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 $\mathrm{P}(A)$. 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 $\mathrm{P}(B|A)$. 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 opportunity suggestions for certain firms that have potential to be realized using their internal capability.

3. Product opportunity identification approach

Identifying potential products that can be elaborated from the existing products a firm already possesses must be a prerequisite for a firm's sustainable growth by seizing an opportunity to dominate the market in advance. This research is motivated in this regard. To explore how to identify product opportunities based on a certain firm's existing internal product portfolio in advance, we present a product opportunity identification model as a procedural framework which consists of 3 steps as shown in Fig. 1: 1) extracting product information from patent database using text mining technique, 2) generating product connection rules employing association rule mining, and 3) identifying product opportunities taking into account a firm's internal product portfolio by measuring potential value of product opportunities.

Our approach can be considered to be similar to TRIZ, known as a theory of inventive problem solving. TRIZ presents general solutions that can be utilized for solving abstracted problems so if we conceptualize practical problems, we can derive new inventions to solve the problems by searching and applying the known general solutions (Altshuller, 1984). Similarly, our approach aims to derive new product implications by constructing meaningful relationships between products and identifying potential products that can be realized from firms' internal capabilities. This conceptual similarity can help our approach to be flexible. For example, TRIZ evolution trends which reveal the patterns of evolution of technology systems show the current status of the systems and then figure out how the systems will evolve in near future (Park et al., 2013; Mann, 2002). Therefore, our approach can be expanded to discover the product evolution trends. If we investigate the trends, we will be able to find out more potential products by combining the firms' internal capabilities with the generic trends.

3.1. Extracting product information

Patents generally contain a written description of the invention of technological elements with illustration examples like the manner and process of utilizing them (Ahuja et al., 2005). In this respect, the patent analysis has been widely used to draw the technological implications and monitor the underlying industrial trends of relative technologies. The fact that patents describe product information that can be realized through the implementation of the invented technologies is worthy of notice since it leads to discover connections between products in a similar technological boundary. These product connections will be a basis for generating meaningful and interesting product relationships in the next step.

For that, a substantial product database has to be established. Product information means a unique name of product that is naturally represented by a set of keywords. These keywords are generally expressed as noun phrases so a product database can be built by extracting noun phrases from textual descriptions in patent documents. However, as the noun phrases do not always indicate the practical product information, it is required to remove noise information by eliminating abnormal noun keywords that cannot be matched to the product dictionary. Moreover, a product name can come in many different forms due to the word spacing and abbreviation. Therefore, it is also necessary to group similar noun keywords together to only obtain useful product information. According to this process, a practical product database (tod.kisti.re.kr) has already been built using U.S. patents by the Korea Institute of Science and Technology Information (KISTI) (Korea Institute of Science and Technology Information (KISTI), 2012) so this research simply employs the existing database. The product database initially contains 118,333 product names and, after the grouping, 57,505 names remain.

Even though the extraction process is based on matching with the product database, the extraction results will include meaningless keywords since there are substantial vague ones in the product database. To remedy this problem, relative studies using text mining techniques have insisted that keyword selection strategies such as using term frequency or document frequency must be incorporated into the keyword extraction process to cut out marginal keywords. These strategies facilitate determining the analysis target keyword sets minimizing human intervention. Therefore, we adopt a constraint strategy using term frequency based on minimum threshold value. Moreover, we only consider compound words not single words to exclude ambiguous product keywords, for example, not just “dispenser” but “drink dispenser” or “foam dispenser”. From this constraint, we obtained 49,884 product names that will be only used for identifying product opportunities in this research. This constraint, however, cannot guarantee to only include meaningful product keywords that are enough to describe the intrinsic feature of products. For example, a “solar cell” is a compound word but it is still ambiguous without the specific information about its raw materials. To overcome this problem, it is necessary to establish a product tree structure which explicitly illustrates upper and lower relationships between product keywords. The lower it goes down in the product tree, the more it clarifies the inherent characteristics of products. Therefore, if we use only the product keywords in the leaf nodes, we can largely eliminate the ambiguous products and consequently derive practical insights about new potential products from the angle of new product planning. However, the product database utilized in this research contains only the list of product keywords so we decide to proceed to identify potential product opportunities under the constraint of using only compound words.

3.2. Generating association rules

This research employs the association rule mining to figure out rules of product connections. To apply it, adjusting its basic principle to meet our research purpose is necessary. The association rule mining exploits

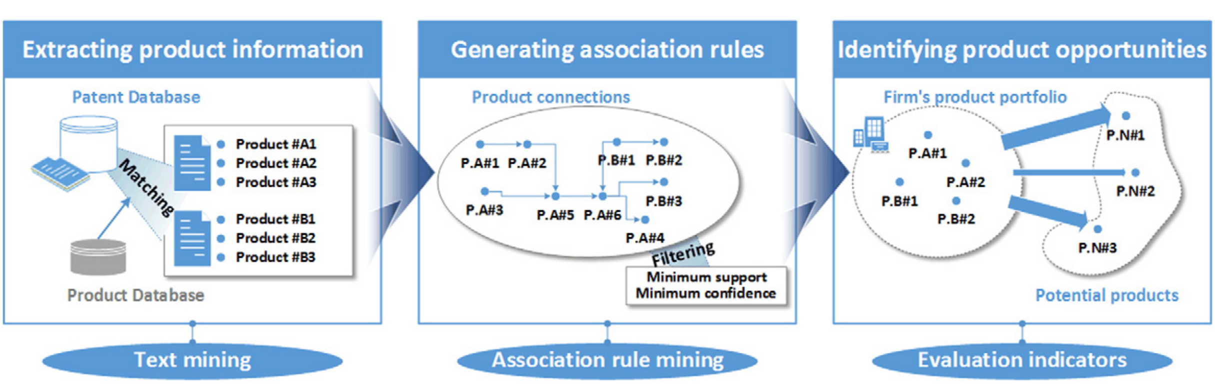

Fig. 1. Procedural framework for product opportunity identification approach.

the information about co-occurrences of items in transactions to generate their relationships. We consider that a patent document describes the invention of technological elements and products that can be developed using the invented technologies. Therefore, we can generate association rules among products using how many times they appear in a same patent document concurrently. From this perspective, a transaction and an item in the association rule mining can be interpreted to a patent document and a product, respectively.

