Journal of Informetrics 11 (2017) 63–79

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations

Lixin Chen

Department of Management and Economics, North China University of Water Conservancy and Electric Power, Zhengzhou, Henan, China

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form 10 April 2016
Accepted 10 April 2016
Available online 21 November 2016

Keywords:
Patent citation
Knowledge linkage
Text similarity
Citing-cited patent pairs
Non-citing-cited patent pairs
Examiner citation
Applicant citation

ABSTRACT

Whether patent citations indicate knowledge linkage is still a controversial issue, which is very important for the widespread use of the patent citation analysis method. We hypothesize that there exists technological knowledge linkage between patents and their citations, and that the linkage can be detected through measuring text similarities between them. To test the hypothesis, we selected citing-cited patent pairs as the observation group and selected patent pairs without citing-cited relationship as the control group. Using the VSM with WF-IDF weighting method, we calculated text similarity values of the two groups. Through comparing text similarity values between the two groups, we validate that in the vast majority of cases text similarity values of citing-cited pairs are much higher than those of non-citing-cited pairs. The study in nano-technology field shows that the above results are the same, although patents in the same technological area are more relevant than in different technological areas. Furthermore, by comparing text similarities between applicant and examiner citing-cited pairs, the results show that in more cases examiner citations indicate knowledge linkage a bit better than applicant citations. Preferably, examiner citations can be regarded as not only the supplement of applicant citations but also the more important technological background and the prior art closely related to the patents. Compared to applicant citations, examiner citations are a good indicator of knowledge linkage rather than an incomplete and noisy indicator. In short, the results suggest that most certainly patent citations can indicate knowledge linkage, and more likely examiner citations can indicate knowledge linkage a bit better than applicant citations, especially for the component of patent claims. Therefore, we accept the hypothesis that patent citations can indicate knowledge linkage, which is the basic assumption of the patent citation analysis method.

1. Introduction

Patent citations, similar to paper citations, are the references of patents. Listing references is an obligation of the patent applicant, who should comply with the legal requirement to disclose the prior art derived from previous patents, and who should supply a complete description of the state of the art in the field of the invention as the technological background (Criscuolo & Verspagen, 2008; Jaffe, Trajtenberg, & Fogarty, 2000). Paper citations can be extensively applied to investigate the linkage of scientific knowledge for tracking science development (Garfield, Sher, & Torpie, 1964), revealing knowledge flow and diffusion (Liu & Rousseau, 2010), mapping science structure (Leydesdorff & Rafols, 2009; Park & Leydesdorff, 2009)

E-mail address: lynnchenlixin@163.com

http://dx.doi.org/10.1016/j.joi.2016.04.018
1751-1577/

and so on. Can patent citations be used to reveal the linkage of technological knowledge? This issue, however, has been in dispute ever since the patent citation analysis method began to be used.

Many researchers believe that patent citations, which are the same as paper citations, can reflect the linkage of technological knowledge. Narin and Olivastro (1988) suggest that patent citations can indicate linkages between companies, between technological areas, and between technology and science. Narin and his cooperators (Carpenter, Cooper, & Narin, 1980; Narin & Olivastro, 1992; Narin & Olivastro, 1998; Narin, Hamilton, & Olivastro, 1995; Narin, Hamilton, & Olivastro, 1997) have used the scientific papers within patent references to measure the linkage between science and technology. In recent decades, patent citations have been extensively used as a proxy for measuring technological knowledge linkage (Callaert, Grouwels, & Looy, 2012; Criscuoloa & Verspagen, 2008; Hu, Chen, Huang, & Roco, 2007; Lo, 2010; Meyer, 2001; Ribeiro, Ruiz, Bernardes, & Albuquerque, 2010; Schmoch, 1997; Tijsen, 2005; Verbeek et al., 2002).

