Study on innovation path recognition based on topics association of science and technology

Haiyun Xu xuhy@clas.ac.cn Chengdu Documentation and Information Center, Chinese Academy of Sciences
Rui Luo luorui@mail.las.ac.cn Chengdu Documentation and Information Center, Chinese Academy of Sciences
Dong Kun dongkun@mail.las.ac.cn Chengdu Library of Chinese Academy of Sciences
Yue Zeng-Hui yzh66123@126.com School of Medical Information Engineering, Jining Medical University
Yan Qi qi.yan@imicams.ac.cn Institute Of Medical Information/Medical Library, CAMS & PUMC
Shu Fang fangsh@clas.ac.cn Chengdu Library of the Chinese Academy of Sciences

1. Introduction

The evolution path of science and technology refers to the development and evolution of innovation content or innovation topic, reflectingthe emergence, diffusion and evolution of technological innovation. There are two innovative elements in the innovation process, one spreads from science to technology, and the other from technology to science. Two innovative elements jointly determine the context and direction of scientific development. Therefore, it is important to study the identification method of the innovation path based on science-technology relevance .

There are two problems in the identification of scientific innovation paths research. First, previous identification methods mainly focus on single innovation element of science or technology, but ignore the intrinsic relevance of science and technology. Second, the quantitative researches on the relevance of science and technology are limited to the numeral feature of literature, which can't reveal the inherent relevance of science and technology. These problems make it difficult to fully grasp the characteristics of scientific innovation, and then affect the accuracy of the path identification of innovation evolution.

This study focuses on both science and technology through bibliometric analysis, thereby exploring the identification method of innovation paths based on the association of science and technology topics. The structure of this paper is as follows: firstly, we define the connotations, concepts and research status of scientific innovation path. Secondly, we discuss the relationship between science-technology based on the existing researches. Thirdly, in order to recognize the innovation evolution path in an interactive perspective, we reveal the interaction of science and technology by analyzing the relevance among science and technology innovation topics in a micro level. At the same time, the genetic engineered vaccine (GEV) is selected as our empirical field. Finally, we summarize the contributions and limitations of this study.

2 Research Content

2.1 Science and Technology Association Model

The development motivation and pattern for the development and evolution of science and technology are different. While the system of science and technology is interactive, the evolution

paths are different. Therefore, it is not appropriate to simply mix the research of scientific and technical topics together. In contrast, in order to maximally reveal the real dynamic mechanism in the science and technology development, we need to analyze the relationship between the two elements from both independent and connected perspectives. In this study, we use scientific papers to represent scientific research and patents represent technical research.

Fig.1 is the schematic diagram of pattern recognition in science-technology interaction. S_1, S_2, S_3... S_n are scientific topics, T_1, T_2, T_3... T_n are technical topics. All the scientific topics and technical topics are distributed on the timeline t, the line thickness indicates the different intensity of topic relevance.

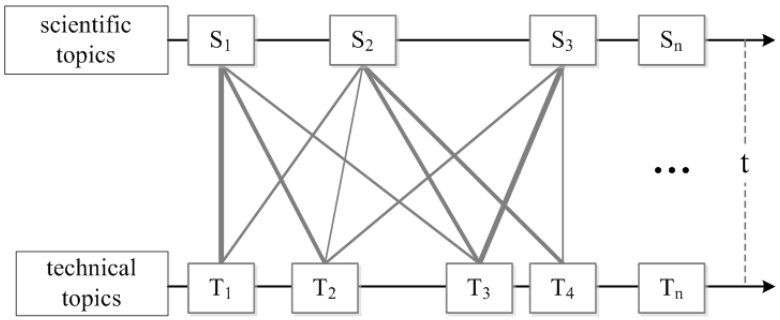

Figure 1 Recognition principle of science-technology interaction pattern

2.2 Science-technology topic relevance of multi-relational fusion

The key technologies in the topic relevant research of science and technology innovation are efficience and accuracy of topics identification and similarity computation, which are both based on text analysis. In this study, a science-technology topic relevance analysis method of multi-relational fusion is used to calculate the topic relevance of science and technology along the time axis, and finally obtain the relevance degree between topic pairs on the two types of innovation paths.

The topic relevance to be fused includes three types: co-terms relevance, authors relevance, and citations relevance. The co-word relevance means that two topics have semantic relevance since they contain the same topic terms. The author relevance means two topics have semantic relevance since they have common authors. The citation relevance means that two topics have semantic relevance because they have citing and cited relations. The calculation method of topic relevance between arbitrary science topic S and arbitrary technology topic T is shown in figure 2. In figure 2, the topic S contains N_S topic terms (blue dots), and the topic T contains N_T topic terms (yellow dots), and the correlation between the topic S and the topic T is a comprehensive calculation based on the topic correlation degree of the co-word relevance, the author relevance and the citation relevance.

