Technological Forecasting & Social Change 148 (2019) 119737

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

# Technology opportunity discovery under the dynamic change of focus technology fields: Application of sequential pattern mining to patent classifications

Jaewoong Choi, Byeongki Jeong, Janghyeok Yoon*

*Department of Industrial Engineering, Konkuk University, 120 Neungdong-ro, Gwangjin-go, Seoul 05029, Republic of Korea*

ARTICLE INFO

*Keywords:*
Technology opportunity discovery
Sequential pattern mining
Uncertainty
Technology intelligence

ABSTRACT

Technology opportunity discovery (TOD), have evolved over time from technology forecasting to an approach based on existing technology capabilities for increased practicality. Unfortunately, TOD studies are still lacking, in that they do not consider the unique direction of the target firm in technology development or the recent trends of technological fields. Consequently, this paper proposes an improved methodology for identifying technology opportunities with less uncertainty by focusing on the target firm's dynamic change of focus technology fields. The proposed approach is as follows: 1) generate a sequence database, containing firms' dynamic change of focus technology fields; 2) explore the frequent sequential patterns from a precedence enterprise (PE) sequence database using the PrefixSpan algorithm to identify the technology candidates from a PE sequence similar with that of the target firm; and 3) evaluate the candidates on technological similarity, business stability, and recency. The results of the proposed approach are expected to help firms discover appropriate technology opportunities by considering both their existing technological capacities and the dynamic change of their focus technology fields. Furthermore, the proposed approach can identify the most appropriate technology opportunities with less uncertainty in real-life business environments by evaluating technological similarity, business stability, and recency.

## 1. Introduction

Business ecosystem is always in a highly volatile climate due to environment uncertainty (Schoemaker, 2015). Superficially, this uncertainty may be caused by environmental factors such as the rapid transition of technology, turbulence in the market, or swift changes in customers' needs (Baruch and Altman, 2016; Burkhardt and Brass, 1990; Jeong et al., 2019; Vahlne and Johanson, 2017). However, the fundamental cause of uncertainty is the difference between the amounts of information possessed and required to implement a task, including sustainable technological development or economic growth (Klir, 2005). From the technology perspective, the information held represents the experts' knowledge, know-how, and experience or a technical document, and the information required for sustainable growth and survival means intelligence, which is also useful in scientific decision making (Mola et al., 2015). With quality technology intelligence to minimize the gap between these two types of information in terms of technology, a group of companies can explore and preempt potential technology faster than the competition (Choi and Park, 2016) and achieve sustainable technological development and growth (Lee et al., 2017a). Here, uncertainty is directly connected to the sustainable growth and development of a firm, meaning that reducing uncertainty is a constant challenge for firms. Scholars refer to this set of activities for reducing uncertainty in terms of technology as technology opportunity discovery (TOD) (Klevorick et al., 1995; Olsson, 2005), from which companies can gain a competitive advantage in terms of identifying emerging technologies and R&D item creation (Ma and Porter, 2015). The initial TOD is typically fulfilled by influential executives and technical experts in the firm by forecasting the market size and promising technologies for the next quarter (Lee et al., 2014). However, this process would soon experience economic/temporal inefficiency and not every technology opportunity proposed as such guarantees economic success growth or technological development because environmental factors are uncontrollable, that is, difficult to predict using only the knowledge and experience of experts (Shibata et al., 2008).

Given these drawbacks, in the era of big data, the accumulation and management of large-scale technical documents such as patents becomes possible (Lee et al., 2009) and numerous data-driven

* Corresponding author.
*E-mail address:* janghyoon@konkuk.ac.kr (J. Yoon).

https://doi.org/10.1016/j.techfore.2019.119737
Received 19 August 2018; Received in revised form 6 August 2019; Accepted 2 September 2019
Available online 07 September 2019

quantitative TOD methodologies have been developed to support experts' scientific decision making. The research stream mainly uses textual information on patents (Tseng et al., 2007). Specifically, Yoon and Kim extracted the subject-action-object (SAO) structure from patent claims, mapped patents in two dimensions based on semantic similarities, and performed outlier detection (Yoon and Kim, 2012). Their study thus reflected the detailed technical content described by patents and provided specific technical insights that are difficult to grasp intuitively in a given technology field. Similarly, Yoon and Magee transformed patent documents into a keyword matrix, developed patent maps using generative topographic mapping, and performed link prediction based on support vector machine to explore technology evolutionary paths and potential candidates for developing new technologies (Yoon and Magee, 2018). On one hand, these studies apprehend the overall trend of a technology area, as well as evidence of emerging technologies, which is not readily visible, based on a large amount of patent bibliographic information. However, because they probe business opportunities at the technology level by focusing only on the textual characteristics of patents in the technology field rather than considering the location of individual firms, a firm cannot be sure that the recommended technology opportunities will derive economic profit or sustainable technological growth in a real-life business environment.

To minimize the uncertainty of the proposed technology opportunities in real-life business circumstances, a series of studies considered the technological position of target firms, including SMEs. As a typical example, Yoon et al. defined the technology capability of a firm with existing technologies or products and derived potential technology opportunities using semantic functional similarities between technologies or products (Yoon et al., 2015). Their research differentiated advantages, in that firms cannot only deal with the patented technologies through the SAO-based semantic analysis, but can also create more specific technology opportunities at product level to be used systematically in an actual business environment. Most of the studies that tried to reduce the uncertainty of technology opportunities based on firms' technology capabilities tended to focus on SMEs because the failure of R&D is more fatal to SMEs in terms of sustainable development and growth than for large firms. Specifically, SMEs lack the ability to shape and detach themselves from their uncertain external environments compared to large firms, thus needing a practical TOD system (North et al., 2001). Consequently, Park and Yoon expressed a firm's own capabilities with the main International Patent Classification (IPC) groups of submitted patents, and determined the technology opportunities from similar competitors through collaborative filtering to reduce the anxiety and uncertainty related to the recommended heterogeneous technology fields (Park and Yoon, 2017). Because this type of research is based on collaborative filtering and IPC information, it requires less preprocessing than previous studies and does not need periodic updates, meaning it can be used practically for various firms. It is also effective in safely establishing the R&D direction in heterogeneous technology fields, particularly for SMEs with relatively scarce economic resources and technological information. However, considering that the firms' focus technology fields may change due to a rapidly changing technological environment, the literature has some limitations, in that it defines the technology capability of the firm from a static perspective. As such, the technology opportunities created in this way may deviate from recent technological trends, by not considering the change in the competence of a precedence enterprise (PE) or the unique technology development direction of the target firm.

The concept "sequence" is used to reflect the temporal order of the firm's focus technology field and define a group of companies that have already undergone technological changes as "precedent enterprises" to ensure the certainty of the recommended technology opportunities for SMEs by determining their precedents. To this end, Korean patents registered since 2000 in the entire technology field were collected and a database was constructed for analysis. The proposed approach consists of four stages. First, the SMEs that receive a recommendation and the PEs that SMEs could benchmark are selected based on the number of registered patents and the R&D nature of the firm. Second, the focus technology fields, in which the firm had registered relatively numerous patents are selected for each year and technological sequences are generated for the IPC main group level by linking the focus technology fields in chronological order. Third, frequent sequential patterns are explored using PEs' sequence database using a PrefixSpan algorithm, and are defined as significant sequential patterns. Finally, PEs' sequential patterns that are similar with the sequence of the target firm are identified, and a group of technology opportunity candidates are derived from these patterns and evaluated on technological similarity, business stability, and recency.

This paper proposes a novel approach, which captures the dynamic changes in the focus technology field through a sequence and discovers technology opportunities, in contrast to prior studies on TOD, where firms' technology capabilities are static. Therefore, the proposed approach can be applied to a new research direction in both academia and business fields. Under the proposed approach, the technological sequence of the target firm, an SME, is used as input and subsequent technology opportunities are discovered, so that the firm's unique direction of technological development can be reflected over time. Additionally, it is the greatest advantage that the proposed approach can reduce the uncertainty of technology opportunities regarding their economic success or technological development in an actual business environment, based on the PEs' experience. In real-life businesses, the proposed approach can support experts' scientific decision making as an advanced form of prior TOD system. Especially for SMEs, where uncertainty is more important than novelty or marketability in decision making regarding the future technology strategy, the proposed approach is expected to be more effective for sustainable growth and development. Using the proposed evaluation process for recency, business stability, and technological similarity between PEs and the target firm, the target firm can adjust the weights of the criteria according to its unique philosophy or purpose.

## 2. Theoretical background

The chronological change of a firm's focus technology field is expressed by the concept "sequence" and significant change patterns are explored through sequential pattern mining to help the target firm discover stable technology opportunities. Therefore, this section describes two theoretical backgrounds, the TOD system and sequential pattern mining.

