Expert Systems with Applications 38 (2011) 12559–12564

Contents lists available at ScienceDirect

Expert Systems with Applications

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

# Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach

Chulhyun Kim [a,1], Hakyeon Lee [b,*], Hyeonju Seol [c,2], Changyong Lee [d,3]

[a] Induk University, Choansangil 14, Nowon-Gu, Seoul 139-749, South Korea
[b] Seoul National University of Science and Technology, 172 Gongreung 2-dong, Nowon-gu, Seoul 139-746, South Korea
[c] Korea Air Force Academy, 335-1, Ssangsu-ri, Namil-myeon, Cheongwon-gun, Chungbuk 363-849, South Korea
[d] Seoul National University, San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742, South Korea

ARTICLE INFO

*Keywords:*
Association rule mining (ARM)
Analytic network process (ANP)
Cross-impact analysis (CIA)
Core technology
Patent co-classification

ABSTRACT

This study proposes a new approach to identifying core technologies from a perspective of technological cross-impacts based on patent co-classification information with consideration of the overall interrelationships among technologies. The proposed approach is comprised of two methods: association rule mining (ARM) and the analytic network process (ANP). Firstly association rule mining (ARM) is employed to calculate the technological cross-impact indexes. Since the confidence measure in ARM is defined as a conditional probability between two technologies, it is adopted as an index for evaluating technological cross-impacts. The technological cross-impact matrix is then constructed with all calculated cross-impact indexes. Secondly, the ANP, which is a generalization of the analytic hierarchy process (AHP), is conducted to produce priorities of technologies with consideration of their direct and indirect impacts. The proposed approach can be utilized for technology monitoring for both technology planning of firms and innovation policy making of governments. A case of telecommunication technology is presented to illustrate the proposed approach.



## 1. Introduction

The characteristics of modern technological changes can be defined as complexity and radicalness. Under such environment, it has become more important to grasp technological trends and advances by analyzing the overall structure of technologies and interaction among them (Lee, Kim, Cho, & Park, 2009). It is considered to be an indispensible activity, in particular, for seeking technological possibilities through technological fusion among various fields of technologies such as IT, BT, and NT. Consequently, there have often been attempts to identify technological structure and relationship.

The identification of technological structure and relationship is mainly conducted through the patent analysis (Trajtenberg, 1990). It is reported that patents contain about 80% of all technological knowledge (Blackman, 1995), and they can be easily accessed and analyzed through various types of public or private databases. Patents are, hence, perceived as useful information for techno-economic analysis and R&D management (Yoon & Park, 2004). A lot of studies have attempted to analyze technological relationships using patent information.

The most commonly used information of patents for analyzing technological relationships is citation information. The basic assumptions of citation analysis are that the knowledge of cited patent is transferred to a citing patent, and there exists a technological linkage between them. Citation analysis is a useful method for identifying technological relationships, and this can be verified with various studies (Basberg, 1987; Hu & Jaffe, 2003; Jaffe & Trajtenberg, 1999; Trajtenberg, 1990; Yoon & Park, 2004). However, they have some shortcomings reported in the literature. First, the average time-lag between citing and cited patents is over 10 years (Hall, Jaffe, & Trajtenberg, 2001). Moreover, since citation analysis only considers citing-cited relationships between individual patents, it is difficult to identify technological relatedness and characteristics from a perspective of technological fields (Yoon & Park, 2004).

To address this limitation, there have been attempts to applying other pieces of information of patents such as co-citation (Lai & Wu, 2005; Stuart & Podoly, 1996), co-word (Courtial, Callon, & Sigogneau, 1993), and keyword vector (Yoon & Park, 2004) for the analysis of technological relationship. They also have, however, their own weaknesses. There is still a time-lag problem in the co-citation analysis. Co-word analysis and keyword vector analysis require qualitative judgments, and therefore have a lack of consistency in the result of analysis.

* Corresponding author. Tel./fax: +82 2 970 6469.
*E-mail addresses:* stddevs@induk.ac.kr (C. Kim), hylee@seoultech.ac.kr (H. Lee), hjseol@afa.ac.kr (H. Seol), akuta7@snu.ac.kr (C. Lee).
[1] Tel./fax: +82 2 950 7608.
[2] Tel.: +82 43 290 6492; fax: +82 43 297 0479.
[3] Tel./fax: +82 2 878 3511.

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.042

On the contrary, the patent analysis with co-classification information has some advantages over the above mentioned measures. Co-classification analysis is to analyze technological relationships based on the fact that patents are classified to some technological classes considering their technological characteristics (OECD, 1994). The assumption of the co-classification analysis is that the frequency by which two classification codes are jointly assigned to a patent document can be interpreted as a sign of the strength of the knowledge relationships, in terms of knowledge links and spillovers (Breschi, Lissoni, & Maleraba, 2003). In contrast to the citation analysis, it is based on the hierarchical technological classification system so that technological relationships can be analyzed not on the level of individual patents but on the various technological levels according to the purpose of studies. Furthermore, errors from the time-lag problem are relatively insignificant since the time of classification information of a patent is equal – patent registration time.

