Computers & Industrial Engineering 107 (2017) 376–387

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Identifying product opportunities using collaborative filtering-based patent analysis

Janghyeok Yoon [a,*], Wonchul Seo [b], Byoung-Youl Coh [c], Inseok Song [c], Jae-Min Lee [c]

[a] Department of Industrial Engineering, Konkuk University, Republic of Korea
[b] Division of Systems Management and Engineering, Pukyong National University, Republic of Korea
[c] Korea Institute of Science and Technology Information, Republic of Korea

ARTICLE INFO

Article history:
Available online 19 April 2016

Keywords:
Product opportunity
Product portfolio
Collaborative filtering
Latent Dirichlet allocation
Text mining
Patent analysis

ABSTRACT

One practical and low-risk approach to product planning for technology-based firms is to identify application products based on their existing product portfolios. Previous studies, however, have tended to neglect the current product development capabilities of target firms and to apply the technical data of specific fields to their methods, thereby failing to quantify a way of identifying various product opportunities. As a remedy, this paper proposes a new multi-step approach to product recommendation. The steps include (1) generating assignee–product portfolio vectors using text mining on a large-scale sample of patents, (2) recommending untapped products for a target firm by using latent Dirichlet allocation and collaborative filtering, (3) producing a visual map based on the promise and domain heterogeneity of the recommended products. To validate the practicability, we applied our approach to a Korean high-tech manufacturer by using all of the patents registered in the United States Patent and Trademark Office database during the period of time from 2009 to 2013. This study contributes to the systematic discovery of new product opportunities across various domains using the existing product portfolios of firms, and could become the basis for a future product opportunity analysis system.

1. Introduction

Faced with a business environment in which an increasing number of competing products rapidly appear and product life cycles shorten, many firms are making efforts to bring new products to market through product planning in order to secure competitiveness and develop sustainably. As a strategy for profit generation from new products and new markets, product diversification involves expansion into new segments of an industry where a firm already is and/or investment in a promising business outside of the scope of the existing business (Ansoff, 1957). One practical and low-risk approach to product planning for the firms that hope to bring new products to market is to expand the building of new product items on the foundation of the firms' existing product portfolios (Yoon, Park, & Coh, 2013; Yoon et al., 2015). Thus, this strategy utilizes existing product development capabilities to help a firm improve the practicality and performance of developing and launching new products, as well as to reduce the level of development risk (Qian, 2002).

Patents have long been considered an up-to-date and valid technical source that reflects current technological advancements as well as contains inventive knowledge with economic value (Yoon, Park, & Kim, 2013). Accordingly, many studies use patent information as raw data for analysis. Regarding product or technology opportunity analysis, prior studies use patents to fit growth curves to forecast the timing of emerging technologies and to estimate the evolutionary stage of new technologies, thereby suggesting strategic directions for technology development (Daim, Rueda, Martin, & Gerdsri, 2006; Liu et al., 2011; Yoon, Park, Kim, Lee, & Lee, 2014). Utilizing patent data within a given technology, some studies develop patent maps and networks to analyze the characteristics of specific patents and patent groups (Yoon & Kim, 2011), to detect the novelty of patents (Geum, Jeon, & Seol, 2013) and to evaluate patent vacuum areas (Lee, Yoon, & Park, 2009; Yoon, Park, & Kim, 2013). Other studies develop methods for new product or technology opportunities by applying idea generation methods, such as morphology analysis (Yoon & Park, 2007), disruptive innovation (Yoon & Kim, 2012) and system evolution patterns (Park, Kim, Choi, & Yoon, 2013; Park, Ree, & Kim, 2013).

Despite the contributions of previous studies, they have some common limitations. First, most studies do not consider the existing product portfolio of a target firm in need of new product opportunities. Rather, the studies identify generic opportunities

* Corresponding author.
E-mail addresses: janghyoon@konkuk.ac.kr (J. Yoon), wcseo@pknu.ac.kr (W. Seo), cohby@kisti.re.kr (B.-Y. Coh), sis@kisti.re.kr (I. Song), jmlee@kisti.re.kr (J.-M. Lee).

http://dx.doi.org/10.1016/j.cie.2016.04.009
0360-8352/

in the context of product domains, so these identified opportunities may not be practically applicable to the target firm. Second, from a methodological perspective, the prior studies are unable to deal with potential product opportunities that lie beyond a given product domain, because they do not incorporate pertinent technical information in various technical fields. In the same vein, further research needs to support product opportunity identification beyond expert knowledge and product domains; this can be realized by handling a large-scale technical data set embracing a wide range of technology areas.

To remedy these limitations, this study proposes a multi-step approach to recommending new application products, which starts from a target firm's existing product portfolio and which utilizes patent-based text mining and collaborative filtering techniques. The proposed approach has four steps: (1) the construction of assignee–product portfolio vectors using text mining of large-scale sample patents, (2) the use of latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003) to identify semantic similarities among the product portfolios of patent assignees, (3) the application of collaborative filtering to identify favorable product items untapped by a target firm and (4) the suggestion of product recommendation portfolio maps for new product opportunities in consideration of the product promise and product heterogeneity. For this study, we construct a database of 1,114,129 patents granted by the United States Patent and Trademark Office (USPTO) between 2009 and 2013. To show how the proposed approach works, we apply the approach to a technology-based firm and recommend practical application product items to the firm.

The contributions of this study are three-fold. First, the proposed approach identifies potential product items from a firm's existing product portfolio. Accordingly, such items could serve as practical inputs in the firm's product planning process. Second, the identified product items are identifiable without restriction of domains, because the proposed approach analyzes large-scale patent data collected from all technology domains. Therefore, the approach could facilitate the creative process of generating new application products and new markets, which product planners do not tend to intuitively consider. Third, the approach contributes to systemizing product opportunity analysis processes and therefore will become a basis for the development of a technology intelligence system that interacts with product experts to identify new application product opportunities.

2. Theoretical background

This study suggests a product opportunity identification approach based on vector space model-based patent text mining, LDA and collaborative filtering, so this section briefly overviews each of these theoretical backgrounds.

2.1. Patent text mining based on the vector space model

A vector space model, or term vector model, is an algebraic model for representing text documents as vectors of identifiers (Salton, Wong, & Yang, 1975). This type of model has been widely used in information retrieval (Turney & Pantel, 2010). In this model, documents are represented as vectors, which are organized by the frequency and weighting of terms. For example, each dimension for a vector corresponds to a separate term, and it is filled with a term weighting, such as term frequency (tf) and term frequency–inverse document frequency (tf–idf). Although the vector space model was initially proposed for information retrieval, its applicability to patent analysis has been proven by incorporating the model into various analytical techniques.

Every patent, whether it is granted or not and whether it has commercial value or not, is a result of research and development (R&D) activity. Therefore, careful analysis of patents can provide information not only about technological competitiveness and the strategic R&D directions of corporations, but also about overall technology trends and opportunities (Park & Yoon, 2014). Use of the vector space model in patent analysis has been tried by various studies that structure technical information to propose analytical approaches. Studies based on the vector space model have attempted to develop patent maps using self-organizing feature maps (Jun, Park, & Jang, 2012; Yoon, Yoon, & Park, 2002), formal concept analysis approach (Lee, Jeon, & Park, 2011) and patent similarities (Chang, Wu, & Leu, 2010; Yoon & Park, 2004), as well as to summarize and classify patent documents (Trappey, Trappey, & Wu, 2009), to identify and examine the possibility of vacuumed technologies using information visualization (Lee, Yoon, & Park, 2009; Son, Suh, Jeon, & Park, 2012) and to analyze patent infringement (Lee, Song, & Park, 2013; Shin & Park, 2005). Other studies have proposed approaches to identify new technology opportunities by incorporating morphology analysis (Yoon, 2008; Yoon et al., 2013; Yoon & Park, 2005, 2007) and developing technology roadmaps for business planning and product designs (Lee, Lee, Seol, & Park, 2008; Lee, Yoon, Lee, & Park, 2009; Yoon, Phaal, & Probert, 2008).

