAM: Convolutional Block Attention Module". Proceedings of the 15th European Conference (ECCV 2018). Munich, Germany: https://doi.org/10.48550/arXiv.1807.06521.