Technological Forecasting & Social Change xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Technological Forecasting & Social Change

journal homepage:

Patent keyword network analysis for improving technology development efficiency

Jinho Choi, Yong-Sik Hwang *

School of Business, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul 143-747, Republic of Korea

ARTICLE INFO

Article history:
Received 24 October 2012
Received in revised form 14 June 2013
Accepted 9 July 2013
Available online xxxx

Keywords:
Patents
Keyword networks
Network analysis
Text mining
Trend analysis

ABSTRACT

The methods of patent analysis are largely divided into network-based patent analysis and keyword-based morphological patent analysis. Both methods have their shortcomings: internal patent information composed of natural languages cannot be analyzed in the network-based patent analysis method, and the correlation between patents cannot be analyzed in the keyword-based morphological patent analysis method. In this research, we analyze the patents of Light Emitting Diode (LED) and wireless broadband fields via a method that incorporates both the network-based patent analysis and the keyword-based patent analysis methods. And by using network indices, we identify the characteristics of the patent keyword network, and also perform a trend analysis to discover how keywords play a significant role in network changes over time. The analysis results indicate that the patent keyword network is sporadic but clustered and shows a clear power law distribution. Further, the inflow keywords are highly likely to tie new connections with other keywords in the existing associated communities. Also, we confirm the fact that, as time passes, the top core keywords of a particular technology field continue to play an important role in the network and that also the rate of technological changes in wireless broadband field is faster than that of LED. Through the proposed analysis, researchers can easily grasp what technology keywords are important in the specific technology field and identify the relations between the essential technology elements; furthermore, this information can be utilized for developing new technologies by combining these technology elements extracted from community analysis.

1. Introduction

Most organizations have come to recognize the importance of technology in their respective industries. As its significance becomes increasingly pronounced, patent disputes between global market-leading IT enterprises also become more complicated. Patent information plays an important role in the process of innovation via assisting the development of inventions and products [1].

A variety of methods for technology prediction have been developed thus far. Representative technology prediction methods are trend impact analysis [2], scientometrics (measuring the science of bibliography) [3], and patent analysis [4]. Patent analysis is especially crucial in predicting technology on the basis of information extracted from patents as it conceives patents as the original source of information for technology and as possessing commercial value. Since most patent data are computerized, this method is particularly advantageous in that it can accurately analyze technical trends in detail [5,6]. In addition, as information communication technology progresses, access to patent data bases is improved, so the practicality of patent analysis is evaluated to be quite positive [7].

The patent analysis method provides comprehension of an important technical field by cumulatively measuring the number of patents for every technical field and by mutually comparing technical fields [8]. This method has the advantage of offering a clear grasp of the important technical fields since it produces numerical results. Due to a methodological

* Corresponding author. Tel.: +82 3408 3175; fax: +82 3408 4310.
E-mail address: yhwang@sejong.ac.kr (Y.-S. Hwang).

0040-1625/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.techfore.2013.07.004

limitation in which only the number of applied patents is known, however, detailed technical trends within a concerned technical field cannot be understood. To surmount this problem, a variety of patent analysis methods have been proposed, including the network based patent analysis method to analyze the relationship between patents from the viewpoint of network [8–11], and keyword-based morphological patent analysis to analyze the content of patented technology [7,12–14]. Network based patent analysis can take a macroscopic look at the overall context of a relevant technical field, since it is able to analyze which patent is important and which patents are interrelated after composing a network, on the basis of the cited relationship between patents [10]. Meanwhile, keyword-based patent analysis can be conducted as a morphological analysis by extracting meaningful technical information from patent contents and has an advantage of understanding detailed information regarding important technical factors mentioned in patents [12].

However, while network based patent analysis has advantages such that it can understand not only influential patents but also mutual relationships between patents in a technical field, the disadvantage is its inability to analyze the technical content of patents since patents are analyzed on an individual basis. Meanwhile, even though keyword based morphological patent analysis can understand core technology information within patents since it analyzes patents on the basis of patent content, it cannot investigate the correlation in core technology information between patents.

