Material performance evolution discovery based on entity extraction and social circle theory*

Jinzhu Zhang[1,*], Wenwen Sun

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

Abstract

Topic evolution analysis describes the emerge, develop, and extinct of topics in a field, which can help researchers understand the history and current situation of the research field. However, the material patent text has a certain domain specificity, and the general entity extraction models cannot extract special entities effectively. Moreover, the belief that topics with high similarity have evolution relationship contradicts the rule of "first the change, then the new topic", which cannot clearly present the dynamic changes and accumulation of topics. Therefore, we design a method to extract the material performance entities accurately and construct dynamic evolution path for material performance topics. Firstly, we propose a material entity extraction model BERT-BiLSTM-CRF, which integrates syntactic dependency analysis and attention mechanism, realizing the accurate extraction of material performance entities. Secondly, we design an algorithm for identifying the evolution relationship between performance nodes based on ring boundaries, which can mine the evolution relationship between performance nodes and existing topics, realizing the dynamic accumulation and change of topics. Finally, we construct the dynamic evolution path of material performance, exploring the complex associations of material performance. Experiments in the field of metal materials confirm that the proposed method can effectively construct the dynamic evolution path of material performance topics, which makes the evolution relationships between topics more abundant and interpretable.

Keywords

Entity extraction, material performance evolution, patent entity relationship

1. Introduction

With the emergence of a large number of patents on materials manufacturing and materials innovation, it has become critical to explore the complex associations and evolutionary trends of material performance. Such exploration can help researchers deepen their understanding of material performance and promotes the invention of new materials [1].

Current researches perspectives on the evolution of material performance are mainly divided into the three perspectives: "performance evolution - microstructure" , "performance evolution - microstructure - manufacture process" , and "performance evolution - manufacture process". The

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); 0009-0007-2773-4679(W. Sun)

specific relationships are shown in Figure 1:

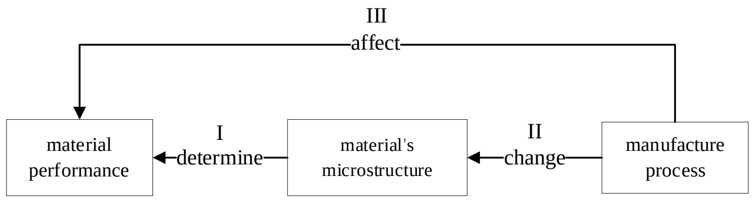

Figure 1: Different Perspectives in the Analysis of The Evolution of Material Performance

Among them, Perspectives I and II [2, 3, 4] involve microstructure, require high levels of expertise and experimental equipment from researchers and readers ; Perspectives III [5] usually proceed in the form of controlling variables, such as varying the temperature of a certain process, then exploring the influence of on the microstructure of the material and the evolution of the material performace, which to some extents restricts the comprehensive understanding and description of the material performance evolution.

In terms of evolution analysis methods, researchers usually use the following approaches:(1) topic identification [6]; (2) topic evolutionary analysis [7, 8]. However, the former is difficult to reveal document semantics effectively, resulting in poor interpretability of the topics [9] . The latter only reflects the relative importance or attention of a topic at a specific time point [10]. Moreover, most of the existing studies are based on topic similarity, believing that there is an evolution relationship between two topics whose similarity above a threshold [11, 12]. But the topic itself is dynamically changing and accumulating, there should be " first the change in material performance, then the new material performance topic ", so the establishment of an evolution relationship based on the similarity between topics is biased.

Therefore, this paper takes the material performance as the research object. Particularly, we propose a method for constructing the topic evolution path by introducing the social circle theory, realizing the dynamic accumulation of material performance topics and the construction of evolution path. Finally, we explore the complex associations in the evolution of material performance on the basis of the dynamic evolution path.

2.Data and method

This paper takes the material performance as the research object. Firstly, we integrate syntactic dependency analysis and attention mechanism to construct the material entity extraction model (BERT-BiLSTM-CRF), we obtaining the performance nodes of each material. And then divided the performance nodes of all materials into time batches by year. After that, we designed an algorithm based on the initial performance topics, to realize the dynamic accumulation of material performance topics and the construction of evolution path.

2.1 Data sources and preprocessing

We considered that to do material evolution, if we collect data at random intervals, it may lead to a lack of completeness and accuracy in the final analysis of the evolution results.So we takes the concept of Germany's "Industrie 4.0" as the background, and selects metal material as an example, which is one of the key foundational materials closely related to this concept. Then we use the Derwent Innovations Index database as the patent data retrieval platform. The patent search expression is "TS=('Metal materials' OR 'Metallic materials' OR 'Metal alloys' OR 'Metal compositions' OR 'Metal-based materials') AND WC=('Materials Science')", with a time interval from 2011 to 2023, where 2011 is the year when the concept of "Industry 4.0" was first introduced. Then, the top 10,000 relevant patent texts were selected as the dataset. In addition, considering the number of patents in each period, we divide it into year batches for material performance evolution.

