Topic Evolution Path and Semantic Relationship Discovery Based on Patent Entity Relationship*

Jinzhu Zhang†
Department of Information Management, School of Economics & Management
Nanjing University of Science and Technology
Nanjing China
zhangjinzhu@njust.edu.cn

Linqi Jiang
Department of Information Management, School of Economics & Management
Nanjing University of Science and Technology
Nanjing China
sufi_jiang@163.com

ABSTRACT

Topic evolution analysis describes the emergence, transition, and extinction of a topic in a technical field, which can help researchers understand the history and current situation of the research field. Current studies is mainly patent text-based methods, which often uses relationships among keywords to construct co-occurrence network and analyses evolution using topic clustering algorithms. However, it didn't consider all the words in the patent and the semantic relationship between them. In addition, the relationships among topics should be more concrete, we should not only find the evolution relationship, but also need to reveal the semantic relationships among topics. Therefore, this paper uses representation learning method to get the semantic representation of each entity/word, and computes the semantic similarity among them to find out pairs of words which are different but with the same meaning in a special context. Moreover, we define multiple semantic relationships among topics, and design a method to use patent entity relationships to obtain the semantic relationships among topics. Experiments in the technical field of UAV transportation have confirmed that the method in this paper can effectively identify the evolutionary relationship between topics and the semantic relationship between topic, Make the evolutionary relationship between topics more abundant and Interpretable. And provide a reference for further enriching and improving the topic evolution analysis method.

KEYWORDS

Topic Evolution Path, semantic relationship between topic, Patent Entity Relationship

ACM Reference format:

Jinzhu Zhang, Linqi Jiang. 2021. Topic Evolution Path and Semantic Relationship Discovery Based on Patent Entity Relationship. In *EEKE'21,September 27-30,2021,Illinois,USA. ACM, New York, NY, USA, 2 pages.* https://doi.org/nnnnnnn/nnnnnnn

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Topic evolution analysis describes the emergence, transition, and extinction of a topic in a technical field, which can help researchers understand the history and current situation of the research field. The result can quickly identify research hotspots, trends and gaps, which is essential to scientific and technological innovation (Liu H,2020).

In the study of topic evolution analysis, topic evolution path and relationship discovery play important role in related research. Current studies could be classified into two classes, including patent citation analysis-based and patent text-based methods (Yu D,2020). This paper focuses on the latter method, which often uses relationships among keywords to construct co-occurrence network and analyses evolution using topic clustering algorithms (No H. J,2015). Then the evolution path and relationship are discovered through comparisons of common keywords among topics in different time series.

However, these common keywords cannot cover the pair of words which are different but with the same meaning in a special context. In addition, the relationships among topics should be more concrete, for example, we should not only find the evolution relationship (i.e., emergence, transition, and extinction), but also need to reveal the semantic relationships among topics (i.e., function-realization or function-area).

Therefore, this paper uses representation learning method (Birunda,2021) to get the semantic representation of each entity/word, and computes the semantic similarity among them to find out pairs of words which are different but with the same meaning in a special context. Moreover, we define multiple semantic relationships among topics, and design a method to use patent entity relationships to obtain the semantic relationships among topics.

2 Data and method

Firstly, a manual labelled dataset is made and a neural network model is trained to extract all patent entities. Then the topic is identified by the clustering method and the evolution path is detected based on semantic similarity among. Finally, a neural network model is trained to extract the relationship between patent entities, and the semantic relationship among topics is discovered through relationships between patent entities

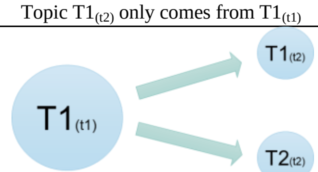
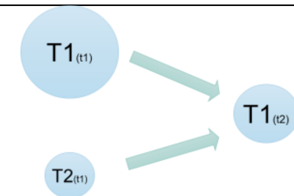
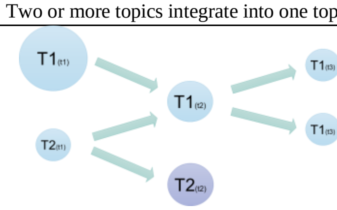
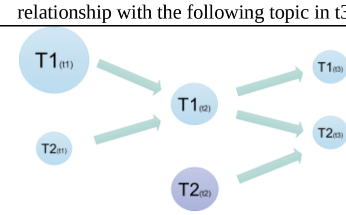
2.1 Data collection

