Journal of Informetrics 15 (2021) 101103

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Inventor profile mining approach for prospective human resource scouting

Jaemin Chung, Namuk Ko, Hyeonsu Kim, Janghyeok Yoon*

Department of Industrial Engineering, Konkuk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 1 June 2020
Received in revised form 31 August 2020
Accepted 30 September 2020

Keywords:
Inventor profile mining
Human resource
Scouting
Technology-based firm
Patent analysis

ABSTRACT

Scouting young and talented human resources with firm-specific domain knowledge has a great impact on performance and sustainable growth among technology-based firms. Previous studies have proposed key researcher identification and recommendation approaches, but few studies have focused on identifying prospective human resources—young and talented people suitable for a firm's technology strategy. Thus, this study proposes an inventor profile mining approach for identifying such human resources. The proposed approach is as follows: 1) collecting patent data related to a target firm and preprocessing candidate inventors' patents; 2) identifying the inventors' technology fields and measuring their invention performance and career; 3) generating performance-career portfolio maps for invention fields to identify prospective inventors aligned with the target firm's technology development directions. We show that this approach can identify prospective inventors through a case study and statistical validation. This approach is expected to be used as a human resources management tool by technology-based firms to help them identify and scout young and talented human resources the most suitable for technology strategies.

1. Introduction

Talented human resources have a positive impact on firm performance because they are valuable and difficult to replace (Buller & McEvoy, 2012). Literatures suggest that activities such as scientific inventions and technological developments representing a firm's innovation are concentrated in small elite groups (Andersson & Berggren, 2007; Göktepe-Hultén, 2008). Moreover, the value of innovation created by a small number of elites is greater than that of the innovation created by average people (Lepak, Takeuchi, & Snell, 2003; Narin & Breitzman, 1995; Oldroyd & Morris, 2012). Therefore, scouting talented human resources effectively is critical to a firm's long-term success (Afzal & Maurer, 2011). In addition, a firm must identify human resources with appropriate capabilities to implement its target strategies (Becker & Gerhart, 1996). Thus, it is important to ensure that human resources have the capabilities relevant to the firm's technical needs and its key business area (Davoodi, Kianmehr, & Afsharchi, 2013; Wang, Ma, Liao, & Du, 2017). In particular, technology-based firms (TBFs) usually need human resources who can implement the firms' technology strategies. Therefore, TBFs should focus on human resources who have technical knowledge and capabilities associated with the firms' technology domain.

From a future-oriented perspective, it is necessary to recruit young and talented human resources who can help a firm innovate. Although young human resources have less domain experience than seasoned and prominent experts, recruiting

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

https://doi.org/10.1016/j.joi.2020.101103
1751-1577/

the young and talented brings the firm several advantages. First, young human resources can adapt to the firm's organizational culture and atmosphere easily (Charan, Drotter, & Noel, 2010). Second, they have the potential to contribute to the firm's long-term growth by developing themselves into well-established human resources (Panagopoulos, Tsatsaronis, & Varlamis, 2017). Finally, recruiting young human resources and then educating them is more cost-effective than recruiting prominent experts (Charan et al., 2010), because the former can not only provide technical knowledge to the firm but can also contribute as potential future leaders.

Recently, access to academic, technical, and web data has become easier, a number of studies based on large amounts of data have appeared to identify and recommend human resources. Some studies recommend experts suitable for system users' needs based on web and academic data (Cifariello, Ferragina, & Ponza, 2019; Sun, Xu, Ma, & Sun, 2015). Other studies analyze academic data to prioritize researchers (Daud, Abbasi, & Muhammad, 2013; Li, Foo, Tew, & Ng, 2009), and identify rising star groups (Daud, Ahmad, Malik, & Che, 2015; Panagopoulos et al., 2017). Unlike the characteristics of data used by these studies, patent data is not only the final products of research and development (R&D) but also the latest technical documents containing plenty of technical information (Franceschini & Maisano, 2012; Guan & Gao, 2009). Also, patent data is suitable data to represent the technical knowledge and capability (Mariani & Romanelli, 2007; No & Lim, 2009; Schettino, Sterlacchini, & Venturini, 2013). From this point of view, there have been some studies analyzing patent data to identify suitable inventors or rising technology stars (Moehrle, Walter, Geritz, & Müller, 2005; Zhu, Zhu, Wang, Cunningham, & Wang, 2019).

Despite the abovementioned studies, there are some limitations in terms of identifying prospective human resources who are young, talented, and strategy-oriented. First, securing appropriate human resources according to a TBF's technology strategy has a great influence on the success or failure of further R&D activities (Becker & Gerhart, 1996). However, prior studies have not sufficiently considered the processes of recommending and discovering human resources tailored for strategy implementation. The studies either identify researchers who are highly similar to a firm's domain or identify rising stars in a particular field. However, R&D strategies are carried out in various directions, such as entering a new technology field or strengthening existing technological capabilities. In addition, the capabilities of human resources required for each strategy would vary (Bercovitz & Feldman, 2007). Therefore, it is important to specifically define human resources' invention fields and then, to find appropriate resources in fields that match the firm's technology strategy. Second, prior studies evaluate the growth trend or potential of human resources' performance without considering the resources' career "youngness." A firm should scout prospective human resources who can provide necessary technical knowledge and contribute to the firm's sustainable growth at a relatively low cost. Thus, in evaluating human resources, it is necessary to use an indicator that measures the resources' youngness. Third, even though R&D developers in TBFs disclose their R&D results as patents, most previous studies evaluate human resources using journal papers or web data. TBFs need to implement their technology strategies through human resources with suitable technical capabilities. As such, it is important to evaluate human resources by analyzing patents representing the resources' technical capabilities and knowledge. Finally, many prior studies use only superficial information such as the number of patents or papers, citations, and co-authorship, whereas, it is necessary to analyze human resources' actual contents.

To remedy these limitations, this study suggests a patent-based inventor profile mining approach to identify prospective human resources; the resources to be identified in this study are inventors. TBFs usually seek human resources with the technical knowledge and capabilities to implement technology strategies. To this end, the proposed approach aims to identify human resources who are young, talented, and suitable for a target TBF's technology strategy based on patent data. We assess inventors' actual invention fields, invention performance, and career. Our approach uses topic modeling to identify sub-technology fields in the target firm's technology domain and to assign invention fields to the inventors. Then, this approach also measures the age-weighted (AW)-index and patent filing years to evaluate the inventors. The proposed approach follows the following steps: 1) collecting and preprocessing patents related to the target firm's existing technologies; 2) identifying and measuring inventors' invention fields, invention performance, and invention career; and 3) generating performance-career portfolio maps for invention fields based on inventors' performance and career to identify prospective inventors.

The contributions of this study are threefold. First, the proposed approach can identify appropriate human resources according to detailed directions of a target firm's technology strategies. Abstracts of patents related to the target firm are used to define the invention field of each inventor. Therefore, through this approach, it is possible not only to grasp the technology fields of the target firm's domain but also to find human resources who could help implement the firm's technology strategies. Second, this approach identifies young human resources based on their invention career. Our approach is thus a method of discovering prospective human resources who can contribute to the firm's sustainable growth and innovation. Finally, since the proposed approach is a quantified method of screening prospective human resources, it can be used as a tool for scouting human resources.

The rest of this paper is organized as follows. Section 2 reviews related studies on recommending and identifying human resources. Section 3 describes the proposed approach. Section 4 illustrates a case study of a specific target firm. Section 5 performs a statistical validation and discusses the study's contributions and implications. Finally, Section 6 provides a conclusion and proposes future research topics.