Although patent text mining approaches using vector space models are controversial in terms of defining keywords and representing technological knowledge (Park, Yoon, & Kim, 2012; Yoon, Choi, & Kim, 2011; Yoon & Kim, 2011), the simple structure and ease of use of the method have enabled various applications for patent text analysis. This study adopts a vector space model to construct assignee–product vectors, which act as inputs for collaborative filtering-based product opportunity recommendation in our proposed approach.

2.2. Latent Dirichlet allocation (LDA)

LDA, which was first presented as a graphical model for topic discovery, is a generative model that allows sets of observations to be explained by unobserved groups (Blei et al., 2003). LDA assumes that documents are made up of words and that the ordering of the words within a document is not important (Misra, Cappé, & François, 2008). In addition, it assumes that every document is represented by a topic distribution and that each topic defines an underlying distribution of words.

The concept behind LDA asserts that documents are represented as random mixtures over latent topics, wherein each topic is characterized by a distribution over words (Wang & Blei, 2011). LDA assumes the following generative process for a corpus D consisting of K topics and M documents each of length N_i:

1. Choose $\theta_i \sim Dir(\alpha)$, where $i \in \{1, \ldots, M\}$
2. Choose $\varphi_k \sim Dir(\beta)$, where $k \in \{1, \ldots, K\}$
3. For each of the word positions i, j, where $j \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, M\}$
   - Choose a topic $z_{ij} \sim Multinomial(\theta_i)$
   - Choose a word $w_{ij} \sim Multinomial(\varphi_{z_{ij}})$

α is the parameter of the Dirichlet prior on the per-document topic distributions, β is the parameter of the Dirichlet prior on the per-topic word distribution, θ_i is the topic distribution for document i (sum of θ_i is 1.0), φ_k is the word distribution for topic k, z_{ij} is the topic for the jth word in document i, and w_{ij} is the specific word.

LDA is known to outperform other dimension reduction techniques when dealing with a large corpus (Blei et al., 2003).

Therefore, many studies have used LDA for the applications of web spam filtering (Bíró, Szabó, & Benczúr, 2008), fraud detection (Xing & Girolami, 2007), human action recognition (Wang, Sabzmeydani, & Mori, 2007), and scientific article and web site recommendation (Das, Datar, Garg, & Rajaram, 2007; Jin, Zhou, & Mobasher, 2005; Krestel, Fankhauser, & Nejdl, 2009; Wang & Blei, 2011).

Although LDA has wide applicability in various fields, it has not yet been used to identify new product opportunities in patent analysis, according to the best of our knowledge. In the present study, we will use assignees as documents and product names appearing in the patent claims of assignees as words. Then, LDA has the capability to provide critical information about topics; for example, product names constituting each topic and topic distributions of assignees.

2.3. Collaborative filtering

Among the most successful approaches to building recommender systems, collaborative filtering is a method that makes automatic predictions (the aspect of filtering) about the interests of a user by collecting preferences or taste information from many users (the aspect of collaborating) (Su & Khoshgoftaar, 2009). The underlying assumption of the collaborative filtering approach is that if a user (u1) has the same opinion on an issue as a person (u2), then u1 is more likely to have the same opinion as u2 on a different issue x than u1 is likely to have the opinion on x of a person chosen randomly. The work flow of collaborative filtering occurs as follows: (1) a user expresses his or her preferences by rating items (e.g., books, movies or CDs) within a system, (2) the system matches this user's ratings against the ratings of other users, and thereby finds people with similar tastes, and (3) utilizing similar users, the system recommends items that the similar users have rated highly but which have not yet been rated by the original user.

The applications of collaborative filtering are various. Collaborative filtering has been used to support cross-selling (Kitts, Freed, & Vrieze, 2000), to alert filtering systems (Sen et al., 2006), to make recommendations to twitter users (Hannon, Bennett, & Smyth, 2010) and to support web news personalization (Das et al., 2007). In terms of using technical data, there are several studies that use collaborative filtering to recommend patent partners (Wu, Sun, & Tang, 2013) and to recommend patents for innovative design collaboration (Trappey, Trappey, Wu, Fan, & Lin, 2013). In particular, a study topically related to our approach collects large-scale user query data on a search engine and proposes a quantified approach to recommending promising technology keywords for small and medium-sized enterprises (SMEs) using co-word analysis, collaborative filtering and regression analysis (Yeo, Kim, Coh, & Kang, 2013).

The assignee–product vectors based on the vector space model can be considered to be the degree of interest in the relevant products of each assignee. Therefore, this study employs collaborative filtering to input assignee–product vectors and predict potential product items untapped by a target firm.

3. Proposed approach

3.1. Databases

The prerequisite for our product opportunity analysis approach is the preparation of product and patent databases. First, the product database is a set that includes feasible product names in patents, and each product name plays a key role to identify product portfolios by assignee. In natural text, however, a product name can be written using slightly different expressions; for example, a product known as "AC–DC converter" can be written as "AC-to-DC converter", "AC/DC converter", or the plural forms of these synonymous variations. Therefore, the product database should include representative product names and their variations, as well as be able to relate each representative name to its variants. This research utilizes the findings of the Korea Institute of Science and Technology Information (KISTI), which is a Korean national government research institute, as the product database. A research project of this institute has produced a product name database that includes all the product names from the USPTO product dictionary (KISTI, 2012). This database has evolved over the last four years and served a technology intelligence system called the technology opportunity discovery (TOD) system (http://tod.kisti.re.kr). All of the product names were validated by examining whether each product name occurs in patent analysis or not. The total number of unique product names is 118,330, and the number of representative product names is 59,914.

The source of data in which to find information about the occurrence of product names is patents, so constructing a large-scale database of patents is necessary. To this end, we used patent data on the USPTO patent bulk download service website (http://patents.reedtech.com). This website updates all of the patents granted to the USPTO database on a weekly basis. In addition, the weekly bulk patent data appears in an XML format, which includes complete information about each patent, such as bibliographic information, descriptions, claims and drawings. Our study developed an application to extract the patent information from the XML files, thereby collecting a total of 1,114,129 patents granted during the period of time from 2009 to 2013.

The databases of product names and patents were stored in our MySQL database. This mega database serves as the foundation on which our proposed approach to identify product occurrences in patent text is based.

3.2. Proposed process for recommending new application products

Building on the product database and patent database, this study proposes an application product recommendation approach consisting of a total four steps (Fig. 1). The steps are: (1) generating assignee–product portfolio vectors by grouping patent–product vectors by assignee, (2) measuring product portfolio semantic similarities between a target firm and other assignees using LDA, (3) rating new product items untapped by the target firm using collaborative filtering, and (4) identifying the promise and domain heterogeneity of the recommended product items using a product recommendation portfolio map. This section describes each step in greater detail.

3.2.1. Generating assignee–product portfolios

The first step of our process generates assignee–product portfolio vectors by identifying occurrences of product names in the patents of each assignee. Although each patent includes various textual sections, the claimed knowledge of the patent – which is composed of key technology or product components – should be specified explicitly in the claim section of the patent. For this reason, this step considers only claim sections to generate assignee–product portfolio vectors. Then, a patent p_m means an invention targeting a list of specific product components and can be structured as an array of integer elements by the vector space model:

$$p_m = [o_{m,1}, o_{m,2}, o_{m,3}, \ldots, o_{m,n}] \tag{1}$$

where n is the product name index and $o_{m,n}$ is an integer, which is the occurrence frequency of product n in patent m. This step structures the patent–product vectors for all patent documents.