### 2.1. Technology opportunity discovery

In a business ecosystem consisting of technology, products, and customers, uncertainty has always been prevalent. Here, uncertainty means technological development and economic success of current technology or products are not guaranteed, even in the near future. Superficially, uncertainty seems to be caused by environmental factors, including the rapid transition of technologies, turbulence of the market environment, and swift changes in customers' needs. However, the fundamental cause of uncertainty is the difference between the amounts of information already possessed and needed to fulfill a task, which include sustainable technological development or economic success. Interpreting this basic cause from a technological viewpoint, the information held means the knowledge and experience of technical experts in a firm, or technical documents such as patents, while the information to complete a task means the intelligence necessary to achieve a firm's technological development and economic growth. To identify and develop such technology intelligence, many attempts have been made, generally called TOD. The early TOD approach is fulfilled by a group of technical experts or influential executives in the firm. Based on information, including knowledge, experience, and know-how, TOD is then implemented to forecast the market size and emerging technologies of the next quarter. Unfortunately, the expert-based TOD is losing reliability and efficiency, since it is not always possible for experts to be aware of a large number of new technical documents or new technologies, although their contribution is still crucial (Shibata et al., 2008). Hence, the TOD approach relying only on experts encountered inefficiency in terms of time and cost, while the uncertainty in the environment is not still addressed.

In response to these drawbacks, as the accumulation and management of massive amounts of technical documents, such as patents and articles, became possible, a large number of studies proposed a quantitative TOD methodology to help experts make scientific decisions more effectively. This research stream attempted to derive technology intelligence, which is difficult to grasp from experts' intuition or experience, from large-scale technical documents. Among the analyzed technical documents, patents have been mainly used because they are the most reliable indicators of technology level or technology intelligence. As a representative study, Yoon and Kim proposed a method to detect new technology opportunities by using SAO-based semantic patent analysis and outlier detection, a method with some excellent features as follows (Yoon and Kim, 2012). First, this study reflected the structural relationships among the technological components of relevant patents and identified the key findings of inventions or expertise of inventors in detail by using SAO-based syntactic analysis. Such results contain information that is hidden in a large amount of patent text and cannot be easily identified using experts' knowledge or experience. Second, the study provided potential technology opportunities at patent level, while most previous studies report only technological directions or promising technology fields. Hence, the results of this study can provide quality technology intelligence to experts in discovering technology opportunities for technology strategy formulation and planning. Similarly, Yoon and Magee suggested an approach to exploring potential technology opportunities, using generative topographic mapping and link prediction for forecasting quantitatively (Yoon and Magee, 2018). This approach provided the detailed features of technology opportunities by extracting and using keywords and citation information from patents, and can thus provide visual results such as the location of patents and the relationships among patents in terms of information. Specifically, using the developed patent map, it is possible to identify technological vacant areas and interpret these areas using the keyword vectors of each predicted cell. Therefore, this approach has the advantage of providing concrete insights into technology planning at the level of specific characteristics. On one hand, Wang et al. attempted to explore potential technology opportunities by applying text mining and an arbitrarily oriented projected clustering algorithm to both patent documents and scientific papers for microalgal biofuels (Wang et al., 2015). Interestingly, by collecting and comparing scientific papers and technological documents such as patents, this study identified a broader domain of technology opportunities, which cannot be represented by patents only. Based on its results, new technology opportunities can be identified from fields with ample scientific research but few technological applications and can be used for constructing R&D strategies.

The main characteristic of the above research stream is forecasting previously unused technologies and the cited studies are preeminent in anticipating technology opportunities that are likely to emerge in the future. Moreover, these studies can identify and visualize the overall trends of specific technology fields, which are elusive for experts' intuition and experience, from large-scale patent bibliographic information. Unfortunately, although these studies can help experts determine technology strategies or plan R&D items by providing them with scientific decision grounds from quantitative data, there are some additional points to be considered for practical applications in a real-life business environment. Above all, since these approaches are inherently based on technology forecasting, they are always accompanied by a risk of error. If the error that the recommended technology opportunities are not adequate for a firm occurs in a real-life environment, it can threaten the firm's survival and growth in a hyper-competitive business ecosystem. Moreover, because these studies typically focus only on patent information in a particular field rather than consider the location of the individual firm in depth, the recommended technology opportunity may not be appropriate, depending on the firm's state (Seo et al., 2016). Again, technology intelligence, which is expressed by the recommended technology opportunities, still involves uncertainty for some firms such as SMEs because these studies guarantee that the technologies are emerging or have potential in a specific technology field, not that the individual firm can perform them successfully in its real-life business environment.

Accordingly, a series of studies that consider a firm's current technology level have been conducted to reduce the uncertainty of proposed technology opportunities in real-life business environments. Specifically, by discovering new technology opportunities based on a firm's existing technology capability, these studies can increase a firm's R&D practicality and significantly reduce the risk on R&D investment (Seo et al., 2016). As a typical case, Yoon et al. attempted to structure information on products, technologies, and functions from a number of patents and identify potential technology opportunities over heterogeneous fields from existing technologies or products (Yoon et al., 2015). This study has made a considerable contribution in that it can improve a firm's R&D practicality and performance and reduce the uncertainty of R&D investment risk by deriving technology opportunities based on current technology and products. Interestingly, the technology opportunities proposed by this study are not restricted to a certain technical field, because they are derived based on functional similarities among technologies or products. Except for defining functions for existing technology and products using SAO analysis and interpreting the proposed technology opportunities, this approach can be an automated system for TOD. Similarly, Seo et al. defined a firm's existing product portfolios as its internal technology capability and identified product opportunity using text and association rule mining (Seo et al., 2016). Their research has the following distinguishing features. Particularly, it proposes only products associated with internal capability as real potential product opportunities for an individual firm, thus being able to reduce the uncertainty of R&D practicality and investment risk significantly. Unfortunately, despite its contribution, this study requires further work, in that it defined a firm's internal capability in terms of only products held by the firm, meaning the proposed opportunities can be rather narrow. Recently, Park and Yoon proposed an improved methodology to determine applicable technology opportunities customized for a target firm using patent classification code and collaborative filtering (Park and Yoon, 2017). First, since this approach defines a firm's technology capacity at specific main group-level IPCs and discovers technology opportunities based on capacity, it can provide firms with a more exact and elaborate interpretation of technology opportunities than prior studies that were conducted at the only product level. Second, this approach can grasp the competitive positions of individual firms by comparing IPC preference scores and can recommend heterogeneous technology opportunities by filtering applicable technology fields from similar firms' technology fields. Additionally, since this approach uses only collaborative filtering and IPC information with less preprocessing, it can be developed into an automated system, which is useful especially for SMEs in the real-life business environments.

Interestingly, these approaches differ in the way they define the technology capacities of individual firms and identify technology opportunities. However, they show less uncertainty regarding R&D investments by trying to determine technology opportunities based on the technology capacities of individual firms. Therefore, it is natural that individual firms prefer these approaches to the prior systems anticipating emerging technologies from the technology field perspective as real users of TOD in a real-life business environment. However, these studies define the technology capacity of a firm from a static perspective, although they recognize that the technology capacity or focus technology fields of a firm may change as a result of a rapidly changing technological environment. Indeed, since these approaches consider less temporal information, such as when similar firms involved the same technology opportunities by registering patents, the technology opportunities offered by these approaches can lag behind the recent trends of technology fields. Moreover, because these approaches often provide the target firm with rather inappropriate technologies, which are too heterogenous from existing technologies to engage in or that have been used in the past, there is the strong indication that the particular and continuous R&D of a firm can be diverted.

### 2.2. Sequential pattern mining

The basic concept of sequential pattern mining was first proposed by Agrawal and Srikant (Agrawal and Srikant, 1995) and developed by Srikant et al. (Srikant and Agrawal, 1996). Sequential pattern mining means to find a set of sequential patterns from a given sequence database, $D_s$, with a given support threshold, *min_sup*. $D_s$ consists of a set of a transactions, which contain the customer's ID and a sequence of elements. A sequence $S$ is a list of the element sets arranged in a time-series order. The sequence $S$ is denoted as $<(i_1), (i_2, i_3), \ldots, (i_n)>$, where an element is an alphabetically ordered set of items or events, $i_n$. Further, a sequence of $k$ elements is called a $k$-sequence or a length-k sequence. A set of sub sequences can be extracted from a sequence $S$, as the so-called Prefix. For example, $<(i_1), (i_2)>$ is the Prefix for $<(i_1), (i_2, i_3), (i_4), (i_5)>$ as $(i_1) \subseteq (i_1)$ and $(i_2) \subseteq (i_2, i_3)$. Next, a sequence database $D_s$ can be projected with a Prefix to mine the sequential pattern containing the Prefix, and the Prefix-based projection is called the Suffix. To project a sequence $<(i_1), (i_2, i_3), (i_4), (i_5)>$ based on a Prefix $<(i_1)>$, the Suffix will be $<(i_2, i_3), (i_4), (i_5)>$. The support for a sequence $S$ is the number of transactions that contained the sequence $S$ in $D_s$. If the support for sequence $S$ is over min_sup, then a positive integer value, $S$, is called a frequent sequential pattern. To determine a completed set of frequent subsequences satisfying the minimum support threshold, several algorithms, including generalized sequential pattern (Srikant and Agrawal, 1996), sequential pattern discovery using equivalence class (Zaki, 2001), frequent pattern-projected sequential pattern mining (Han et al., 2000), and Prefix-projected sequential pattern mining (PrefixSpan) (Han et al., 2001), were proposed, and the basic structure of the algorithm is as follows (Fig. 1).

Basic definitions, including Prefix, Suffix, Projected database, and Support, are borrowed from sequential pattern mining and the PrefixSpan algorithm is applied to determine a set of frequent sequential patterns from PEs' sequence database. The procedure of the PrefixSpan algorithm, shown in Algorithm 1, is as follows. First, we scan the sequence database to identify all frequent items, which are length-1 sequential patterns satisfying the given *min_sup*. Second, we project the database according to each frequent item and divide into different subsets. Based on these projected databases, we continue to find the frequent length-1 sequential patterns with the same corresponding prefix, referring to the frequent length-1 sequential patterns. We then repeat identifying the frequent length-(k + 1) sequential patterns from a projected database with every length-k sequential pattern.