2. Approaches for human resource recommendation and identification

R&D is one of the most important activities for the innovation of TBFs and requires a higher level of creativity and complexity than other firm activities (Mairesse & Sassenou, 1991). Therefore, TBFs need to find human resources with extraordinary technical talent to ensure R&D success (Ernst, Leptien, & Vitt, 2000). However, each human resource has a different knowledge level, and the knowledge required by each firm also varies (Panagopoulos et al., 2017). Thus, a firm must ensure that the knowledge possessed by human resources can meet the firm's technical needs (Naeem, Khan, & Afzal, 2013). Using academic and web data, a number of studies have been conducted to recommend human resources suitable for the needs of system users and firms. Davoodi et al. (2013) proposed a method for finding experts with knowledge related to system users' needs using expert information and Wikipedia data collected from the web. They constructed a semantic-based document-concept similarity matrix representing the semantic similarity between the system users' needs and expert profiles. Suitable experts were then recommended through a clustering model and centrality measures. Their method contributes to research on how to find groups of experts with the knowledge desired by the system users through semantic relationships. Yang et al. (2013) proposed a system for recommending experts and collaborative groups required for national R&D. Firms need experts to help with paper writing and project consulting in specific subject areas when conducting R&D. Thus, the study investigated how to find experts suitable for firms' target areas based on information regarding Korean national R&D and R&D collaboration. They proposed an algorithm that can calculate expert knowledge levels and provided expert knowledge map services for the R&D subject area. Similarly, Wang et al. (2017) recommended researchers for university–industry collaboration based on a firm's needs. They identified a set of candidate researchers and then recommended appropriate researchers for industrial R&D projects based on a contextual trust analysis model. Their approach offers a new perspective on promoting university–industry collaboration and provides a new way for firms to access more researchers based on a social network platform and recommendation system.

The aforementioned studies all propose recommending experts who have the knowledge desired by system users or firms. They suggest various ways through which firms may access experts, but they do not consider the youngness in the careers of the candidate human resources from a human resource management (HRM) perspective. Firms have two alternatives in recruiting human resources: 1) recruiting young human resources who may have the potential required for the firm's sustainability; 2) recruiting mature human resources who have a long career and possess the knowledge currently required by the firm (Charan et al., 2010). In some cases, the former offers greater benefits from a future-oriented perspective. Young human resources incur lower recruitment costs than mature experts and easily adapt to the organizational cultures and atmospheres. Moreover, young human resources who grow into well-established human resources after recruitment can contribute as potential leaders who can help drive firm success (Daud et al., 2013). Therefore, firms should identify human resources by taking their youngness into consideration. In this regard, several studies were appeared to find rising stars using academic data. These studies defined the rising star as a researcher who currently has relatively low research results but will become prominent (Li et al., 2009).The PageRank algorithm proposed by Brin and Page (1998) was used to identify rising stars (Daud et al., 2017; Liu, Bollen, Nelson, & Van de Sompel, 2005; Zhang et al., 2016a, 2016b). These studies all assume that competent researchers may have many social links with other researchers. Thus, these studies recommended researchers who have little research achievement but are likely to become prominent experts. Other studies used supervised machine learning techniques to find rising stars in co-author networks. (Daud et al. (2015) classified rising stars through models such as the Maximum Entropy Markov Model and Classification, and Regression Trees based on the researchers' influence, journal information, and co-author information. Nie et al. (2019) labeled researchers based on whether their impact score had increased and classified rising stars through binary classification models. Panagopoulos et al. (2017) also detected rising stars through an unsupervised machine learning technique. They clustered researchers for each period based on researchers' performance, impact, and sociability features. Then, the study analyzed the characteristics of the clusters over time to obtain a group of rising stars. This study is an early attempt to identify rising stars based on feature changes over time by integrating researchers' quantity-, impact-, and collaboration-related features.

In a similar view to the above-mentioned expert recommendation and rising star identification studies, two studies have sought to analyze inventors based on patents. Patents are the final product of technological R&D and are documents containing specific technical and scientific information about inventions (No & Lim, 2009). Therefore, analyzing the patents held by inventors allows us to define their technical knowledge and experience (Agiakloglou, Drivas, & Karamanis, 2016; Ferrucci & Lissoni, 2019; Kiehne & Krill, 2017; Mariani & Romanelli, 2007; Schettino et al., 2013). Indeed, Moehrle et al. (2005) proposed an approach to define the profiles of inventors within a firm in order to form R&D teams. They calculated similarities in the Subject-Action-Object structures of patent text data and clustered inventors with similar technical fields by constructing an inventor competence map using multidimensional scaling. Their approach can be used to find inventors who can complement a firm's current technical capabilities or who operate in technology fields completely different from the firm's current capabilities. By contrast, Zhu et al. (2019) proposed a method of detecting early-career inventors who are expected to perform well. They identified rising technology stars by extracting indicators related to inventors' technology performance, sociability, and tech-diversity. They then clustered the inventors by timestamps and analyzed how the cluster features changed over time. Their study considered the diversity of the technical field in evaluating inventors and is an early attempt to identify promising inventors based on patent data.

Identifying appropriate human resources for TBFs requires integrating all the perspectives adopted by the studies discussed above. First, human resources suitable for a TBF should have the technical capabilities required for the firm's target

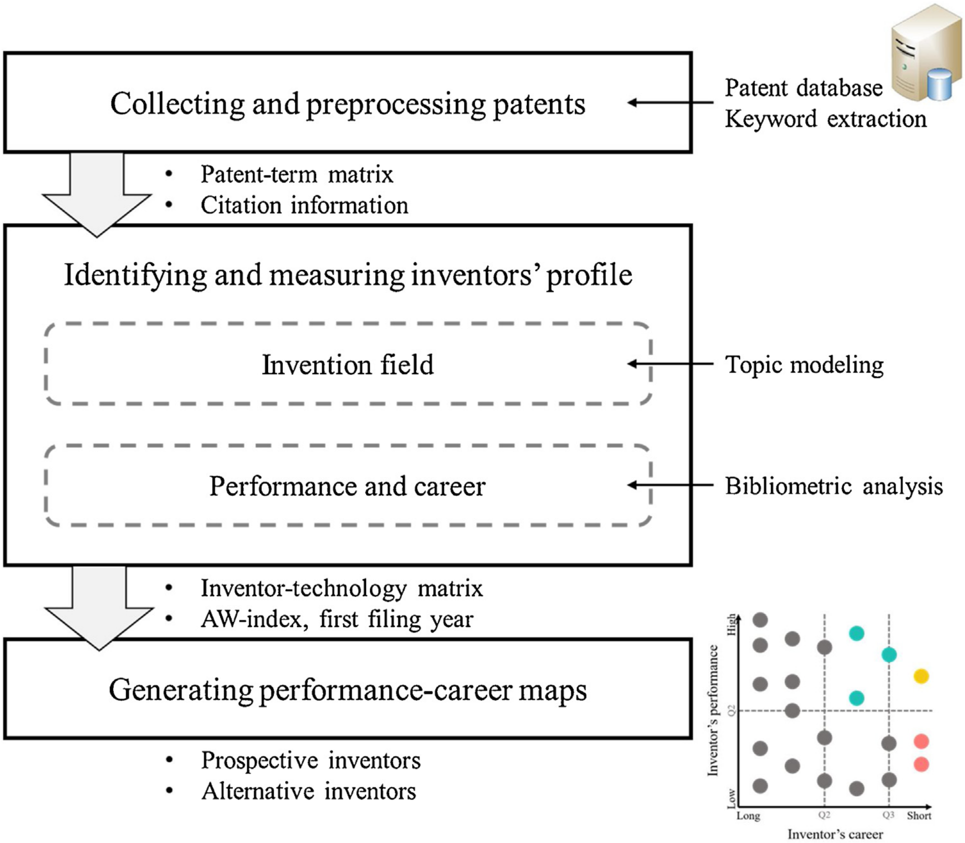

Fig. 1. Overview of proposed approach.

technology development strategy. Firms may establish several strategic directions, such as discovering new technology areas or strengthening existing ones. Therefore, new human resources recruited to assist in the implementation of the firm's technology strategies should have the appropriate technical knowledge and capabilities for each strategy. Second, firms should recruit and train young human resources who can contribute to the firm over the long term. However, academic and patent data do not include an indicator of the youngness of human resources. Thus, it is necessary to define an indicator, such as an invention career, that can replace youngness as a measure. Finally, TBFs need to use patent data to identify human resources with technical capabilities. Human resources in TBFs generally disclose their technology development results through patents. Thus, it is difficult to represent inventors' technical capabilities through only journal papers or web data. This study proposes an approach to identify prospective inventors that addresses all the above considerations. The prospective inventors defined in this study not only possess the technical knowledge and capabilities suitable for a firm's technology strategy but also have excellent invention performance and potential despite being early in their careers.