Next, this step generates assignee–product vectors by grouping the patent–product vectors. This grouping is necessary because our approach aims to recommend new application products to a target

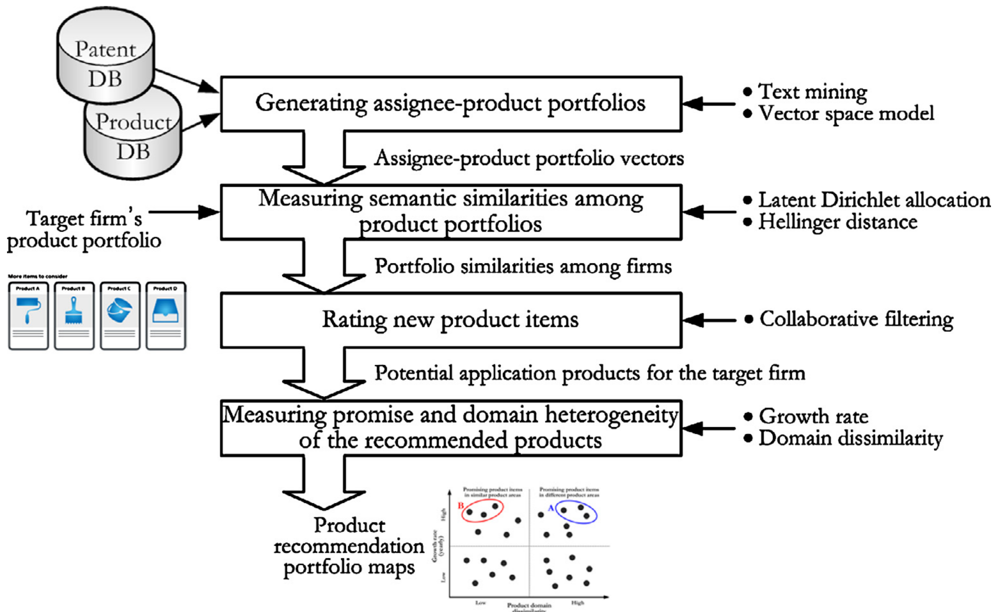

Fig. 1. Overall process.

firm by taking into consideration the product portfolios of other assignees. The bibliographic section of a patent contains information about patent assignees, such as assignee names, addresses and nationalities, so the bibliographic section helps recognize identical assignees and facilitate the generation of assignee–product portfolio vectors from the patent–product vectors.

When grouping the patent–product vectors of an assignee into an assignee–product vector, we have to carefully decide how to represent the occurrence frequency of a product name within an assignee's patents as the level of interest that the assignee has in the product. However, unlike user rating scores, the frequency of the occurrence of a product name within an assignee's patents does not directly represent the assignee's preference for the product. Rather, the frequent occurrence of a product name could be considered an implicit feedback. Accordingly, this step assumes that an assignee has an interest in a product or that the assignee endeavors to develop a product if the total occurrence frequency of the product name in the assignee's patents is more than a threshold value γ. Then, we can represent this level of interest as a 0 or 1. For example, if γ is set to 10, then only products with 10 or more total occurrence frequencies in an assignee's patents have values of 1. The assignee–product portfolio vector of an assignee a_i can be written as:

$$a_i = [h_{i,1}, h_{i,2}, h_{i,3}, \ldots, h_{i,n}] \quad (2)$$

$$h_{i,n} = \begin{cases} 1, & \text{if } sum_{p_m \in P_i}(o_{m,n}) \geqslant \gamma \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where P_i indicates the patent set of assignee i, $sum_{p_m \in P_i}(o_{m,n})$ means the total occurrence frequency of product name n in patent p_m of assignee i, and $h_{i,n}$ has a binary value of 0 or 1. In this way, this step generates assignee–product vectors for all patent assignees.

3.2.2. Measuring semantic similarities between the product portfolios of a target firm and other assignees

This step is intended to calculate semantic product portfolio similarities between pairs of firms for a target firm in need of application product opportunities. Therefore, the first input for this step is assignees' product portfolio vectors, each of which can be represented as an assignee–product vector; this conversion is easily achieved by assigning element values of 0 or 1 by Eqs. (1) and (2). Then, this step constructs an assignee–product portfolio matrix, including assignee–product vectors of the target firm and other assignees. This matrix A can be written as:

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_i \\ a_{target\_firm} \end{bmatrix} \quad (4)$$

To identify semantic similarities between the portfolios of pairs of firms, this step applies LDA to the assignee–product portfolio matrix. As a result of LDA application, we can obtain the probability distribution of topics for each assignee. Then, the probability vectors of topics for assignees are used to calculate the semantic similarities among the product portfolios of assignees. Given probability distributions P and Q of two assignees' topics, the Hellinger distance can be used to quantify the similarity between two probability distributions. For discrete probability distributions $P_{target\_firm} = (p_1 \ldots p_k)$ and $Q_{assignee} = (q_1 \ldots q_k)$, the Hellinger distance $H(P,Q)$ is defined as:

$$H(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum\nolimits_{i=1} (\sqrt{p_i} - \sqrt{q_i})^2} \quad (5)$$

This measure satisfies the property $0 \leqslant H(P,Q) \leqslant 1$, and $H(P,Q)$ becomes 0 if the two probability distributions are identical. Consequently, the semantic similarity between the product portfolios of two assignees $sim(P,Q)$ is defined as:

$$sim(P,Q) = 1 - H(P,Q) \quad (6)$$

3.2.3. Rating new application products untapped by the target firm

The objective of this step is to predict the rating scores of product items that have not yet been developed by the target firm. Building on the semantic similarities between product portfolios of the target firm and other firms, the product rating score can be obtained by the collaborative filtering method that normalizes item rating scores by top-k firms similar to the target firm.

Therefore, the equation for predicting the rating score of an untapped product item for the target firm $r_{c,s}$ can be written as:

$$r_{c,s} = \frac{1}{\mu}\sum_{c' \in C} sim(c,c')(r_{c',s}) \tag{7}$$

$$\mu = \sum_{c' \in C} sim(c,c') \tag{8}$$

where C is the set of top-k assignees that are the most similar to the target firm c, $sim(c,c')$ is the product portfolio similarity between the target firm and an assignee within C, and s is the product item untapped by the target firm. After this calculation process, this step ultimately arranges the untapped product items in descending order of rating scores. Top products with the highest rating scores could be recommended as potential products for the target firm.

3.2.4. Identifying the promise and domain heterogeneity of recommended products

This step sorts out the types of product items recommended by the previous step. By taking into consideration the promise and domain heterogeneity of the recommended products, this step outputs a map, termed the product recommendation portfolio map. Because the occurrence trend of product name in patents can represent the future promise of the product item, we first consider the yearly increase rate of the occurrence of product names in all patents. Second, the domain heterogeneity of the recommended application products to the target firm's existing products can be a way to find new product items in different technology areas. In this paper, we define the technology domain information of a product using a set of international patent classification (IPC) codes of the patents that contain the product name. In a similar way, if a target firm has a product portfolio with five existing products, then we can define the portfolio's technology domain using the set of IPC codes for these five products. Accordingly, the domain heterogeneity $d\_hetero(EPP, RP)$ of the existing product portfolio (EPP) of a target firm and a recommended product (RP) can be formulated by modifying the Jaccard coefficient:

$$d\_hetero(EPP, RP) = 1 - \frac{n[S_{ipc}(EPP) \cap S_{ipc}(RP)]}{n[S_{ipc}(EPP) \cup S_{ipc}(RP)]} \tag{9}$$

where $S_{ipc}(X)$ is the set of IPC codes of a product or product portfolio X and $n[Y]$ is the number of elements in set Y.

Using these two indexes, a two-dimensional portfolio map is ultimately generated (Fig. 2). In the portfolio map, the oval areas in blue and red are promising and favorable for the target firm, respectively. First, the oval in blue (A) represents a product area that consists of recommended product items that have high growth potential and exist in product domains relatively different from the domain of the target firm's existing product portfolio. Therefore, the blue oval may include new application product opportunities outside of the scope of the target firm's existing business. Second, the oval in red (B) represents a product area that consists of recommended product items that have high growth potential and exist in product domains that are similar to the target firm's existing product portfolio. Therefore, recommended products in the red oval have a possibility of being application products that can be developed using the target firm's existing technologies and facilities.

4. Illustration and discussion

This section applies our proposed approach to a specific target firm, thereby recommending potential application product opportunities for the firm. Then, we examine the technical feasibility of developing the recommended products.

