



The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis



Farshad Madani*, Charles Weber

Portland State University, USA

ARTICLE INFO

Article history:

Received 16 June 2015

Received in revised form

12 May 2016

Accepted 16 May 2016

Keywords:

Technology mining

Patent mining

Bibliometrics analysis

Keyword network analysis

Cluster analysis

CiteSpace

ABSTRACT

Text mining methods allow researchers to investigate technical documents (tech mining) and specifically explore patents for valuable information (patent mining). To review literature and analyze the evolution of patent analysis and patent mining methods, bibliometrics analysis and keyword-based network analysis is applied on 143 papers extracted from the 'Web of science' database. Bibliometrics analysis was applied to determine top players researching in patent mining. Applying cluster analysis on the keyword network shows three main stages of patent analysis evolution. Also, it is discussed how patent mining is evolutionized in terms of information retrieval, pattern recognition and pattern analysis.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Patent databases are an important source of information for innovators [1–3], R&D engineers [4,5], corporate executives [6–9], and policy makers in technology latecomer countries [10–12]. Innovators need information on prior art, in order to assess whether their inventions are commercially viable [13]. R&D engineers, who are trying to solve a particular technical problem, want to identify patents that may contribute to the solution of their problem [4,5]. Corporate executives, who are looking for a technology that fits their product strategy, will make use of patent searches to identify how and where they can gain access to the desired technology [6–9]. Policy makers in technology latecomer countries tend to conduct patent analyses, in order to identify particular gaps in the capabilities of their national innovation systems [10–12]. In all the above instances, patent databases serve as a critical source of information upon which policy decisions are based.

For patent databases to be helpful in decision making, the information that they provide must be accurate, presented in a comprehensible format and delivered in a timely manner. This can only be done if the users of patent databases have access to

capabilities in keyword extraction, pattern recognition and pattern analysis. These crucial aspects of modern text mining have thus become an integral component of decision making, both at the strategic and tactical levels.

Patent citation analysis and even established statistical techniques like Term Frequency-Inverse Document Frequency (TF-IDF) for patent keyword analysis do not provide the user of patent databases with an understanding of the content and the context of the patent. The user cannot determine whether a patent contains relevant prior art, unless he/she actually reads the patent. This process is highly inefficient. Tens of millions of patents reside in the databases of the world's major patent offices. The innovator may take years to identify all patents that are relevant. In order for patent databases to be useful for the abovementioned stakeholders, the processes for extracting and analyzing relevant information must be highly efficient. Researchers in academia have made significant progress in the area of applying text mining for keyword extraction [14] and pattern recognition [15–18]. However, the field of pattern analysis is still in its infancy by comparison.

Due to advances in natural language processing, text mining methods and tools have become increasingly available in many different research areas including technology management where scholars try to extract useful information and textual patterns from technical documents, particularly patents. Applying text mining methods in technical documents is named 'tech mining' or

* Corresponding author.

E-mail addresses: fmadani@pdx.edu, farshad.madani@gmail.com (F. Madani).

'technology mining', and for patent analysis purposes, it is named 'patent mining'. Porter, as one of the pioneers in technology mining, has defined 'tech mining' in his book [19, p. 19] as: *'the application of text mining tools to science and technology information, informed by understanding of technological innovation processes.'* Therefore, tech mining has two significant characteristics: 1) using 'text mining tools' and 2) applying these tools to 'technology management'. As shown in Fig. 1, the number of published papers and the number of citations in the tech mining area illustrates a hyperbolical progression; there is a jump in the number of publications after 2005 and a huge rise in the number of citations after 2012.

Given the rapid evolution of patent mining, it is not clear how patent mining has been developed and how the scholars are trying to apply different methodologies to expand this research area. Few papers that shed light on this area and find answers to the above-mentioned questions have been published to date. Abbas et al. [20] have reviewed 22 articles published in the field of patent analysis, and they have provided a general taxonomy of techniques for patent analysis. Also, in an editorial note [21], Porter and Chiavetta investigated six papers published in The First Global Tech Mining (GTM) Conference. They report four main analytics tools which are bibliometrics, data mining, network analysis and cluster analysis. In addition, they reveal eight application areas including emerging technologies and technology dynamics (trend analyses), technology forecasting, roadmapping and foresight, R&D management, engineering industries, science and technology (S&T) indicators, evolutionary economics, technology assessment and impact analysis, as well as science, technology and innovation policy studies.