An association rule of product connections consists of two sets of product items, conditional products as a source part of the rule and consequent products as a target part. This rule implies that there is a potential opportunity to proceed to the consequent products using similar technological elements that are employed in developing the conditional products. Thus, from this rule, a firm can acquire ideas about potential products that will be realized through the connections between other products. However, the potential products may not always be directly related to new product opportunities since the firm's internal capabilities of realizing the potential products are critical to uncover valuable opportunities. Moreover, the potential products should be new for firms. Therefore, this research does not aim to generate association rules for connecting products by simply employing the association rule mining, but try to identify new product opportunities by measuring potential value of product opportunities taking into account a firm's internal capabilities. How to measure the potential value is presented in the next step.

3.3. Identifying product opportunities

A procedure for identifying product opportunities consists of two sub-steps: 1) exploring the novelty of potential products uncovered by the association rules and 2) measuring the potential value of product opportunities. The association rules generally reveal a set of products by dividing conditional and consequent ones that have directed connections in a similar technological boundary. To be practical product opportunity, the reveled consequent products should be new to a target firm that is currently seeking ideas for new product development. However, product information in the firm's patent documents is also utilized when employing the association rule mining, so some items that are already in the firm's existing product portfolio can be regarded as potential products. In this case, those results must be meaningless to the target firm. Therefore, it should be examined whether the consequent products of the association rules are already in the firm's existing product portfolio. The second sub-step is to measure potential value of the consequent products that are considered potential after exploring the novelty. It is meaningless to suggest a huge list of potential products since firms generally have limited resources to attempt to realize all of them. The potential value of each product should be calculated to identify practical product opportunities.

To measure the potential value, we present an evaluation indicator reflecting three aspects: the confidence measure of the association rules, the importance of the conditional products in a target firm and the importance of the consequent products in all the other firms except the target firm. The first aspect implies the relative weight of product connections in the association rules. If the confidence measure is high enough, it will be straightforward to realize the consequent products from the conditional products. The second aspect means how much capability the certain firm possesses to deal with the conditional products that will facilitate the development of the consequent products. If the firm has relatively high capability, it can be thought that it has a strong basis for realizing the consequent products. The final aspect indicates how many firms are accustomed to dealing with the consequent products. If there are many firms, the products can be considered have a low novelty value although they are new to the target firm. To define a single evaluation indicator by simply synthesizing these three aspects, it should be assumed that their relative variability is similar. However, this assumption cannot be guaranteed so we first normalize these values before aggregating them together. Consequently, for a certain firm A, the potential value of a consequent product k can be calculated as:

$$\textit{potential value}_k = Nor\left(\sum_{i=1}^{N}\left(conf_{i,k} \times weight_i\right)\right) \times Nor(\alpha_k) \times \beta_k \quad (1)$$

$$weight_i = \frac{num(P_A(i))}{num(P_A(\cdot))} \quad (2)$$

$$\alpha_k = 1 - \frac{\sum_{J \neq A} num(P_J(k))}{\sum_{J \neq A} num(P_J(\cdot))} \quad (3)$$

$$\beta_k = \begin{cases} 1, & \textit{if } num(P_A(k)) < threshold_\beta \\ 0, & \textit{otherwise} \end{cases} \quad (4)$$

$$Nor(v) = \frac{v - \min(\cdot)}{\max(\cdot) - \min(\cdot)} \quad (5)$$

where $conf_{i,k}$ is the confidence value of the association rule that indicates a connection from the product i to the product k and N is the number of conditional products that have the product k as a consequent part in the connections. $weight_i$ is the importance of the product i as a conditional item for the product k in the firm A. It is calculated by the ratio of the number of patents that contain the information of the product i to the size of the firm A's patent portfolio as shown in Eq. (2) where $num(P_A(\cdot))$ means the number of patens the firm A holds and $num(P_A(i))$ means the number of patents that contain the product i within the firm A's patent portfolio. α_k is the extent of the usage of the product k in all the other firms except A. It is measured by the ratio of the number of patents that contain the information of the product k to the size of the patent portfolios of all the other firms except A as shown in Eq. (3) where $\sum_{J \neq A} num(P_J(\cdot))$ means the number of patents all the other firms except A hold and $\sum_{J \neq A} num(P_J(k))$ means the number of patents that contain the product k in the patent portfolios of all the other firms except A. When measured in this way, however, the higher value means the lower potential so α_k is defined as 1 minus this measured value since the value is between 0 and 1. The last parameter, β_k, means the novelty of the product k for the firm A. If the product is already included in the firm's product portfolio, its value is 0, otherwise 1. It is, however, inappropriate to say that the certain product is not fresh to the firm A when it has appeared only one time in the firm's patents. Therefore, we use a certain threshold value to determine the value of β_k. $Nor(v)$ indicates the result of the min-max normalization. Note that, the last parameter has the value of 0 or 1 so it needs not be normalized.

4. Illustration

4.1. Product information extraction

The existing product database developed by KISTI contains 57,505 product names. Among them, we only use the product names represented by compound words not single words since they can indicate specific products. From this constraint, 49,884 names remain and these refined products will be only considered for the product opportunity analysis. And then, we collect patents granted in the United States Patent and Trademark Office (USPTO) during 2009 and 2013 to analyze the extent of the usability of the products in various technological areas. The total number of the collected patents is more than 1,000,000. A patent includes a number of textual descriptions such as summary of the invention, description of the preferred embodiments, experiment examples and claims. Among them, the claims describe the essential features of the invention with additional details so they have been generally recognized as a core component of patents (Lee et al., 2013). This research, therefore, explores which products have melted into the claimed knowledge by extracting product names from these claim sections. A total 810,276 patents appear that they deal with the refined products. It is, however, inappropriate to believe that if a certain product appears more times in patent *A* than in patent *B*, the patent *A* is more likely to be related to that product. If an appearance frequency of the product exceeds a predetermined threshold value, then it should be thought that the patent sufficiently reflects the product in its claimed knowledge of the invention. The refined products appear 5.9 times on average per patent so we use 6 as the threshold value to clarify whether a patent deals with a certain product as a part of the essential features of its invention. From that, we acquire 395,980 patents that include one or more refined products. These refined patents and products will be incorporated into the product opportunity identification process addressed in subsequent steps. It indicates that we use only about 36% of all available patent data. Therefore, it is required to show that this research does not lose the generality in spite of using only a few patent data sets. Comparing the distribution of all available patents and that of the refined patents by International Patent Classification (IPC) can be a good way to show the generality. Fig. 2 depicts the comparison of the distributions by IPC section. They seem similar on the whole but the refined patents are mainly classified into the section H (electricity) while all available patents are largely classified into the section G (physics). It can be because the refined patents only include ones in which the information about the practical products is explicitly described. To compare the both distributions in depth, we measure the Pearson's correlation coefficient. One patent can be classified into multiple IPCs so if we calculate the total number of patents by IPC, a patent with multiple IPCs must be redundantly counted. It means that the sum of the total number of patents by IPC will exceed the number of all available patents. As shown in Table 1, the correlation coefficients are statistically significant at the 0.01 level. Note that, the number of the refined patents can never be larger than the number of all available patents in each IPC section and sub-class so if we measure the correlation coefficient, it will naturally show a positive linear relationship. Nonetheless, the fact that the correlation coefficients are over 0.9 indicates that there are extremely strong relationships between the refined patents and all available patents. Therefore, although we use only a few patent data sets, it can be thought that this research does not lose the generality since the refined patents sufficiently represent all available patents.