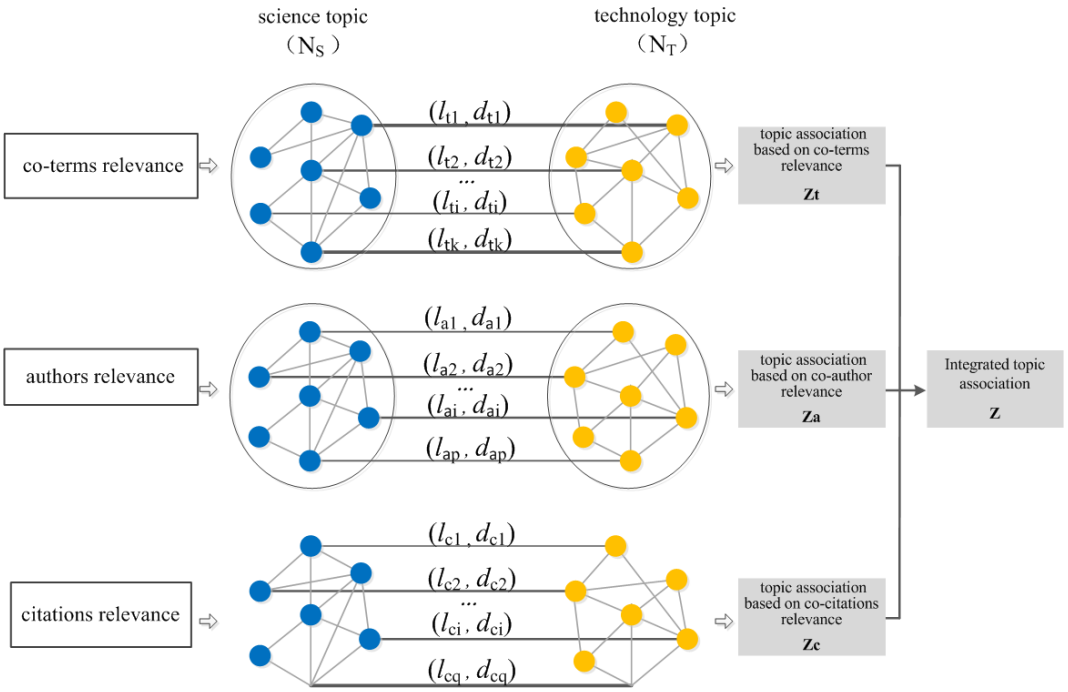

Figure 2 Calculation method of topics relevance

In this study, the degree calculation of topic association based on co-terms relevance is taken as an example to illustrate the acquisition of the single topic relevance. As shown in figure 2, the topic *S* and the topic *T* contain some common topic terms, so there are multiple common term relevance paths between the two topics: $l_{common\ 1}$, $l_{common\ 2}$... $l_{common\ k}$, and the corresponding path weight is $d_{common\ 1}$, $d_{common2}$... $d_{common\ k}$. Then, the Z_{common}, which means the degree of co-term relevance between the topic *T* and the topic *S* is calculated by (1):

$$Z_t = \frac{k}{N_S N_T} \sum_{i=1}^{k} d_{ti} \qquad (1)$$

Where $\sum_{i=1}^{k} d_{ti}$ is the summation of co-term relevance; $N_S N_T$ is the path amount involved topic relevance between the topic *T* and the topic *S*. In this paper, if the topic *T* and the topic *S* are related, then there will be the path of topic relevance. The function of $\frac{k}{N_S N_T}$ is to prevent too much topics through changing the absolute value of topic relevance the degree to a relative value,which can help to avoid the absolute value is too high. D_{ti} depends on the frequency of the topics connected by $l_{commom\ i}$. If there are *X* topic S are connected with topic T by $l_{common\ i}$, then the value of *X+Y* is $d_{common\ i}$.

Similarly, we take the same way to get the topic association between authors and citations. Finally, the degree of fusion topic is calculated by (2):

$$Z = \alpha \frac{k}{N_S N_T} \sum_{i=1}^{k} d_{ti} + \beta \frac{p}{N_S N_T} \sum_{i=1}^{p} d_{ai} + \gamma \frac{q}{N_S N_T} \sum_{i=1}^{q} d_{ci} \qquad (2)$$

In (2), three parameter *α, β, γ* represent the weight coefficient of co-term relevance, authors relevance and citations relevance, and the parameter is determined comprehensively depend on expert judgment based on the specific meaning of research field and the topics relevance effect.

3. Empirical analysis and conclusion

In this study, the Web of Science database of Thomson Reuters was selected to search the scientific papers, and the Derwent Innovations Index database was used to search the patent documents. Papers and patent documents were collected in the GEV field for a period as analysis data sets, search date was January 6, 2018, publication year was up to 2017, and the paper record was 4146 and patent is 4050. According to the above analysis method, the related paths of science and technology topics were obtained and analyzed.

Comparative analysis shows that compared to a single path of science or technology innovation, innovation paths associated with science and technology topics can quantitatively reveal science and technology interactions from a micro level and more fully demonstrate the development laws and evolution characteristics of technological innovation, thus eliminating the current one-sidedness of only one innovative elements in the innovative path.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant No. 71704170), the China Postdoctoral Science Foundation funded project (Grant No. 2016M590124), and the Youth Innovation Promotion Association, CAS (Grant No. 2016159).