To conduct a literature review that is as comprehensive as possible, we deploy a methodology that investigates all published patent mining papers in the Web of Science database. We applied bibliometrics analysis to recognize the main papers, authors, universities, and journals. More importantly, we applied cluster analysis on a keyword network, which was extracted from the abstracts of the papers. CiteSpace [22], a free Java application for visualizing and analyzing citations and contents in scientific literature, is applied as the main analysis tool to figure out, detect and visualize emerging trends. CiteSpace is developed by Chaomei Chen whose research is 'information visualization' [23–25]. By applying co-citation network analysis, CiteSpace enables us to identify co-citation clusters and trace how the trend of researches has been developed [24]. The main techniques implemented in the software are spectral clustering and feature selection algorithms [24].

Visualization of the results is the main characteristic of CiteSpace, which assists more analysts to make sense of the trends and evolution [23]. Information visualization in this software is beyond just visualizing graphical displays. This method deploys cognitive, social and collaborative activities [25].

There are some papers in which the authors used CiteSpace as the main tool for bibliometric analysis. Tonta and Darvish [26] used CiteSpace in their research to do social network analysis (cluster methods and centrality measures), and co-occurrence analysis on authors and journals to reveal the social structure of a discipline in terms of collaboration among scientists. In another study, Dhami and Olsson [27] applied CiteSpace to analyze the clusters of co-citations network to figure out the evolution of personal conflict theories. Furthermore, CiteSpace is deployed to study co-citation patterns from 1987 to 2006 to disclose the overall evolution of S&T Policy [28].

2. Data collection

To extract the right papers from Web of Science, it is important to apply keywords that refer to the concept of tech mining. To do so, we consider Porter's definition [19, p. 19] to build the concept of patent mining based on two pillars: 1) 'purposes' and 2) 'applications' of tech mining. This approach helps cover all papers related directly and indirectly to patent mining papers. Since there are many other terminologies used in patent mining that have not used known patent mining terminologies but applied 'text mining' tools for 'technology management purposes'.

To build a reliable keyword list, several tools and techniques were applied. As mentioned above, possible keywords that refer to the concept of patent mining are listed. To figure out the keywords, reviewing publications of renowned authors is a quick trick. For example, the publications of Alan Porter as one of the pioneers in 'tech mining' are beneficial to make a preliminary list of keywords. But there are many authors who use their own keywords, so 'Keyword planner', an option of 'Google Adwords', is utilized to figure out what keywords people have been looking for in Google while they search for 'patent mining'. For example, 'tech mining', 'text mining', and 'text classification' are the most relevant and applied keywords. To test the initial list of keywords, it is necessary to examine two aspects: 'relevance', and 'applicability'. In respect to relevance, this aspect is questioned if the keyword is discussed in both 'patent mining' papers and other research papers.

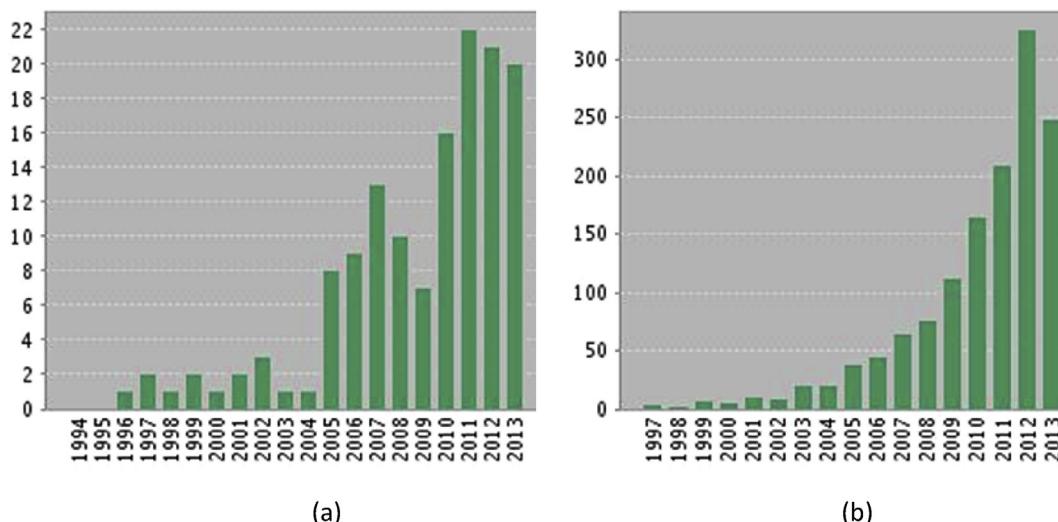


Fig. 1. a) published papers, b) Citations.

Download English Version:

<https://daneshyari.com/en/article/37764>

Download Persian Version:

<https://daneshyari.com/article/37764>

[Daneshyari.com](https://daneshyari.com)