2.2 Method for extracting the material entities

Under the background of the continuous improvement of the manufacture process, the processing method of materials is constantly progressing. At the same time, the change in the manufacture process of the material will bring about changes in the material performance [13].

Therefore, we defined the performance entity and manufacture process entity of metal material. In addition, we refer to the relationship shown in Fig. 1, establishing the causal relationship between the two for subsequent analysis. (Among them, the manufacture process entity and causal relationship will be used in our next step of exploring the reasons for performance evolution, so it is rarely involved in this paper.)

Then, considering the content of material patents contains a large number of technical terms, material components, we constructed an entity extraction model (BERT-BiLSTM-CRF) by combining syntactic dependency analysis and attention mechanism. The combined use of these methods provides a more accurate extraction of the material performance entities and manufacture process entities from patent contents, providing a basis for subsequent material analysis and research.

2.3 Method for constructing dynamic evolution path for material performance topics

In this part, we first define six evolution types, then we designed an algorithm for identifying the evolution relationship of performance nodes based on ring boundaries. Finally, we present the detailed process of the method for constructing dynamic evolutionary path of material performance topics.

2.3.1 Social Circle Theory

Social circle theory suggests that the social circle formed around a person reflects the closeness of his or her social relationships. That is, a person's intimate social circle usually consists of relationships with a high degree of relevance; then followed by the normal friends circle and the strangers circle. In addition, there may exist such a part of people in the sea of people: they are temporarily outside your normal friends circle, but there are certain similarities between each other, and they may become your friends or even intimate friends in the future, so this paper defines them as potential friends. Therefore, centered on the individual, their affinity rank order is: intimate friends, normal friends, potential friends, strangers, and the position belongs to: within the intimate friends circle, outside the intimate friends circle within the normal friends circle, outside strangers circle within the normal friends circle, outside the strangers circle, specifically as shown in Figure 2.

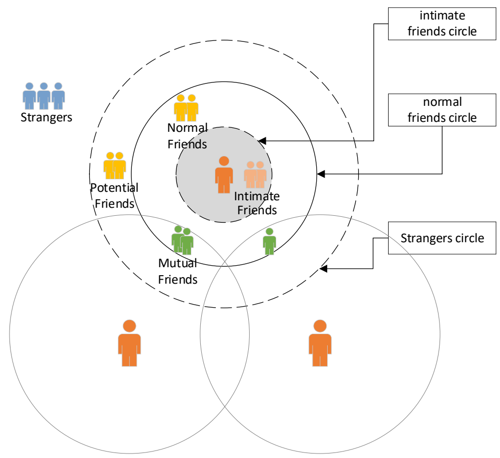

Figure 2: Social Circle Theory (See Appendix A for Detailed Picture)

We refer to this theory and combine it with existing research to define six evolution types. And propose an algorithm for identifying the evolution relationship of performance nodes, the details are shown in parts 2.3.2 and 2.3.3 respectively.

2.3.2 Definition of evolution types

This paper defines six evolution types based on existing studies. Among them, the four types of develop, evolve, emerge, and fuse are derived from four different social relationships in the social circle theory, and the two types of extinct and split refer to the existing studies to ensure the diversity of evolution types. In addition, this paper also improves the fuse and split types, by further refining the different contributions of each theme in them, which helps to consider the dynamic interactions between themes in more detail. See Appendix B for details.

2.3.3 Identifying the evolution relationship of performance nodes based on ring boundaries

We refer to social circle theory and improve the model proposed by Zhang et al. [14], proposing algorithm for identifying the evolution relationship of performance nodes based on ring boundaries, the specific algorithm and its correspondence are shown in Figure 3. Firstly, for the existing performance topics, the centroid of each topic is calculated, and the maximum Euclidean distance between each topic's patent and its centroid is taken as the topic boundary. The topic boundary is extended outward by a ratio less than 1 to obtain the outer ring boundary and shrunk inward by the same ratio to obtain the inner ring boundary. After several comparison tests, we finally set the ratio in this study to 0.2.