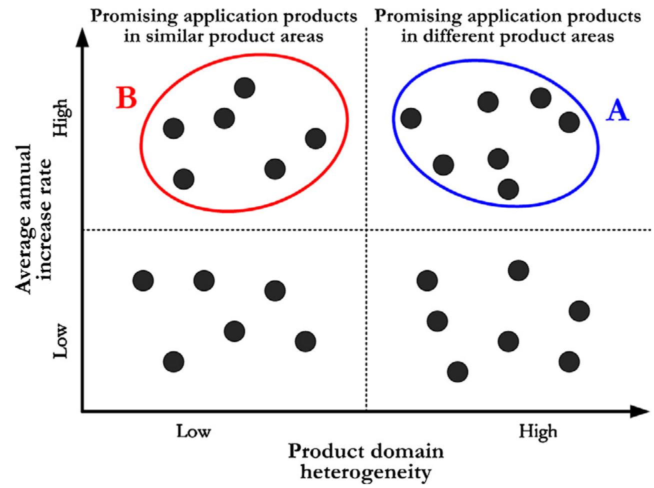

Fig. 2. Product recommendation portfolio maps.

4.1. Defining a target firm

Our approach requires the definition of the product portfolio of a target firm. The product portfolio for a real firm can be defined manually by using the 59,914 product names, although the real firm does not have many patents. However, we currently do not have in-depth information for such a firm, so for convenience our case study selected "Dongjin Semichem Co. Ltd." as a target firm among the assignees in our database. Dongjin Semichem Co. Ltd. has 37 patents registered in the USPTO database. In the category of SMEs, this firm is a Korean manufacturer that produces materials for flat panel displays and semiconductors, materials for alternative energy sources, and various foaming agents. By analyzing the occurrence frequencies of product names in the patents of this target firm, its product portfolio was found to have a total of 49 unique product names in different patents (Table 1). Some example product names are color filters, photosensitive polymers, condenser lens, dry films, light absorbers, fuel cells and methanol fuels.

4.2. Recommending application product opportunities for the target firm

To generate assignee–product portfolio vectors, we constructed patent–product vectors for all patents using product and patent databases. For each patent–product vector, we used only the product names that appear 10 or more times in all of the patents (among a total of 59,914 product names), because we considered product names with too low occurrence frequencies to have less value as product opportunities. Statistical analysis showed that each of the 1,114,129 patents contains an average of approximately 10.51 products in the text of its claims. Using the patent–product vectors, we generated assignee–product portfolio vectors. In doing so, we excluded two types of assignees, including firms with few patents and universities. First, only firms with five or more patents were used because the number of patents by assignee shows whether or not the assignee steadily performs technology development activities. Second, universities are not normal firms and their patents belong to technology domains that are too diverse for our purposes.

As a result, we extracted a total of 591,923 valid patent–product vectors. For these patent–product vectors, the number of unique valid product names was 17,227 (28.8% of the total products) and the number of unique firms was 16,195. Then, by setting a

Table 1
Product names appearing in the target firm's patents.

The product portfolio of the target firm (49 products)
Color filter, solar cell, emitting diode, ethylene glycol, photosensitive resin, photosensitive compound, light absorber, fuel cell, photosensitive polymer, membrane-electrode assembly, polymer binder, chemical mechanical polishing slurry, acrylate resin, epoxy resin, ester compound, ethyl cellulose, dry film, filter panel, dipropylene glycol, diethylene glycol, condenser lense, insulation film, concave mirror, block copolymer, black pigment, backlight unit, attitude sensor, convex lens, methacrylate monomer, thin film transistor, siloxane compound, protection film, polymer resin, polymer film, polymer electrolyte membrane, polyimide resin, polyethylene glycol, indium tin oxide, methanol fuel, flat panel, light guide plate, lcd display panel, lcd display, iron salt, ultraviolet light stabilizer, insulating substrate, acetic acid, heterocyclic compound, photo initiator

threshold value $\gamma = 10$, we generated a total of 16,195 assignee–product vectors. Each assignee–product vector had dimensions of 17,227 product names; each dimension had a value of 1, if the product name that corresponds to the dimension appeared in the patent of the assignee.

To identify the product portfolio similarities between this target firm and all the other firms, we applied LDA to the assignee–product portfolio vectors. Currently, various tools exist for LDA application, including JGibbLDA (http://jgibblda.sourceforge.net), GenSim (http://radimrehurek.com/gensim) and NetMiner4.0 (http://www.netminer.com). Using NetMiner4.0, we fitted the model with a random walk Markov chain Monte Carle (MCMC) method based on Gibbs sampling (Casella & George, 1992). To define the number of topics, we referred to recent international standard industry classifications (http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27) and set the number of topics to 80. As the options for MCMC, Dirichlet hyper-parameters alpha and beta were set to 0.626 and 0.1, respectively; the recommended alpha value in LDA is 50 divided by the number of topics (Griffiths & Steyvers, 2004), and smaller values of beta produce more specific topics. For a Gibbs sampler, we set the number of iterations = 100, burn-in = 10 and sample-lag = 1.

As a result, we obtained the topic distribution probability vector of each assignee over the 80 topics (Table 2). For example, firms in the industry of digital games and entertainment were found to belong to topic 1, with keywords such as computing device, touch screen, computer output device, number generator and game device (Appendix A). Next, we were able to compute the semantic product portfolio similarities between the target firm and all the other firms using Eqs. (5) and (6). For the collaborative filtering, we selected the top 50 firms that are the most similar to the target firm in terms of product portfolios (Table 3).

Building on the semantic similarities between pairs of the product portfolios, we applied Eqs. (7) and (8) of collaborative filtering to recommend the products untapped by the target firm (Table 4). According to the recommendation results, application products, such as polarizing plates, polyester resin, protective film and optical film, were high-ranked with rating scores of more than 0.2. That is, polarizing plates could be a good candidate for further product development. A polarizing plate or polarizer is an optical filter that allows light of a specific polarization to pass and blocks waves of other polarizations. From a technical perspective, we found that such a product can be developed by applying the target firm's existing capability of producing color filters. In addition, we found polyester resins as another recommended product; polyester resins are used in sheet molding compounds and they can be technically produced by applying the same manufacturing processes utilized for color filters and methanol fuel cells.

Although some of the recommended products may serve as fresh ideas for new product planning, a recommendation alone does not provide information about the promise and domain heterogeneity of potential products. Therefore, we computed the domain heterogeneity of each recommended product to the target firm's product portfolio using Eq. (9), and identified the average annual increase rate in occurrence of the recommended product name in patents during the period of time from 2009 to 2013 using

Table 2
Part of assignee-topic distributions, sorted by topic 1 (digital games and entertainment).