3. Proposed approach

This study proposes an approach for identifying prospective inventors. The proposed approach identifies young and talented human resources with the capabilities suitable for technology strategies of a target TBF. Our approach consists of three steps (Fig. 1): 1) we collect and preprocess patents related to the target firm's major technology domain; 2) we identify inventors' invention field and measure their invention performance and career; 3) we generate performance-career portfolio maps based on the firm's technology strategy. The following sections describe the approach in detail.

3.1. Collecting and preprocessing patents

Fig. 2 shows an overall process of collecting and preprocessing patent data in this approach. The approach first requires collecting initial patent data. As this study aims to identify prospective inventors, it is necessary to use patent data with high-quality inventor disambiguation. Unlike patent offices in other countries, the United States Patent and Trademark Office (USPTO) provides improved inventor identification information that can be used to distinguish inventors (Li et al., 2014). Therefore, the initial patent data used in this study are drawn from the USPTO. The initial data can be collected using

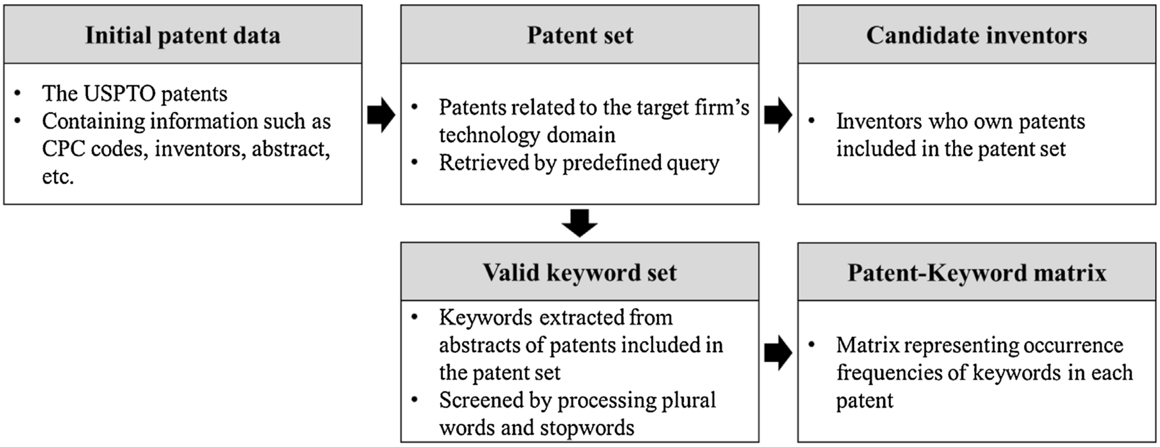

Fig. 2. Process of collecting and preprocessing patent data.

PatentsView (http://www.patentsview.org), a patent visualization and analysis platform supported by the USPTO, which provides access to regularly updated US patents. Thus, we can access to the latest patent information, such as inventors, applicants, cooperative patent classification (CPC) codes, citations, and abstracts, through PatentsView.

This study aims to identify suitable human resources for a target firm. Therefore, a patent set and candidate inventors should be selected carefully. The patent set is a set of patents related to the target firm's technology domain and can be retrieved via a predefined query from the initial patent data. CPC codes or technical keywords are often used to represent a firm's technical domain (Kim & Bae, 2017; Moreira & Wichert, 2013; Yang et al., 2015); thus, these can be used to define the query. After obtaining the patent set, we can create a list of candidate inventors who own patents included in the patent set. The candidate inventors are inventors with technical knowledge related to the target firm and are thus candidates for prospective inventors.

Our approach uses keywords extracted from patent abstracts to determine the candidate inventors' invention fields. A patent document includes textual data such as title, abstract, description, and claims. Unlike other textual data, the abstract summarizes the essential contents of the invention (Park, Yoon, & Kim, 2012). Therefore, keywords extracted from the patent abstracts can indicate an inventor's technical knowledge and domain (Xu et al., 2012). Such keywords can be extracted using natural language processing tools such as TextBlob (https://textblob.readthedocs.io/en/dev/), RAKE (https://github.com/aneesha/RAKE), and spaCy (https://spacy.io/). By using these tools, noun phrase keywords extracted from patent abstracts and term frequency (TF) value of keywords can be obtained. However, the excluded keywords include many plural words (e.g., "portable devices," "digital images"), numeric units (e.g., "1inch," "15mhz"), and meaningless words (e.g., "different types," "different method"). Thus, we can obtain a valid keyword set by lemmatizing plural keywords to singular and removing stopwords that lacked technical meaning. After that, we obtained the final TF value of valid keywords by summing TF values of keywords that have been changed into the same form by lemmatizing for each patent. Finally, we generate a patent-keyword matrix for the next step through the patent-keyword-TF values.

3.2. *Identifying and measuring inventors' profile*

In this section, we identify and measure the candidate inventors' profile in order to evaluate them. The profile defined in this step is divided into two categories: 1) invention field, and 2) invention performance and career. The former represents the technology field of an inventor's R&D experience, and the latter represents how talented and young an inventor is. The subsections below describe in detail how to identify and measure each profile.

3.2.1. *Invention field*

This subsection adopts Latent Dirichlet Allocation (LDA) proposed by Blei, Ng, and Jordan (2003) to identify the invention fields of the candidate inventors. We identify sub-technology fields related to the technology domain of the target firm through topic modeling based on the patent-keyword matrix constructed in the previous step. A topic modeling is a statistical model that determines latent topics by semantically clustering keywords in a document set (Ko, Jeong, Choi, & Yoon, 2017). A document consists of a set of keywords, combinations of keywords mean topics of the documents, and one document can have one or more topics (Jeong, Yoon, & Lee, 2019). Using the topic modeling, we can identify topics in each document based on the occurrence frequency of keywords included in the document and infer keywords associated with each topic (Blei & Lafferty, 2009). Among the various topic modeling algorithms, this approach uses LDA-based topic modeling, which performs better than other algorithms (Blei et al., 2003).

The topic modeling outputs a document-topic distribution matrix and topic-keyword distribution matrix by inputting a document-keyword matrix and the number of topics. The patent-keyword matrix obtained in the previous step is used

Inventor	Patent #	Tech. 1	Tech. 2	Tech. 3
Inv. 1	1	0.75	0.04	0.21
Inv. 1	2	0.80	0.02	0.18
Inv. 2	3	0.20	0.01	0.79
Inv. 3	4	0.01	0.88	0.11
Inv. 1	5	0.09	0.03	0.88
Inv. 2	5	0.09	0.03	0.88
Inv. 3	6	0.16	0.77	0.07
Inv. 2	7	0.08	0.04	0.88

Inventor-Patent-Technology matrix

Summation by inventor

	Tech. 1	Tech. 2	Tech. 3
Inv. 1	1.64	0.09	1.27
Inv. 2	0.37	0.08	2.55
Inv. 3	0.17	1.65	0.18

Inventor-Technology matrix

Fig. 3. Concept of computing inventor-technology matrix.

as the document-keyword matrix, one of the topic modeling's input values. The number of topics—the other input—can be defined through techniques such as the elbow method, the information criterion method, and the information theoretic method (Jeong & Yoon, 2017). For example, the elbow method is a technique of deriving average cosine similarity between topics according to the number of topics and selecting a point at which the similarity is stabilized as the number of topics increases. These techniques have the advantage of being able to scientifically select the number of topics in which descriptions between topics are well separated. This study performs the topic modeling based on patent abstract data containing a brief technical description of the invention. Therefore, each topic derived from the topic modeling can be interpreted as a sub-technological field, and thus an inventor's invention field. Consequently, we denote the document-topic distribution matrix and topic-keyword distribution matrix as a patent-technology distribution matrix and technology-keyword distribution matrix, respectively.

The technical names of the sub-technology fields are labeled based on the contributions of the keywords and patents in the two distribution matrices derived from the topic modeling. The candidate inventors' invention fields are then defined based on the patent-technology distribution matrix. As a patent has more than one inventor, the patent-technology distribution matrix can be extended to an inventor-patent-technology matrix. Then, an inventor-technology matrix can be obtained by summing the contributions based on the inventors from the extended matrix (Fig. 3). The inventor-technology matrix represents the technical contribution each inventor has made to the field. The invention field of each candidate inventor can then be defined by selecting the technology field with the greatest contribution in the inventor-technology matrix. A sub-technology field similar to the target firm's strategy is found through a qualitative process using the technology-keyword distribution matrix. Then, a set of inventors corresponding to the sub-technology field is obtained through the inventor-technology matrix. In the next step, a performance-career portfolio map is generated based on the inventor set.

