Science-technology correlative topics detection based on VOSviewer

Introduction

Research on the relationship between science and technology is not only a description of objective phenomena, but also an important work to understand the laws of science and technology innovation to support the decision-making of science and technology development. The most common method to uncover science-technical knowledge linkages is the co-occurrence analysis or citation analysis between the academic papers and patents. Some existed indicators such as SL(Science Linkage), TL (technology linkage) mostly base on the citations and not deeply into the content relevance level, so some scholars suggested a new integrative index called Science-Technology Linkage (STL) base on the research topic analysis of thesis and patent. The LDA topic modelling was used to extract the topics, and the terms (in a specific probability) of different topics are somewhat repetitive that make the topic analyzing mistily and equivocally. It needs to be improved, and it is well known that, the VOSviewer focusing on visualization of bibliometric networks supports text mining and topic clusting and visualization, so it is deserved to explore the feasibility of detecting and visualizing the science-technology correlative topics based on VOSviewer in a special way.

Methodology and Process

Theoretical hypothesis

The term map of VOSviewer visualizes the topics being covered by specific set of publications, and the overlay term map can show the partial characteristic of the topics being covered by the subset of publications.We call the former as global map, and the overlay term map as partial map. For example, the global map can be the macro-level research areas or the scientometric research area, and the partial map can be specific institutions's research strengths or focus. Following this idea, we regard science and technology as two specific institutions and then investigate their respective focus and the correlative ones. The global map is constructed based on the collection of domain-specific science and technology publications, and then overlay maps are used to review which subjects science is biased on and which subjects technology is biased on. From the perspective of research topic, it can be seen that whether the topic is biased towards scientific research or technological development in a specific period. The topics (or the clusters) with more coincident or similar score-colored nodes in the two local graphs can be regarded as the correlative topics of science and technology.

Topic Detection

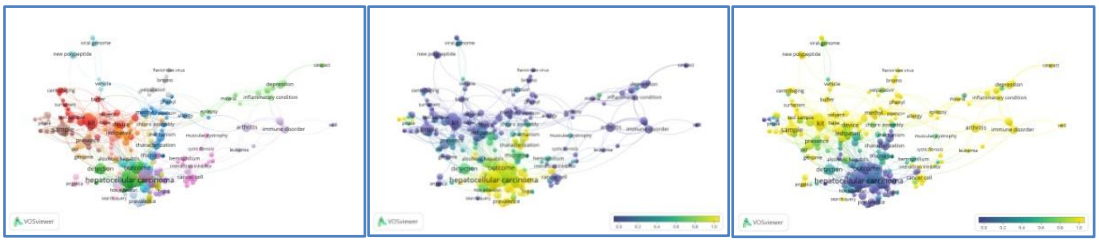
In the overlay visualization of a map, each cluster contains different amount of items, that is word or phrase, and each item has respective score on which can be colored based. Actually, the item score of the partial map is related to the probability of the item appearing in the particular object's focus. Roughly speaking, in the science partial map, the higher the score of a item, the higher the frequency of the item in paper publication than patent literature, and the more high score item in a cluster, the more science deflective the topic. Therefore, when we define a cluster as a topic, the topic score can be obtained by calculating the mean value of all item scores, the formula is:

$$\mathrm{Sc_i} = \frac{\sum_{\mathrm{j}}^{\mathrm{Nt_i}} \mathrm{St_j}}{\mathrm{Nt_i}}$$

Where Sc_i denotes the score of a Cluster i, St_j denotes the score of term j in Cluster i and Nt_i denotes the number of terms in the cluster. If the score come from the papers' partial map, the item is scientific bias; if it is from the patents' partial map, the item is technical bias. The number of clusters and their items of the two partial maps are same as that of the global graph, so each cluster can obtain these two skewness values. Therefore, we believe that for a cluster (topic), the closer the two skewness values are, the more significant it is that researchers have carried out much theoretical research explorations and technical development simultaneously on this topic and achieved fruitful results. Such a topic can be regarded as the science-technology correlative topic.

Case Study

We randomly selected 422 papers and 500 patents of the HCV field published in 2017. Only the title was extracted to constitute the corpus file, and thesaurus was established in the process. The global map and two partial overlay maps were obtained as follows:



(a) (b) (c)

Fig.1 global map(a), science overlay map(b), and technology overlay map(c)

A total of 19 clusters were obtained using the default parameters. Figure (a) shows all the clusters, and the yellow nodes in figure (b) and (c) respectively show the themes that the paper collection and patent collection focus on. It can be seen that there are great differences between science and technology, as well as some overlaps as the blue nodes shown. According to the above, the correlative topics and their skewness values are shown in the table 1:

Tab.1 correlative topics and related data

Cluster NO.	Items (as the blue nodes shown on the Fig.1(b)and(c))	SSV	TSV
16	Specificity, susceptibility, pcr, genome ,polymorphism, …	0.53	0.47
3	multiple myeloma, occluding, yellow fever virus, Mir, assembly,…	0.69	0.31

Note: SSV indicates the Science-skewness value, TSV indicates the technology-skewness value.

Conclusions

The empirical test demonstrates the feasibility of this approach and the possible advantages compared with the LDA model. Firstly, the relatively more well-defined clusters and scored topics can be achieved more easily, and moreover, through the interpretation of the themes we believe that the overlay map reflects the related themes and separate themes of the scientific and technical communities in a specific field well. However, the formula maybe too simple, so it can only be used as a preliminary exploration. The data size and corpus source need to be expanded and the term cleaning need to be enhanced for the further research on this idea in the future.

Acknowledgments

The work described in this paper was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (Grant no. 2017PT63008) and 2017 "PUMC Youth Fund" project of Chinese Academy of Medical Sciences (Grant no. 2017330008).