**Algorithm 1.** PrefixSpan algorithm (Aloysius and Binu, 2013)

**Input**: A sequence database, $D_s$, and the minimum support threshold, *min_sup*
**Output**: The complete set of frequent sequential patterns
**Parameters**: α: a sequential pattern; l: the length of α; $D_s|\alpha$: the α-projected database, if $\alpha \neq <>$; otherwise, the sequence database, $D_s$
**Method**: PrefixSpan
1. Scan $D_s|\alpha$ once, find the set of frequent items $I_f$ so that
a) $I_f$ can be assembled to the last element of α to form a sequential pattern; or
b) $<I_f>$ can be appended to α to form a sequential pattern.
2. For each frequent item $I_f$, append it to α to form a sequential pattern $\alpha'$ and output $\alpha'$.
3. For each $\alpha'$, construct $\alpha'$-projected database $D_p|\alpha'$, and call PrefixSpan $(\alpha', l+1, D_s|\alpha')$.

## 3. Methodology

A novel approach is proposed, which can identify a set of technology candidates with more stability and less uncertainty in a real-life business environment based on a firm's dynamic change of the focus technology field. Specifically, the proposed approach expresses each firm's transitional change of the focus technology fields as a sequence, discovers significant sequences for the target firm using sequential pattern mining, and then recommends the best technology candidates

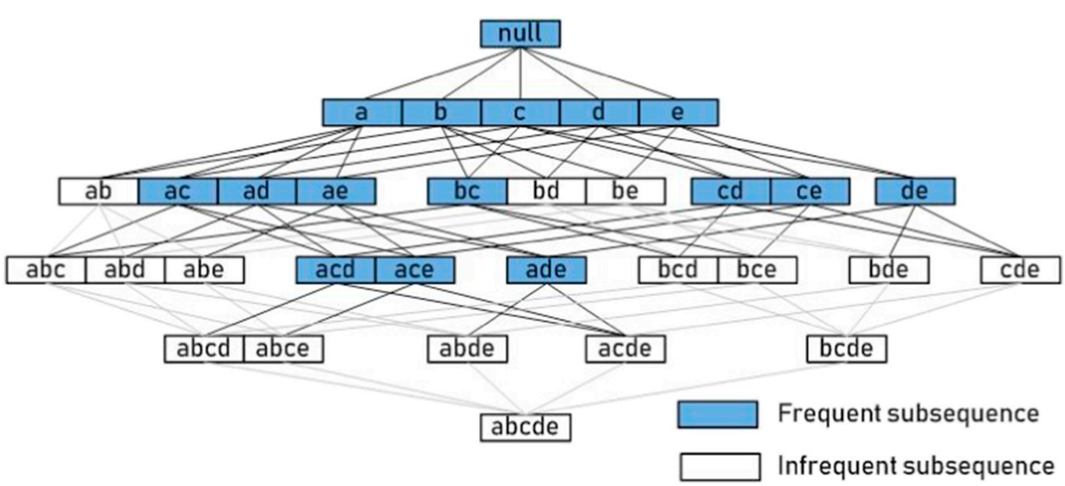


Fig. 1. Basic structure of sequential pattern mining.

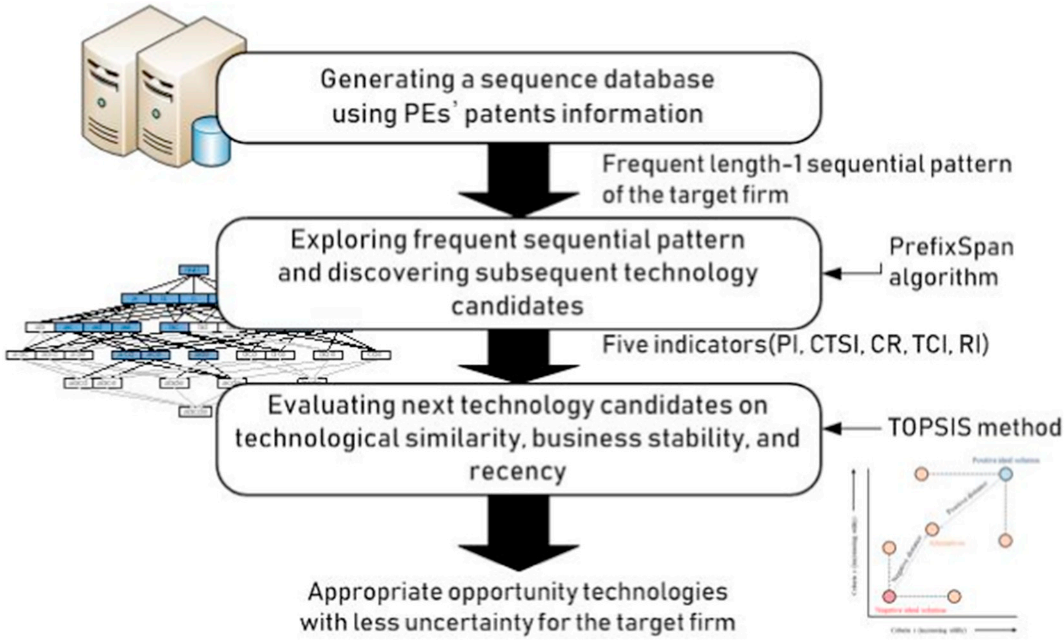


**Fig. 2.** Overall process of the proposed methodology.

using the TOPSIS method. The proposed approach consists of three steps: 1) generating a sequence database using PE patents information; 2) exploring frequent sequential patterns and discovering technological candidates from the database; 3) evaluating technological candidates based on technological similarity; business stability and recency (Fig. 2).

### 3.1. Generating a sequence database using PE patents information

The first step defines SMEs and PEs. The latter SMEs can benchmark in establishing R&D strategies and discovering technology opportunities based on their technology capacities. Here, the proposed approach measures a firm's technology capacity using the number of registered patents. The proposed approach expresses a patent as a list of patent IPC main groups that is ordered alphabetically and screened for redundancy. For example, a patent can be denoted as a list (G02F 1, H01L 27, H01L 29, H01L 31) or a set of IPCs that the patent belongs to. The generated list can be used an element of the firm's technological sequence and interpreted as one technology field in use. Additionally, the list can reflect the convergence or heterogeneity of specific technology fields. Next, to clearly grasp the time-varying focus on the field of technology of each firm, the focus technology fields are selected by year based on the concentration ratio (CR) of each element, CR is calculated as:

$$\mathrm{CR}(e_i) = \frac{\textit{the number of patents registered belonging to } e_i}{\textit{the number of patents registered in year of } e_i}.$$

Next, the elements with a ratio greater than 0.5 are selected as cases where the applicant has focused relatively heavily on that technology field in that year (e.g., if a firm $a$ registered 10 patents in 2007 and seven of the patents belong to (G02F 1, H01L 27, H01L 29, H01L 31), the CR of the field will be 0.7). Then, this step generates a set of sequences for a firm by connecting the selected focus technology field by year. As such, a firm can have multiple sequences because there can be two elements with CRs of 0.5 in a certain year.

### 3.2. Exploring frequent sequential patterns and discovering technology candidates from the database

As SMEs can often face challenges in establishing next technology strategies due to relatively small liabilities and limited resource in real-life business environments (Hlady Rispal and Servantie, 2017), PEs with a similar technological change can provide practical examples. Therefore, this step identifies PEs that have experienced a change in focus technology fields similar with the change of the target firm, and determines the technology fields that PEs have chosen after the change. For the PE sequence database constructed in the previous step, the most frequent length-1 sequential pattern of the target firm is used as a Prefix by applying the PrefixSpan algorithm. Here, the Prefix can be set according to the purpose or viewpoint of the target firm. A set of sequences longer than a length-2 sequence and each support value of them can be obtained. Next, this step projects the sequence set of PEs with the second Prefix, which is the second element in the sequential pattern of the target firm. Finally, two results are derived: 1) PEs similar with the target firm and 2) the PEs' selected sequential patterns longer than length-2. From a set of selected sequences of PEs, this step also searches third or subsequent elements of the sequences and defines the elements as a set of technology candidates.

**Table 1**
Indicators for candidate evaluation.

| Evaluation perspective | Indicators | Comparison object |
|---|---|---|
| Technological similarity between the target firm and PE | Parallax index (PI) | The *i*th elements of the target firm and PE |
| | Focus technology similarity index (FTSI) | The *i*th elements of the target firm and PE |
| Business stability in a real-life business environment | Concentration ratio (CR) | A subsequent technology candidate |
| | Technology capability index (TCI) | The *i*th elements of PE |
| Recency from the evaluation point | Recency index (RI) | A subsequent technology candidate |

### 3.3. Evaluating technology candidates on technological similarity, business stability, and recency

The third step calculates five indicators to evaluate the set of technology candidates and derives a recommendation index for each technology candidate. That is, the technology candidates are evaluated on technological similarity, business stability, and recency, which are intertwined with uncertainty and thus expected to facilitate the technical decision making of the target firm. The indicators listed in Table 1 are calculated by comparing the PE sequence ($S^{PE} = < e_1^{PE}, e_2^{PE}, e_3^{PE} >$) with the target firm's sequence ($S^{target} = < e_1^{target}, e_2^{target}, e_3^{target} >$).