3.2.2. Performance and career

In this subsection, the candidate inventors' invention performance is measured through bibliometric analysis. The candidate inventors' invention careers are also measured based on the filing dates of their registered patents. A firm needs to scout young and talented human resources. Therefore, invention performance and career, representing an inventor's technical talent and youngness, are essential evaluation indicators. In the previous subsection, we used the patent set related to the target firm's technology domain. In this subsection, on the other hand, the performance and careers of the candidate inventors are measured using all of their registered patents. If we measure the performance and career with patents corresponding to the technology domain of the target firm, we can see the level of capability and experience in the domain. However, the main goal of this subsection is to reveal how talented the inventor is, regardless of the technology field, and how young the inventor is. Thus, we calculate the performance and career measure of each inventor based on all of his registered patents.

Many studies on researcher analysis use the h-index to measure research performance (Bornmann & Marx, 2011; Franceschini & Maisano, 2012; Patel et al., 2013). The h-index, first proposed by Hirsch (2005), is a representative indicator that quantifies the research output of individual researchers (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009) . This index can be calculated based on the number of publications produced by a researcher and the number of their citations. For example, a researcher with five or more publications with at least five citations would have an h-index of 5. Since the h-index considers both the quantity and quality of publications, it can measure a researcher's capability more effectively than simple measures such as the number of citations or publications (Bornmann, Mutz, Hug, & Daniel, 2011).

However, the h-index is not sensitive to recent research performance and does not assign high scores to researchers with relatively few publications and citations (Rousseau & Leuven, 2008). Therefore, the h-index is unsuitable for evaluating researchers with short research careers (Jin, Liang, Rousseau, & Egghe, 2007). To overcome this disadvantage, several other indicators have been proposed, such as the AW-index, the age-weighted citation rate (AWCR), and the age-weighted citation

Table 1
Comparing AW-index and h-index of two inventors.

Inventor A					Inventor B				
Patent #	# of citation	Age of patent	AW-index	h-index	Patent #	# of citation	Age of patent	AW-index	h-index
1	70	5	5.1478	4	1	15	25	3.4053	8
2	10	4			2	13	23		
3	18	4			3	26	20		
4	5	2			4	35	20		
5	3	1			5	10	18		
					6	12	15		
					7	8	12		
					8	10	12		
					9	24	10		
					10	17	8		

rate per author (AWCRpA) (Velmurugan & Radhakrishnan, 2015). These indicators not only take into account the number of citations but also consider the age of each publication (Jin et al., 2007). The AW-index tends to be used more often than other age-dependent indicators to evaluate researchers who have more recent research results and shorter research careers (Cucchetti et al., 2013; Khey, Jennings, Higgins, Schoepfer, & Langton, 2011; Long, Boggess, & Jennings, 2011). The AW-index of a researcher can be calculated through the equation as follows:

$$AW\ index = \sqrt{\sum_{j} \frac{citation_j}{age\ of\ article_j}} \tag{1}$$

where $citation_j$ is the number of received citations of article j, one of the articles the researcher published; $age\ of\ article_j$ denotes the difference between the current year and the publication year of article j.

We apply the AW-index to patent data to measure the performance of an inventor—one of the profile we want to derive from this subsection. The $citation_j$ is replaced by the number of received citations patent j, one of the patents the inventor owned. Also, $age\ of\ article_j$ is replaced by the age of patents denoting the difference between the current year and the filing year of patent j. The prospective inventors defined in this study have short invention careers but are talented human resources. Therefore, the AW-index may better represent the prospective inventors' invention performance than the h-index. For example, Table 1 shows a comparison between the two inventors' AW-index and h-index values. Inventor A and Inventor B have 5 and 10 patents, respectively. Inventor A has fewer citations than Inventor B but has only recently registered patents. On the bases on youngness and talent, Inventor A deserves a better evaluation than Inventor B. However, the h-index of Inventor A is lower than that of Inventor B. On the other hand, the AW-index of Inventor A is higher than that of Inventor B because Inventor A has excellent recent patents. This example shows that the AW-index can measure the performance of prospective inventors more accurately.

The career of an inventor, the other profile derived from this section, is measured by the first filing year. To identify prospective inventors, we must be aware of the youngness of each candidate inventor. However, though patent data include information such as inventors' names and addresses, but not the actual ages. Therefore, an indicator that could replace the age of individual inventors should be defined. The earliest filing point of an inventor's registered patents is the time when the inventor's technically certified knowledge and capability are first released. Thus, the first filing year can be regarded as the point at which the inventor's career began. Consequently, based on the assumption that inventors who started their careers recently are likely to be young, we use the first filing year to evaluate youngness. We derive the first filing year through the earliest filing year of the registered patents owned by each inventor to represent the invention career of the candidate inventors.

3.3. Generating performance-career portfolio maps

In this section, prospective inventors (i.e., young and talented human resources suitable for the technology strategy of the target firm) are identified. We suggest a map, denoted as a "performance-career portfolio map," based on two indicators that represent each map dimension (Fig. 4). The two indicators are invention performance and invention career, as derived from the previous section. The map is generated according to the target technology strategy of the firm based on the sub-technology field identified through the topic modeling. For example, to identify human resources useful for enhancing a capability related to the firm's present focal technology domain, the firm would first select a sub-technology field similar to the firm's existing capability based on expert judgment. Then, the firm would generate a map for candidate inventors in the selected sub-technology field. On the other hand, to identify human resources useful for advancing into a new technology field, the firm would generate a map for candidate inventors in the sub-technology field the firm wishes to enter.

The candidate inventors on the map can be divided into four groups by dividing the performance-career portfolio map's invention performance into the second quartile (Q_2^p), and the invention career into the second (Q_2^c) and third quartiles (Q_3^c). The first group is represented by the nodes in blue on the map. The blue nodes indicate inventors whose invention

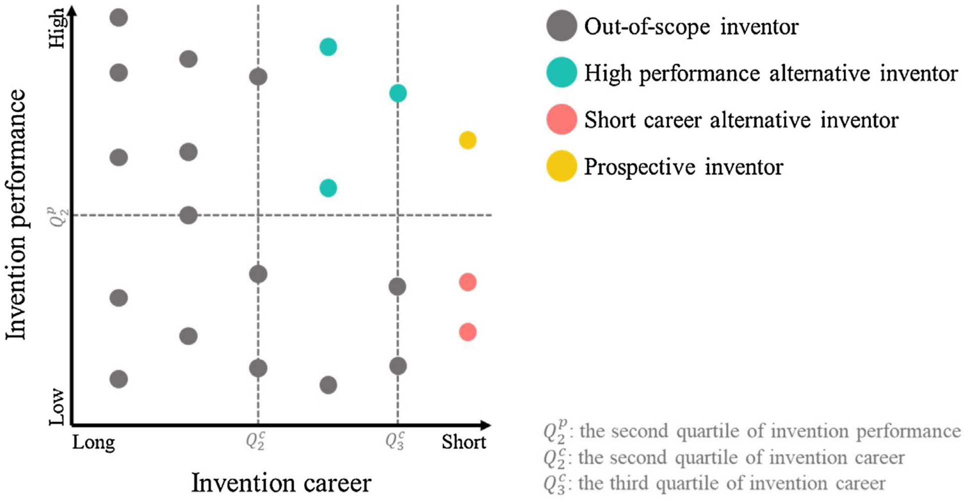

Fig. 4. Performance-career portfolio map.

performance is higher than Q_2^p and whose invention career is shorter than Q_2^c and longer than Q_3^c. The inventors in this group are high-performance alternative inventors who are not very young but have had significant invention experience. The second group is represented by the red nodes on the map. These are inventors whose invention performance is lower than Q_2^p and whose invention career is shorter than Q_3^c. These inventors are short career alternative inventors who are young and have some invention experience. The third group is represented by the yellow nodes. The inventors in this group have invention performance higher than Q_2^p and invention careers shorter than Q_3^c. These inventors are prospective inventors and are thus the young and talented inventors targeted in this study. The final group, indicated by the grey nodes, comprises inventors with insufficient performance or excessively long careers. These inventors are thus out-of-scope inventors that this study does not aim for. The target firm can therefore first look at the yellow nodes, representing prospective inventors on the map. The firm could also consider the alternate inventors to obtain secondary alternatives.

