Technological Forecasting & Social Change 105 (2016) 94–104

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Product opportunity identification based on internal capabilities using text mining and association rule mining

Wonchul Seo [a], Janghyeok Yoon [b], Hyunseok Park [c], Byoung-youl Coh [d,*], Jae-Min Lee [d], Oh-Jin Kwon [d]

[a] Division of Systems Management and Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 608-737, Republic of Korea
[b] Department of Industrial Engineering, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 143-701, Republic of Korea
[c] Engineering Systems Division, Massachusetts Institute of Technology, 265 Massachusetts Avenue, Building N52-373B, Cambridge, MA 02139, USA
[d] Technology Opportunity Research Team, Korea Institute of Science and Technology Information, 66, Hoegi-ro, Dongdaemun-gu, Seoul 130-741, Republic of Korea

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form 3 December 2015
Accepted 15 January 2016
Available online 16 February 2016

Keywords:
Product opportunity identification
Internal capability
Text mining
Association rule mining
Patent analysis

ABSTRACT

Identifying new product opportunities must be a prerequisite for a firm's sustainable growth since it can help create new market segments. In this regard, a number of studies have attempted to suggest systematic methods to discover new technological opportunities. However, from these methods, it is difficult to figure out which products can come into the market as a result of the technological opportunities. Moreover, they have tried to measure generic potential values without considering a specific target firm so it is hard to judge whether the discovered opportunities are technically feasible to the target firm. These problems tend to reduce the practicality of the discovered technological opportunities. Therefore, this paper proposes a systematic approach to identify potential product opportunities by reflecting the target firm's internal capabilities. The capabilities are inherently unobservable so we need to figure out substitutes for the firm's capabilities. The existing products belonging to a firm can be generally a basis for developing new products. The firm is already good at dealing with the existing products so we consider the firm's existing product portfolios its internal capabilities. We first extract product information from patent database using text mining technique, and then generate product connection rules represented as directed pairs of products. Finally, we evaluate potential value of product opportunities taking into account a firm's internal capabilities. An empirical study is conducted to show the applicability of the presented approach using patents granted in the United States Patent and Trademark Office during 2009 and 2013. We expect that our approach can facilitate product-oriented R&D by presenting a front-end model for new product development and deriving feasible product opportunities according to the target firm's internal capabilities. Moreover, the presented systematic approach can be a basis for an R&D planning system that can help R&D planners in performing product-oriented technology planning activities.

1. Introduction

Identifying new technological opportunities have been crucial to preserve a firm's sustainable growth since it can help secure innovative and advanced technologies (Park et al., 2013). Resulting from this importance, there have been various studies on how to systematically discover technology opportunities: identifying opportunities using keyword-based morphology analysis (Yoon and Park, 2005), drawing a technology map using principal component analysis and defining evaluation indicators to measure potential value of vacant areas on the map (Lee et al., 2009), and detecting signals of technology opportunities analyzing outliers on the similar technology map (Yoon and Kim, 2012).

They have presented systematic methods to discover technology opportunities, but they have only focused on suggesting conceptual implications of technologies. How to create new products using the implications has not been presented. Therefore, they cannot show which products can come into the market as a result of the discovered implications. This problem eventually leads to lower reliability of the technology opportunities. Moreover, the previous methods have dealt with measuring generic potential values without considering the internal capabilities of firms that are willing to invest on research and development (R&D). The generic potential values of technologies can only indicate that it is possible to dominate the future technological landscape only if the potential technologies are developed as expected. Therefore, it should be considered whether the opportunities are feasible from the perspective of a firm's internal capabilities. The only potential technologies associated with the internal capabilities can be thought as the real opportunities for firms (Song et al., 2012).

The literature discussed two types of innovations: radical innovations and incremental innovations (Chandy and Tellis, 1998). Among

* Corresponding author.
E-mail addresses: wcseo@pknu.ac.kr (W. Seo), janghyoon@konkuk.ac.kr (J. Yoon), parkhs@mit.edu (H. Park), cohby@kisti.re.kr (B. Coh), jmlee@kisti.re.kr (J.-M. Lee), dbajin@kisti.re.kr (O.-J. Kwon).

http://dx.doi.org/10.1016/j.techfore.2016.01.011
0040-1625/

them, incremental innovations generally involve the improvements in existing product lines (Rubera and Kirca, 2012). Therefore, from the perspective of incremental innovations, the existing products belonging to a certain firm can be generally a basis for developing new products (Rubera and Kirca, 2012). The firm is already good at dealing with the existing products so we can consider the firm's existing product portfolios its internal capabilities. In this regard, we present a systematic method for identifying product opportunity based on the internal product portfolios. The method basically uses patent data to derive product implications by formulating relationships between products. As a patent has been thought as a straightforward proxy for identifying the level of technologies due to its feature as an up-to-date reliable source of technological intelligence (N. Ko et al., 2014; Yoon and Kim, 2011), patent data are widely used to analyze technological trend and landscape such as depicting technological knowledge flows between technology classes (Lai and Wu, 2005; Stuart and Podolny, 1996), exploring technological opportunities (Yoon et al., 2014), analyzing interdisciplinarity of technology fusion (N. Ko et al., 2014), and measuring the extent of technology impact for R&D planning (S.-S. Ko et al., 2014). The use of patent analysis facilitates firms to analyze the technological trends systematically so that they can discover new technology opportunities (Curran and Leker, 2011; Chen and Chang, 2012). The patent analysis-based studies generally tend to focus on a technology-oriented perspective since the patents mainly deal with the technological contents. Many industry fields heavily depend on the technological elements in most cases but there must be other fields relying on the theoretical principles rather than technological aspects such as construction, chemistry, physics, and mathematics. To encompass these industries, the analysis on the contents of academic papers and commercial reports has to be conjoined with the patent analysis (Kim et al., 2009). Despite the restrictions of using only patent data, the target of our research is to identify new opportunities of practical products that can be utilized in the market. Therefore, this research would like to build on the patent analysis. It indicates that our method may not work very well at the perspective of some industries that largely depend on the theoretical principles.

