Technological Forecasting & Social Change xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Analysis of patent documents with weighted association rules

Serkan Altuntas [a,*], Turkay Dereli [b,1], Andrew Kusiak [c,2]

[a] Bayburt University, Department of Industrial Engineering, Bayburt, Turkey
[b] Gaziantep University, Department of Industrial Engineering, Gaizantep, Turkey
[c] The University of Iowa, Department of Mechanical and Industrial Engineering, Iowa City, IA, USA

ARTICLE INFO

Article history:
Received 1 March 2014
Received in revised form 20 August 2014
Accepted 21 September 2014
Available online xxxx

Keywords:
Patent analysis
Weighted association rules
Patent citation
Data mining

ABSTRACT

Analysis of relationship among various technologies may result in technology maximizing profits. Weighted association rules are proposed to determine dependencies across technologies. The proposed approach recognizes uneven importance of patents and technology class in terms of their technological impact and commercial significance. The proposed analysis approach is illustrated with a case study.

1. Introduction

Patent information has been used to study technological developments, trends, and technology potential (Zhang, 2011; Pilkington et al., 2009) as well as decision making in research and development (Thorleuchter et al., 2010). Innovation trends can be analyzed by using patents (Engler and Kusiak, 2008). Information from patent data allows companies to avoid investing in obsolete technology (Wang et al., 2012) and it enhances strategic planning (Abraham and Moitra, 2001). Patent analysis offers key information concerning technology environment (Porter and Cunningham, 2005) and addresses component technologies (Trappey et al., 2012). Therefore, patent analysis is suitable for determining relationship among different technologies. Details of patent analysis approaches are provided in Porter and Cunningham (2005), Ernst (2003), Breitzman and Mogee (2002), and Griliches (1990).

The International Patent Classification (IPC) system developed following the Strasbourg Agreement in 1971 (http://www.wipo.int/classifications/en/) for classification of patents. Based on this system, each patent is assigned at least one IPC code according to its scope. The number of patents associated with an IPC code can be considered as an indicator of technology development in the area covered by these patents (Dereli et al., 2011).

In this paper, four weighted association rule based algorithms: *binary weighted association rule mining* (BWARM), *weighted association rule mining* (WARM) which includes two different approaches, namely WARM_1 and WARM_2, and *boolean weighted association rule mining* (BOWARM) are proposed to analyze relations among technologies. Weights are used to reflect the technological impact and commercial importance of patents for WARM_2 and to reflect the technological impact and commercial importance of technology class for BWARM, WARM_1, and BOWARM. Three algorithms (BWARM, WARM_1, and BOWARM) consider the number of patents included in IPC codes to compute the weight (see Tables 1 and 2). A high number of patents covered by an IPC code indicate high interest in the technology represented by this IPC code. The fourth algorithm (WARM_2) uses the number of patent citations to compute the weight (see Table 3). In this paper, the weighted association rule algorithm considers the importance of patents and technology class.

Yoon and Kim (2012) and Choi and Park (2009) discussed the citation-based patent analysis. A large number of citations of

* Corresponding author. Tel.: +90 507 769 69 63; fax: +90 458 211 11 72.
E-mail addresses: saltuntas@bayburt.edu.tr (S. Altuntas), dereli@gantep.edu.tr (T. Dereli), andrew-kusiak@uiowa.edu (A. Kusiak).
[1] Tel.: +90 342 360 10 23; fax: +90 342 360 19 22.
[2] Tel.:+1 319 335 5934; fax: +1 319 335 5669.

http://dx.doi.org/10.1016/j.techfore.2014.09.012
0040-1625/

Table 1
Transactions.

Patent no	Items (IPC codes)
USX	B07C, C02F, E01B, F03B
USY	F03B, F03G, B07C, C02F
USZ	H01B, B07C, C02F

a patent indicate that the examined patent is an important invention leading to technology improvement (Deng et al., 1999). Citations are largely used to approximate the value of invention and thus the value of knowledge contained in a patent (Gay et al., 2005). Tseng (2009) pointed out to validation studies demonstrating a positive relationship between the patent citations and importance of technology. Carpenter et al. (1981), Trajtenberg (1990), Albert et al. (1991), Narin (1993), Karki (1997), Harhoff et al. (1999), Trajtenberg (1999), Jaffe et al. (2000), Harhoff et al. (2003), Hall et al. (2005), Gay et al. (2005), and Schneider (2007) discussed association between patent citations and their importance. Therefore, the fourth algorithm (WARM_2) uses the number of citations of a patent in determining the importance of invention.