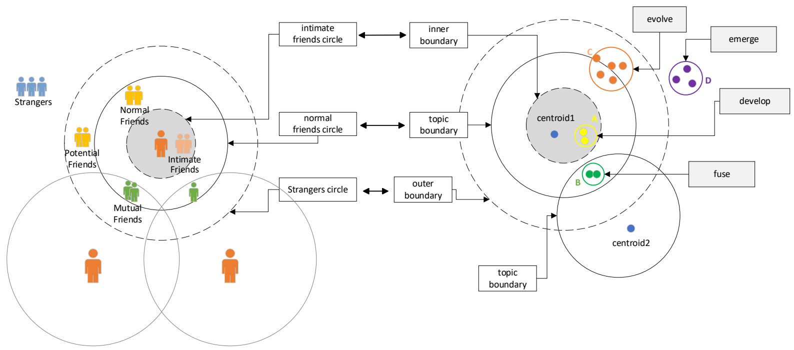

Figure 3: Algorithm for Identifying the Evolution Relationship of Performance Nodes based on Ring Boundaries (See Appendix C for Detailed Picture)

Subsequently, for material performance nodes in subsequent batches, the Euclidean distance between each performance node and the centroids of existing performance topics is calculated separately to identity the evolution relationship between them. The specific rules are as follows:

a) **develop**: inside the inner ring boundary
b) **evolve**: outside the inner ring boundary and inside the outer ring boundary
c) **emerge**: outside the outer ring boundary
d) **fuse**: boundary intersection

Then, hierarchical clustering is introduced to obtain the different types of performance topics in the batch, and merge similar topics that exceed a threshold (we set it to 0.8 in this paper). For a topic in the previous batch, if its number of topics obtained more than two in this batch, the evolution type is considered as split. Furthermore, in the construction of the evolutionary path, if a topic has no evolution relationship with the following topics, we consider it as extinct type.

2.3.4 Construction of the dynamic evolution path for metal material performance topics

The construction of the dynamic evolution path of metal material performance mainly includes the following steps, which are shown in Figure 4. Firstly, after the extraction of performance entities of each material, we get the performance node of each material, and then, all material performance nodes are divided into time batches according to the year. Secondly, the K-Means algorithm is used to cluster the first batch of data to obtain the initial performance topics.

Subsequently, for performance nodes in subsequent batches, the algorithm for identifying the evolution relationship of performance nodes (see Section 3.3.3 for the specific algorithmic process) is used to identify their evolution relationships with each performance topic. Then, hierarchical clustering is introduced to obtain the performance topics of different evolution types in this batch, and merge similar topics that exceed a threshold. Finally, incremental iterations are carried out in the above manner to obtain the material performance topics at different year batches, thereby achieving the dynamic construction of the material performance evolution path. (see Appendix D for the entire picture, where Cluster stands for topic)

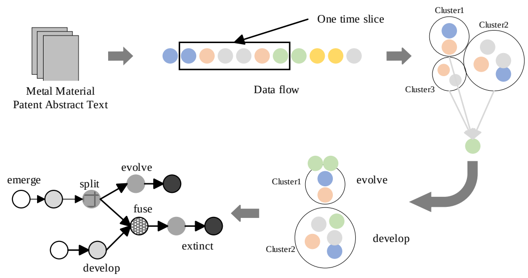

Figure 4: Process of Constructing The Evolution Path of Metal Material Performance

3.Result

In the construction of the evolution path of metal material performance, we use examples from the years 2021-2023 to obtain the results. Specifically, the number of performance clusters in 2021, 2022, and 2023 are 66, 55 and 60. The result of the evolution path are shown in Figure 5, where yellow, red, and green represent the years 2021, 2022, and 2023 respectively. (see Appendix E for the entire picture, where Cluster stands for topic)

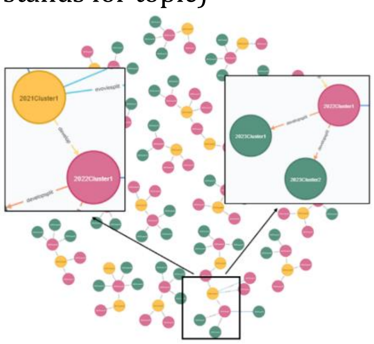

Figure 5: Evolution Path of Metal Material Performance from 2021 to 2023

From the results of the evolution path above, it can be observed that in 2022, Cluster1 developed from the Cluster1 in 2021. The contents of the two Clusters are as follows: ['high excellent low corrosion', 'resistance good powder strength temperature'], ['high excellent alloy low mechanical', 'strength corrosion resistance process good']. It is not difficult to find that both Clusters focus on improving the corrosion resistance and mechanical performance of metal materials, which aligns with the practical application requirements of metal materials [15].

In 2022, Cluster1 further developed and split into Cluster1 and Cluster2 in 2023. The contents of the three Clusters are as follows: ['high excellent low corrosion', 'resistance good powder strength temperature'], ['good surface heat', 'resistance layer wear low'], and ['resistance good base layer wear', 'layer wear surface low heat']. As can be seen, the 2023 Cluster1 maintained the original corrosion resistance and wear resistance performance of Cluster1 in 2022, and further improved the surface heat performance of metals. This may be achieved through improvements in material technology and alloy additions, thus enhancing the performance in practical applications.