The presented method first extracts product information from patent documents using text mining, and then generates association relationships between products employing association rule mining. Association rule mining as one of the data mining techniques reveals specific relationships among several items exploring their co-occurrences in a huge database (Kim et al., 2011). Finally, the method quantitatively identifies product opportunities for a certain firm taking into account its existing product portfolios. For the quantitative analysis, an evaluation indicator is presented to measure potential value of each individual product. The indicator reflects whether each product is strongly associated with the firm's existing internal product portfolios. An empirical study is conducted to show the applicability of the presented method. We expect that our research can facilitate product-oriented R&D which is more market friendly than general R&D by identifying product opportunities according to the firm's internal product portfolios. Furthermore, the presented systematic method can be a basis for an R&D planning system that aids planning directors in performing product-oriented technology planning activities.

2. Groundwork

2.1. Product opportunity identification

For organizational competitiveness, it has become the most important aspect of R&D to find out new product items that appreciated by customers (Wolff and Pett, 2006; Kowang et al., 2014). R&D planners are extremely keen to draw critical factors that have high impact on the success of developing new products (Chen and Chen, 2009). Among various factors, the front-end of new product development has been considered the most influential in terms of its effect on the entire process of product development (Oliveira and Rozenfeld, 2010). Various front-end models have been presented and they commonly point out the importance of a product opportunity identification activity since it must be a starting point for product development processes (Koen et al., 2001; Cooper, 2001). Technology opportunity can be realized through product development so identifying product opportunities can be judged as including the discovery of technology opportunities (Yoon et al., 2014). Therefore, this paper focuses on how to systematically identify product opportunities based on the direct relationships between products themselves.

2.2. Text mining

Patent analysis for exploring technological trend and landscape mostly utilizes bibliometric data of patent documents such as patent classification codes and citation information (Narin, 1994; Trajtenberg et al., 1997). However, these approaches using the bibliometric data naturally exclude the technological meanings implied in patent documents (N. Ko et al., 2014; Choi et al., 2013). To remedy this problem, various studies have tried to encompass the technological contents described in patent documents into the patent analysis processes by adopting text mining techniques (Choi et al., 2012). Text mining generally deals with how to extract latent knowledge from unstructured textual descriptions (Yoon et al., 2014). Applying text mining techniques to patent analysis enables to reveal technological details, implications, and trends and subsequently it can be helpful make R&D strategies and investment policies (Tseng et al., 2007). In this paper, we utilize the text mining techniques to elicit product information from textual descriptions in patent documents and systematize the information to be employed in the phase of generating rules of product connections.

2.3. Association rule mining

As one of the data mining techniques, association rule mining helps find out intriguing relationships among items in a huge database (Kim et al., 2011; Shih et al., 2010). It has an assumption that there must be some hidden relationships between purchased items in transactions (Agrawal et al., 1993; Kuo et al., 2011). Therefore, we can come to see and understand customers' purchasing behavior as long as we take out the hidden relationships. Resulting from this feature of association rule mining, it has been applied in various research areas such as assigning products in retail (Ahn, 2012), predicting potential defects in software development (Song et al., 2006), identifying core technologies from the perspective of technological cross-impacts (Kim et al., 2011), and designing convergent product concepts based on the combination of association rule mining and decision tree (Lee et al., 2012).

To determine relevance of mined rules, two measures, support and confidence, are used. The support measure evaluates the probability that items occur in transactions (Shih et al., 2010). For an example rule $A \rightarrow B$, the support measure of A means the probability that item A occurs in all transactions so it can be formulated as $P(A)$. The support measure of B has similar meaning. To determine the usefulness of the mined rules, both support measures of A and B should be considered (Kim et al., 2011). The confidence measure evaluates the conditional probability that consequent items of the mined rule occur in transactions given that conditional items have already occurred in the same transactions (Shih et al., 2010). For the same example rule $A \rightarrow B$, it can be formulated as $P(B|A)$. A typical algorithm to generate association rules is the apriori algorithm (Agrawal et al., 1993). It first draws item sets that have support measure exceeding a pre-defined minimum support threshold, and then generates association rules among the item sets that have high confidence value than a pre-defined minimum confidence. In this paper, we utilize the association rule mining to generate interesting rules among products that create meaningful product connections. Moreover, we present an evaluation indicator to measure potential value of each product taking into account the firm's existing internal product portfolios. From that, we can make product