Table 1
Summary of descriptive statistics and correlation matrix.

	By IPC section (n = 8)				By IPC sub-class (n = 629)			
	Mean	S.D.	1	2	Mean	S.D.	1	2
1. All available patents	156,813	137,409	1		2290	8112	1	
2. Refined patents	56,265	53,292	0.923[a]	1	714	2496	0.966[a]	1

[a] Statistically significant at the 0.01 level.

4.2. Association rule generation

In this step, we first count co-occurrence frequency of products in a same patent and as a result of this we derive 128,596 co-occurrence cases of pairs of products. And then, we generate association rules using the co-occurrence information applying the association rule mining. To make the generated rules meaningful, we should examine the usefulness and the certainty of the rules (Kim et al., 2011). The usefulness can be explained by the support measure, the probability that a product appears in patents, so it is calculated by the ratio of the appearance frequency of the product to the total number of patents. The certainty can be described by the confidence measure, the conditional probability that a consequent product appears in patents given that a conditional product has already appeared in the same patents, so it is calculated by the ratio of the co-occurrence frequency of the two products to the appearance frequency of the conditional product. To generate association rules, minimum threshold value for the support and confidence measure should be predetermined. The lower minimum support value would result in a larger number of rules by using many more products including that their appearance frequencies are significantly low. Similarly, the lower minimum confidence value would result in generating many more rules including that their certainties are too low. To generate enough rules based on the analysis on diverse product connections, we set minimum support value and minimum confidence value as 0.1% and 0.5%, respectively. Applying the a priori algorithm, a typical algorithm for association rule mining, leads to generate 4239 association rules (Table 2).

4.3. Product opportunity identification

This step identifies product opportunities using the generated association rules. The rules present a number of derivable products from certain conditional products. Therefore, the consequent products in the rules with high confidence must be potential opportunities to a target firm that possesses enough capabilities in developing the conditional products although it must be examined whether the consequent ones are new to the target firm. We measure potential value of the consequent products using the confidence value of the rules, weight, α and β. All parameter values except the confidence are different on each firm since the potential value of a product must be different according to which firm tries to deal with the product. Therefore, we need to select one target firm to illustrate the applicability of our approach. Canon Inc. has granted the third most patents in the USPTO during 2009 and 2013, is a multinational corporation specialized in the imaging and optical products including digital cameras, camcorders, printers, and medical

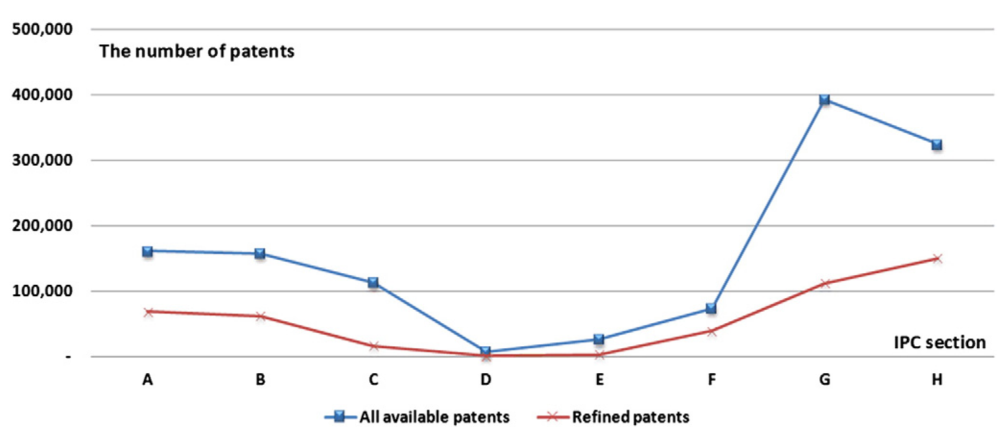

Fig. 2. Distributions of all available patents and the refined patents by IPC section.

Table 2
Part of generated association rules (top 20).

Conditional product	Frequency (support)	Consequent product	Frequency (support)	Confidence
RFID reader	408 (0.10%)	Radio frequency identification tag	709 (0.18%)	39.22%
Semiconductor film	832 (0.21%)	Insulation film	3091 (0.78%)	36.42%
Transmitting device	403 (0.10%)	Receiving device	679 (0.17%)	36.23%
Conductive film	957 (0.24%)	Insulation film	3091 (0.78%)	34.27%
Semiconductor memory	1979 (0.50%)	Memory cell	3016 (0.76%)	33.10%
LCD display panel	445 (0.11%)	LCD display	2907 (0.73%)	30.79%
Intake valve	432 (0.11%)	Exhaust valve	442 (0.11%)	29.40%
Silicon nitride	517 (0.13%)	Insulation film	3091 (0.78%)	29.01%
Exhaust valve	442 (0.11%)	Intake valve	432 (0.11%)	28.73%
Semiconductor package	491 (0.12%)	Semiconductor die	504 (0.13%)	24.64%
Semiconductor die	504 (0.13%)	Semiconductor package	491 (0.12%)	24.01%
Intake valve	432 (0.11%)	Combustion engine	2218 (0.56%)	23.15%
Objective lens	849 (0.21%)	Optical disk	946 (0.24%)	22.85%
Radio frequency identification tag	709 (0.18%)	RFID reader	408 (0.10%)	22.57%
Nonvolatile memory	1910 (0.48%)	Memory cell	3016 (0.76%)	21.94%
Flash memory device	398 (0.10%)	Memory cell	3016 (0.76%)	21.86%
Memory cell	3016 (0.76%)	Semiconductor memory	1979 (0.50%)	21.72%
Backlight unit	431 (0.11%)	LCD display	2907 (0.73%)	21.58%
Receiving device	679 (0.17%)	Transmitting device	403 (0.10%)	21.50%
Light guide plate	509 (0.13%)	LCD display	2907 (0.73%)	20.83%

equipments. We believe that Canon Inc. is suitable for the illustration of our case study due to its focused businesses.