However, some researchers (Jaffe et al., 2000) hold the opposite opinions and suggest that patent citations indicate the technological knowledge linkage incompletely. Patent citations look like paper citations, but they are different in many respects (Meyer, 2000; Michel & Bettels, 2001). According to US patent laws, not only the applicant is required to provide references when filing a patent application, but also the patent examiner need to add references for judging patentability through searching the related prior art during the patent examining period.[1] The survey by Jaffe et al. (2000) shows that inventors respond that one half patent citations indicate no knowledge spillover, since addition of citations by the patent examiner is unknown to the inventor and has no effect on the invention. They suggest that the inventor citations, instead of the total citations, should be taken as an indicator of knowledge flow. Li and Meng (2010) insist that simply applying patent citation data to indicate knowledge linkage is both conceptually and technologically illogical and unreasonable. Li, Chambers, Ding, Zhang, and Meng (2014) further point out that only scientific papers self-cited by the inventor should be used for measuring linkage between science and technology, and the examiner citation and non-self-citation by inventor papers should be excluded due to excessive "noise".

Obviously, it is a complex and controversial issue whether patent citations can indicate knowledge linkage or not. The disputations mainly focus on the differences of the two types of patent citations that are respectively added by applicants and examiners (Alcacer & Gittelman, 2006; Alcacer, Gittelman, & Sampat, 2009). Jaffe et al. (2000) regard examiner citations as an incomplete and noisy indicator for measuring knowledge flow and diffusion. However, basing their study on European Patent Office search reports, Criscuoloa and Verspagen (2008) find that there are mainly two kinds of examiner citations: documents of particular relevance that restrict patent application claims, accounting for 36%; and references related to technological background, accounting for 62%. Patent attorneys anticipate citations most likely to be added by examiners, so that examiner and applicant citations may come to resemble each other closely (Alcacer & Gittelman, 2006). That means examiner citations are relevant to applicant citations, which are used for disclosing prior art and describing technological background. Li et al. (2014) suggest that examiner citations can indicate knowledge linkage logically because the citing behavior of examiner is regulated by patent laws.

Despite many limitations, criticisms and disputations, by using the patent citation analysis method, many researchers have revealed technological knowledge flow, diffusion and transfer (Chen & Hicks, 2004; Hu & Jaffe, 2003; Nelson, 2009; Park & Suh, 2013), have traced the technological trajectories and the technological frontiers (Epicoco, 2013; Érdi et al., 2013; Martinelli, 2012; Mina, Ramlogan, Tampubolon, & Metcalfe, 2007), have mapped the technological knowledge domain to display its structure and its relation (Lai & Wu, 2005; Wang, Zhang, & Xu, 2011; Weng & Daim, 2012; Yeh, Sung, Yang, Tsai, & Chen, 2013), have explored the technology and patent classification (Shih & Liu, 2010), and even have evaluated patent, technology and innovation ability at the levels of organization, region and country (Albert, Avery, Narin, & McAllister, 1991; Carpenter & Narin, 1983; Lanjouw & Schankeman, 2004; Verspagen, 2000; Wartburg, Teichert, & Rost, 2005; Yeh et al., 2013; Yoon & Park, 2004).

Similar to the paper citation analysis method, the patent citation analysis method must be provided with a basic condition or assumption: patent citations can reflect knowledge linkage. Undoubtedly, the assumption is an essential and crucial issue for the widely use of the patent citation analysis method. Only based on this condition, can technological knowledge flow and diffusion be disclosed, can the technological frontiers be traced, can the technological knowledge structure be mapped, and can technology innovations be evaluated by using the method. Therefore, we pose a hypothesis that patent citations can indicate knowledge linkage: patents are similar to or are relevant to their citations in technological knowledge, and that the linkage can be detected through measuring text similarities between them. Through the empirical research, we try to gather supporting evidence for the hypothesis.

2. Data

The United States Patent and Trademark Office (USPTO) provides all kinds of patent documents freely. We constructed the origin database through downloading the full text documents of USPTO patents that were granted during the 1976–2013 period, with over 300G storage capacity in all. In our research, only utility patents, which are the most common of all patents, were taken as the study object. There are more than four million full text documents of utility patent in the origin database.

[1] Applicant citations are submitted on the filing date (the priority date), which is the submission date of a patent application, whereas examiner citations are added during the patent examining period, after the filing date but before the granting date (the date of authorizing the patent).