Research Policy 44 (2014) 195–205

Contents lists available at ScienceDirect

Research Policy

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

The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems

Anthony Breitzman, Patrick Thomas*

1790 Analytics LLC, 130 North Haddon Avenue, Haddonfield, NJ 08033, United States

ARTICLE INFO

Article history:
Received 6 March 2013
Received in revised form 19 June 2014
Accepted 23 June 2014
Available online 16 July 2014

Keywords:
Technology
Emergence
Patent
Citation
Forecasting

ABSTRACT

Emerging technologies are of great interest to a wide range of stakeholders, but identifying such technologies is often problematic, especially given the overwhelming amount of information available to analysts and researchers on many subjects. This paper describes the Emerging Clusters Model, which uses advanced patent citation techniques to locate emerging technologies in close to real time, rather than retrospectively. The model covers multiple patent systems, and is designed to be extensible to additional systems. This paper also describes the first large scale test of the Emerging Clusters Model. This test reveals that patents in emerging clusters consistently have a significantly higher impact on subsequent technological developments than patents outside these clusters. Given that these emerging clusters are defined as soon as a given time period ends, without the aid of any forward-looking information, this suggests that the Emerging Clusters Model may be a useful tool for identifying interesting new technologies as they emerge.

1. Introduction

This paper discusses a model designed to help support the efforts of researchers and analysts attempting to locate emerging technologies. This model, named the Emerging Clusters Model, is specifically designed to locate emerging, high-impact technologies in close to real time, rather than retrospectively. That is, it attempts to identify what *is* emerging, not what *has* emerged. The Emerging Clusters Model is based on advanced patent citation techniques, developed to overcome the time lags associated with traditional approaches to patent citation analysis.

Earlier generations of the Emerging Clusters Model have been discussed in previous papers (Thomas and Breitzman, 2006; Chang and Breitzman, 2009). Those earlier generations of the model were largely exploratory, and covered limited patent systems and time periods. The current generation of the model, discussed in this paper, is much more ambitious, and covers multiple patent systems over an extended time period. This generation of the model is also much more flexible, and accounts for differences in referencing practices across patent systems, and changes in these practices over time. As a result, the model described in this paper is specifically designed to be extensible, both in terms of patent systems covered, and time periods examined.

This paper also describes the first large-scale, longitudinal test of the Emerging Clusters Model. The papers describing earlier generations of the model report promising results from individual years, but point to the need for a more general test of the model covering multiple years. This paper describes such a test, covering results from Emerging Clusters selected each year between 1980 and 2006 (the most recent year for which sufficient data are available to track the impact of patents in these clusters, as discussed later).

2. Background

In the study of innovation, a great deal of attention is paid to emerging technologies. Such technologies have the potential to be highly generative, and may open up whole new areas of technology and science. This potential often draws interest from various organizations. These include government agencies looking to fund promising new ideas, corporations hoping to gain a foothold in rapidly emerging fields, and investment institutions seeking returns from early investments in key innovators.

There is a long history of research related to identifying and characterizing emerging technologies, dating back to when Schumpeter (1912) coined the term 'creative destruction' to describe the emergence of new technologies, which spawned new industries while destroying old ones. More recently, Christensen

* Corresponding author. Tel.: +1 856 216 1790x12; fax: +1 856 216 2262.
E-mail addresses: abreitzman@1790analytics.com (A. Breitzman), pthomas@1790analytics.com (P. Thomas).

http://dx.doi.org/10.1016/j.respol.2014.06.006
0048-7333/

and Bower (1996) used the term 'disruptive technology' to describe a new development that disrupts the status quo in an existing technology.

Researchers have examined a range of issues related to emerging and disruptive technologies, and there are a number of well-established methods in this area. These include approaches to understanding how new technologies emerge, and deriving tools and frameworks such as TRIZ (Altshuller, 1999) to assist inventors and organizations in developing such technologies. They also include methods for forecasting future developments in a given technology. For example, the Delphi Method (Linstone and Turoff, 1975) combines opinions from a panel of experts using a systematic, iterative process, based on the idea that forecasts combining multiple expert opinions are likely to be more accurate than forecasts from a single source. Another well-established approach often used in technology forecasting is the Bass Model (Bass, 1969). This was originally developed to model new product adoption, but it has also been used in technology forecasting. It models the numbers of 'innovators' and 'imitators' in a given emerging technology over time, and forecasts diffusion of the technology based on these numbers.