4. Case study: *Suprema*

In this section, a case study is conducted for a specific target firm to show how the proposed approach is applied. We select *Suprema*, a biometrics-security firm specialized in fingerprint and face recognition technology, as the target firm. *Suprema* is a technology-based small and medium-sized enterprise (SME) with 28 USPTO-registered patents offering products and solutions in various areas, such as access control, time and attendance management, and mobile authentication. Implementing technology strategies requires firms to acquire human resources with relevant technical knowledge. In particular, SMEs must select human resources carefully because their human and material resources are limited. Thus, we consider *Suprema*, a technology-based SME, would be suitable for illustrating our inventor profile mining approach.

4.1. Collecting and preprocessing patents

This case study starts with the collection of initial USPTO patent data. Through PatentsView, we obtained 7,144,406 patents filed since 1960. Based on the initial patent data, we defined a query to retrieve a patent set related to the technical domain of the target firm. The firm has technologies related to collecting, processing, and managing biometric information, as well as security-related technologies such as access management and recognition algorithms. Of the 28 registered patents owned by the target firm, 21 have the G06 K subclass CPC code. The subclass G06 K represents technologies for the recognition of data, presentation of data, record carriers, and handling record carriers. This subclass accurately represents the major technology domain of the target firm. Thus, we simply defined the query to retrieve patents containing the subclass G06 K.

Based on the predefined query, 102,303 patents corresponding to the patent set were gathered. The collected patent data included information such as inventors, abstracts, application dates, and citations. A total of 102,306 candidate inventors who own patents included in the patent set were also gathered. The quantitative trend of the patent set and the candidate inventors are shown in Fig. 5. The graph on the left of the figure shows the number of registered patents by year, and the graph on the right shows the number of inventors filing a patent for the first time by year. As the figure shows, both graphs began to increase gradually from 1970 onward, and the number of registered patents and new inventors increased explosively in the last decade.

Keywords were extracted from the abstracts of all patents in the patent set using spaCy, a natural language processing tool that can extract keywords from a large set of documents with high accuracy and speed. A total of 157,013 keywords were extracted. After that, we changed plural keywords to singular and removed stopwords. We also removed keywords that

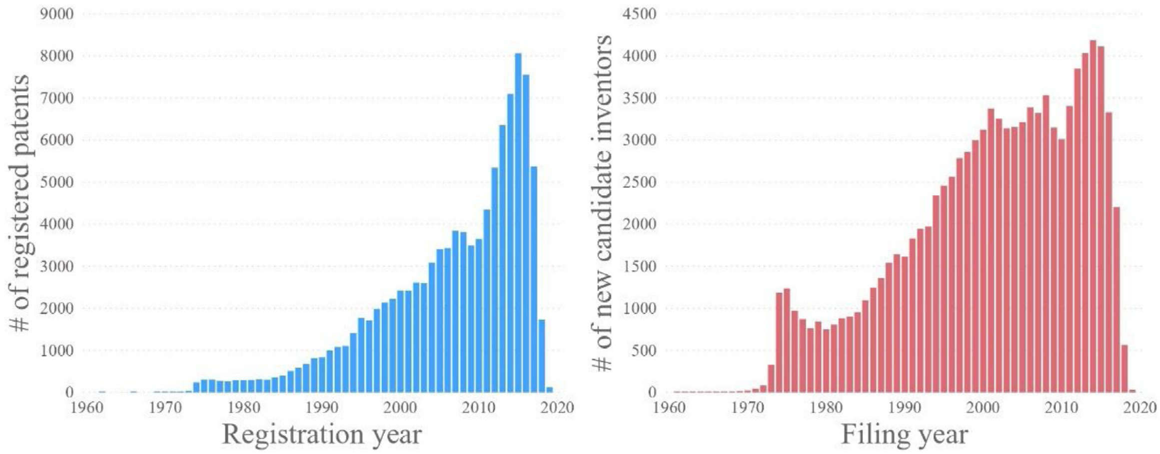

Fig. 5. Annual trends of the number of registered patents and new inventors.

Table 2
Twelve technologies and their major keywords.

Tech. #	Technology description Major keywords
1	Image processing based on color features data point, color information, input image, image feature, infrared light
2	Reading and tracking process of barcode and media data bar code, print medium, media content, video image, identification data
3	Image formation, feature extraction and object recognition print data, image information, pixel data, feature information, image signal
4	Wireless communication and video/image data processing video data, IC card, video frame, image frame, wireless communication, IC chip
5	Image processing circuit and method image data, digital data, real time, processing circuitry, captured image data
6	Recognition and generation method of data based on video contents input data, video content, face image, captured image, laser beam
7	Constructing data for application to machine learning training data, electrical signal, feature data, computer program, attribute information
8	Decoding and recognition of characters or objects character recognition, object detection, decodable indicia, optical character recognition
9	Object and attribute information identification in print, image, and fingerprint data image processing, identification information, printing data, fingerprinted data
10	System for sharing and transmitting location information feature point, computer program product, location information, feature value
11	Pattern recognition of sensor-based optical data pixel value, sensor data, light beam, pattern recognition, visible light, output data
12	Etc. digital image, coded data, feature vector, image quality, reference image

had document frequency values less than 1. Through this process, 39,433 keywords were selected for a valid keyword set. In addition, 78,006 patents including keywords in the valid keyword set were selected. Finally, we generated a patent-keyword matrix with the TF values of each keyword included in the valid keyword set.

4.2. Identifying and measuring inventors' profile

4.2.1. Invention field

The study performed LDA-based topic modeling using the patent-keyword matrix. We used the elbow method, a technique of selecting a point where similarities between topics enter a stable state, to acquire an appropriate number of topics (Jeong et al., 2019; Wang, Liu, Ding, Liu, & Xu, 2014). Through the elbow method, we performed topic modeling by designating the number of topics as 12. After that, we checked that keywords with high contributions of each topic were not duplicated to ensure that the number of topics was set appropriately. Finally, we manually labeled the name of the 12 detailed sub-technology fields (inventors' invention fields of the G06 K subclass) through the technology-keyword distribution matrix and the patent-technology distribution matrix (Table 2). For example, Tech. 8 was labeled "Decoding and recognition of characters or objects" because it had major keywords such as "character recognition," "object detection," "decodable indicia," and "optical character recognition." As another example, Tech. 10 had major patents such as "Location tracking system

Table 3
Part of inventor-technology matrix.

Inventor ID	Tech. 1	Tech. 2	Tech. 3	Tech. 4	Tech. 5	Tech. 6
6068966-1	0.041667	0.041667	0.041667	0.041667	0.041667	0.541667
6069401-1	0.041667	0.041667	0.041667	0.041667	0.041667	0.041667
6069538-5	0.760945	0.332374	0.532375	15.39031	1.657374	0.332375
6069564-1	0.027778	0.027778	0.027778	0.027778	0.027778	0.027778
6069564-3	0.027778	0.027778	0.027778	0.027778	0.027778	0.027778

Table 4
AW-index and invention start year of inventor 7798394-3.

Patent ID	(1) # of citation	(2) Filing year	$(3) = B - (2)$ Age of patent	$(4) = (1) \div (3)$ Age-weighted citation	$(5) = \sqrt{\sum(4)}$ AW-index	$(6) = \min((2))$ First filing year
7798394	8	2006	14	0.5714	3.1278	2006
8694315	17	2013	7	2.4286		
8770476	2	2010	10	0.2000		
9070129	69	2008	12	5.7500		
9117212	3	2014	6	0.5000		
9330386	2	2014	6	0.3333		

*$B(base\ year) = 2020$.

conveying event information based on administrator authorizations," "Notification system for occurrences of group events based on zone and location of mobile devices," and "Multi-level privilege notification system operated based on indoor location information received from a location information sources," so it was labeled "System for sharing and transmitting location information."