First, the parallax index (PI) reflects the technological similarity in terms of time. The PI is derived by calculating the difference in the registration year of patents that belong to $e_i^{PE}$ and $e_i^{target}$ for each corresponding element. In the previous step, the proposed approach regards registering many patents in a specific field of technology as focusing on that field. Therefore, the gap of registration year between the PEs' patents and the target firm's patents for each corresponding element, such as $e_i^{PE}$ and $e_i^{target}$, is interpreted as the temporal gap of the focus time on each technology field. Specifically, the PI is defined as:

$$\mathrm{PI}(e_i) = \text{the registration year difference between patents belonging to } e_i^{PE} \text{ and } e_i^{target}.$$

Using the PI, the relative technology development level of the target firm in comparison with PEs' level can be identified. In other words, if the PI is large in $e_1$, there would be a gap in the point when the PE and the target firm each had focused on technology field $e_1$, which means that the target firm is technically lagging behind the PEs by that difference (Fagerberg, 1987). That is, the larger the PI, the more incorrect the recommendation, given that the chronological continuity of the elements in the sequential pattern can be cut off by a large temporal interruption. In short, the recommendation derived from the antiquated changing pattern from $e_1^{PE}$ to $e_2^{PE}$ cannot be ensured because the temporal gap with the target firm is too large and it cannot be guaranteed that the recommended technology candidate also preserves technological continuity (Guo et al., 2013). The index is an important indicator in considering the continuity of technological development, which is influential for the sustainable growth and development of firms, but is often overlooked by the prior TOD systems based on the technology capability (Park and Yoon, 2017). Since the proposed approach expresses the firms' technology capabilities from a dynamic perspective and considers detailed time information, the temporal changes in the firms' capability can be reflected and the continuity of technological development can be preserved.

The focus technology similarity index (FTSI) of $e_i$ measures the technological similarity between $e_i^{PE}$ and $e_i^{target}$ from an IPC code perspective, along with PI in terms of time. To compute the FTSI of $e_i$, the IPC information of patents belonging to $e_i^{RE}$ and $e_i^{target}$ is used. First, the classification similarity between two IPCs i and j ($CS_{ij}$) is defined as follows, in line with prior studies (Kyebambe et al., 2017; Nakamura et al., 2015):

$$CS_{ij} = \begin{cases} 0 \ \mathit{if} \ i \ \mathit{and} \ j \ \mathit{differ\ in\ section} \\ 0.25 \ \mathit{if} \ i \ \mathit{and} \ j \ \mathit{differ\ in\ section\ but\ not\ class} \\ 0.5 \ \mathit{if} \ i \ \mathit{and} \ j \ \mathit{differ\ in\ class\ but\ not\ subclass} \\ 0.75 \ \mathit{if} \ i \ \mathit{and} \ j \ \mathit{differ\ in\ subclass\ but\ not\ main\ group} \\ 1 \ \mathit{if} \ i \ \mathit{and} \ j \ \mathit{share\ the\ same\ subclass\ and\ main\ group} \end{cases}.$$

Subsequently, the FTSI of $e_i$ is calculated, based on the classification similarity between $e_i^{RE}$ and $e_i^{target}$, a list of patent IPC main groups. Here, the IPC main group of patents is used as an item of $e_i$, where a patent may belong to one or more IPC main groups.

$$\mathrm{FTSI}(e_i) = \frac{\sum_{j=1}^{N_R} \sum_{k=1}^{N_T} CS_{jk}}{N_R N_T},$$

where $N_R$ and $N_T$ are the number of items of $e_i^{PE}$ and $e_i^{target}$, respectively. Therefore, a high FTSI($e_1$) means the target firm and the PE focus on similar technology fields, and a high average FTSI between the target firm and the PE means that the technical competencies of the target firm and the PE are similar at that time. Therefore, this implies that the target firm has the capacity to successfully perform the subsequent technology opportunity selected by the PE. Further, a technology candidate with a high FTSI can drive the target firm toward sustainable growth and development.

The CR of the technology candidate is a measure of the proportion of patents belonging to the candidate to the total patents registered in that year by the candidate, the formula being already defined in the first phase. Specifically, a technology candidate with a high CR value means that the PE registered a relatively large number of patents in the technology field, similar with the candidate (Park and Yoon, 2017; Yoon et al., 2015). Additionally, the candidate has been validated through the case of the PE in a real-life business environment, its business stability being guaranteed for the target firm (Lee et al., 2017b). Similarly, the uncertainty of the candidate is low enough that the target firm is more likely to develop and grow sustainably with the candidate. Therefore, the CR of each technology candidate is used as an indicator of business stability in a real-life business environment.

To evaluate the business stability in terms of the PEs' technology capability, the technology capability index (TCI) is used. In the situation where a firm's focus for a given field of technology changes from $e_1$ to $e_2$, to the technology candidate, the TCI of $e_i$ is a quantitative indicator of how much technology capacity the firm has for the $i$th element. Specifically, the proposed approach uses the number of patents that had been registered in the field $e_i$ by the year of $e_i$, for the TCI of $e_i$. Therefore, an element with high TCI value denotes that the firm had high technology capability in the technology field $e_i$ by registering numerous patents (Yang et al., 2014; Yoon and Magee, 2018). Consequently, a technology candidate with a high TCI average value can have the empirical grounds to change to the candidate as per the example of the PE and guarantee the stability of the candidate because TCI refers to the degree to which the target firm can trust the case of the PE. As such, the TCI of each technology candidate is also used for measuring the business stability of the target firm's business environment. A high TCI and CR guarantee that the candidate would be appropriate to help the target firm grow sustainably and survive in an intense business environment.

Finally, the recency index (RI) is a measure of how up to date a technology candidate is. RI is calculated as the difference between the time when a set of patents of the candidate were registered by the PE and the evaluation time. RI quantifies the extent to which the candidate is lagging behind recent technology trends. Since most prior methodologies based on firms' technology capabilities explore and recommend technology opportunities from a static perspective, their results may make the target firm miss the technology opportunities that reflects a recent trend of utilized technology fields (Park and Yoon, 2017). As a result, the target firm can lag behind the latest trends. In most fields of technology, SMEs, including the target firm, are followers of recent trends and attempt to catch up, while PEs are usually fast followers or first movers for these trends (North et al., 2001; Savioz and Blum, 2002). Therefore, it is necessary for SMEs to discover technology candidates that deviate less from recent trends by using the RI.

After extracting five indicators from the sequences of PEs and the target firm, the proposed approach applies the technique for order performance by similarity to ideal solution (TOPSIS) to the indicators to derive the recommendation index. TOPSIS, first proposed by Hwang and Yoon (Hwang and Yoon, 1981), is a typical multi criteria decision making method. The main concept of TOPSIS is how to choose the alternative (same as our candidate) with the shortest distance from the positive ideal solution (PIS) and the farthest distance from the negative ideal solution (NIS) (Chen, 2000). TOPSIS generates a matrix that consists of the indicators and the candidate. Next, vector normalization is applied to the matrix and the same weights are given to the indicators. Here, the weights can be adjusted by the target firm according to its purpose or philosophy (Krohling and Pacheco, 2015). From the weighted matrix, PIS and NIS are identified for each indicator. The fact that PI and RI are the indicators with a negative impact and the remaining indicators have a positive impact should be noted. For each candidate, the distances from both PIS and NIS are calculated and the relative closeness to PIS, $\xi_i$, is defined as:

$$\xi_i = \frac{d_i^-}{d_i^+ + d_i^-}.$$

The proposed approach ranks the candidates by $\xi_i$ value and identifies the best candidates as the ones with high $\xi_i$ values. A high $\xi_i$ value of a candidate means it is close to PIS. Therefore, the best candidates have high FTSI, CR, and TCI and low PI and RI. The best candidates imply that their technology fields have less uncertainty for the target firm regarding sustainable technological development and growth. The reason for that can be described as follows. First, because the best candidates with low PI and high FTSI are derived from the PEs that are technologically similar with the target firm, there is a good chance that the target firm performs successfully, while maintaining a continuous technological development direction. Second, since the best candidates with high CR and TCI are validated by PEs, the target firm can implement the candidate technologies successfully in a real-life business environment. Finally, the best candidates with high RI are less risky for the target firm to develop as related technologies, because these candidates reflect the latest trends of the related technology fields, meaning the target firm can keep up with recent trends by utilizing these candidates in R&D activities.

## 4. Case study

### 4.1. Data

Firm X, an SME, is selected for conducting a case study to determine how effective the proposed approach is for SMEs in reality. Before applying the proposed approach, data on the patents registered at the Korean Intellectual Property Office (KIPO) database from 2000 to 2016 (1,678,751) are collected by using the KIPRIS plus open API service. The relationship between applicants and their patents using applicants' unique IDs is established. According to the aforementioned definition, SMEs with between 20 and 50 patents and PEs with 50–100 patents are selected from all applicants. Here, such categorization was the results of referring to some previous studies dealing with firms using patent information (Lee et al., 2017a; Park and Yoon, 2017). Also, the cut-off values were determined at the end of a thorough discussion with patent experts. In addition, since the proposed approach is customized for technology-based firms with continuous technological development directions, the applicants who hold more than 100 patents are regarded as conglomerates trying to develop their technologies in various directions and excluded from the analysis. In addition, applicants such as the research institute of universities and governments are also removed from the sample to accurately determine technology opportunities for firms. Finally, 933 and 3681 applicants are selected and classified as PEs and SMEs, respectively (Table 2).

