

From cooperation to citation or from citation to cooperation?

Introduction

With the development of science and technology, science and technology are more and more penetrating each other, and the intersection and integration of science and technology are often opportunities for major scientific and technological innovation. Therefore, the research on the relationship between science and technology is a topic of common concern in academic and industrial circles. The existing research has done a lot of work on the exploration of this relationship, among which the most common practice is the bibliometrics method which regards the paper literature as the representative of science and the patent literature as the representative of technology and analysis and reveals the relationship by using the citation between the two or the co-occurrence of categories/subject words.

In the previous study (Qi, 2019, 2021), we proposed several measurement indicators based on citation-relation among paper and patent literature, which can reveal the characteristics and trends of knowledge flowing from science to technology and from technology to science. Based on these literature data and citation relationship information, we plan to further observe the relationship between research institutions corresponding to the literature and its evolution over time. In particular, because the scholars believe that industry-university-research cooperation is also a vital manifestation of science-technology association, we mainly focus on the two relations of cooperation(CO) and citation(CI) between institutions, analyze their changes over time, and explore the possible causal logic between the two relations.

Observation and discovery

This paper mainly observes the citing papers and cited patents. We first give some definitions. When one scientific paper refers to a patent, we think that there is a many-to-many reference relationship (such as A_{11} - P_{11} , A_{11} - P_{12} , A_{12} - P_{11} , A_{12} - P_{12} , A_{13} - P_{11} , A_{13} - P_{12}) between the institutions to which the author belongs (such as A_{11} , A_{12} , A_{13}) and the patent application institutions (such as P_{11} , P_{12}), and we name this reference relationship as "cross-border citation". If there is more than one author institution of the cited paper, there is a cooperative relationship between these institutions (for example, A_{11} * A_{12} , A_{11} * A_{13} , A_{12} * A_{13}). It should be specially stated that this paper focuses on the relationship between science and technology from the perspective of knowledge flow. Therefore, we only focus on the technical attribute of the patent and do not pay attention to the ownership of its rights. Thus, the institution here is a patent application institution rather than a patentee institution.

We used the data from the HCV field based on previous research. Through repeatedly data cleaning and verification, we found that among 33,086 documents related to HCV research from 2008 to 2017 included in the Web of Science database, 590 scientific papers referenced 1,223 patents.

Through case analysis, we found the following phenomena:

Phenomenon 1: in an earlier year, the paper (A1) cited the patent (P1), and then one paper institution and one patent application institution cooperated to publish a paper in a later year. Before that, there was no cooperation between them.

For example, a paper published in 2008 and entitled "2-Arylbenzimidazoles as Antiviral and Antiproliferative Agents-Part 1" cited the patent entitled "A series of compounds of Formula I are disclosed which are useful in treating viral hepatitis C" applied in 2002. Two of the author institute, the University of Cagliari and Bristol Myers Squibb Co, published two articles in 2013, entitled "Sheehan Suicidality Tracking Scale (S-STTS): Reliability, convergent and discriminative validity in young Italian adults" and "Dasatinib plus Capecitabine for Advanced Breast Cancer: Safety and Efficacy in Phase I Study CA180004".

This phenomenon leads us to speculate and think—can the "cross-border" reference between institutions promote the follow-up cooperation (co-author)? Or, how to measure the "ability" or "probability" of "cross-border" reference relations triggering the inter-agency cooperation?

At the same time, it needs to be specially pointed out that such cooperation is different from the cooperation between institutions that have not applied for patents. It has the characteristics of industry-university-research cooperation and is more conducive to promoting innovative knowledge flowing from the theoretical stage to the application stage. Therefore, we believe that such cooperation is also valuable. If there is a definite causal logic or "triggering effect" between cross-border citation and cooperation, such institutes can be recommended for each other as potential partners by using this cross-border citation relationship.

Phenomenon 2: in an earlier year, two institutions cooperated to publish papers (A2), and in a later year, the papers published by one institution cited the patents applied by the other institution (P2). Before that, there was no "cross-border" reference between the two.

For example, GlaxoSmithKline and Icagen Inc cooperated in two articles, but the research topic was not HCV. The first article was published in June 2003, and the second article was published in October 2015. The cross-border citation occurred in 2012. The paper entitled "Synthesis and antiviral activity of novel HCV NS3 protease inhibitors with P4 capping groups" published by GlaxoSmithKline cited the patent "Cyclohexyl inhibitors of potassium channel function" applied by Icagen Inc in 2003.