One of most widely used data mining algorithms is the association rule algorithm. It extracts relationships between various objects (here IPC codes). Details of association rule algorithms are provided in Hipp et al. (2000), Zhao and Bhowmick (2003), Kotsiantis and Kanellopoulos (2006), and Forsati and Meybodi (2010). Jun (2011a,b) and Kim et al. (2011a,b) applied classic association rules to patent documents. The classic association rules cannot capture meaningful relationships among the IPC codes because the importance of patents varies. To address this limitation, weighted association rule algorithms are applied. The approach proposed in this paper finds associations between different technologies and determines possible technology clusters by using patent information, namely the number of patent, IPC code, issue data and the number of citation (see Figs. 1 and 2).

The remainder of this paper is organized as follows. The literature on associations among technologies and applications of data mining to patent analysis is reviewed in Section 2. Three patent analysis algorithms are proposed in Section 3. The application of the proposed approach to patent analysis is presented in Section 4. The limitations of the proposed approach are summarized in Section 5. Conclusions and future research directions are provided in Section 6.

2. Literature review

The literature surveyed in this paper is grouped into studies of relationships among various technologies and mining patent information.

Table 2
Weighted items database.

ID	Item	Number of patents	Weight
1	B07C	100	0.1 (100/1000)
2	C02F	200	0.2
3	E01B	250	0.25
4	F03B	75	0.075
5	F03G	75	0.075
6	H01B	300	0.3
Total		1000	1

2.1. Literature on relationships among technologies

A forecasting technique, Cross-Impact Analysis (CIA), has been frequently used to examine relationship among different technologies. CIA uses a cross-impact matrix representing causal relationships between events (Jeong and Kim, 1997). It determines a conditional probability of an event given that other events have or have not occurred (Cho and Kwon, 2004). Details of CIA can be found in Dalkey (1971–1972). Jeong and Kim (1997) presented a cross-impact model based on fuzzy logic for qualitative analysis of technologies. They constructed a cross-impact matrix with fuzzy relations. They also obtained answers to questions from experts responding with Certain (C), Very strong (VS), Strong (S), Medium (M), Weak (W), Very Weak (VW), and None (N) to find key technologies through the technological impact relationships. Choi et al. (2007) used CIA to find associations among information and communication technologies. They collected patent data from the USPTO using a self-developed code. In their study, a cross impact network was suggested to analyze complex relationships among technologies. Thorleuchter et al. (2010) proposed a quantitative CIA approach to forecast impact among technologies and defense applications.

In addition to these studies, Yoon and Park (2004) applied text mining for patent analysis to illustrate relationships among patents as a visual network. Lee et al. (2009a,b) proposed analytic network process (ANP) approach to find interactions between technologies. They conducted a case study of telecommunication technologies based on patent citation data to show application of the proposed approach. Furthermore, No and Park (2010) used citation numbers to develop trajectories of technology fusion for nanobiotechnology. In their study, they found relations (citation networks) for nanobiotechnology at the patent class level. Kim et al. (2011a,b) analyzed patents to identify relationships between technology-based services and Information and Communication Technologies (ICTs). They conducted patent citation analysis to find the technological linkage between technology-based services and ICTs.

2.2. Literature on data mining in patent document

Data mining offers tools for analysis of data and knowledge discovery. The result provided by data mining algorithms may range from providing a general understanding of the nature of the data to accurate modeling and prediction (Kusiak and Smith, 2007). Researchers applied data mining approaches such as association rules, text mining, clustering, classification, and time series analysis to patents. The number of studies on association rules applied to patent documents is limited. Jun (2011a) used association rules to analyze patents related to the technologies of database theory and its applications. Each patent was considered as a transaction and each classification code of a patent was considered as an item. Jun (2011a) computed the value of support, confidence and lift to find association among codes of technology. Jun (2011b) combined association rule mining and self organizing maps to forecast the vacant technology of image and video technology as well. Patent data retrieved from the USPTO database was examined by association rule mining. Self-organizing maps were used to classify the vacant technology of image and video technology. Kim et al. (2011a,b) applied association rule mining, ANP, and Cross-Impact Analysis to determine interactions among technologies. Conditional probabilities

Please cite this article as: Altuntas, S., et al., Analysis of patent documents with weighted association rules, Technol. Forecast. Soc. Change (2014), http://dx.doi.org/10.1016/j.techfore.2014.09.012