Next, the PE sequences are generated by connecting their focus technology fields (CR of at least 0.5) by year to represent the temporal change of focus technology fields. The initial sequences of PEs are 3889 whose minimum length is 1 and the maximum is 16. Here, since short sequences do not have a significant impact on finding frequent sequential patterns, a separate screening task was not carried out. Table 3 shows some of the initial sequence set of the PEs. Indeed, we stored the registration year, the number of patents, the CR value per each sequence's element, and the sequence length in the sequence database.

### 4.2. Target firm's technology sequence

Firm X is selected for the case study to show how effective the proposed approach is and how practical the results are in a real-life business environment. As previously mentioned, firm X is an SME, with relatively scarce resources in terms of experts or facilities and thus competitively lagging. Specifically, firm X is engaged in the semiconductor, flat panel display, and IT convergence fields with 34 registered patents in the KIPO database. Table 4 shows details on its main patents with relatively high CR values, such as registration year, IPCs, title of patent, and CR values. Firm X tried to supply display, semiconductor products such as active-matrix organic light-emitting diode display, front open unified pod, chemical delivery supply system by the registered 34 patents, including "Medicinal fluid mixing feeder," "3-dimensional image device," "LCD line real time monitoring system," "Encoding apparatus," and "Dispenser type nozzle device." Firm X has registered most patents on the IPC (H01L-021) "Process or apparatus specially adapted for the manufacture or treatment of semiconductor or solid-state devices or of parts thereof" according to World Intellectual Property Organization, such as "Method for working wet station equipment," "Controlling device of gas temperature," "Door of vertical tube cleaner having drain guide," "Carrousel apparatus for vertical tube cleaner," and "Apparatus for recycling fluid for cleaning semiconductor substrate." Interestingly, in terms of each IPC's CR value, firm X has mainly engaged in various IPCs such as (H01L-021), (B01F-015), (B05B-001, B05B-011, H01L-051), (G02B-005, G02F-001), and (H01L-021, H01L-051). Therefore, firm X's focus technology fields have been gradually changing over time.

Therefore, based on firm X's main patents, a set of sequences are generated to articulate the temporal change of its focus technology fields. For example, $<$ (H01L-021), (B01F-015) $>$ means that firm X's focus technology fields had changed from (H01L-021) to (B01F-015), which is about "Accessories for mixers," by registering relatively more patents such as the "Medicinal fluid mixing feeder" in 2007. As another example, $<$ (H01L-021), (G02F-001) $>$ means that firm X had originally focused on (H01L-021) but changed to IPCs related to (G02F-001), by registering relatively many patents, such as the "Apparatus and method for bonding substrate" in 2014. The IPC (G02F-001) represents "Devices or arrangements for the control of the intensity, color, phase, polarization or direction of light arriving from an independent light source, e.g. switching, gating, or modulating; Non-linear optics."

### 4.3. Results

For identifying the technology opportunities of firm X, the proposed approach is as follows: 1) explore PEs' frequent patterns that are similar with firm X's sequential patterns; and 2) discover technology candidates based on these patterns and evaluate the candidates using five

**Table 2**
Selection of applicants.

| Type | Description | Number of applicants |
|---|---|---|
| PEs | A set of firms that have high-level technology capability, and so can be subject to benchmarking | 933 |
| SMEs | A set of firms that have scarce information and resource for discovering technology opportunities | 3681 |

**Table 3**
Some of the initial sequences.

| Sequence id | Initial sequence | Length |
|---|---|---|
| 119980001162-1 | < (B60K-020), (B21K-021, C09D-101), (C09D-127, C09D-007), (B62D-005, B60K-020), (B60K-020), (B60K-020), (B60T-007), (B60K-020, F16H-059) > | 9 |
| 119980001162-2 | < (B60K-020), (B21K-021), (F16H-059), (B62D-005), (B60K-020), (B60K-020), (B60K-020), (B60T-007), (B60K-020, F16H-059) > | 9 |
| 119980004196-3 | < (H01H-009), (H02B-013), (H01H-033), (H01H-009), (H01H-033), (F16K-003, F16K-051, H01L-021), (G21F-009), (H01H-033), (H01H-067, H01H-009), (H01H-033) > | 10 |
| 119980004196-8 | < (H01H-009), (H02B-013), (H01H-009), (H01H-009), (H01H-033), (H01L-021), (G21F-009), (H01H-033), (H01H-067, H01H-009), (H01H-033) > | 10 |
| 119980009236-2 | < (C11C-003), (C07F-009), (A23L-001, A61K-008), (A61K-008, A61Q-019), (C07F-009, C08K-005), (E02F-009-, F15B-013, F15B-020), (A61K-008, A61Q-019) > | 7 |

**Table 4**
Firm X's main patents.

| Registration year | IPCs | Title of patent | CR |
|---|---|---|---|
| 2006 | (H01L-021) | Method for working wet station equipment | 1 |
| 2007 | (B01F-015) | Medicinal fluid mixing feeder | 1 |
| 2008 | (H01L-021) | Controlling device of gas temperature; Door of vertical tube cleaner having drain guide; Carrousel apparatus for vertical tube cleaner | 0.75 |
| 2014 | (B05B-001, B05B-011, H01L-051) | Dispenser type nozzle device | 0.5 |
| 2014 | (G02B-005, G02F-001) | Apparatus and method for bonding substrate | 0.5 |
| 2015 | (H01L-021) | Apparatus for recycling fluid for cleaning semiconductor substrate; Mask jig assembly | 0.5 |
| 2016 | (H01L-021, H01L-051) | Apparatus and method for cleaning mask of organic light emitting diode | 1 |

**Table 5**
PE sequential patterns.

| PEs | Number of patents | Sequential pattern |
|---|---|---|
| PE 1 | 56 | < (H01L-021), (G02F-001) > |
| PE 2 | 91 | < (H01L-021), (G01R-001, G02F-001) > |
| PE 3 | 53 | < (H01L-021), (B65G-049, G02F-001, H01L-021) > |
| PE 4 | 97 | < (H01L-021), (G02F-001, H01L-021) > |
| PE 5 | 65 | < (B82Y-040, H01L-021), (B82Y-040, G02F-001, G06F-003) > |
| PE 6 | 89 | < (H01L-021), (G02F-001) |
| PE 7 | 51 | < (G02F-001, H01L-021), (G02F-001) > |
| PE 8 | 69 | < (G03F-007, H01L-021), (G02B-005, G02F-001) > |
| PE 9 | 97 | < (H01L-021), (G02F-001) > |

indicators on technological similarity, business stability, and recency to assist with decision making in a real-life business environment.

First, we explored the PEs' frequent sequential patterns from the initial set of 3889 sequences using PrefixSpan algorithm. As a result, 4,342,048 sequential patterns were identified, but only 1,202,647 sequential patterns with 4 or more support values and 3 or more lengths are used as a valid sequential pattern set. The cut-off value of support was consulted with several patent analysis-related researchers. In addition, the length of sequential patterns was set to 3 to find next technology candidates. Next, comparing the firm X's sequential patterns with the valid frequent sequential patterns, < (H01L-021), (G02F-001) > was identified to be also frequent in the PEs' sequential patterns. Indeed, other sequential patterns were excluded from the next analysis, since they were not included in the valid sequential patterns or were of low frequency. Finally, 9 PEs were identified to have undergone technological changes similar with the firm X's one. Table 5 represents the PEs and their sequential patterns. Among the 9 PEs, PE 9 with 97 patents also experienced the similar change of < (H01L-021), (G02F-001) > over two years. PE 9 focused on the (H01L-021) field, related to the manufacture or treatment of semiconductor or solid-state devices, by registering patents such as "Apparatus for picking up semiconductor chips and method of picking up semiconductor chips," "The identifier imprinting apparatus and imprinting method of the semiconductor wafer," and "Processor and processing method, and robot device." Interestingly, PE 9 changed its focus technology field in 2007 to the (G02F-001) field, by registering a relatively large number of patents, including "Substrate laminating apparatus and method thereof and substrate detecting apparatus" and "Part mounter." Here, it can be expected that PE 9 may be used to identify the future technology opportunities of firm X. As another example, PE 2 had undergone changes in its focus technology fields such as < (H01L-021), (G01R-001, G02F-001) > , similarly to firm X. Similarly, PE 2 focused on the (H01L-021) field, by registering patents such as "Probe units for inspecting panel type inspected objects," but its focus technology field changed to the (G01R-001, G02F-001) field of "Devices or arrangements for the control of the intensity, color, phase, polarization or direction of light arriving from an independent light source" and "Details of instruments or arrangements of the types covered by groups" (G01R-005, G01R-013) or (G01R-031). Because PE 2 had undergone similar changes as firm X, its focus technology fields after the change are expected to give firm X quality insights into discovering technology opportunities with less uncertainty.