Firm names	Topic-1	Topic-2	Topic-3	Topic-4	Topic-5	Topic-6	
Bally Gaming Inc.	0.2838	0.0054	0.0054	0.0550	0.0034	0.0098	...
Multimedia Games Inc.	0.2732	0.0066	0.0089	0.0200	0.0070	0.0080	...
Scientific Games International Inc.	0.2542	0.0061	0.0098	0.0159	0.0074	0.0072	...
SHFL Entertainment Inc.	0.2121	0.0088	0.0115	0.0130	0.0091	0.0090	...
Shuffle Master Inc.	0.2087	0.0070	0.0112	0.0086	0.0074	0.0111	...
WMS Gaming Inc.	0.2075	0.0035	0.0045	0.1001	0.0026	0.0091	...
Digideal Corporation	0.1893	0.0092	0.0103	0.0153	0.0094	0.0103	...
Aruze Gaming America Inc.	0.1836	0.0077	0.0080	0.0094	0.0069	0.0131	...
Walker Digital LLC	0.1753	0.0055	0.0079	0.0111	0.0054	0.0084	...
Scientific Games Holdings Limited	0.1648	0.0100	0.0107	0.0194	0.0097	0.0098	...
Video Gaming Technologies Inc.	0.1581	0.0101	0.0107	0.0098	0.0098	0.0107	...
IGT	0.1566	0.0048	0.0039	0.1041	0.0021	0.0062	...
Cork Group Trading Ltd.	0.1552	0.0087	0.0093	0.0113	0.0087	0.0093	...
Aristocrat Technologies Inc.	0.1537	0.0101	0.0103	0.0101	0.0099	0.0135	...
Universal Entertainment Corporation	0.1507	0.0082	0.0051	0.0172	0.0040	0.0066	...
CFPH LLC	0.1491	0.0065	0.0101	0.0105	0.0065	0.0091	...
Konami Gaming Inc.	0.1454	0.0096	0.0096	0.0169	0.0089	0.0101	...
Zynga Inc.	0.1345	0.0074	0.0083	0.0126	0.0073	0.0094	...
Spielo International Canada ULC	0.1319	0.0107	0.0098	0.0125	0.0095	0.0112	...
Leap Forward Gaming	0.1302	0.0110	0.0105	0.0120	0.0108	0.0130	...
Olympian Gaming LLC	0.1209	0.0114	0.0121	0.0110	0.0104	0.0114	...
GTECH Rhode Island Corporation	0.1202	0.0079	0.0105	0.0120	0.0075	0.0095	...
GTECH Corporation	0.1184	0.0095	0.0104	0.0188	0.0095	0.0146	...
Electronic Arts Inc.	0.1158	0.0086	0.0090	0.0146	0.0080	0.0105	...
Fortunet Inc.	0.1154	0.0110	0.0113	0.0098	0.0098	0.0127	...

Table 3
Part of product portfolio similarities between the target firm and its top 50 similar firms.

Rank	Firm name	Portfolio semantic similarities
1	Kimoto Co. Ltd.	0.9388
2	Arisawa MFG Co. Ltd.	0.9366
3	Dexerials Corporation	0.9351
4	Ifire Ip Corporation	0.9346
5	Chisso Petrochemical Corporation	0.9342
6	JNC Corporation	0.9298
7	E Ink Corporation	0.9236
8	Sipix Imaging Inc.	0.9229
9	Tomoegawa Co. Ltd.	0.9219
10	Asahi Kasei Kabushiki Kaisha	0.9203
11	Chisso Corporation	0.9199
12	Toppan Printing Co. Ltd.	0.9177
13	Sumitomo Chemical Co. Ltd.	0.9162
14	JNC Petrochemical Corporation	0.9158
15	Toray Advanced Materials Korea Inc.	0.9156
16	Nissha Printing Co. Ltd.	0.9114
17	Sumitomo Metal Mining Co. Ltd.	0.9096
18	Nippon Chemical Industrial Co. Ltd.	0.9095
19	Dongwoo Fine Chem Co. Ltd.	0.9086
20	Tosoh Corporation	0.9083
	...	...
Sum of portfolio similarity values		45.3904

the geometric mean. As a result, we obtained the top 50 product recommendations with rating scores of 0.1 or higher, as well as their domain heterogeneity and product-name average annual increase rate in the collection of patents (Table 4); the number of type A products was 8 and the number of type B products was 19.

For an overall understanding of the product recommendation result, we generated a product recommendation portfolio map (Fig. 3), which displays the types of recommended products in terms of promise (average = 0.094) and domain heterogeneity (average = 0.075).

From the perspective of the target firm, new application product opportunities in promising and heterogeneous domains (Area A) were found to be polyester resin, touch screens, gas barrier films, glass powder, ammonium salt, polyvinyl alcohol, acid ester and titanium oxide. Currently, the target firm has a lot of experience in producing polymer compounds because the firm has developed photovoltaic polymers and color filters. These existing products are photosensitive products based on the manufacturing capabilities of polymers, which include thin sheets of polycarbonate or polyester. Therefore, among the identified product opportunities, polymer materials, including polyester resin, polyvinyl alcohol and acid ester, might be the most promising candidates for potential application products for the target firm. In fact, the target firm has used and developed various photosensitive polymer compounds similar to polyester resin, polyvinyl alcohol and acid ester to manufacture photovoltaic cells, color filters and thin films. In addition, the potential for gas barrier films was extrapolated from the existing portfolio of fuel cell and methanol fuel products because the manufacturing processes of these products include membrane manufacture technology; a membrane is a selective barrier that allows the passage of certain constituents and retains other constituents found in a liquid. As found in the target firm's product portfolio, some products, including membrane-electrode assembly and polymer electrolyte membrane, were the target products that the firm intended to develop in its patents.

Application products in Area A may require some degree of change to the firm's existing manufacturing processes, while potential application products in Area B are likely to be produced without much change to its manufacturing processes and facilities. Various application product opportunities in Area B, such as sugar alcohol, transparent film, plastic substrate, light emitting devices and optical film, were found to be within a very similar domain to that of the target firm's existing portfolio. The target firm's current capabilities of producing photovoltaic cells and color filters can be directly extended to the production of transparent film and optical film from a technical perspective. We concluded that light emitting devices could be a good opportunity for further product development, because the target firm had dealt with several products, including light color filter, guide plate, lcd display, lcd panel and ultraviolet light stabilizer, which are directly related to light emitting devices. Sugar alcohol seems to be another direct application product because the target firm likely has existing

Table 4
Application products recommended for the target firm.

Recommended products	Rating	Domain heterogeneity	Average annual increase rate	Type
Polarizing plate	0.3220	0.0000	0.0169	
Curing agent	0.2814	0.0366	0.1273	B
Polyester resin	0.2622	0.2166	0.1892	A
Acrylic resin	0.2405	0.0238	0.1346	B
Protective film	0.2207	0.0178	0.0760	
Polymerizable monomer	0.2204	0.0467	0.1304	B
Silane coupling agent	0.2193	0.0152	0.1286	B
Optical film	0.2012	0.0118	0.1374	B
Transparent film	0.1818	0.0439	0.1530	B
Acid ester	0.1813	0.0755	0.1624	A
Ammonium salt	0.1800	0.0861	0.1073	A
Conductive film	0.1619	0.0062	0.0034	
Amino acid	0.1617	0.8410	0.0726	
Organic pigment	0.1606	0.0216	0.0406	
Mineral acid	0.1595	0.0995	0.0503	
Resin component	0.1595	0.0275	0.1406	B
Propylene glycol	0.1594	0.0421	0.1605	B
Curable resin	0.1420	0.0365	0.0644	
Electroluminescent device	0.1418	0.0035	−0.0173	
Ink jet	0.1413	0.3845	0.0255	
Polyolefin resin	0.1409	0.0127	0.1425	B
Metal film	0.1406	0.0245	0.0384	
Acrylate monomer	0.1398	0.0364	0.1491	B
Acid anhydride	0.1393	0.0495	0.0923	
Titanium oxide	0.1392	0.0752	0.1520	A
Polymerization initiator	0.1384	0.0220	0.2099	B
Metal foil	0.1215	0.0598	0.1176	B
Plastic substrate	0.1214	0.0403	0.1118	B
Compensation film	0.1214	0.0000	−0.2337	
Crystal device	0.1213	0.0050	−0.0144	
Organosilicon compound	0.1209	0.0341	0.0000	
Flexible printed circuit	0.1208	0.0352	0.0786	
Silicon nitride	0.1206	0.0150	0.0495	
Gas barrier	0.1205	0.0939	0.1096	A
Resin film	0.1204	0.0448	−0.0247	
Organic resin	0.1204	0.0235	0.0209	
Sugar alcohol	0.1201	0.0681	0.2574	B
Printed circuit	0.1200	0.3715	0.0369	
Dicarboxylic acid	0.1198	0.0553	0.1328	B
Polyvinyl alcohol	0.1197	0.0806	0.1779	A
Phosphoric acid	0.1197	0.0677	0.1151	B
Retardation film	0.1196	0.0097	−0.0543	
Glass powder	0.1192	0.0876	0.1560	A
Liquid crystal film	0.1014	0.0000	−0.1199	
Touch screen	0.1011	0.2118	0.4491	A
Light emitting device	0.1009	0.0200	0.2409	B
Ethylene oxide	0.1008	0.0667	0.1349	B
Conductive adhesive	0.1008	0.0064	0.0805	
Organic semiconductor	0.1007	0.0066	0.0682	
Magnesium oxide	0.1001	0.0710	0.1711	B