First, the weight value means how familiar the target firm is with the conditional products that would be a source of technological knowledge to realize the consequent products. This value is calculated by the ratio of the number of patents that include the conditional product to the total number of patents the target firm holds as shown in Eq. (2). As a result, we obtain the top 20 conditional products for Canon Inc. (Table 3). The target firm possesses the technological capabilities including image and information processing, printing, recording, image detecting and sensing, light emitting device, and color filter to manufacture the imaging and optical products. Although a conditional product has high weight value, its potential value of product opportunity may not be high when the confidence of the rule from the conditional product to the consequent product is relatively low. The low confidence indicates that the strength of the connection between the two products is not much high so even if a firm is very familiar with the conditional product, it cannot facilitate the realization of the consequent product. Therefore, we apply the weight value conjointly with the confidence of the association rules. Moreover, they are measured for the conditional products, so it is necessary to total up the values for each consequent product since the consequent products are only regarded as the evaluation target of the product opportunity.

Table 3
Weight value of conditional products for Canon Inc. (top 20).

Conditional product	Frequency	Weight
Image processing apparatus	596	4.23%
Information processing apparatus	472	3.35%
Printing apparatus	300	2.13%
Imaging apparatus	247	1.75%
Communication apparatus	173	1.23%
Recording apparatus	134	0.95%
Image sensor	88	0.63%
Pickup device	75	0.53%
Light emitting device	73	0.52%
Communication device	41	0.29%
Transmitting unit	39	0.28%
Detecting device	38	0.27%
Sensing device	35	0.25%
Insulation film	31	0.22%
Management unit	30	0.21%
Output terminal	28	0.20%
Color filter	27	0.19%
Supply device	27	0.19%
Computer output device	24	0.17%
Piezoelectric element	23	0.16%

Second, the α value means how much the other firms are not concerned with the consequent products. The high α value indicates that the related products are not prevalent in the market so there can be market novelty. This value is calculated by 1 minus the ratio of the number of patents that include the consequent product to the total number of patents all the other firms hold as shown in Eq. (3). The total number of products which take up the consequent part of all the generated association rules is 1211, and they are regarded as the evaluation target of this α measure.

Third, the β value (0 or 1) means how much the consequent products are new to the target firm. If the value is 0, the degree of their novelty is relatively low. Therefore, only the products with $\beta = 1$ can be considered for identifying potential opportunities. The evaluation target of this β measure is same with the α measure. As shown in Eq. (4), the β parameter can only have the value of 0 or 1 depending on a threshold value ($threshold_\beta$). To determine an appropriate threshold value, we construct a scree plot of the number of products by appearance frequency of products (Fig. 3) and then conduct a scree test which is one of the most used graphical strategies to determine the proper number of components to retain (Cattell, 1966; OuYang and Weng, 2011). The number of products sharply decreases when the appearance frequency moves from 0 to 1. Therefore, this research uses the value 1 as the threshold value for the β measure.

Finally, the potential value of the consequent products is derived by aggregating the results of the above measures. To present a single evaluation indicator by simply multiplying them together, their relative variability should be similar. However, the coefficients of variation of first and second values are 3.824 and 0.001, respectively. It means that their relative variability is very high so the simple multiplication cannot be reasonable. Therefore, we first normalize these values and then aggregate them together. From that, we obtain the top 20 potential products for Canon Inc. (Table 4). Image processing program shows the most potential value of product opportunity. From the perspective of the target firm who possesses the capabilities of imaging and image processing, it will be a good opportunity to develop software systems that can control and analyze images in various research and industry fields including medicine, geography, and applied photonics. Contrast medium as a substance used to improve the contrast of structures within human body will also be a potential product since the target firm is good at dealing with radiation detecting and sensing devices. Similarly, the function of backlight unit is to emit light to the liquid crystal panel and the typical light source includes Cold Cathode Fluorescent Lamp

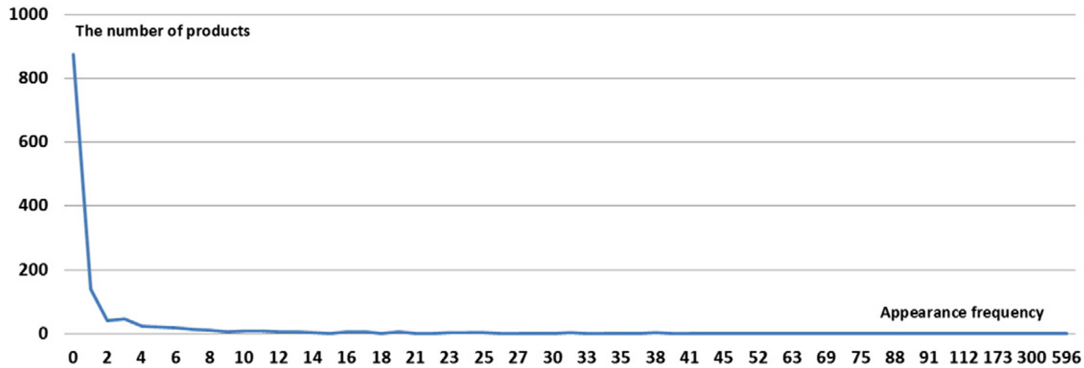

Fig. 3. Scree plot of the observed number of products.

(CCFL), Hot Cathode Fluorescent Lamp (HCFL), and Light Emitting Diode (LED). This product can be another opportunity to the target firm since it has technologies for light emitting device and color filter. Therefore, the identified product opportunities can become the potential application products for the target firm, Canon Inc., which can be developed by the similar technologies the firm had already possessed.

5. Discussions

5.1. Real value-based sub-indicator β

This paper dealt with how to identify new product opportunities by reflecting the existing internal capabilities of target firms. To do that, we first extracted product names from the massive patents using text mining, and then generated product connection rules represented as directed pairs of conditional and consequent products employing association rule mining. Finally, we defined a measurement indicator to evaluate potential value of consequent products from the perspective of a target firm so that it could identify new product opportunities. Product opportunities have different meanings according to who adopts these opportunities. Thus, we explored a target firm's internal capabilities based on its existing product portfolio and identified potential products that can be realized using the existing capabilities. If the identified products have already been widely covered by the other firms, the degree of their market novelty must be subject to decline so we refined the potential products in accordance with the novelty. Moreover, we figured out whether the refined products were new to the target firm since if the products were already in the existing product portfolio of the target firm, they should not be considered opportunities. We put these criteria together to form a measurement indicator to evaluate potential value of product opportunities.