In the second stage, a set of IPCs on which nine PEs had focused on after the change are identified. Each technology field that the PEs concentrated on after the change, including < (H01L-021), (G02F-001) > , is evaluated as a technology opportunity based on technological similarity, business similarity, and recency. Here, the values in the same IPC can vary depending on which firm focused on the IPC and when. For example, if PE 3 and PE 5 had focused on IPC (G06F-003) in 2004 and 2006, respectively, including < (H01L-021), (G02F-001) > after the change, IPC (G06F-003) can be evaluated differently for each PE. This is because of the technological and environmental situations under which PE 3 and PE 5 experienced the change of focus technology fields, which are evaluated using the proposed indicators. Considering FTSI as an example, which is calculated by comparing the PE's sequential pattern with that of firm X's, PE 3's sequential pattern of < (H01L-021), (B65G-049, G02F-001, H01L-021) > and PE 5's sequential pattern of < (B82Y-040, H01L-021), (B82Y-040, G02F-001, G06F-003) > are somewhat different, meaning the values of their FTSI indexes vary. After the evaluation, the TOPSIS method is applied to the evaluation results, and Table 6 shows part of the normalized matrix.

Based on the normalized matrix, the best candidate and the worst

**Table 6**
Normalized matrix of technology opportunity candidates.

| Technology opportunity candidates | PEs | CR | PI ($e_1$) | PI ($e_2$) | RI | TCI ($e_1$) | TCI ($e_2$) | FTSI ($e_1$) | FTSI ($e_2$) |
|---|---|---|---|---|---|---|---|---|---|
| (B08B-001, B08B-003) | PE 9 | 0.222 | 0.072 | 0.122 | 0.143 | 0.224 | 0.438 | 0.171 | 0.195 |
| (H01L-021) | PE 4 | 0.148 | 0.048 | 0.122 | 0.163 | 0.463 | 0.088 | 0.171 | 0.097 |
| (B60L-011, H02J-007) | PE 2 | 0.222 | 0.048 | 0.163 | 0.041 | 0.122 | 0.336 | 0.171 | 0.123 |
| (G09F-009, H01L-051) | PE 6 | 0.111 | 0.048 | 0.122 | 0.123 | 0.106 | 0.190 | 0.171 | 0.195 |
| (H01L-021) | PE 6 | 0.111 | 0.048 | 0.122 | 0.123 | 0.106 | 0.190 | 0.171 | 0.195 |
| (G06F-003) | PE 3 | 0.022 | 0.048 | 0.122 | 0.020 | 0.075 | 0.146 | 0.171 | 0.064 |
| (G02F-001) | PE 7 | 0.166 | 0.144 | 0.082 | 0.123 | 0.021 | 0.277 | 0.085 | 0.195 |
| (H01L-031) | PE 6 | 0.117 | 0.096 | 0.122 | 0.123 | 0.096 | 0.190 | 0.171 | 0.195 |
| (B65G-049, G01N-021, G02F-001, H01L-021) | PE 1 | 0.111 | 0.048 | 0.204 | 0.204 | 0.096 | 0.204 | 0.171 | 0.195 |
| (G02F-001) | PE 1 | 0.111 | 0.048 | 0.204 | 0.204 | 0.096 | 0.204 | 0.171 | 0.195 |

candidate for each indicator are identified. Here, RI, and PI are cost criteria, while TCI, CR, and FTSI are benefit criteria. Next, the distances from the best and the worst candidates for each technology opportunity candidate are respectively calculated. As the last step of TOPSIS, the final recommendation index, called similarity in TOPSIS, is derived and the set of technology opportunity candidates are ranked based on this index. As a result, the proposed approach provides a ranking of technology opportunity candidates, but the ranking may vary depending on the weight value of the indicators. Table 7 shows the high ranked technology opportunity candidates, and a numerical analysis of their indicator values is conducted to detail the interpretation of the ranking.

A candidate of (B08B-001, B08B-003) from PE 9 is ranked first, which is closest to the best candidate and farthest from the worst candidate, as shown as Fig. 3. This is because the candidate had overwhelmingly high CR and TCI values. Its high CR value of 1.00 indicates that PE 9 made relatively more technological investment in the IPCs of (B08B-001, B08B-003) by registering many patents after the change of its focus technology field, < (H01L-021), (G02F-001) > . In terms of TCI, it implies that PE 9's focus was changed to the IPCs of (B08B-001, B08B-003), although the PE had had a high level of technology capacity in (G02F-001)-related fields. High CR and TCI values mean that the business stability of the candidate in a real-life business environment is assured in PE 9's case. Overall, the candidate (B08B-001, B08B-003) from PE 9 has high values for the business stability and technological similarity indicators, although its recency is relatively low. This means that the uncertainty of the candidate is relatively low in terms of firm X's technological environment. Next, a candidate of (B60L-011, H02J-007) from PE 3 is rated highly, with a recommendation index of 0.568. The recency indicator value is evidently high among the candidate's indicators, because the candidate (B60L-011, H02J-007) is the most recently selected technology field by PE 2. Therefore, the candidate is most likely to reflect recent trend in related technology fields.

While the candidate (B60L-011, H02J-007) is derived from PE 2 with low technological similarity, the recency and business stability indicators are rather high, meaning candidate (B60L-011, H02J-007) can be an adequate technology opportunity for firm X to advance into new technology fields. In case of candidate (G06F-003) from PE 3, the recency is noticeably high with a value of 2.00, meaning the candidate can help firm X catch up to PEs in the position of fast follower in the technology competition. Meanwhile, the candidate's business stability is somewhat lacking for firm X, since PE 3's technology capacity in IPCs (H01L-021), (G02F-001) is lower than other PEs' and the other indicators have average levels. Therefore, candidate (G06F-003) would be adequate to reflect a trend of related technology fields and to catch up to the technical level of PEs. Candidate (B65G-049, G01 N-021, G02F-001, H01L-021) from PE 1 is also high evaluated in terms of the RI and FTSI indicators. The high RI values imply that PE 1 concentrated on IPCs (B65G-049, G01 N-021, G02F-001, H01L-021) in 2016, relatively more recently than the other PEs. Therefore, the candidate is expected to be up to date in related technology fields. For FTSI, firm X and PE 1 experienced a similar change in focus technology fields, although at different times. Candidate (B65G-049, G01 N-021, G02F-001, H01L-021) is thus applicable to firm X's existing technology capacity. On the other hand, PE 6's candidates, such as (G09F-009, H01L-051), (H01L-031), have a high FTSI because PE 6's focus technology fields changed from IPC (H01L-021) to (G02F-001), while firm X's focus technology fields changed from IPC (H01L-021) to (G02B-005, G02F-001). Therefore, since PE 6 had undergone the most similar change with firm X in terms of IPC, PE 6's candidates can be good options for firm X.

Most prior studies tend to emphasize new or heterogenous technology fields as the only technology opportunities, but this is not reliable since the target firm or SMEs with less information and resources can misunderstand such results. On the other hand, around 45% of the candidates belong to homogenous technology fields such as (H01L-021), (G02F-001), (H01L-021, G02F-001), while the rest include heterogenous technology fields. The proposed approach also evaluates the candidates belonging to the former, rather than only the latter, because the candidates of homogenous technology fields are also selected by PEs after the change in focus technology fields. Again, the technology fields including existing technology capacity cannot be new technology opportunities, but such homogeneous technology fields are also

**Table 7**
Top ranked technology opportunity candidates.

| Technology opportunity candidates | PEs | The distance to the best candidate | The distance to the worst candidate | Recommendation index ($\xi_i$) | Rank |
|---|---|---|---|---|---|
| (B08B-001, B08B-003) | PE 9 | 0.273 | 0.579 | 0.679 | 1 |
| (H01L-021) | PE 4 | 0.400 | 0.549 | 0.579 | 2 |
| (B60L-011, H02J-007) | PE 2 | 0.373 | 0.490 | 0.568 | 3 |
| (G09F-009, H01L-051) | PE 6 | 0.462 | 0.393 | 0.460 | 4 |
| (H01L-021) | PE 6 | 0.462 | 0.393 | 0.460 | 4 |
| (G06F-003) | PE 3 | 0.505 | 0.398 | 0.441 | 6 |
| (G02F-001) | PE 7 | 0.501 | 0.388 | 0.436 | 7 |
| (H01L-031) | PE 6 | 0.471 | 0.360 | 0.430 | 8 |
| (B65G-049, G01 N-021, G02F-001, H01L-021) | PE 1 | 0.501 | 0.371 | 0.426 | 9 |
| (G02F-001) | PE 1 | 0.501 | 0.371 | 0.426 | 9 |

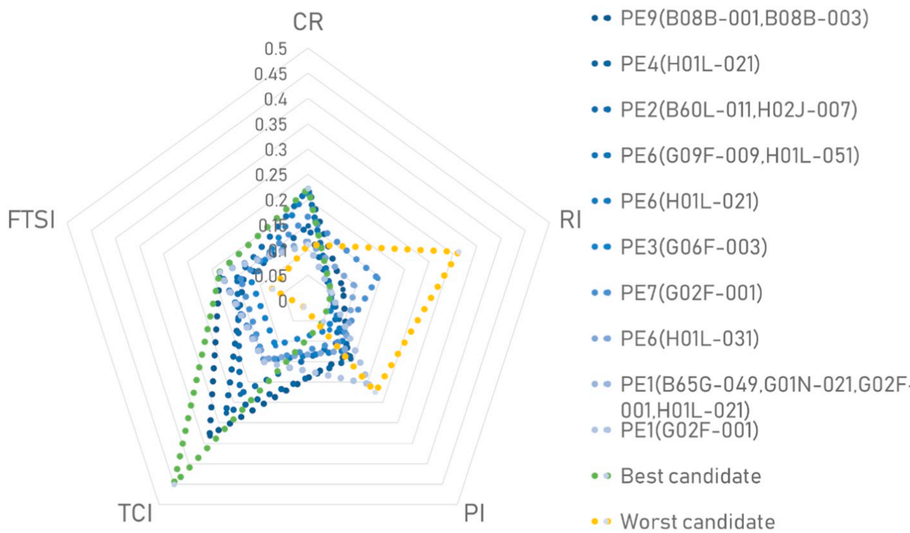


**Fig. 3.** Top ranked candidates' positioning graph.

considerable options for firm X in terms of R&D planning. Indeed, some PEs focused on the homogeneous technology fields and developed their existing technology capacities. In addition, Fig. 4 compares the candidates' average indicator values belonging to the homogenous and heterogenous technology fields. However, the values of homogenous and heterogenous fields candidates are similar, meaning both can be good options for the target firm as technology opportunities.

