Computers & Industrial Engineering 98 (2016) 289–299

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Generating patent development maps for technology monitoring using semantic patent-topic analysis

Mujin Kim, Youngjin Park, Janghyeok Yoon *

Department of Industrial Engineering, Konkuk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form 16 February 2016
Accepted 5 June 2016
Available online 9 June 2016

Keywords:
Patent development map
Technology monitoring
Patent map
Bibliometrics
Latent Dirichlet allocation
Topic analysis
3D printing technology

ABSTRACT

Patent development maps (PDMs) are a useful visual and monitoring tool for technology-trend identification, and therefore proper technology planning, because they provide an overall understanding of a technology's historical development and current stage. The rapid increase in technical data, however, has made it costly and time-consuming to monitor the technology development progress manually. Although some studies have suggested how to identify development paths among patents, little attention has been paid to synthetic consideration of the two core factors for PDMs: (1) the succession relationship among patents in terms of technological content and (2) the technological taxonomies of individual patents. Therefore, this paper suggests a semantic patent topic analysis-based bibliometric method for PDM generation.

The method consists of (1) collecting and preprocessing patents, (2) structuring each patent into a term vector, (3) identifying the technological taxonomies of patents by applying latent Dirichlet allocation, and (4) visualizing the development paths among patents through sensitivity analyses based on semantic patent similarities and citations. This method is illustrated using patents related to 3D printing technology. This method contributes to quantifying PDM generation and, in particular, will become a useful monitoring tool for effective understanding of the technologies including massive patents.

1. Introduction

With the move towards a globalized technological environment, firms are competing for new technologies and securing intellectual property rights to assist in their technological competitiveness. R&D in such environments regards patent analysis before a new project as an essential prerequisite. It is reported that up to 30% of all R&D expenditures is wasted on redeveloping existing inventions (European Commission (EC) 2007). In particular, patents, as the most prolific and up-to-date technology source, contain up to 80% of recent technical information worldwide, because most patent applications are published within 18 months after their first filing, irrespective of their country of origin. These statistics suggest that monitoring current and historical patent advancements is important when developing new technologies.

Patent development maps (PDMs) are one useful tool for monitoring technology. They describe the development relationships among patents within a given technology domain over time. PDMs, as the output of prior art searches, have been widely used in industry, and they include two typical core components (Fig. 1): (1) Technological succession relationships among patents over time and (2) the technological taxonomies of the patents (Yoon & Choi, 2012). First, the technological succession relationship indicates the knowledge link between a former patent and its succeeding patents; thus, it shows the development paths among them. Second, the technological taxonomies in PDMs indicate the subtopics constituting a given technology, and each patent is assigned to one of the taxonomies. Therefore, PDMs with these two core components can provide R&D planners and researchers with an appropriate understanding of the current stages and historical progress of a technology, and thereby assist effective R&D planning from an evolving technological perspective (Choi & Park, 2009).

Customary approaches for PDM development rely on experts creating them manually. However, the rapid increase in the number of global patents has made it difficult to construct PDMs in this manner (Yoon, Park, & Kim, 2013). This problem, in particular, grows more serious in the case of rapidly evolving technologies, such as emerging and high technologies (Yoon, Park, Kim, Lee, & Lee, 2014). Thus, some studies have suggested how to construct patent maps in the forms of network and positioning maps; they are largely grouped into citation-based and content-based approaches.

* Corresponding author.
E-mail address: janghyoon@konkuk.ac.kr (J. Yoon).

http://dx.doi.org/10.1016/j.cie.2016.06.006
0360-8352/

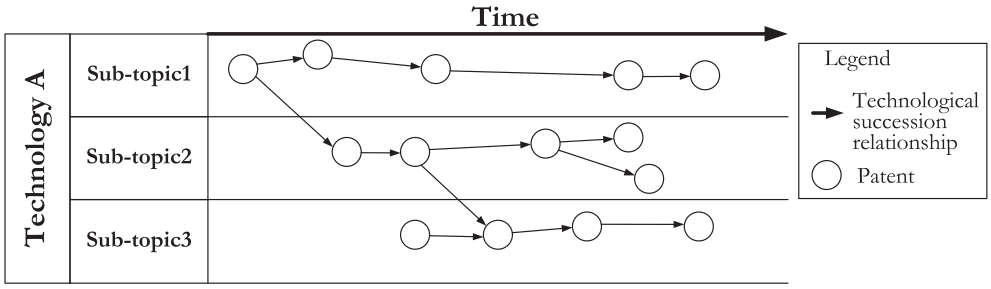

Fig. 1. PDM schematic.