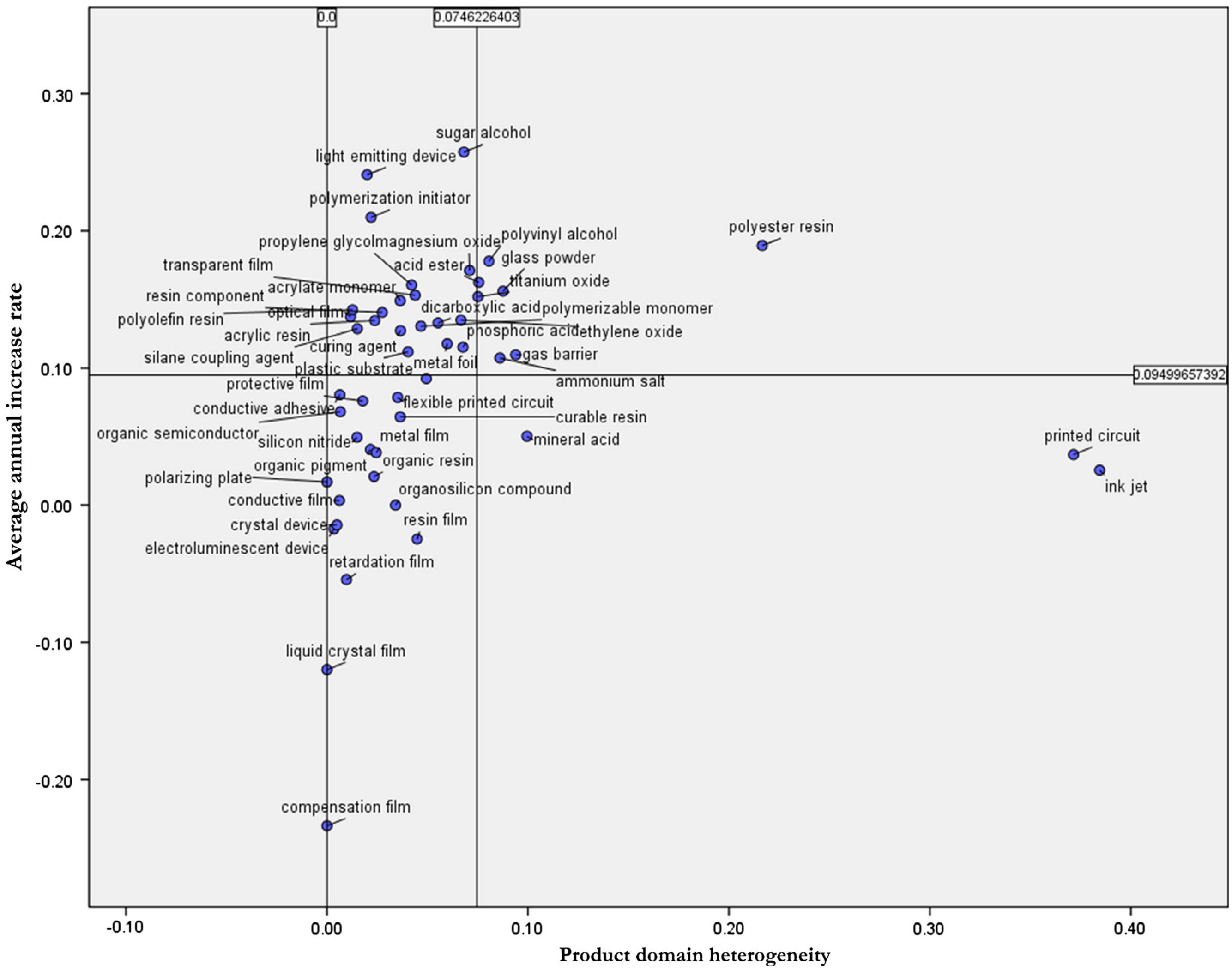

Fig. 3. A product recommendation portfolio map for the target firm (products "amino acid" and "touch screen" were excluded for effective display).

capabilities of processing various types of alcohol from methanol fuel manufacture processes.

In this case study, we applied our recommendation approach to a high technology-based target firm and were able to identify potential application products for the firm. For some selected application products, we examined their technical feasibility of being developed by utilizing the target firm's existing product portfolio.

4.3. Discussion

Identifying new application products using a technology-based firm's existing product capabilities is a useful approach to product planning, in that it allows practical and customized product planning by leveraging the firm's manufacturing capabilities, processes and facilities. To this end, this paper proposed a new approach to product opportunity recommendation, the working mechanisms of which are the vector space model, LDA and collaborative filtering.

First, the proposed method modeled a technology-based firm's existing capabilities using a product portfolio, or a vector of product names that appear in a comprehensive set of patents from the firm. Generally, the input for collaborative filtering is user preference scores for specific items. In our case, however, the occurrence frequency of product names in a firm's patents is an implicit feedback and does not directly represent the firm's product preferences. In fact, it is difficult to gauge the preferences of each firm for products from patent text data alone. For this reason, we simplified the modeling process of assignee–product vectors after careful consideration by technology experts. We assumed that a firm has an interest in a product or endeavors to develop the product if the total occurrence frequency for the product in the firm's patents is larger than a threshold value $\gamma = 10$. This allowed the conversion of each assignee–product vector into a vector from, with element values of 0 or 1.

Second, to identify semantic similarities between pairs of product portfolios, this research employed LDA. Basically, two assignee–product vectors can be simply computed using similarity measures such as Jaccard coefficients and cosine coefficients. However, these simple measures cannot successfully capture the hidden relationships among product names of the assignee–product vectors herein. Several techniques, including latent semantic analysis and principal component analysis, currently exist to identify latent factors in high-dimension vectors and to measure their semantic similarities on a latent-factor dimension. Although these techniques allow semantic similarity computation between two high-dimension vectors, they do not provide sufficient information about each latent factor. In contrast, LDA provides information about topics, as well as the product names consisting each topic and topic distributions of firms. Due to these advantages of LDA, our approach adopted LDA to identify semantic similarities between the portfolios of a target firm and other firms.

Third, in order to recommend new application products to a target firm, our approach utilized collaborative filtering. As made apparent by many previous studies, an important issue of collaborative filtering is the cold-start problem (Bobadilla, Ortega, Hernando, & Bernal, 2012). This means that collaborative filtering may not work well for firms with the insufficient number of patents. Therefore, after careful consultation with technology experts, our analysis included only firms with five or more patents in order to mitigate the cold-start problem raised by the assignee–product vectors. As a result, each assignee was found to have an average of 36.55 products. In fact, there is no right answer in determining the number of patents that will effectively exclude invalid firms. Accordingly, this process should be conducted with care, because it directly affects the recommendation results by collaborative filtering.

5. Concluding remarks

This study proposed an approach to the identification and recommendation of new application products based on a firm's existing product portfolio. As the building blocks of the approach, this study uses the vector space model, LDA and collaborative filtering. In terms of the specific steps of the approach, each firms' product portfolio is structured as an assignee–product vector using the occurrence information of product names in a comprehensive sample of the firm's patent text, then semantic similarities between the product portfolios of pairs of firms are computed by applying LDA to assignee–product vectors, and finally, new application products untapped by the target firm are recommended using collaborative filtering. The proposed approach provides a visual map, called the product recommendation portfolio map, which classifies the recommended application products in terms of product heterogeneity and promise. The functionality of the approach was demonstrated herein using all of the patents registered in the USPTO between 2009 and 2013, together with the relevant firms. The proposed approach contributes to exploring new product opportunities across various domains using the existing product portfolios of firms and thereby stands to assist product planners in the ideation processes of identifying new application products and new markets, which product planners may not consider intuitively.