## 5. Discussion

This paper presents a different TOD approach, reflecting a temporal change of a target firm's focus technology fields to help the firm find appropriate technology opportunities. In explaining the results, the proposed approach provides a set of technology opportunity candidates and their evaluation scores on business stability, technological similarity, and recency. These results not only provide practical assistance to a firm exploring technology opportunities, but also help identify the PEs that underwent similar changes. Further, this section provides specific descriptions of technology opportunities in terms of related patents and application fields. Specifically, the proposed approach handled patents related with the proposed technology opportunities and consulted with patent experts. The experts pool consists of 3 current patent attorneys and 10 patent analysis researchers (3 professors and 6 graduate students). The patent attorneys have mainly focused on SME promotion and technology business consulting for about 12 years. The professors have tried to quantitatively analyze a large amount of patent data for discovering technology opportunities and detecting emerging technologies for about 10 years. Also, the graduate students have been involved in similar research for about four years. With the experts, we discussed on the similarity between the individual technology opportunities and the firm X's technology capacity, the feasibility and the reason why the PEs selected the technology opportunities, based on the actual patents and technologies. Therefore, the following description is expected to help the target firm obtain specific insights on technology opportunities.

There are many heterogeneous technology fields identified in the results. First, the technology opportunity (B08B-001, B08B-003) is related to "Cleaning by methods involving the use of tools, brushes, or analogous members" and "Cleaning by methods involving the use of presence of liquid or stream." There are many patents including "Cleaning apparatus and cleaning method using the same," "Roll wiper cleaning apparatus," "Graphite oxide cleaning device and cleaning method using the same," and "Apparatus cleaning metal member for manufacturing of muffler." Specifically, the technology opportunity can be developed using IPC (B08B-003/08): "The liquid having chemical or dissolving effect" and the IPCs of (B08B-001/04) about "Using rotary

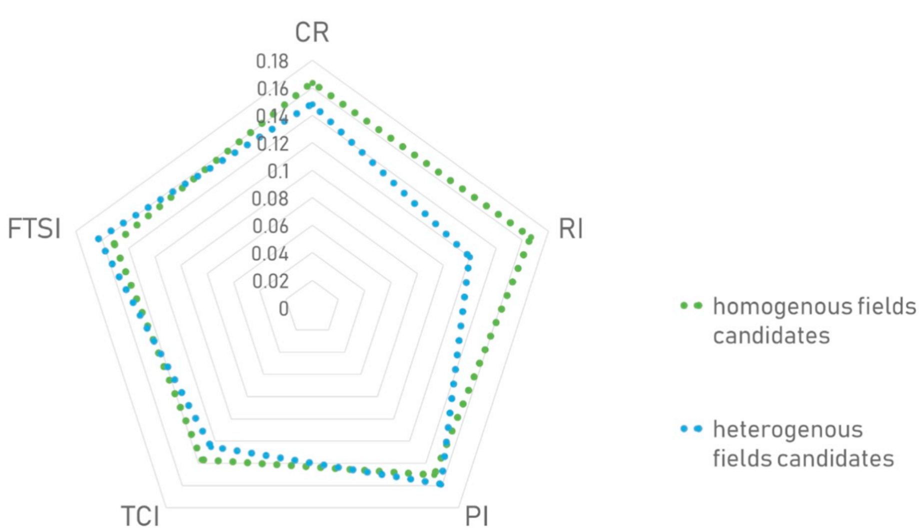


**Fig. 4.** Comparison of evaluation average values in heterogenous and homogenous fields.

operative members." In addition, according to the experts' suggestion on the technology opportunity, "Molding attaching device for vehicle glass and method of the same," "Method of transmitting and receiving broadcast signals, and broadcast reception device using said method," and "Washing apparatus for solar photovoltaic module" are proposed as application technologies. Next, IPC (B60L-011, H02J-007) is related to "Electric propulsion with power supplied within the vehicle" and "Circuit arrangements for charging or depolarizing batteries or for supplying loads from batteries." As an example, a patent on "Charging device, charging system and method of managing for charging" consists of IPC (B60L-011/18), that is, "Using power supplied from primary cells, secondary cells, or fuel cells," and IPC (H02J-007/00). Additionally, the experts devised specific technologies, such as "Directional sound source filtering apparatus using microphone array and controlling method thereof," "System and method for providing contents using history information of different types of terminals," "Apparatus and method for adjusting position of electric bus battery exchanger," as application technologies of the technology opportunity. In the case of (G09F-009, H01L-051), that is, "Indicating arrangements for variable information in which the information is built-up on a support by selection or combination of individual elements" and "Circuit arrangement for charging or depolarizing batteries or for supplying loads from batteries," the experts proposed a set of application technologies such as "Electric bus and electric bus battery exchanging system," "Variable message sign," "Message sign apparatus with waterproofing structure," and "Intelligent and versatile sign-board for showing information." On the other hand, IPCs (B65G-049, G01 N-021, G02F-001, H01L-021) are also evaluated highly and represent a combination of existing technologies and technologies related to "Conveying systems characterized by their application for specified purposes not otherwise provided for" and "Investigating or analyzing materials by the use of optical means." Based on the IPCs, the experts proposed "Moving body apparatus, exposure apparatus, exposure method, and device manufacturing method," "Stage drive method and stage drive apparatus, exposure apparatus, and device producing method," "Substrate holding device, exposure apparatus having same, exposure method, method for producing device, and liquid repellent plate," and "Pattern forming apparatus, mark detecting apparatus, exposure apparatus, pattern forming method, exposure method and device manufacturing method" as applicable technology fields.

Meanwhile, there are some technology fields similar with the existing technology fields of firm X, such as (H01L-021), (G02F-001). At first, IPC (G06F-003), that is, "Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit," is high evaluated. For this IPC, the experts proposed a set of application technologies such as "Touch panel and touchscreen apparatus including the same," "Touch sensing apparatus and data processing method thereof," "Apparatus and method for sensing capacitance, and touchscreen apparatus," and "Electronic pen, and system and the method for inputting using the same." On the other hand, IPC (H01L-031) is also highly evaluated and related to "Semiconductor devices sensitive to infra-red radiation, light, electromagnetic radiation of shorter wavelength, or corpuscular radiation and specially adapted either for the conversion of the energy of such radiation into electrical energy or for the control of electrical energy by such radiation; Processes or apparatus specially adapted for the manufacture or treatment thereof of parts thereof." From IPC (H01L-031), the experts devised "Solar generating apparatus using the solar light and solar heat," "Group-wise floating solar power system and method for tracking solar," "Solar cell module structure having snow removal unit and method for controlling the snow removal unit," and "Photovoltaic generation device comprising cleaning device" as application technology fields.

In order to show the feasibility of the proposed methodology, we selected the firm X, one of technology-based SMEs, for the above case study. Based on the firm X's patents, we identified the temporal change of focus technology which the firm had undergone, found the PEs that experience similar changes, and finally discovered and evaluated technology opportunities. This approach can be effective in that it can provide stable and guaranteed technology opportunities for SMEs who are in the position of followers in technology competition situations. Furthermore, based on sequence, which is a basic concept of the proposed methodology, it would be possible to represent the temporal change of focus technology in terms of country or industry and discover significant sequential patterns. However, technological changes in nations, industries, or conglomerate are expected to have diverse directions, and thus additional work would be required to refine valid sequences or sequential patterns. Therefore, the attempt to apply the proposed methodology from a broad perspective should be considered in further works.

## 6. Concluding remarks

In terms of technology, uncertainty in real-life business environments has been prevalent but remained unresolved in prior TOD studies. This uncertainty can be caused by the gap between the amount of information held and that needed to achieve sustainable growth and survival, also called technology intelligence. Although many prior studies tried to reduce this gap by deriving technology intelligence from technical documents such as patents, there are improvements that can be made to the way of identifying technology opportunities mainly because prior studies do not consider that the firm's focus technology fields can change over time in tandem with rapidly changing technologies or market environments. In response, this study proposes an improved approach to identify technology opportunities with less uncertainty, by comparing PEs' similar sequences to a target firm's dynamic change of focus technology fields by applying a PrefixSpan algorithm based sequential pattern mining. Based on PEs' similar sequences, a set of technology opportunity candidates are identified and evaluated in terms of business stability, technological similarity, and recency.