Citation-based studies consider the citations between two patents as knowledge flows (Gress, 2010); therefore, they visualize the knowledge flows (Choi & Park, 2009; Hung & Wang, 2010). Despite their simplicity and ease of use, however, the studies are limited in their ability to identify substantive succession relations among patents in PDMs because they neglect the patent contents; rather, citation-based approaches are a good aid to monitoring the overall trend of widely ranging technologies using large-scale patents.

On the other hand, content-based studies employ text-mining techniques to measure the content similarity between pairs of patents. By exploiting such similarity information, the studies have suggested patent networks (Chang, Wu, & Leu, 2010; Yoon & Kim, 2011; Yoon & Park, 2004) or two-dimensional positioning maps (Bergmann et al., 2008; Yoon & Kim, 2012). However, they have paid little attention to PDM generation that combines the substantive succession relationship among patents with technological taxonomy identification.

As noted above, prior studies used citations or content similarities to generate patent maps in the forms of networks and positioning maps, but, despite their usefulness, have not sufficiently addressed the quantified development of PDMs. Therefore, this study suggests a method of generating PDMs by combining patent citations and semantic patent-topic analysis. The method consists of (1) collecting and preprocessing the patent data of a given technology, (2) structuring each patent into a term vector, (3) identifying technological taxonomies through semantic patent-topic analysis based on latent Dirichlet allocation (LDA), and (4) visualizing the semantic succession relationships among the patents assigned to one of the taxonomies. This method is illustrated using 3D printing technology patents. This method contributes to quantifying the PDM generation process. In addition, as a monitoring tool for a technology's current stages and historical paths, the method will help technology experts understand high technologies; in particular, those which contains massive patents.

The organization of this paper is as follows: First, we present a brief overview of the groundwork, followed by the proposed method. Then, we use 3D printing technology as an illustrative example, and conclude with a discussion and further research.

2. Background

This paper describes how to construct PDMs based on semantic patent topic analysis; therefore, this section presents a brief overview of the technological development path identification, followed by LDA-based topic analysis studies.

2.1. Related work on technological development paths

Visualizing technological development paths is an effective method of providing an overall understanding of the historical stages of a specific technology over time. According to the authors' best knowledge, only a few methodological studies have been conducted to identify and visualize technological development paths. One citation-based study proposed a novel measure called the forward citation node pair by multiplying the number of forward citations of the two linked patents, and then visualizing the relationships among flash memory system patents into a network (Choi & Park, 2009). The study was an initial attempt to identify technological development paths, but its limitation lay in its lack of consideration of the technological content. Another limitation was that each former patent could have a succession relationship with only one of the later patents that cited the former patent.

In light of technological content, a semantic keyword network was suggested to visualize major technology topics and their relationships over time (Kim, Suh, & Park, 2008). The semantic network approach was interesting, but it focused on the chronological technology keywords over time. It did not deal with large-scale patents, their succession relationships, or their technological taxonomies.

Some commercial patent services, including PIAS (www.kipris.or.kr) and WINTELIPS (www.wintelips.com), currently provide patent-based technology development matrices. PIAS and WINTELIPS use international patent classification codes as technological taxonomies and simply allocate each patent into a matrix cell by classification code and year. WINTELIPS also visualizes a network that is composed of patents citing or cited by a patent, but its service logic does not consider the patents' technological content.

Visualizations by prior studies and commercial services are limited in providing the core PDM components. Therefore, our quantified PDM generation method has advantages in that it can identify semantic technological succession relations among patents and extract sub-topics within a given technology.

2.2. LDA in patent analysis

Latent Dirichlet allocation (LDA) is a generative topic model which finds latent topics in a text corpus, based on the assumption that authors generally write documents with respect to specific topics (Blei, Ng, & Jordan, 2003). Using the LDA process, a document is represented as a mixture of topics that produce words with certain probabilities (Fig. 2). Unlike latent semantic analysis (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990), the

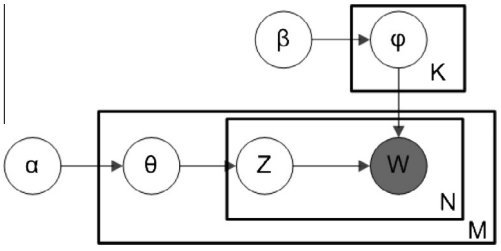

Fig. 2. LDA concept (Blei et al., 2003).