We expect that this study will make both academic and practical contributions to relevant fields. From a methodological view, our approach quantifies the process of identifying product opportunities. Product or technology opportunity identification processes in prior studies have depended heavily on the intuition of experienced technology experts, while our study uses objective data – namely, patents – to recommend new application products for a target firm by exploiting collaborative filtering. Our quantified process could be implemented into integrated software systems, thereby efficiently supporting experts in identifying new product opportunities. From an industrial perspective, our approach could be a customizable tool for SMEs. Most SMEs are aware of the need for product development to stay competitive, but they suffer from insufficient information and human resources to generate new ideas (Yeo et al., 2013). When this is the case, one effective product strategy for the SMEs will be to identify application products that make full use of their existing facilities and manufacturing processes. In this regard, our approach will help SMEs identify practical and low-risk product items by exploiting their existing product portfolio.

Despite the contributions of this study, challenges for future research still remain. First, the proposed approach simplifies the existing product portfolios of firms as assignee–product vectors with element values of 0 or 1. In fact, the precise definition of product portfolios is a key to furthering our approach. Accordingly, a topic for future research to develop is how to identify the preferences of firms for products using fuzzy logic and machine learning techniques. Second, our approach was applied to one example target firm, but it has the potential to be applicable to real-world firms in a variety of technology domains. Therefore, further research should develop application examples in different technology domains by defining existing product portfolios of the real-world firms. Finally, because the proposed approach quantifies the identification of new product opportunities, implementing the approach into an integrated software system is a natural next step as well as a fascinating topic for future research.

Acknowledgement

This paper was supported by Konkuk University in 2013.

Appendix A. 80 topics and their main contributing keywords (i.e., product names)

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword	# of assignees
Topic-1	computing device	touch screen	computer output device	number generator	game device	97
Topic-2	control valve	temperature sensor	pressure sensor	water outlet	inner tube	55
Topic-3	computing device	energy storage	temperature sensor	web server	emitting diode	7
Topic-4	host computer	random access memory	peripheral component	nonvolatile memory	memory controller	272
Topic-5	pressurized fluid	fluid passage	sealing ring	valve actuator	valve housing	318
Topic-6	signal generation device	low pass filter	processing device	digital signal processor	pass filter	99
Topic-7	pressure sensor	carbon steel	steel strap	steel sheet	gas supply	88
Topic-8	current sensor	power system	dc/dc converter	energy storage	dc power supply	299
Topic-9	mounting bracket	mounting plate	support element	woven fabric	electric motor	62
Topic-10	solar cell	carbon nanotube	silicon nitride	indium tin oxide	mass spectrometer	158
Topic-11	emitting diode	printed circuit	light emitting device	light sensor	incandescent light bulb	158

Appendix A (*continued*)

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword	# of assignees
Topic-12	communication apparatus	network interface controller	communication terminal	mobile radio	signal transmission	198
Topic-13	lcd display	color filter	thin film transistor	light emitting device	lcd display panel	260
Topic-14	dispenser apparatus	storage container	container can	plastic container	beverage container	305
Topic-15	machine tool	servo motor	robotic arm	outer ring	force sensor	84
Topic-16	optical waveguide	optical receiver	optical detector module	optical filter	optical transmitter	225
Topic-17	print head	ink jet	printing device	printing apparatus	feed roller	190
Topic-18	printed circuit	emitting diode	computing device	monitoring device	plug jack	6
Topic-19	hydrogen sulfide	active carbon	reactor vessel	distilling column	solid catalyst	393
Topic-20	temperature sensor	heating device	gas supply	heating apparatus	water outlet	83
Topic-21	digital video	digital audio	film frame	media device	remote control device	254
Topic-22	vehicle seat	vehicle door	instrument panel	seat frame	seat assembly	349
Topic-23	image sensor	image processor	imaging device	acquisition device	image processing apparatus	139
Topic-24	networking device	user device	application server	proxy server	communication device	596
Topic-25	pulse generator	signal generation device	temperature sensor	computer readable memory	motion detector	8
Topic-26	signal transmission	output terminal	signal generation device	switching circuit	nonvolatile memory	13
Topic-27	electrical lead	gas generator	carbon fiber	aircraft control system	aircraft wing	90
Topic-28	imaging lens	optical glass	lens holder	zoom lens	lens barrel	154
Topic-29	position sensor	magnetic sensing device	switch device	pressure sensor	plug jack	6
Topic-30	cleaning apparatus	guide rail	printing plate	printing press	cleaning fluid	41
Topic-31	epoxy resin	inorganic filler	flame retardant	block copolymer	curing agent	496
Topic-32	fluid reservoir	pump device	fluid container	fluid passage	peristaltic pump	205
Topic-33	output terminal	operational amplifier	differential amplifier	charge pump	switching circuit	456
Topic-34	electric motor	planetary gear	sun gear	drive gear	motor shaft	309
Topic-35	polymer film	piezoelectric transducer	printed circuit	electrical insulator	mounting device	6
Topic-36	elongate shaft	invasive surgical device	cutting blade	inflatable balloon	delivery catheter	354
Topic-37	transmission device	coding unit	synchronization unit	transmitting device	extraction unit	114
Topic-38	linear actuator	computing device	amino acid	processing device	cooling fluid	2
Topic-39	mounting bracket	cam follower	motor assembly	torsion spring	mounting plate	233
Topic-40	drilling fluid	drill string	pressurized fluid	tungsten carbide	drilling rig	163
Topic-41	computing device	support element	temperature sensor	pressure sensor	polyvinyl chloride	4
Topic-42	guide rail	transfer device	conveying device	transport device	feeding device	308
Topic-43	mounting plate	fuel cell	computer output device	support element	contact plate	1
Topic-44	wiring board	storage part	insulation film	electrode pad	resin film	166
Topic-45	supply device	connection plate	screw thread	metal housing	vacuum pump	5
Topic-46	combustion engine	cylinder head	control valve	hydraulic pump	hydraulic circuit	333
Topic-47	silicon nitride	silicon substrate	contact pad	semiconductor wafer	bond pad	381
Topic-48	phase detector	logic circuit	memory circuit	memory controller	interface circuit	256
Topic-49	printed circuit	communication device	processing device	receiving device	temperature sensor	8
Topic-50	beam splitter	imaging device	objective lens	image sensor	optical filter	450
Topic-51	computing device	server computer	web browser	web server	search engine	810
Topic-52	switching valve	detecting device	supply pipe	screw shaft	position detector	124

(*continued on next page*)

Appendix A (*continued*)

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword	# of assignees
Topic-53	amino acid	therapeutic agent	light chain	polyethylene glycol	antibody fragment	1085
Topic-54	energy storage	temperature sensor	plug jack	printed circuit	electric motor	13
Topic-55	support element	actuating device	coupling device	gear wheel	leaf spring	328
Topic-56	coupling device	measurement device	electric motor	processing device	support device	3
Topic-57	radio frequency identification tag	rfid reader	identification tag	communication device	identification device	201
Topic-58	touch screen	proximity sensor	lcd display	audio device	digital camera	170
Topic-59	sensing device	optical sensor	sensor device	position sensor	sensor element	32
Topic-60	pressure sensor	electronic controller	molded plastic	emitting diode	computer output device	7
Topic-61	acetic acid	amino acid	hydrochloric acid	acid ester	mineral acid	744
Topic-62	control valve	air filter	filter housing	annular seal	outer housing	173
Topic-63	local oscillator	power amplifier	bandpass filter	noise amplifier	transmit antenna	319
Topic-64	printed circuit	nonvolatile memory	computing device	processing device	output terminal	17
Topic-65	flexible printed circuit	connecting arm	connecting plate	torsion spring	heat dissipation unit	152
Topic-66	computing device	communication device	time stamp	central processing unit	communications device	81
Topic-67	gas supply	nitrogen gas	oxygen gas	semiconductor wafer	substrate holder	217
Topic-68	temperature sensor	pressure sensor	measurement device	monitoring device	sensing device	33
Topic-69	amino acid	body fluid	fluorescent dye	sample container	flow cell	229
Topic-70	plug jack	female connector	male connector	plug connector	connector housing	298
Topic-71	fuel cell	alkaline earth metal	electric current collector	lithium ion	phosphoric acid	317
Topic-72	mounting plate	threaded fastener	top rail	wall panel	guide rail	190
Topic-73	internal thread	bone screw	screw thread	insertion tool	invasive surgical device	283
Topic-74	sensor element	pressure sensor	monitoring device	temperature sensor	piezoelectric element	122
Topic-75	printed circuit	emitting diode	temperature sensor	image sensor	light emitting device	15
Topic-76	polyethylene glycol	ethylene glycol	ethylene oxide	acrylate monomer	block copolymer	618
Topic-77	emitting diode	computing device	plug jack	pressure sensor	monitoring device	5
Topic-78	rear wheel	vehicle wheel	speed sensor	rear axle	brake system	250
Topic-79	plug jack	snap fit	cam follower	return spring	flexible arm	73
Topic-80	plug jack	polyvinyl chloride	amino acid	cooling fluid	hollow tube	29

References

Ansoff, H. I. (1957). Strategies for diversification. *Harvard Business Review, 35*(5), 113–124.