The proposed approach mainly addresses the target firm's dynamic change of focus technology fields. Since this technological change is often overlooked in prior studies, their results are likely to be inappropriate and create uncertainty in the business environment. Therefore, our approach can be considered an improved TOD approach that provides technology opportunities with relatively little uncertainty and high practicality in a real-life business environment. Moreover, as analyzed in the discussion section, technology opportunities can offer quality insights in planning subsequent R&D items or business strategies in real-life business environments. Prior studies emphasized heterogeneous technology fields as technology opportunities, but this can be mistaken as that the target firm needing to enter heterogeneous fields where uncertainty is still high. On the other hand, since the proposed approach identifies technology opportunities from PEs and evaluates them objectively even if they belong to homogenous fields, business stability is likely to be high in a real-life business environment. Further, the proposed approach needs limited expert intervention, except for the weighting process in the evaluation, meaning it can be developed as an automated system in practice. The automated system can increase the time/cost efficiency for SMEs with scarce resources and information in business decision making. That is, SMEs can obtain their results simply by accumulating patent data and updating the sequence database thanks to the high reproducibility of the proposed methodology. Additionally, PEs' technological sequence can provide various results by adjusting the Prefix according to philosophy or purpose of the target firm. From a national aspect, the sequence database can be generated not only from applicants' or SMEs' perspective, but also from a national or large enterprise perspective for identifying and comparing holistic technology flows.

Despite these contributions, this study is not without limitations.

First, in defining SMEs and PEs based on each firm's technology capacity, the proposed approach only uses the number of patents registered by each firm. The thresholds for categorizing the applicants to SMEs and PEs with the number of patents have been set by the experts based on previous studies and statistical data. Of course, the number of patents is a simple and intuitive concept, having been used in several studies. In further studies, the measure of technology capacity should be improved by considering firms' various nature such as financial information, market share, technology valuation results in addition to the number of patents. Additionally, the way to define the firm's focus technology fields per year should be enhanced because the proposed approach tends to circumscribe the focus technology fields by using only the technology fields in which the firm had registered relatively many patents in a certain year. As a result, weak technological change may not be identified through IPC information, but this point can be addressed by using all involved technology fields, increasing the efficiency of the sequential pattern mining algorithm or applying text mining to patent data in defining the temporal change of focus technology field in further studies. In conclusion, the aim of the proposed approach is to inform the firm on appropriate technology opportunities with less uncertainty based on the target firm's technological change and similar PE cases. Therefore, the proposed approach can help SMEs follow the leading firms in the technology competition and achieve sustainable development and growth.

## Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045768). This work is supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry and Energy (MOTIE, 20174010201490).

## References

Agrawal, R., Srikant, R., 1995. Mining sequential patterns. In: Data Engineering, 1995. Proceedings of the Eleventh International Conference on. IEEE, pp. 3–14.
Aloysius, G., Binu, D., 2013. An approach to products placement in supermarkets using PrefixSpan algorithm. Journal of King Saud University-Computer and Information Sciences 25, 77–87.
Baruch, Y., Altman, Y., 2016. The ecosystem of labor markets and careers. People Strategy 39, 16.
Burkhardt, M.E., Brass, D.J., 1990. Changing patterns or patterns of change: the effects of a change in technology on social network structure and power. Adm. Sci. Q. 104–127.
Chen, C.-T., 2000. Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets Syst. 114, 1–9.
Choi, S., Park, H., 2016. Investigation of strategic changes using patent co-inventor network analysis: the case of samsung electronics. Sustainability 8, 1315.
Fagerberg, J., 1987. A technology gap approach to why growth rates differ. Res. Policy 16, 87–99.
Guo, B., Gao, J., Chen, X., 2013. Technology strategy, technological context and technological catch-up in emerging economies: industry-level findings from Chinese manufacturing. Tech. Anal. Strat. Manag. 25, 219–234.
Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U., Hsu, M.-C., 2000. FreeSpan: frequent pattern-projected sequential pattern mining. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 355–359.
Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M., 2001. Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth. In: Proceedings of the 17th International Conference on Data Engineering, pp. 215–224.
Hlady Rispal, M., Servantie, V., 2017. Business models impacting social change in violent and poverty-stricken neighbourhoods: a case study in Colombia. Int. Small Bus. J. 35, 427–448.
Hwang, C.-L., Yoon, K., 1981. Methods for multiple attribute decision making. In: Multiple Attribute Decision Making. Springer, pp. 58–191.
Jeong, B., Yoon, J., Lee, J.-M., 2019. Social media mining for product planning: a product opportunity mining approach based on topic modeling and sentiment analysis. Int. J. Inf. Manag. 48, 280–290.
Klevorick, A.K., Levin, R.C., Nelson, R.R., Winter, S.G., 1995. On the sources and significance of interindustry differences in technological opportunities. Res. Policy 24, 185–205.
Klir, G.J., 2005. Uncertainty and Information: Foundations of Generalized Information Theory. John Wiley & Sons.
Krohling, R.A., Pacheco, A.G., 2015. A-TOPSIS–an approach based on TOPSIS for ranking evolutionary algorithms. Procedia Computer Science 55, 308–317.
Kyebambe, M.N., Cheng, G., Huang, Y., He, C., Zhang, Z., 2017. Forecasting emerging technologies: a supervised learning approach through patent analysis. Technol. Forecast. Soc. Chang. 125, 236–244.
Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29, 481–497.
Lee, Y., Kim, S.Y., Song, I., Park, Y., Shin, J., 2014. Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. Scientometrics 100, 227–244.
Lee, K., Go, D., Park, I., Yoon, B., 2017a. Exploring suitable technology for Small and Medium-Sized Enterprises (SMEs) based on a hidden Markov model using patent information and value chain analysis. Sustainability 9, 1100.
Lee, J., Kim, C., Shin, J., 2017b. Technology opportunity discovery to R&D planning: key technological performance analysis. Technol. Forecast. Soc. Chang. 119, 53–63.
Ma, J., Porter, A.L., 2015. Analyzing patent topical information to identify technology pathways and potential opportunities. Scientometrics 102, 811–827.
Mola, L., Rossignoli, C., Carugati, A., Giangreco, A., 2015. Business intelligence system design and its consequences for knowledge sharing, collaboration, and decision-making: an exploratory study. International Journal of Technology and Human Interaction (IJTHI) 11, 1–25.
Nakamura, H., Suzuki, S., Sakata, I., Kajikawa, Y., 2015. Knowledge combination modeling: the measurement of knowledge similarity between different technological domains. Technol. Forecast. Soc. Chang. 94, 187–201.
North, D., Smallbone, D., Vickers, I., 2001. Public sector support for innovating SMEs. Small Bus. Econ. 16, 303–317.
Olsson, O., 2005. Technological opportunity and growth. J. Econ. Growth 10, 31–53.
Park, Y., Yoon, J., 2017. Application technology opportunity discovery from technology portfolios: use of patent classification and collaborative filtering. Technol. Forecast. Soc. Chang. 118, 170–183.
Savioz, P., Blum, M., 2002. Strategic forecast tool for SMEs: how the opportunity landscape interacts with business strategy to anticipate technological trends. Technovation 22, 91–100.
Schoemaker, P., 2015. Strategic approaches to managing uncertainty. In: Decision Strategies &.
Seo, W., Yoon, J., Park, H., Coh, B., Lee, J.-M., Kwon, O.-J., 2016. Product opportunity identification based on internal capabilities using text mining and association rule mining. Technol. Forecast. Soc. Chang. 105, 94–104.
Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation 28, 758–775.
Srikant, R., Agrawal, R., 1996. Mining sequential patterns: Generalizations and performance improvements. In: International Conference on Extending Database Technology. Springer, pp. 1–17.
Tseng, Y.-H., Lin, C.-J., Lin, Y.-I., 2007. Text mining techniques for patent analysis. Inf. Process. Manag. 43, 1216–1247.
Vahlne, J.-E., Johanson, J., 2017. The internationalization process of the firm—a model of knowledge development and increasing foreign market commitments. In: International Business. Routledge, pp. 145–154.
Wang, M.-Y., Fang, S.-C., Chang, Y.-H., 2015. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels. Technol. Forecast. Soc. Chang. 92, 182–195.
Yang, C.S., Wei, C.P., Chiang, Y.H., 2014. Exploiting technological indicators for effective technology merger and acquisition (M&A) predictions. Decis. Sci. 45, 147–174.
Yoon, J., Kim, K., 2012. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. Scientometrics 90, 445–461.
Yoon, B., Magee, C.L., 2018. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. Technol. Forecast. Soc. Chang. 132, 105–117.
Yoon, J., Park, H., Seo, W., Lee, J.-M., Coh, B., Kim, J., 2015. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. Technol. Forecast. Soc. Chang. 100, 153–167.
Zaki, M.J., 2001. SPADE: an efficient algorithm for mining frequent sequences. Mach. Learn. 42, 31–60.

**Jaewoong Choi** is a researcher in Department of Industrial Engineering of Konkuk University, Korea. His research topics include patent intelligence, technology opportunity discovery, and social media mining for product and technology planning.

**Byeongki Jeong** is a researcher in Department of Industrial Engineering of Konkuk University, Korea. His research topics include technology intelligence, patent mining, technology opportunity identification, and social media mining for future planning.

**Janghyeok Yoon** is an associate professor in Department of Industrial Engineering of Konkuk University, 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 opportunity identification, technology roadmapping and technology convergence, and business intelligence, including social media mining.