This paper uses the Derwent Innovations Index database as the patent data retrieval platform and the data is retrieved on July 2019. The patent search expression is “IP = B64* AND TI= (((un-manned OR automatic OR autonomous OR remotely piloted OR nonhuman) AND (aircraft OR "aerial vehicle" OR airship* OR drone OR plane OR aircraft* OR airplane OR aerobat* OR aerostat*)) OR "UAV"”, the time interval is from 2008 to 2020. A total of 4507 patents with title, abstract, patent application date and other features are retrieved and processed as the data source. It is divided into different time series for evolution analysis considering the number of patents in each time period.

2.2 Discovery of Topic Evolution Path Based on Semantic Similarity Among Entities

It has following four steps for semantic similarity among entities. Firstly, a subset about training and testing set is made where the entities are manually labelled. Secondly, a BiLSTM-CRF (Lample, G.2016) model is trained on this dataset and evaluated through quantitative indicators. After 100 training iterations, the accuracy of the model exceeded 90% and became stable. Similarly, the loss dropped to below 5.8 and stabilized. Thirdly, this model is used to detect entities on all patens of each patent. Finally, K-Means is applied for clustering topics and get entities of each topic. Patent documents contain a lot of long professional vocabulary. Compared with commonly used LDA, the identified patent entity will not lose professional information. Fourthly, a word representation learning method is applied and the semantic similarity among entities of each topic could be calculated.
Based on semantic similarity among entities, a similarity threshold is determined, in which two different entities could be treated as the same meaning if the similarity is higher than threshold. We define five topic evolution patterns, including development, division, integration, extinction, and emerging. They are shown in Table 1, in which the size of the circle represents the number of entities under each topic. Moreover, $T1_{(t1)}$ and $T2_{(t2)}$ means the topic T1 and T2 at time t1 and t2 respectively.

Table 1: Five patterns of topic evolution path

Evolution pattern	Expression and illustration
development	 Topic $T1_{(t2)}$ only comes from $T1_{(t1)}$
division	 Topic $T1_{(t1)}$ is divided two or more topics in t2
integration	 Two or more topics integrate into one topic
extinction	 $T2_{(t2)}$ is extinct which has no evolutionary relationship with the following topic in t3
emerging	 $T2_{(t2)}$ is treated as an emerging topic which has no evolutionary relationship with previous topic.

2.3 Discovery of Semantic Relationship Among Topics

Firstly, we predefine five types of semantic relationships among patent entities, which are shown in Table 2. Secondly, we manually label a small dataset for training and testing with predefined relationships. Thirdly, we train a OpenNRE (Han, X., 2019) model on this dataset and evaluate it through quantitative indicators. Fourthly, the model is used to predict all relationships among entities. Finally, the semantic relationship between two topics is determined based on the semantic relationship among all pairs of entities.

Table 2: Semantic relationship among topics

Semantic relationship	Expression and illustration
Mechanical relationship(**M**)	refers to the containment relationship, position relationship, etc. of some mechanical parts.
Efficacy relationship(**E**)	refers to the relationship of efficacy enhancement
Function-Area relationship(**FA**)	refers to the application field or function field of certain machines or systems.
Function-Realization Relationship(**FR**)	refers to some patented devices or systems that realize certain functions.
Control relationship(**C**)	refers to a certain control relationship.

3 Result

We set semantic similarity to 0.7 where a pair of entities with similarity higher than it is considered to have the same meaning. In the discovery of topic evolution path, the results are obtained from 2015-2016 as an example. There are six topics in 2015 and eight topics in 2016, where the evolution probability is shown in Table 3.