Bíró, I., Szabó, J., & Benczúr, A. A. (2008). Latent Dirichlet allocation in web spam filtering. In *Proceedings of the 4th international workshop on adversarial information retrieval on the web* (pp. 29–32). ACM.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems, 26*, 225–238.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*(3), 167–174.

Chang, P.-L., Wu, C.-C., & Leu, H.-J. (2010). Using patent analyses to monitor the technological trends in an emerging field of technology: A case of carbon nanotube field emission display. *Scientometrics, 82*(1), 5–19.

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change, 73*(8), 981–1012.

Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web* (pp. 271–280). ACM.

Geum, Y., Jeon, J., & Seol, H. (2013). Identifying technological opportunities using the novelty detection technique: A case of laser technology in semiconductor manufacturing. *Technology Analysis & Strategic Management, 25*(1), 1–22.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America, 101*(suppl. 1), 5228–5235.

Hannon, J., Bennett, M., & Smyth, B. (2010). Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on recommender systems* (pp. 199–206). ACM.

Jin, X., Zhou, Y., & Mobasher, B. (2005). A maximum entropy web recommendation system: Combining collaborative and content features. In *Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining* (pp. 612–617). ACM.

Jun, S., Park, S. S., & Jang, D. S. (2012). Technology forecasting using matrix map and patent clustering. *Industrial Management & Data Systems, 112*(5), 786–807.

KISTI (2012). In *Design and construction of a knowledgebase for technology opportunity discovery*. Daejun, Korea: KISTI.

Kitts, B., Freed, D., & Vrieze, M. (2000). Cross-sell: A fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities. In *Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 437–446). ACM.

Krestel, R., Fankhauser, P., & Nejdl, W. (2009). Latent Dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on recommender systems* (pp. 61–68). ACM.

Lee, C., Jeon, J., & Park, Y. (2011). Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach. *Technological Forecasting and Social Change, 78*(4), 690–702.

Lee, S., Lee, S., Seol, H., & Park, Y. (2008). Using patent information for designing new product and technology: Keyword based technology roadmapping. *R&D Management, 38*(2), 169–188.

Lee, C., Song, B., & Park, Y. (2013). How to assess patent infringement risks: A semantic patent claim analysis using dependency relationships. *Technology Analysis & Strategic Management, 25*(1), 23–38.

Lee, S., Yoon, B., Lee, C., & Park, J. (2009). Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change, 76*(6), 769–786.

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

Liu, J. S., Kuan, C.-H., Cha, S.-C., Chuang, W.-L., Gau, G. J., & Jeng, J.-Y. (2011). Photovoltaic technology development: A perspective from patent growth analysis. *Solar Energy Materials and Solar Cells, 95*(11), 3130–3136.

Misra, H., Cappé, O., & François, Yvon. (2008). Using LDA to detect semantically incoherent documents. In *Proceedings of the twelfth conference on computational natural language learning* (pp. 41–48). Manchester, United Kingdom: Association for Computational Linguistics.

Park, H., Kim, K., Choi, S., & Yoon, J. (2013). A patent intelligence system for strategic technology planning. *Expert Systems with Applications, 40*(7), 2373–2390.

Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert Systems with Applications, 40*(2), 736–743.

Park, H., & Yoon, J. (2014). Assessing coreness and intermediarity of technology sectors using patent co-classification analysis: The case of Korean national R&D. *Scientometrics, 98*(2), 853–890.

Park, H., Yoon, J., & Kim, K. (2012). Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics, 90*(2), 515–529.

Qian, G. (2002). Multinationality, product diversification, and profitability of emerging US small- and medium-sized enterprises. *Journal of Business Venturing, 17*(6), 611–633.

Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM, 18*(11), 613–620.

Sen, S., Geyer, W., Muller, M., Moore, M., Brownholtz, B., Wilcox, E., & Millen, D. R. (2006). FeedMe: A collaborative alert filtering system. In *Proceedings of the 2006 20th anniversary conference on computer supported cooperative work* (pp. 89–98). ACM.

Shin, J., & Park, Y. (2005). Generation and application of patent claim map: Text mining and network analysis. *Journal of Intellectual Property Rights, 10*(3), 198–205.

Son, C., Suh, Y., Jeon, J., & Park, Y. (2012). Development of a GTM-based patent map for identifying patent vacuums. *Expert Systems with Applications, 39*(3), 2489–2500.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence, 2009*, 4.

Trappey, A. C., Trappey, C., & Wu, C.-Y. (2009). Automatic patent document summarization for collaborative knowledge systems and services. *Journal of Systems Science and Systems Engineering, 18*(1), 71–94.

Trappey, A. J., Trappey, C. V., Wu, C.-Y., Fan, C. Y., & Lin, Y.-L. (2013). Intelligent patent recommendation system for innovative design collaboration. *Journal of Network and Computer Applications, 36*(6), 1441–1450.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research, 37*(1), 141–188.

Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 448–456). ACM.

Wang, Y., Sabzmeydani, P., & Mori, G. (2007). Semi-latent Dirichlet allocation: A hierarchical model for human action recognition. In *Human motion – Understanding, modeling, capture and animation* (pp. 240–254). Springer.

Wu, S., Sun, J., & Tang, J. (2013). Patent partner recommendation in enterprise social networks. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 43–52). ACM.

Xing, D., & Girolami, M. (2007). Employing latent Dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters, 28*(13), 1727–1734.

Yeo, W., Kim, S., Coh, B.-Y., & Kang, J. (2013). A quantitative approach to recommend promising technologies for SME innovation: A case study on knowledge arbitrage from LCD to solar cell. *Scientometrics, 96*(2), 589–604.

Yoon, B. (2008). On the development of a technology intelligence tool for identifying technology opportunity. *Expert Systems with Applications, 35*(1), 124–135.

Yoon, J., Choi, S., & Kim, K. (2011). Invention property-function network analysis of patents: A case of silicon-based thin film solar cells. *Scientometrics, 86*(3), 687–703.

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.

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.

Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change, 72*(2), 145–160.

Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *Engineering Management, IEEE Transactions on, 54*(3), 588–599.

Yoon, B., Park, I., & Coh, B.-Y. (2013a). Exploring technological opportunities by linking technology and products: Application of morphology analysis and text mining. *Technological Forecasting and Social Change*.

Yoon, J., Park, H., & Kim, K. (2013b). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics, 94*(1), 313–331.

Yoon, J., Park, Y., Kim, M., Lee, J., & Lee, D. (2014). Tracing evolving trends in printed electronics using patent information. *Journal of Nanoparticle Research, 16*(7), 1–15.

Yoon, J., Park, H., Seo, W., Lee, J.-M., Coh, B.-Y., & Kim, J. (2015). Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework. *Technological Forecasting and Social Change, 100*, 153–167.

Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: Application of text mining. *R&D Management, 38*(1), 51–68.

Yoon, B. U., Yoon, C. B., & Park, Y. T. (2002). On the development and application of a self-organizing feature map-based patent map. *R&D Management, 32*(4), 291–300.