Among the various techniques using the information of patent co-classification, technological cross-impact analysis (CIA) has been used to identify core technologies and interrelationships between technologies by analyzing the cross-impact between technologies quantitatively based on patent classification data (Choi, Kim, & Park, 2007). In the patent-based CIA, the cross-impact index between two technologies is calculated with the probabilities based on the patent co-classification information to analyze the impact between technologies. However, it is subject to some limitations. Firstly, it is nearly impossible to calculate cross-impact indexes without developing a computer program because it requires a huge amount of calculation with patent data. Secondly, regarding to the identification of core technologies, the previous patent-based CIA does not take into account the overall interrelationships among technologies, only considers the relationships between two technologies.

In response, this paper proposes a new approach to identifying core technologies from a perspective of cross-impacts based on patent co-classification information with consideration of the overall interrelationships among technologies. The proposed approach is comprised of two methods: association rule mining (ARM) and the analytic network process (ANP). At first, association rule mining (ARM) is employed to calculate technological cross-impact indexes. ARM is one of the representative data mining techniques for exploring vast database. Since the confidence measure in ARM is defined as a conditional probability between two technologies and is of the same formula with cross-impact index, it is adopted as the index of evaluating technological cross-impacts. Then, the technological cross-impact matrix is constructed with all calculated cross-impact indexes. Secondly, the ANP, which is a generalization of the analytic hierarchy process (AHP), is employed to identify core technologies based on the cross-impact matrix. Since the ANP is capable of measuring the relative importance that captures all the indirect interactions in a network, the derived limit priority indicates the importance of a technology in terms of impacts on other technologies, taking all the direct and indirect influences into account (Lee et al., 2009).

The remainder of the paper is organized as follows: Section 2 deals with methodological background including CIA, ARM, and the ANP. The proposed approach is explained in Section 3, and illustrated in Section 4. The paper ends with conclusions in Section 5.

## 2. Methodological background

### 2.1. Cross-impact analysis (CIA)

The changing or evolving process of a system could be regarded as a set of some events. Since they interact with each other, the occurrence of a specific event takes an effect on the probability of other events' occurrence. Therefore, it is impractical to forecast the probability of an event's occurrence without considering the occurrence of other events. Like social systems, the technological change or progress occurs as a result of the occurrence of various events. For example, the development of mobile phone has to do with that of technologies such as mobile network, memory, and liquid crystal display. When technological events occur as a result of the interactions with each other, an impact of each event of interest on other events is called the cross-impact (Jeong & Kim, 1997; Lee, 1997; Schuler, Thompson, Vertinsky, & Ziv, 1991). Accordingly, CIA has been used as a technique for forecasting the emergence of new technologies and to identify the interrelations between technologies by defining the emergence of new technologies as event occurrences (Choi et al., 2007).

The general process of CIA is as follows: (1) define the events to be included in the analysis, (2) estimate the initial probability of each event, (3) estimate the conditional probabilities for each event pair, (4) perform a calibration run of the cross-impact matrix, (5) evaluate the results. In conventional CIA, the step (2) and (3) require the experts' subjective judgments based on their domain knowledge and therefore inconsistent estimates may result. Further, in the step (4), the two kinds of probabilities derived from the former steps should be adjusted because of the intuitive estimation. To overcome these shortcomings of the conventional CIA, a patent-based CIA was proposed which analyzes cross-impacts between technologies quantitatively based on patent classification data (Choi et al., 2007).

In this study, the cross-impact of technology '*A*' on the technology '*B*' is defined as the conditional probability $P(B|A) = N(A \cap B)/N(A)$ where $N(A)$ refers to the total number of patents classified in technology *A*, and $N(A \cap B)$ indicates the number of patents classified in both technology *A* and *B*.

### 2.2. Association rule mining (ARM)

ARM is one of the data mining techniques to search for interesting relationships among items in large database. An association rule stands for the co-occurrence of two items, and indicates that if two items occur together frequently they have a strong association relationship (Han & Kamber, 2001). ARM has mainly been applied to firm activities, especially to marketing (Liao & Chen, 2004). It has also been used in various areas such as bioinfomatics (Creighton & Hahash, 2003; Oyama, Kitano, Satou, & Ito, 2002), medicine (Ca & Jiang, 2003), and finance (Hsieh, 2004).

The three measures for evaluating the rule interestingness are support, confidence, and lift. Their brief descriptions are shown in Table 1. The typical procedure of ARM consists of two steps (Agrawal, Imielinski, & Swami, 1993): (1) search for frequent itemsets – to create all item combinations over the threshold value of support, (2) generate association rules – to select itemsets over the threshold value of confidence or lift among the frequent itemsets found in (1). The step (1) is a very time consuming job and the most representative technique for this is Apriori algorithm (Agrawal & Srikant, 1994).

### 2.3. Analytic network process (ANP)

The ANP is a generalization of the AHP (Saaty, 1996). As mentioned before, the AHP decomposes a problem into a hierarchy in which each decision element is considered to be independent; thus, it cannot accommodate interrelationships among elements. The ANP extends the AHP to problems with dependence and feedback. It allows for more complex interrelationships among decision elements by replacing the hierarchy in the AHP with a network. Thus, the ANP produces priorities or relative importance of