We extended the patent-technology distribution matrix to an inventor-patent-technology matrix and summed contribution values based on inventors. Finally, we obtained an inventor-technology matrix and then assigned major invention fields to each inventor. Table 3 shows a part of the inventor-technology matrix. For example, the major technology field of inventor 6069538-5 is Tech. 4. This inventor has patents titled "Wireless IC device and component for wireless IC device" and "Method for determining existence of wideband impedance matching circuit in a wireless IC device system." As these patents are for technologies related to wireless IC devices capable of wireless data transmission and reception, the technical capability of this inventor is similar to Tech. 4.

4.2.2. Performance and career

We computed the AW-index and first filing year to measure the invention performance and career of the candidate inventors. The AW-index of each candidate inventor was computed through Eq. 1. We used the patent set containing the G06 K subclass to identify the candidate inventors' invention fields. However, it is difficult to represent inventors' overall performance and careers if only patents included in the patent set are used. Therefore, it is better to use all the candidate inventors' registered patents. Thus, we measured the candidate inventors' invention performance and career using all the registered patents.

Table 4 shows the process and result of calculating the AW-index and the first filing year of inventor 7798394-3. We first counted the number of citations of all registered patents owned by the inventor. Then, we calculated patent age as the difference between the base year and the patent filing year, setting the base year as 2020. Next, the number of citations was divided by the age of each patent to obtain the age-weighted citation. Finally, we obtained the AW-index by taking the square root of the sum of all the age-weighted citations of the inventor. We also obtained the first filing year by selecting the minimum value of the filing years of the registered patents owned by the inventor.

4.3. Discovering prospective inventors by performance-career portfolio maps

We generated performance-career portfolio maps for the candidate inventors in each technology field. Before generating the maps, we screened the inventors because there were too many candidate inventors in each technology field. Generating a map using all candidate inventors would reduce the map's readability and make it difficult to identify prospective inventors. Therefore, we built maps only for candidate inventors with more than five registered patents, an AW-index greater than 0, and with a first filing year between 2010 and 2019. PatentsView data for which we have collected patent data may not provide 100 % complete inventor identification information. Thus, we actually checked that there were no disambiguation issues for candidate inventors plotted on the maps.

We describe three examples of identifying prospective inventors for some of the 12 sub-technologies derived above. Table 5 shows a brief summary of the examples. As shown in the table, we describe scenarios for the three technology strategy directions of the target firm as examples. The first example is a scenario in which the target firm wants to expand the technology area through its existing technology. In this scenario, the target firm selects a technology field that can apply its technology capabilities. The selected technology field is similar to the capabilities of the target firm, but probably not

Table 5
Summary of examples in this section.

Example	Tech. #	Scenario
1	3	Extending technology areas based on existing technology capabilities
2	4	Entering a new technology area
3	9	Enhancing the existing technology capability

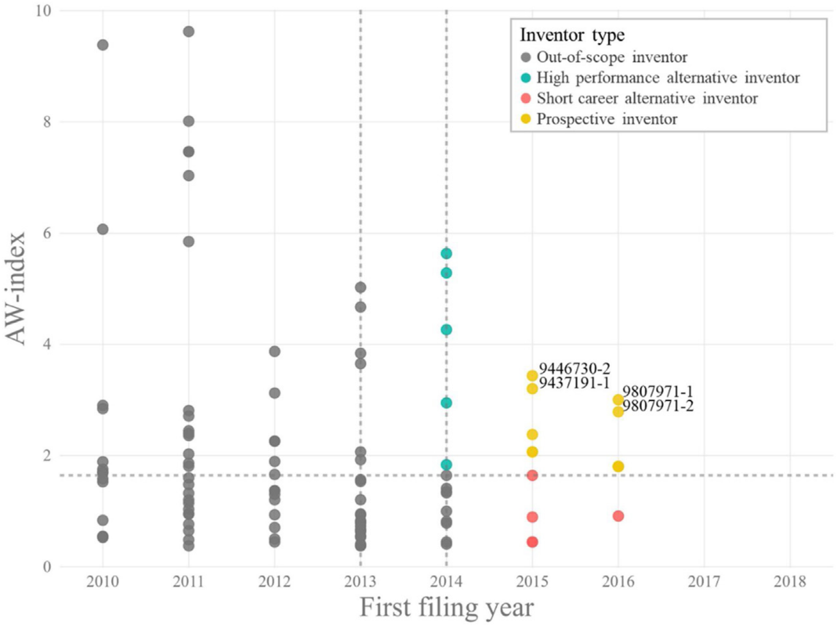

Fig. 6. Performance-career portfolio map of Tech. 3.

the major technology domain of the firm. The second example is when the target firm wants to enter a new technology area. In this case, the target firm may select a completely different technology field from its major technology domain. The last example is a scenario when the target firm wants to strengthen its existing technology capability. Unlike the previous example, the firm may select a technology field most similar to its existing capability. In the subsequent paragraphs, we explain how to identify human resources through performance-career portfolio maps for each scenario.

As a first example, if the target firm wants to extend technology areas based on its existing technology capabilities such as image formation and recognition, the firm can focus on inventors assigned to Tech. 3. Fig. 6 shows a performance-career portfolio map of this technology. There are 114 inventors on the map, of which 17 are prospective inventors. Among the prospective inventors, 9446730-2 and 9437191-1, with the first and second highest AW-index values, possess recognition and identification technology through image information that can be acquired inside the vehicle. Their AW-index values are 3.4327 and 3.1990, respectively. In addition, both have the same first filing year, 2015. Meanwhile, 9807979-1 and 9807971-2 have technologies related to vision systems that can detect specific parts of livestock. They have AW-index values of 3.000 and 2.7839, respectively, and started their careers in 2016. All of these inventors have AW-index values greater than Q_2^p (1.647) and first patent filing years shorter than Q_3^c (2014). Therefore, if the firm seeks human resources in order to develop capabilities related to Tech. 3, the prospective inventors marked with yellow nodes in Fig. 6 are the best choice.

The second example concerns Tech. 4. Fig. 7 shows a performance-career portfolio map of Tech. 4. If the target firm wants to implement a technology strategy for wireless communication, the firm can focus on inventors assigned to Tech. 4. Because wireless communication technology is different from the target firm's existing capabilities, it can be considered that the firm intends to newly enter the wireless communication field. There are two technologies, namely wireless communication, and video/image processing. Therefore, the firm should look at the invention contents of each inventor. For example, among the prospective inventors of Tech. 4, 9602795-3 has technologies for creating video and recommending events of user interest through video clips, and 9578279-1 has technologies for removing low-quality video frames and detecting discontinuities on the video frames. On the other hand, 9355285-2 has technologies for wireless transmitter-receivers for point-of-sale and antenna systems for NFC object readers, and 9988193-2 has contact card-charging technologies for transaction cards. Thus, if the firm wants to scout human resources related to wireless communication technology, the firm can select inventors such as 9355285-2 or 9988193-2. The target firm can therefore quantitatively reduce the number of candidate inventors in the targeted technology field through a performance-career map and then qualitatively evaluate the inventors' detailed invention contents to find the most suitable inventors.

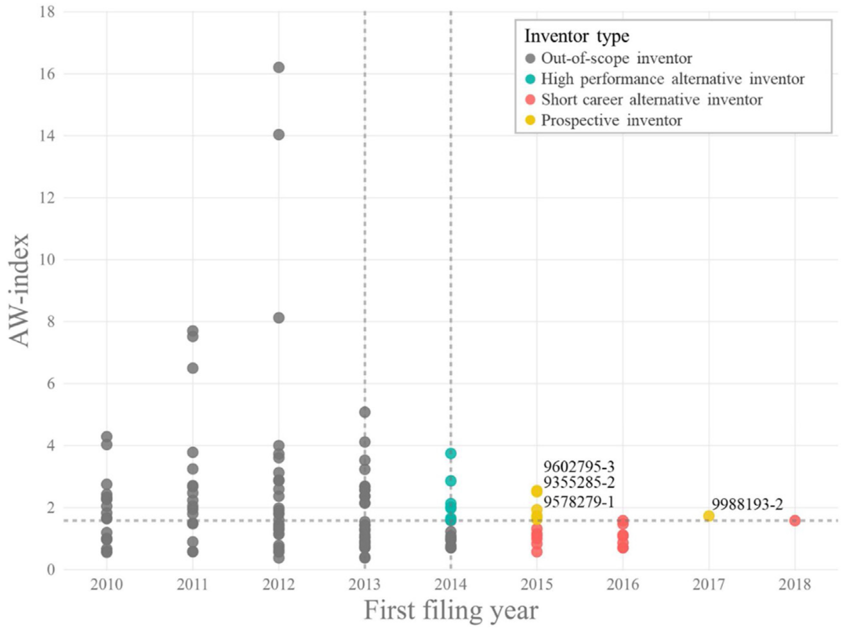

Fig. 7. Performance-career portfolio map of Tech. 4.