For the last criteria, examining whether the potential products were new to the target firm, we choose binary indicator variable, β as shown in Eq. (4). The β value is 0 if and only if the potential product was already included in the target firm's product portfolio. We assumed that the product is in the product portfolio when the number of patents that include the product exceeds the threshold value. The result could vary depending on the threshold value. Therefore, we now try to use real value-based β rather than binary-based one. For a target firm *A*, the β_real_k can be calculated as:

$$\beta\_real_k = 1 - \frac{num(P_A(k))}{num(P_A(\cdot))} \tag{6}$$

where $num(P_A(\cdot))$ is the number of patens the firm *A* holds and $num(P_A(k))$ is the number of patents that include the product *k* within the firm *A*'s patent portfolio. Fig. 4 shows a product opportunity portfolio map for Canon Inc. The horizontal axis means the β_real and the vertical axis means the potential value of product opportunities excluding the original binary-based β.

Many additional products have appeared in the portfolio map. These ones had not been included in the potential product list shown in Table 4 since their original βs were all 0. The vertical axis shows which products can be realized using the internal capabilities of the target firm regardless of whether they are new to the firm or not. The high value products including information processing apparatus, imaging device, printing apparatus, image processing apparatus, and color filter seem to be already within the scope of the target firm's business since their weight values are very high (Table 3). From this result, we can discuss the appropriacy of our approach twofold. First, using the potential value excluding β could suggest valuable products that the target firm had already been concerned with. That means our approach before applying the criteria of whether the suggested products are new to the target firm showed appropriate results by reflecting the target firm's concerns and interests well. Second, the products with high potential value in Fig. 4 were regarded as already being implemented by the target firm since most of them were listed in Table 3 that represented the products that have extremely high weight value for the target firm. The weight value means how familiar the target firm is to the relevant products so it can be regarded that the firm already has a lot of experience of dealing with these products. Therefore, if we use the binary value-based β, we can eliminate the meaningless opportunities that cannot be novel to the target firm. It indicates that our approach using the binary value-based β could show the properly screened opportunities that are new to the target firm. It does not, however, mean that the portfolio map analysis is useless. This map can allow the target firm to recognize different

Table 4
Potential value of product opportunities for Canon Inc. (top 20).

Potential product	$\sum(conf \times weight)$	α	β	Potential value
Image processing program	0.292136	0.991878	1	0.289764
Contrast medium	0.072842	0.987344	1	0.071920
Semiconductor memory	0.116840	0.525123	1	0.061355
Storage part	0.058461	0.976577	1	0.057092
Communication circuit	0.058384	0.963355	1	0.056244
Camera module	0.055705	0.925009	1	0.051527
Transport belt	0.048140	0.986966	1	0.047512
Identity card	0.045470	0.984511	1	0.044765
Material container	0.042519	0.993578	1	0.042246
Wireless communication module	0.040307	0.982244	1	0.039591
Magnetic gradient coil	0.034521	0.989233	1	0.034149
Halftone screen	0.032093	0.993389	1	0.031881
Backlight unit	0.028249	0.922743	1	0.026067
Communications device	0.034840	0.730072	1	0.025436
Print substrate	0.024070	0.996978	1	0.023997
Printing module	0.024070	0.996600	1	0.023988
Transceiver module	0.024484	0.974122	1	0.023850
Filter substrate	0.020270	0.983000	1	0.019925
Light guide plate	0.022254	0.891764	1	0.019846
Separation filter	0.019726	0.996600	1	0.019659

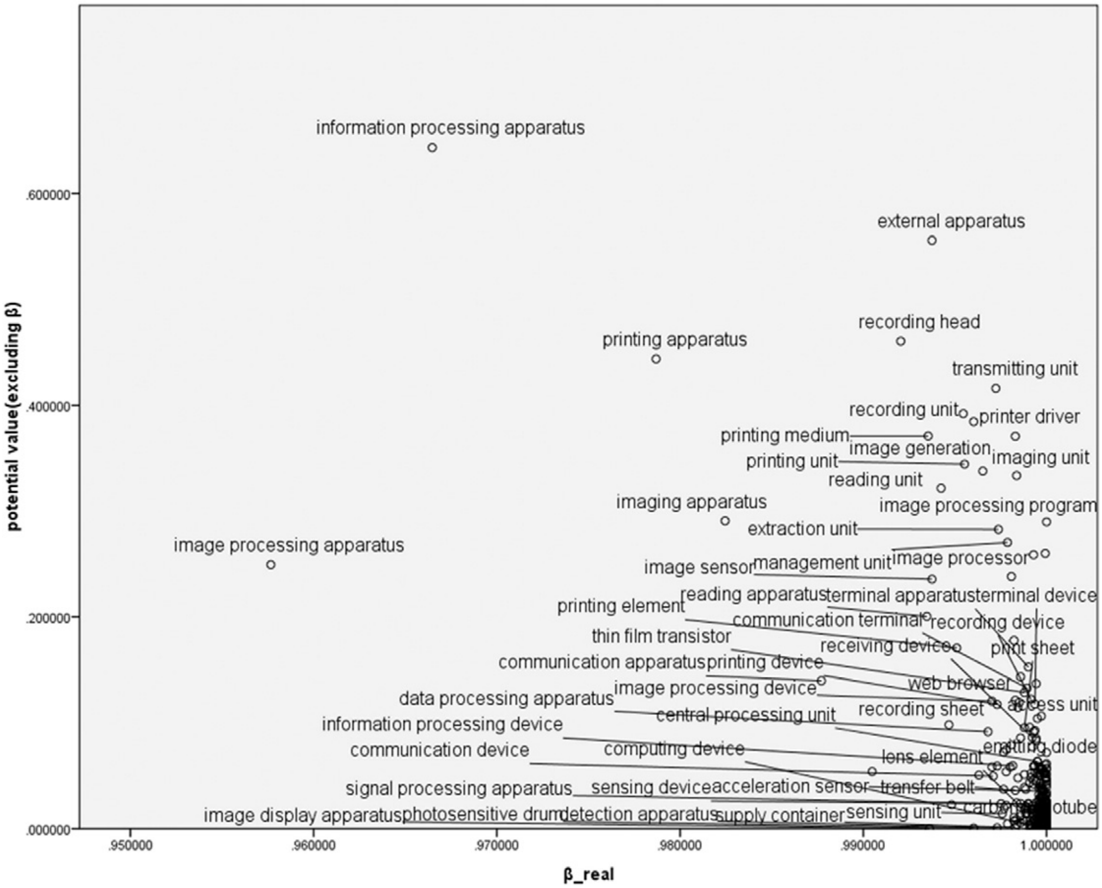

Fig. 4. Product opportunity portfolio map for Canon Inc.

opportunities from those found by our approach by incorporating its own judgment about the novelty of potential products.