This phenomenon also leads us to speculate and think: can the co-author promote the subsequent "cross-border" reference? Or, how to measure the "ability" or "probability" of inter-agency cooperation leading to "cross-border" reference relations?

A cross-border citation is an important form and embodiment of the science-technology association. For example, if a paper is cited by a patent, it can be inferred that the corresponding scientific knowledge has been used as the research basis, theoretical support, or reference by the technology to a certain extent; and if a patent is cited in a paper, the technical knowledge has been used as the basis or material for scientific research. Due to the past cooperation, the institutions have established a deeper understanding. Therefore, when one party conducts technology research and development, the other party can track it in time, resulting in "cross-border" references. If it can be found and demonstrated that some cooperation can promote the future exchange of science and technology, the characteristics of such cooperation can be used to recommend potential partners for the institution, or extend to analyze the development process of some research topics.

Phenomenon 3: cross border self-citation

An organization appears in both citing papers and cited patents. For example, "proteome analysis identifies the role of heat stress in production of genetic HCV in Huh7 cells harboring incident

HCV" and "comparative proteomics analysis of Huh7 cells harboring incident HCV" both cited the patent WO03040356, and the institute lists both include the enterprise of Yang Daiying Biotechnol Inst. This phenomenon reflects the closest relationship between science and technology, however, there is no accurate data on the specific proportion, so this paper will not make more analysis for the time being.

Preliminary design of indicators

For phenomenon 1, we preliminarily designed several simple indicators to evaluate the contribution of "cross-border citation" in promoting follow-up cooperation. There are two problems here. First, there may be such a phenomenon—there is a cooperation relationship (first cooperation) between the two institutions before the occurrence of "cross-border citation", and there is a cooperation (second cooperation) after the occurrence of "cross-border citation". In this case, it is not easy to judge how much of the role "cross-border citation" plays. Perhaps the second cooperation originates from the first cooperation, and it may also be that "cross-border citation" has created new opportunities for cooperation. Second, we found that the cooperation between the two institutions in the later stage can be about a broader theme, so there may be three situations, including the "cross-border citation" of the two institutions leading to the cooperation on the topic(TC), in the same field(FC), or different fields(DC) in the later years. Theoretically, citation makes the greatest contribution in the case of TC, and the cooperation on the same subject after "cross-border citation" to a certain extent reflects the closeness of inter-institutional relations with both technological research and theoretical research capabilities, thus accelerating the innovation process. The simplest formula is:

$$\text{Contribution of a cross-border citation: } \text{Cont}_{1111} = |A_{11} - P_{11}| / |A_{11} * P_{11}| * \alpha$$

Where $|A_{11} - P_{11}|$ denotes the number of cross-border citation relationships between A_{11} , one author institution of the citing papers, and P_{11} , one application institution of the cited patent, $|A_{11} * P_{11}|$ denotes the number of papers (or patent applications) jointly published by the two, and α is the adjustment coefficient of the numerical range or the weight coefficient of different scenarios.

It is possible that, there are multiple cross-border citations corresponding to a pair of one paper and one patent, which will trigger much cooperation. In this case, multiple values of "Cont" can be synthesized as a measure at the document level. Further, it can be expanded to the measurement of a literature set, and the literature set can be about the same subject or different subjects from the same domain so that the characteristics of this logical association can be observed from the perspective of the domain.

Conclusions

At present, only manual analysis and interpretation of some data have been carried out, and more case analysis needs to be conducted on a larger range of data sets. And based on this, more comprehensive design, improvement, and empirical verification of indicators and their formulas will be continued. Secondly, the above case analysis and index design are all aimed at the situation of "citing paper and cited patent". We will continue to investigate the situation of "citing patent and cited paper" in the future.

Acknowledgments

This work was supported by the Basic Scientific Research Project of Chinese Academy of Medical Sciences (2017PT63008 and 2017330008).

References

- Qi, Y. (2019). Research on the Analytical Method of Science-technology Relationship Evolution in Specific Fields Based on Patent Citation Analysis——Taking HCV as an Example. *Library and Information Service*,63(23): 97-105.(In Chinese)
- Qi, Y. (2021). Analytical Method of Science-technology Relationship Evolution in Specific Fields Based on Citation Analysis. GTM 2021.