4. Conclusion

This paper proposes an algorithm for identifying the evolution relationship of performance nodes based on ring boundaries, which can not only realize the dynamic accumulation and construction of metal material performance evolution path, but enrich and improve the topic evolutionary analysis method. Currently, we are combining the manufacture process entities of each material to further analyze the causes of the evolution of material performance in depth, and to better understand the evolution trends and the changing patterns of material performance.

Acknowledgments

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. KYCX23_0632).

References

[1] Kaushal Jha, Suman Neogy, Santosh Kumar, R N Singh, & G K Dey. Correlation between microstructure and mechanical properties in the age-hardenable Cu-Cr-Zr alloy. Journal of Nuclear Materials: Materials Aspects of Fission and Fusion, 2021, 546, 152775. doi: 10.1016/j.jnucmat.2020.152775.

[2] Enyu Guo, Guohua Fan, & Tongmin Wang. Microstructure Evolution Mechanism of Metallic Materials: Progress on In Situ Studies Using Synchrotron Radiation Source. Failure Analysis and Prevention, 2021, 16(01): 1-14+91. doi: 10.3969/j.issn.1673-6214.2021.01.001.

[3] ZhiGuo Liu, Ri You, & Jintao Wu. Study on relationship between dam concrete material performance evolution and its pore structure. WATER RESOURCES AND HYDROPOWER ENGINEERING, 2016, 47(11): 25-28+35.

[4] Yinzhi Li, Gang Liu, Xinzhe Chen, Zhenhua Su, Lianwu Yan, & Yingbiao Peng. Effects of Sintering Temperature and Holding Time on the Microstructure and Properties of Ti(C,N)-Based Cermets[J]. Packaging Journal, 2023, 15(03): 37-45.

[5] Yang Liu, Baoliang Liu, Shi Kai, Xiaoyuan Han, Yi Xia, & Jianzhao Shang. Effects of different zirconia-based raw materials on properties of MgO-Al-C materials. REFRACTORIES, 2023, 57(01): 59-64. doi: 10.3969/j.issn.1001-1935.2023.01.013.

[6] Lee J W, & Han D H. Data Analysis of Psychological Approaches to Soccer Research: Using LDA Topic Modeling. Behavioral Sciences, 2023,13(10), 43-47. doi: 10.3390/bs13100787.

[7] Parlina A, Ramli K, & Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. Information, 2020, 11(2): 69. doi: 10.3390/info11020069.

[8] Zibiao Li, & Li Zhang. Evolution of Patented Technology Topics of Steel Materials Based on LDA Model. SCIENCE AND TECHNOLOGY MANAGEMENT RESEARCH, 2020, 40(24): 175-183. doi: 10.3969/j.issn.1000-7695.2020.24.023.

[9] Behrouzi S, Sarmoor Z S, Hajsadeghi K, & Kavousi K. Predicting scientific research trends based on link prediction in keyword networks. Journal of Informetrics, 2020, 14(4): 101079. doi: 10.1016/j.joi.2020.101079.

[10] Wu Z, Xie P, Zhang J, Zhan B, & He Q. Tracing the Trends of General Construction and Demolition Waste Research Using LDA Modeling Combined With Topic Intensity. Frontiers in Public Health, 2022, 10(3), 22-26. doi: 10.3389/fpubh.2022.899705.

[11] Jinxia Liu, Qianqian Hou, Jing Du, Fuhou Chai, & Li Zhang. Hot Topics Evolution Research of Emerging Field Under the Sub Topics and Vocabulary Related Perspective. Journal of Intelligence. 2023,42(3):123-129. doi:10.3969/j.issn.1002-1965.2023.03.017.

[12] Reza T H S, Sadegh A, & Soroush T. An embedding approach for analyzing the evolution of research topics with a case study on computer science subdomains. Scientometrics, 2023,128(3): 11-16. doi: 10.1007/s11192-023-04642-4.

[13] Qing-Miao Hu, & Rui Yang. The endless search for better alloys. Science, 2022, 378: 26-27. doi:10.1126/science.ade5503

[14] Zhang Y, Zhang G, Zhu D, & Liu J. Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics. Journal of the Association for Information Science and Technology , 2017, 68(8): 1925-1939. doi: 10.1002/asi.23814.

[15] Y.B. Lei, Z.B. Wang, B. Zhang, Z.P. Luo, J. Lu, & K. Lu. Enhanced mechanical properties and corrosion resistance of 316L stainless steel by pre-forming a gradient nanostructured surface layer and annealing. Acta Materialia. 2021, 208, 116773. doi: 10.1016/j.actamat.2021.116773.