5.2. Evaluation using independent sub-indicators

This approach measures the potential value of product opportunities with a single evaluation indicator that combines several independent sub-indicators together. Although multiple indicators are actually applied in the measurement process, a single rank based on a comprehensive result should be provided to facilitate firms to discern appropriate opportunities that are feasible and compatible with their situations. To synthesize the sub-indicators, it is mandatory to understand the theoretical relationship between them but it is impossible at this time, so this approach normalized the values of sub-indicators and then simply multiplies them together. In this way, we cannot preserve the individual meaning of each sub-indicator. Therefore, it must be useful to derive meaningful insights for the product R&D planning by examining the values of independent sub-indicators individually. This section explores the meanings of sub-indicators and then discusses about the insights with a case of a target firm, Canon Inc.

Potential products based on the raw data of the sub-indicators for the target firm are summarized in Table 5. The confidence value means the strength of the relationships between products and the weight value indicates the degree of the firm's internal capabilities for realizing the potential products. These values were already presented in Tables 2 and 3 so this section only deals with the result of the aggregation of these values that can reveal the candidates of potential products practicable from the internal capabilities. The product opportunities based on the internal capabilities can be candidates for not only the internal R&D but also the external R&D since the opportunities can be strategically utilized in transferring the technological knowledge the target firm possesses to other firms to aid their R&D activities. Several products related to the imaging technologies are presented as top candidates including imaging device, image sensor, and imaging unit. Canon Inc. is specialized in the imaging and optical products so this result can be seen as reasonable. It reveals that Canon Inc. has a chance to contribute to the market by transferring its own technological knowledge about the imaging products to other firms who want to penetrate the relevant market. This chance can offer additional profit to Canon Inc. It is, however, not examined whether they are red ocean products that are already prevalent in other firms therefore they may not be included in the final list of potential products. All of the top 20 potential products do not appear in the result of Table 4. It means that they are eliminated by the restrictions of α and β. Second, the α value explains the market novelty of potential products by examining whether they are prevalent in the market or not regardless of the internal capabilities. It can reveal the ideas of blue ocean products without considering the possibility of realization and the needs of the market. Therefore, it is significant that the value can provide suggestions about the product R&D planning in a long-term perspective. This research has mainly focused on the short-term perspective in that it seeks potential products that can be realized right away based on the internal capabilities. Most of products with high α values seem not to be directly related to the target firm's R&D areas except photovoltaic inverter which is one of the core parts in a solar photovoltaic power generation system. This product can also be a potential opportunity in that the target firm has technological capabilities to deal with transmitting unit, piezoelectric element, and detecting and sensing devices as shown in Table 3. Finally, the β value

Table 5
Potential products using sub-indicators for Canon Inc. (top 20).

Potential product	$\sum(conf \times weight)$	Potential product	α	Potential product	β
Information processing apparatus	0.004928	Fireproof wall	0.999995	Image processing program	1
External apparatus	0.002862	Held power tool	0.999995	Semiconductor memory	1
Imaging device	0.002654	Large-sized battery pack	0.999992	Contrast medium	1
Printing apparatus	0.002598	Semiconductor package module	0.999990	Storage part	1
Recording head	0.002503	Smoothing reactor	0.999990	Communication circuit	1
Transmitting unit	0.002270	Disposable tip	0.999990	Camera module	1
Image processing apparatus	0.002149	Controlled rectifier	0.999987	Computing device	1
Recording unit	0.002087	Exhaust silencer	0.999987	Transport belt	1
Printer driver	0.001953	Electronic video display	0.999987	Identity card	1
Printing medium	0.001944	Passive electronic Component	0.999987	Material container	1
Image sensor	0.001933	Network file server	0.999985	Wireless communication module	1
Image generation	0.001928	Battery pack system	0.999985	Communications device	1
Imaging apparatus	0.001925	Electrical accessory	0.999982	Magnetic gradient coil	1
Printing unit	0.001793	Magnetic tape apparatus	0.999982	Halftone screen	1
Imaging unit	0.001783	Power train control module	0.999982	Backlight unit	1
Reading unit	0.001713	Pneumatic gun	0.999982	Combustion engine	1
Management unit	0.001527	Machine device	0.999982	Transceiver module	1
Color filter	0.001519	Interior mirror	0.999980	Print substrate	1
Communication device	0.001484	Photovoltaic inverter	0.999980	Printing module	1
Insulation film	0.001482	Metal diaphragm	0.999980	Light guide plate	1

describes the internal novelty of potential products by exploring whether they are already addressed in the target firm or not. The value is between 0 and 1 and only if the value is exactly 1, the total potential value in Table 4 can be non-zero. Therefore, in Table 5, we display the same result with that in Table 4. This value enables to ascertain what the current status of the target firm's product portfolio is.

Examining the raw data result of each sub-indicator can also facilitate exploring the potential products. For example, if we obtain the candidates of potential products practicable from the internal capabilities by using the result of the aggregation of the confidence and weight values and then qualitatively evaluate the degree of their market and internal novelty, we can make a thorough investigation into the potential product opportunities in a broader scope.

5.3. Feasibility and usability of this research

Confirming that the revealed potential product opportunities from our approach are feasible and useful to the target firm can be suitable to verify the contribution of this research. In the case study, Canon Inc. was assumed as the target firm since we believed that it is suitable for the case study due to its narrow business scope specialized in the imaging and optical products. We have tried to acquire feedback of the result of the case study from the relevant experts but failed to do that. As an alternative, we have decided to conduct another case study using Samsung Electronics Co., Ltd. and then to interview a few experts who are related to the R&D activities to confirm whether it is feasible or not since it might be relatively easy to contact them and get some feedback. In fact, Samsung Electronics Co., Ltd. may not be appropriate to apply our approach because it deals with a variety of products including mobile phones, household appliances, display panels, and semiconductor chips. Despite of this drawback, we proceed to new case study in that it can be possible to get proper feedback. The result of the new case study is summarized in Table 6. We have interviewed several experts about the feasibility of the result. They commonly agreed with the necessity of the 1st and 3rd potential product items. Organic resin (1st one) is one of the most used transesterification catalysts in the biodiesel manufacturing process. They insisted that the firm is currently emphasizing the importance of bio technology for a sustainable growth and therefore there is a large amount of potential internal and external demand for the catalytic products that can be a basis for developing future biofuel sources. Retardation plate (3rd one) is a key part of Organic Light Emitting Diode (OLED) display products. One problem the OLED display has is that the expression of a black color and contrast can easily be deteriorated because of the reflection of external light. The experts argued that the retardation plate can be a candidate solution for the problem in that it suppresses the reflection. OLED display is also one of the anchor products of the firm so the identified opportunity can become the potential application product. However, the experts did not accept the other items because they are quite vague and non-specific. It indicates that we have to elaborate our approach by establishing a product tree structure that can help to describe the unique nature of products clearly as we already mentioned above.