In addition to these long-established approaches to technology forecasting, there have also been many recent research efforts in this area. These include strategic planning frameworks such as technology roadmapping (Phaal et al., 2004), which provides a means for organizations to track the development of potentially disruptive technologies. They also include methods such as forecasting innovation pathways (FIP) for forecasting the likely future development of selected technologies regarded as potentially emergent (Robinson et al., 2013). FIP provides a ten-step approach to generating such forecasts, incorporating both document analysis and expert opinions. Daim et al. (2006) also suggest a combination of document and expert analysis in the generation of forecasts in three specific technologies – fuel cells, food safety, and optical storage.

Alongside these research efforts, there have also been specific government sponsored programs directed toward forecasting technological emergence. Notable among these is the European PromTech project, which endeavors to locate emerging technologies via analysis of scientific literature (Roche et al., 2010; Schiebel et al., 2010). Technological emergence has also attracted attention from researchers applying methods drawn from other subjects. For example, Yu and Lee (2013) used self organizing maps (SOM), which are artificial neural networks, to identify promising emerging technology clusters. Meanwhile, Kim et al. (2012) carried out technology forecasting via text mining, with the data modeled quantitatively using decision trees.

Despite this widespread interest in emerging technologies, identifying such technologies is far from straightforward. This is especially true given the widespread electronic communication and publication of information, which has meant that the sheer scale of information on a particular subject can be overwhelming. Researchers and analysts have to search through this mass of information to locate interesting emerging technologies. Their task is not aided by the fact that the number of truly emergent technologies is dwarfed by the number of mature, mundane, or failed technologies.

One way of reducing the information load associated with locating emerging technologies is to focus on a single unit of analysis, with patents being a popular option. Patents are not always easy to work with, and they have several shortcomings, notably that technologies may be held as trade secrets rather than patented. Having said this, patents represent a vast technology-related data source. According to Kahaner (1996), more than 75% of information contained in US patents is never released anywhere else. Other data points also highlight the increasing importance attached to patents, both at the company level and the country level. For example, the recent smart phone 'patent wars' between Apple, Google, Samsung and others have seen companies in the smart phone industry build patent portfolios as rapidly as possible, both through internal invention and through external acquisition (Carrier, 2012). Meanwhile, the emergence of China as a major economic force has been reflected in patent filings from Chinese inventors. There was a fourfold increase in Chinese-invented US patents between 2006 and 2011 (Source: US Patent & Trademark Office), and a threefold increase in patent filings at the European Patent Office from Chinese inventors over the same time period (Source: European Patent Office). There has also been a rapid growth in patenting at the Chinese Patent Office, with domestic filings increasing threefold between 2006 and 2010, and foreign filings almost doubling over the same time period (Source: State Intellectual Property Office of the People's Republic of China).

While focusing on patents reduces the amount of information to be processed, patents in themselves still represent a very large data source. Over 253,000 utility (i.e. invention) patents were granted in 2012 in the US system alone, bringing the total number of US patents granted to well over eight million. In the search for emerging technologies, most of these patents will be of little interest, perhaps because they describe incremental advances on mature technologies, or because they describe technologies with relatively low potential. The purpose of the Emerging Clusters Model is to sort through this mass of patents, and identify the relatively small number with a strong, and developing, impact. As such, it represents a 'big data' approach to technology forecasting, involving the mining of large data sets in order to extract meaningful patterns and information.

3. Overview of model

The Emerging Clusters Model uses an advanced form of patent citation analysis, a widely used technique for assessing the impact of patents on subsequent technological developments. The basic idea behind patent citation analysis is that highly cited patents (i.e. patents cited as prior art by many later patents) tend to contain technological information of particular importance. As such, they form the basis for many new innovations, and so are cited frequently by later patents. This does not mean that every highly cited patent is important, or that patents cited infrequently are necessarily trivial. However, numerous validation studies have revealed the existence of a strong positive relationship between patent citations and measures of technological importance and commercial value, although there have been dissenting voices (Wang, 2007). Useful overviews of such validation studies can be found in Breitzman and Mogee (2002), Sampat and Ziedonis (2004), and Hsieh (2011).

From a technological perspective, one early validation study (Carpenter et al., 1981) found that patents related to IR 100 invention awards were cited twice as often as peer patents. Meanwhile, using surveys, Albert et al. (1991) demonstrated that patents identified by industry experts as important were cited more frequently by later patents than patents identified by experts as less important. Trajtenberg (1990) also reported a close association between citation-based patent indices and independent measures of the social value of innovations. More recently, Blind et al. (2009) reported higher citation rates for patents that protect technological advancements, versus patents filed largely for defensive or bartering purposes; while Barberá-Tomás et al. (2011) demonstrated how patent citations are a valid representation of knowledge flows within a network.

From an economic perspective, numerous studies have revealed a positive relationship between patent citations and economic value. At the level of individual patents, Harhoff et al. (1999)