Table 3: Evolution probability among topics (%)

2015/2016	$T0_{(t2)}$	$T1_{(t2)}$	$T2_{(t2)}$	$T3_{(t2)}$	$T4_{(t2)}$	$T5_{(t2)}$	$T6_{(t2)}$	$T7_{(t2)}$
$T0_{(t1)}$	5.1	14.4	**32.4**	12.4	0	**20.5**	16.5	0
$T1_{(t1)}$	**40.0**	**21.9**	**24.5**	7.2	5.4	10.6	1.4	6.8
$T2_{(t1)}$	4.8	**24.7**	18.2	**20.5**	0	14.0	19.3	0
$T3_{(t1)}$	**100**	11.2	12.9	0	**100**	1.6	0	**100**
$T4_{(t1)}$	3.4	**23.3**	18.4	13.5	0	17.9	**24.3**	0
$T5_{(t1)}$	5.0	14.0	8.6	**34.1**	0.	**23.0**	16.2	0

As shown in the Table 3, the pairs of topics with probability more than 20% are emphasized with bold. They mean that there is a certain evolutionary relationship between them. For example, $T1_{(t1)}$, $T2_{(t1)}$, $T4_{(t1)}$ are integrated into $T1_{(t2)}$. $T1_{(t1)}$ is divided into $T0_{(t2)}$, $T1_{(t2)}$ and $T2_{(t2)}$. $T3_{(t1)}$ is developed into $T7_{(t2)}$ because these two topics are almost the same.

Then, we obtain the semantic relationships between topics which have an evolution path, which is shown in Table 4. The abbreviations of semantic relationships are illustrated in Table 2.

Table 4: Semantic relations of topic evolution

2015/2016	$T0_{(t2)}$	$T1_{(t2)}$	$T2_{(t2)}$	$T3_{(t2)}$	$T4_{(t2)}$	$T5_{(t2)}$	$T6_{(t2)}$	$T7_{(t2)}$
$T0_{(t1)}$			E			E		
$T1_{(t1)}$	M	E	E					
$T2_{(t1)}$		E		FA				
$T3_{(t1)}$	M				M			M
$T4_{(t1)}$		E					E	
$T5_{(t1)}$				FA		FA		

According to the results in Table 4, the semantic relationship between $T1_{(t1)}$ (UAV communication medium and function) and $T0_{(t2)}$ (camera device and image transmission system) is “Mechanical relationship". It shows that patents on communication media and functions of UAVs have developed over time, and a large part of the research topics have been split into image information transmission, which is very consistent with the development in the field.

In addition, the semantic relationship among $T2_{(t1)}$ (UAV functional module), $T5_{(t1)}$ (UAV kinetic energy device) and $T3_{(t2)}$ (UAV application field) is "Function-Area relationship". Obviously, a large part of patent applications in the field of UAV transportation are related to UAV system applications, including geological survey, forest fire prevention, water resources inspection and protection, environmental science and ecology, agriculture and other key technologies used in industries.

4 Conclusion

This paper proposes a method for discovery of Topic Evolution Path and Semantic Relationship among topics Based on Patent Entity Relationship. The result could prove the effectiveness of this method and could enrich and improve the topic evolution analysis method. In the next step, we would like to apply other neural network models that may do better in patent entity relationship extraction and compared with baseline method for deep analysis.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 71974095).

REFERENCES

Liu, H., Chen, Z., Tang, J. et al. Mapping the technology evolution path: a novel model for dynamic topic detection and tracking. Scientometrics 125, 2043–2090 (2020). https://doi.org/10.1007/s11192-020-03700-5

Yu, D., Xu, Z., & Wang, X. (2020). Bibliometric analysis of support vector machines research trend: a case study in China. International Journal of Machine Learning and Cybernetics, 11(3), 715-728.

No, H. J., An, Y., & Park, Y. (2015). A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining. Technological Forecasting and Social Change, 97, 181-192.

Birunda, S. S., & Devi, R. K. (2021). A Review on Word Embedding Techniques for Text Classification. In Innovative Data Communication Technologies and Application (pp. 267-281). Springer, Singapore.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.

Han, X., Gao, T., Yao, Y., Ye, D., Liu, Z., & Sun, M. (2019). OpenNRE: An open and extensible toolkit for neural relation extraction. arXiv preprint arXiv:1909.13078..