Table 6
Potential value of product opportunities for Samsung Electronics Co., Ltd. (top 5).

Potential product	$\sum(conf \times weight)$	α	β	Potential value
Organic resin	0.019776	0.981111	1	0.019402
Optical head	0.014200	0.963166	1	0.013677
Retardation plate	0.013083	0.990933	1	0.012964
Mirror assembly	0.012685	0.955232	1	0.012118
Steering element	0.010626	0.987533	1	0.010494

6. Conclusions

This paper proposed a systematic approach to identify potential product opportunities. To explore how to identify product opportunities based on a firm's internal capabilities, we first extracted product information from the massive patent sets using text mining technique, and then generated association rules between products employing association rule mining. Finally, we measured potential value of product opportunities with a measurement indicator that enables to identify potential products that can be realized using a target firm's capabilities.

Preoccupying new opportunities must be a prerequisite for a firm's sustainable growth by dominating the market in advance. In this regard, a number of studies have attempted to discover technology opportunities. They have commonly focused on the suggestion of conceptual implications of technologies. Through this approach, it is hard to figure out which products can come into the market as a result of the technological implications and subsequently it tends to reduce reliability of the technology opportunities. Moreover, they have tried to measure generic potential values without considering a specific target firm that is willing to conduct R&D. The potential value of a product must vary according to who adopts the product opportunity. Therefore, it should be judged whether the opportunity is feasible from the perspective of a firm's internal capabilities. The only potential products associated with the internal capabilities can be thought as the real opportunities for individual firms. 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.

Despite the contribution, further research issues still remain. First, we have only concentrated on the technical perspective for product opportunities. However, there must be other requirements for successful new product development including economic valuation and evaluation of the degree of internal capabilities for realizing the potential products. Therefore, a topic to develop a comprehensive method for identifying product opportunities by aggregating these points should be conducted in the further research. Second, we have described the firm's internal capabilities using only the products that the firm is concerned with. The internal capabilities cannot be fully represented by the existing products. Therefore, further research should be directed to determine the features of technological capabilities and make connections to the products by employing the existing studies about the technology DNA. Third, it is necessary to build a support system to increase the industrial usability of our approach. The approach contributes to the quantitative analysis of the product opportunity identification so it will be straightforward to implement the support system. Fourth, this research has focused on the identification of potential product opportunities by reflecting firms' internal capabilities which had already been accumulated from the past. It means that our approach is using the past data to foresee the future opportunities. However, for the foresight of the future opportunities, the future information should be incorporated. Therefore, we can concretize the presented approach by investigating the product evolution trends in accordance with the TRIZ technology evolution trends based on the conceptual similarity between ours and TRIZ. Finally, we only considered compound words not single words to exclude ambiguous product keywords. This constraint, however, is not enough to extract purely meaningful product keywords that facilitate product planning activities by describing the intrinsic feature of products. To remedy this problem, it is necessary to establish a product tree structure, not a simple product list. The product tree explicitly illustrates upper and lower relationships between product keywords. The lower it goes down in the product tree, the more it clarifies the inherent characteristics of products. Therefore, using only the product keywords in the leaf nodes can help to describe the unique nature of products clearly. This approach will foster our research in that it can derive practical product opportunities from the angle of new product planning.

Acknowledgments

This work was supported by the Pukyong National University Research Fund in 2013 (CD-2013-1342). This work was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2014R1A1A1005317). This work was partially supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2015R1A1A1A05027889) and performed as a subproject of KISTI's project Establishment of Information Analysis Intelligence Systems funded by the Ministry of Science, ICT and Future Planning.

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. The ACM SIGMOD International Conference on Management of Data(ACM SIGMOD '93), Washington DC, USA.

Ahn, K.-I., 2012. Effective product assignment based on association rule mining in retail. Expert Syst. Appl. 39 (16), 12551–12556.

Ahuja, G., Coff, R.W., Lee, P.M., 2005. Managerial foresight and attempted rent appropriation: insider trading on knowledge of imminent breakthroughs. Strateg. Manag. J. 26 (9), 791–808.

Altshuller, G.S., 1984. Creativity as an Exact Science: The Theory of the Solution of Inventive Problems. Gordon & Breach Science Publishing, New York.

Cattell, R.B., 1966. The Scree test for the number of factors. Multivar. Behav. Res. 1 (2), 245–276.

Chandy, R.K., Tellis, G.J., 1998. Organizing for radical product innovation: the overlooked role of willingness to cannibalize. J. Mark. Res. 35 (4), 474–487.

Chen, Y.-S., Chang, K.-C., 2012. Using the entropy-based patent measure to explore the influences of related and unrelated technological diversification upon technological competences and firm performance. Scientometrics 90 (3), 825–841.

Chen, Y.-S., Chen, C.-C., 2009. A model of factors moderating the relationship between new product development and company performance. Int. J. Soc. Behav. Personal. 37 (8), 1043–1049.

Choi, S., Kim, H., Yoon, J., Kim, K., Lee, J.Y., 2013. An sao-based text-mining approach for technology roadmapping using patent information. R&D Manag. 43 (1), 52–74.

Choi, S., Park, H., Kang, D., Lee, J.Y., Kim, K., 2012. An sao-based text mining approach to building a technology tree for technology planning. Expert Syst. Appl. 39 (13), 11443–11455.

Cooper, R.G., 2001. Winning at New Products: Accelerating the Process from Idea to Launch. 3 ed. Perseus, Cambridge, USA.

Curran, C.-S., Leker, J., 2011. Patent indicators for monitoring convergence—examples from nff and ict. Technol. Forecast. Soc. Chang. 78 (2), 256–273.