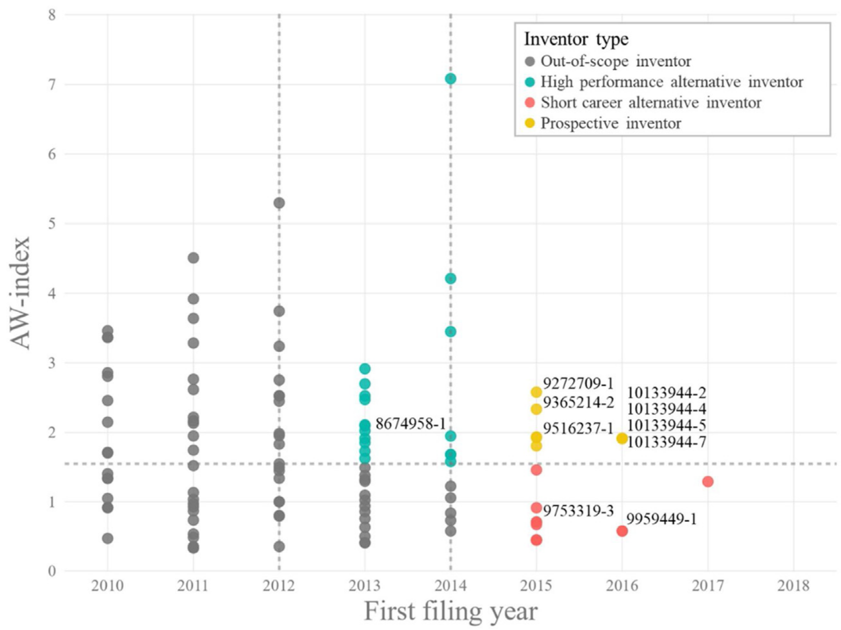

Fig. 8. Performance-career portfolio map of Tech. 9.

The last example is when the target firm wants to identify inventors useful for strengthening the firm's major technology domain. In this case, the firm can focus on candidate inventors of Tech. 9. This technology field is most similar to the major technology domain of the target firm. However, although the firm's main domain is only relevant to fingerprint data, Tech. 9 targets various data sources, such as images, prints, transactions, and fingerprints. Thus, the firm must screen inventors with technical knowledge of fingerprint data. Fig. 8 shows a performance-career portfolio map of Tech. 9. The prospective inventors of Tech. 9 have technologies for detecting and predicting the movement of vehicle occupants (10133944-2, 10133944-4, 10133944-5, 10133944-7), identifying traffic light conditions (9272709-1, 9365214-2), and identifying objects in images (9516237-1). They are excellent inventors with relatively high invention performance and short invention careers. They also have technical knowledge related to object and attribute identification. However, they do not have technology development experience with fingerprint data. Thus, if the firm has not been able to find suitable prospective inventors, alternative

Table 6
Number of inventors by group according to performance indicators.

	# of total prospective inventors	# of total short career inventors
AW-index based result	138	167
h-index based result	64	241

inventors with high performance or short careers can be considered. The alternative inventors are still superior to the out-of-scope inventors because they satisfy either the youngness or talent criterion. Among the alternative inventors, 8674958-1 is not very young but has relatively high performance. He also has 10 registered patents related to fingerprint identification, such as capacitive fingerprint sensors, accurate fingerprint imaging, fingerprint sensor arrays, and noise-removed fingerprint detection circuits. Meanwhile, 9753319-3 and 9959449-1 have lower invention performance than the prospective inventors but are young enough and have more than five patent registrations. For example, 9753319-3 has 11 registered patents related to fingerprint recognition devices and methods for screens and touch buttons, 9959449-1 has 33 registered patents such as a fingerprint identification method and device, a terminal unlocking method, fingerprint template improvement, and an update method. Therefore, the firm should first consider the prospective inventors in the targeted technology field. However, if no suitable prospective inventors are identified, the firm can then consider alternative inventors with high performance or short careers.

5. Discussion

5.1. Comparison between AW-index and h-index

In this study, the AW-index was used as an indicator to reasonably reflect the invention performance of young and talented inventors. In Section 3.2.2, we stated that the AW-index is more appropriate than the h-index to evaluate the performance of inventors with less invention experiences. However, there may still be questions about whether the AW-index really yields better results than the h-index. This section compares an AW-index based result and an h-index based result with the same data used in the case study. To do this, we calculated the h-index of the candidate inventors, and built performance-career portfolio maps only for candidate inventors selected based on the similar criteria as the case study: candidate inventors with more than five registered patents, an h-index greater than 0, and with a first filing year between 2010 and 2019.

We mentioned that there are cases where the AW-index is more rational than the h-index for inventors who have a small number of recent patents with a large number of citations. In fact, we can find examples through the comparison results based on the two performance indicators. For example, inventor '9609288-4' has three registered patents. The two patents he filed on March 11, 2016 have 31 and 23 citations, respectively, and the other patent filed on March 17, 2017 has one citation. His AW-index is 3.7193. It is true that despite his little experience, he made a pretty good inventions, so he is a prospective inventor of Tech. 6. However, his h-index is 2, and looking at the h-index based results, he is a short career inventor. Similarly, inventor '10223614-1', one of the prospective inventors of Tech. 1, has an AW-index of 3.3912. On the other hand, he is a short career inventor since his h-index is 2. Table 6 is a table representing the number of prospective inventors and the number of short career inventors derived from the AW-index based result and the h-index based result. According to the actual examples and the table, the h-index based result may have underestimated the performance of 74 prospective inventors who have enough technical talent.

In scouting human resources, there is a trade-off between the number of human resources to evaluate and the qualitative effort of experts. If there are too many screened human resources, the effort of experts is also high. On the other hand, if there are not so many screened human resources, there is a possibility that excellent human resources will be missed instead of requiring less expert effort. Therefore, the approach of this study should be able to screen an appropriate number of young and talented inventors. From this point of view, the h-index based result screened relatively too few prospective inventors. For example, in the case of Tech. 4 described as the second example in Section 4.3, there were a total of seven prospective inventors in the AW-index based result. However, as shown in Fig. 9, there is no prospective inventor in Tech. 4 in the h-index based result. In this case, HRM experts have to invest a lot of effort because they have to evaluate too many alternative inventors. Tech. 4 is a particularly bad case, but the h-index based result generally screened fewer prospective inventors than the AW-index based result. As a result, the AW-index better represents the invention performance of young and talented inventors and screens an appropriate number of prospective inventors compared to the h-index.