In order to resolve the limitations of both network-based patent analysis and keyword-based morphological patent analysis, a technology analysis that integrates both analyses is proposed in this research. The result of the proposed analysis can be used as information for technology prediction. In other words, this method constructs a keyword network after extracting important technology information through text mining on the basis of individual patents and then conducts community network analysis for the keyword network. To identify the unique characteristics of the patent keyword network, density [15], clustering coefficient [16], and degree [17], indices of the each network are analyzed. Furthermore, trend analysis, which identifies recent important keywords over time, is conducted by using frequency, degree, and normalized betweenness[1] [18,32] indices.

These methods enable us to understand detailed information on the core technology factors of each patent, analyze the interrelationships between them, and propose to experts in the relevant field a combination of specific technology factors to be developed in the future. This paper is composed of five main sections. In the subsequent section, prior research related to this research is introduced. In Section 3, the data used in this research are introduced, and the processes of extracting meaningful technology keywords for each patent through text mining and analyzing the communities based on the extracted keywords are proposed in order. In Section 4, the analytical results are presented, and the validity and effectiveness of this research are verified on the basis of these results. Then, the characteristics of the patent keyword network are analyzed based on major network analysis indices including density, clustering coefficient, degree, etc. Furthermore, trend analysis results are presented in order to identify important keywords that can offer insight into the research topic in each field. Finally, in Section 5, the significance and limitations of our research and further research issues are presented.

[1] Denoted as "nBetweenness," the normalized betweenness centrality is the betweenness divided by the maximum possible betweenness and is presented in percentage.

2. Literature review

In this section, existing analysis methods used for patent analysis and the methods used in this research are introduced.

2.1. Patent analysis

A patent consists of the content of technical embodiments, technology classification codes, cited information, and owner information. Technology change trends, technology levels, and commercial values can be understood through the analysis of the component factors. Thus, patent analysis provides important information to relevant persons in charge of R&D, technology policies and/or technology strategies [8,10]. As mentioned earlier, the patent analysis method is largely divided into the network-based and the keyword-based methods.

2.1.1. Network-based patent analysis

Network-based patent analysis is composed of a citation network constructed by citation information, and is analyzed from the viewpoint of the network [8]. In citation networks, patent documents are considered as nodes, and the link represents that a patent has cited other patents.

Yoon & Park [10] proposed an exploratory process of generating a patent network and conducting the ensuing quantitative analysis. They suggested the overall process of developing patent network by integrating text mining and network analysis. Wartburg et al. [8] analyzed the technology cluster group by creating a network of cited patents. Through the analysis of mutual relationships between patents, network-based patent analysis provides information on which patent is influential in a specific technical field and an understanding of the macroscopic flow of the technical field. Li et al. [9] utilized the field of nanoscale science and engineering patent citation information to identify knowledge transfer patterns. In the paper, network topological analysis, core network analysis, and critical node analysis were conducted. As a result, the research found that the patent citation network doesn't have efficient knowledge diffusion capability in terms of average path lengths. Also, the degree distribution of the citation network exhibited that only relatively few patents play a core role. Verspagen [11] traced the technological trajectories in the perspective of patent citation network in fuel cells research. The search path node pair (SPNP) was chosen as an indicator for finding the main path that corresponds to the main flow of ideas. It was found that the technological trajectories by citation paths are selective and cumulative, and also found that the development in part of the citation network moves quite naturally between different levels of aggregation. While the network based patent analysis has advantages such that it can show not only influential patents but also mutual relationships between patents, the network-based patent analysis cannot scrutinize the technical contents of each patent since the method

Please cite this article as: J. Choi, Y.-S. Hwang, Patent keyword network analysis for improving technology development efficiency, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.07.004