Kim, C., Kim, H., Han, S.H., Kim, C., Kim, M.K., Park, S.H., 2009. Developing a technology roadmap for construction R&D through interdisciplinary research efforts. Autom. Constr. 18 (3), 330–337.

Kim, C., Lee, H., Seol, H., Lee, C., 2011. Identifying core technologies based on technological cross-impacts: an association rule mining (arm) and analytic network process (anp) approach. Expert Syst. Appl. 38 (10), 12559–12564.

Ko, N., Yoon, J., Seo, W., 2014a. Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. Expert Syst. Appl. 41 (4), 1955–1963.

Ko, S.-S., Ko, N., Kim, D., Park, H., Yoon, J., 2014b. Analyzing technology impact networks for R&D planning using patents: combined application of network approaches. Scientometrics 101 (1), 1–20.

Koen, P., Ajamian, G., Burkart, R., Clamen, A., Davidson, J., D'Amore, R., Elkins, C., Herald, K., Incorvia, M., Johnson, A., Karol, R., Seibert, R., Slavejkov, A., Wagner, K., 2001. Providing clarity and a common language to the fuzzy front end. Res. Technol. Manag. 44 (2), 46–55.

Korea Institute of Science and Technology Information (KISTI), 2012T. Design and Construction of a Knowledgebase for Technology Opportunity Discovery. KISTI, Daejun, Korea.

Kowang, T.O., Rasli, A., Long, C.S., 2014. A triangulated approach to the assessment of important constructs in new product development process. Life Sci. J. 11 (6), 330–338.

Kuo, R.J., Chao, C.M., Chiu, Y.T., 2011. Application of particle swarm optimization to association rule mining. Appl. Soft Comput. 11 (1), 326–336.

Lai, K.-K., Wu, S.-J., 2005. Using the patent co-citation approach to establish a new patent classification system. Inf. Process. Manag. 41 (2), 313–330.

Lee, C., Song, B., Park, Y., 2012. Design of convergent product concepts based on functionality: an association rule mining and decision tree approach. Expert Syst. Appl. 39 (10), 9534–9542.

Lee, C., Song, B., Park, Y., 2013. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. Tech. Anal. Strat. Manag. 25 (1), 23–38.

Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29 (6), 481–497.

Mann, D., 2002. Hands on Systematic Innovation. Creax Press, Leper, Belgium.

Narin, F., 1994. Patent bibliometrics. Scientometrics 30 (1), 147–155.

Oliveira, M.G., Rozenfeld, H., 2010. Integrating technology roadmapping and portfolio management at the front-end of new product development. Technol. Forecast. Soc. Chang. 77 (8), 1339–1354.

OuYang, K., Weng, C.S., 2011. A new comprehensive patent analysis approach for new product design in mechanical engineering. Technol. Forecast. Soc. Chang. 78 (7), 1183–1199.

Park, H., Ree, J.J., Kim, K., 2013. Identification of promising patents for technology transfers using triz evolution trends. Expert Syst. Appl. 40 (2), 736–743.

Rubera, G., Kirca, A.H., 2012. Firm innovativeness and its performance outcomes: a meta-analytic review and theoretical integration. J. Mark. 76 (3), 130–147.

Shih, M.-J., Liu, D.-R., Hsu, M.-L., 2010. Discovering competitive intelligence by mining changes in patent trends. Expert Syst. Appl. 37 (4), 2882–2890.

Song, M.J., Lee, J.-G., Park, J.-M., Lee, S., 2012. Triggering navigators for innovative system design: the case of lab-on-a-chip technology. Expert Syst. Appl. 39 (16), 12451–12459.

Song, Q., Shepperd, M., Cartwright, M., Mair, C., 2006. Software defect association mining and defect correction effort prediction. IEEE Trans. Softw. Eng. 32 (2), 69–82.

Stuart, T.E., Podolny, J.M., 1996. Local search and the evolution of technological capabilities. Strateg. Manag. J. 17 (1), 21–38.

Trajtenberg, M., Henderson, R., Jaffe, A., 1997. University versus corporate patents: a window on the basicness of invention. Econ. Innov. New Technol. 5 (1), 19–50.

Tseng, Y.-H., Lin, C.-J., Lin, Y.-I., 2007. Text mining techniques for patent analysis. Inf. Process. Manag. 43 (5), 1216–1247.

Wolff, J.A., Pett, T.L., 2006. Small-firm performance: modeling the role of product and process improvements. J. Small Bus. Manag. 44 (2), 268–284.

Yoon, B., Park, Y., 2005. A systematic approach for identifying technology opportunities: keyword-based morphology analysis. Technol. Forecast. Soc. Chang. 72 (2), 145–160.

Yoon, B., Park, I., Coh, B.-Y., 2014. Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. Technol. Forecast. Soc. Chang. 86, 287–303.
Yoon, J., Kim, K., 2011. Identifying rapidly evolving technological trends for R&D planning using sao-based semantic patent networks. Scientometrics 88 (1), 213–228.
Yoon, J., Kim, K., 2012. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. Scientometrics 90 (2), 445–461.

Wonchul Seo is an assistant professor in Division of Systems Management and Engineering, Pukyong National University, Republic of Korea. He earned his PhD degree from Pohang University of Science and Technology. He has industrial experiences in companies such as Samsung Electronics Co. His research interest lies in technology intelligence, product opportunity identification, and service management.

Janghyeok Yoon is an assistant professor in Department of Industrial Engineering of Konkuk University, Republic of Korea. He earned his PhD degree from Pohang University of Science and Technology. He has industrial experiences in companies and research institutes such as LG and Korea Institute of Intellectual Property. His research interest is in technology intelligence-related topics, including technology forecasting, technology roadmapping, and technology convergence.

Hyunseok Park is a postdoctoral associate in Massachusetts Institute of Technology, USA. He earned his PhD degree from Pohang University of Science and Technology. His research topics include patent intelligence, patent information abstraction, and technology planning.

Byoung-youl Coh is a researcher of the Korea Institute of Science and Technology Information. He earned his PhD degree from Seoul National University. His theme of study has involved scientometrics, technology opportunity discovery, and R&D selection.

Jae-Min Lee is a researcher of the Korea Institute of Science and Technology Information. He earned his PhD degree from Seoul National University. He has industrial experiences in Samsung, and his main research theme includes scientometrics, technology opportunity discovery, and knowledge science.

Oh-Jin Kwon is a researcher of the Korea Institute of Science and Technology Information. He earned his PhD degree from University of Seoul. His main research theme includes scientometrics, data mining, data science.