5.2. Statistical validation of proposed approach

The approach proposed in this study can screen prospective inventors who are more likely to become experts in the future. This approach assumes that such prospective inventors will contribute to a firm for a long time because they are now young and talented and are highly likely to grow into well-established inventors. Thus, this section validates whether prospective inventors are indeed likely to become well-established as well as the excellence of prospective inventors. The validation was performed in two ways: by 1) performing a *t*-test of inventors' average invention performance after five years, and 2)

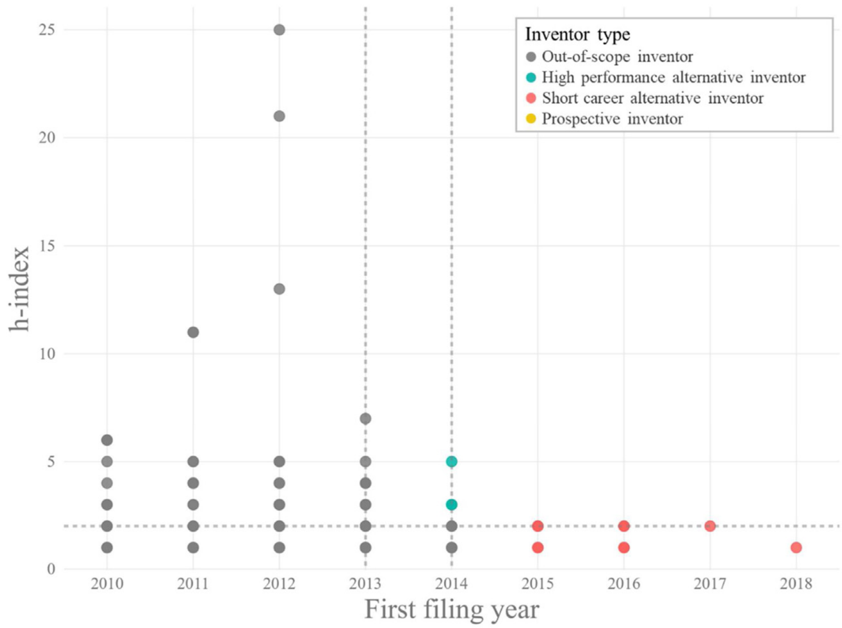

Fig. 9. Performance-career portfolio map of Tech. 4 based on h-index.

Table 7
Summary of t-test results.

Case	i	X	$N_{X,i}$	$\mu_{X,i}$	Std. error	t-statistic	p-value	DF	95 % of confidence interval	
									Lower	Upper
1	2010	P	76	6.77	0.40	9.16	0.00	94	3.06	4.76
		S	179	2.85	0.14					
2	2015	P	131	4.84	0.26	11.69	0.00	146	2.60	3.66
		S	177	1.71	0.07					

*$N_{X,i}$: the total number of inventors by group (X) in year i.
*Significance level (α) = 0.05.

comparing the numbers and ratios of inventors who showed the highest invention performance for each technology field after five years. We compared prospective inventors to short career inventors because they had similar invention careers.

For three base years (2010, 2015, and 2020), we measured the candidate inventors' performance and identified their inventor types for the same target firm as that considered in the case study. The AW-index values of the inventors who started their invention in the last 10 years of each base year were calculated. For example, if the base year is 2010, the AW-index for 2010 would be calculated for inventors who started their invention between 2000 and 2009. Similarly, if the base year is 2015, the AW-index for 2015 would be calculated for inventors whose first filing year is between 2005 and 2014. The AW-index for the 2020 base year is the same as in the case study results. After the candidate inventors' AW-index was measured for all base years, prospective inventors and short career alternative inventors for each technology field were identified for each base year.

5.2.1. Comparison of average performance after five years

As a first validation, we verified that the prospective inventors generally performed better than short career alternative inventors in the next base year. We conducted two t-tests to statistically compare invention performance mean values of the prospective inventors and the short career alternative inventors after five years. The first test compared the mean values of the 2015 AW-index of the two inventor groups in 2010. The second compared the mean values of the 2020 AW-index of the two inventor groups in 2015. In both cases, two-tailed t-tests for unequal sample size and unequal variance were used. The null hypothesis of the first case was $\mu_{P,2010} = \mu_{S,2010}$, while the alternative hypothesis was $\mu_{P,2010} \neq \mu_{S,2010}$. The null hypothesis of the second case was $\mu_{P,2015} = \mu_{S,2015}$ and the alternative hypothesis was $\mu_{P,2015} \neq \mu_{S,2015}$. Here, $\mu_{X,i}$ denotes the mean value of the AW-index for prospective inventors (P) and short career alternative inventors (S) in year i after five years. As shown in Table 7, both cases had very small p-values compared to the significance level of 0.05. Therefore, the results of the two t-tests indicated a statistically significant difference between the mean values of invention performance

Table 8
Number and ratio of inventors in 2010 who belong to the Top N in 2015.

N	Prospective inventors ($N_{P,2010} = 76$)		Short career alternative inventors ($N_{S,2010} = 179$)	
	# of top N inventors	Ratio of top N inventors	# of top N inventors	Ratio of top N inventors
10	28	0.37	11	0.06
30	63	0.83	37	0.21
50	72	0.95	70	0.39

*$N_{X,i}$: the total number of inventors by group (X) in year i.
* *Ratio of top N inventors = # of top N inventors / $N_{X,2010}$.*

Table 9
Number and ratio of inventors in 2015 who belong to the Top N in 2020.

N	Prospective inventors ($N_{P,2015} = 131$)		Short career alternative inventors ($N_{S,2015} = 177$)	
	# of top N inventors	Ratio of top N inventors	# of top N inventors	Ratio of top N inventors
10	52	0.40	0	0.00
30	115	0.88	34	0.19
50	131	1.00	101	0.57

*$N_{X,i}$: the total number of inventors by group (X) in year i.
* *Ratio of top N inventors = # of top N inventors / $N_{X,2015}$.*

between the two inventor groups after five years. In other words, these results can support that the prospective inventors identified through our approach will perform better than other inventors with similar invention careers.

5.2.2. Comparison of number and ratio of top inventors after five years

As a second validation, we verified that the prospective inventors are more likely to become top-ranked inventors than the short career alternative inventors in the next base year. For each inventor group, we obtained the number of top N inventors, which is the number of inventors who belonged to the top N of invention performance in the target base year for each technology field among the inventors of the previous base year. For each group, we then calculated the ratio of top N inventors by dividing the number of top N inventors by the total number of inventors in the previous base year. We computed the number and ratio in two cases: 1) the number and ratio of inventors in 2010 among the top N in 2015 (Table 8) and 2) the number and ratio of inventors in 2015 among the top N in 2020 (Table 9). We compared the two groups based on these ratios because each group had a different number of inventors. As shown in both tables, the ratio values of the prospective inventors were all greater than those of the short career alternative inventors. These results can support the point that the prospective inventors identified through the proposed approach are more likely to attain the highest invention performance than the alternative inventors with similar invention careers.

5.3. Academic contributions and managerial implications

This study presented an approach for identifying human resources suitable for a target TBF's technology strategy based on patent data. TBFs must understand the technological contents and capabilities of human resources when scouting resources to assist in the implementation of the firms' technology strategies. From a forward-looking perspective, TBFs should recruit and train young human resources to obtain experts who can continuously contribute to the firms' long-term development. To that end, our approach can identify human resources with a short career but excellent invention performance in the target firm's focal technology field. This study offers several academic contributions and managerial implications.

From the academic point of view, this study proposed a patent-based human resource identification approach that takes both the performance and career of inventors into consideration. For firms' human resource scouting results to be effective, new human resources should be young inventors with the potential to become experts. In addition, especially for TBFs, new resources should have experience in patent registration resulting from technology development. However, prior studies have limitations in identifying human resources suitable for TBFs. For example, most studies do not properly assess the technical capabilities or experience of human resources. The studies also do not reflect the youngness of human resources. Therefore, the present study is unique in three ways. First, this study used patent data rather than academic or web data. Human resources in TBFs generally disclose their R&D results as patents. However, several studies have used only academic or web data to assess human resources. By using patent data, our approach can evaluate both the R&D experience and performance of human resources. Second, this study considered the youngness of human resources. It is difficult to represent this factor through indicators such as the degree of performance improvement or potential. For this reason, we considered the youngness of human resources through their first patent filing year, enabling our approach to identify human resources who are more likely to be young from an invention career perspective. Finally, this study used the AW-index to measure young human resources' performance effectively. Indicators such as the number of citations and the h-index do not clearly reflect the performance of young human resources with relatively little invention experience. However, the AW-index used in this study can help identify prospective inventors who have performed well despite having little invention experience.

Overall, our approach allows TBFs to identify young and talented human resources that can contribute over the long term suitable for the firms.

From the managerial point of view, our approach can be used by managers as a tool to monitor human resources according to a firm's technology development strategy. This managerial implication can be explained from two perspectives. First, our approach allows firms to identify different human resources depending on the targeted technology strategy. Generally, TBFs plan and execute technology strategies in various directions, such as entering new technology fields or strengthening their existing technologies. Thus, it is necessary to accurately determine the technology fields of the candidate human resources and identify those with the capabilities appropriate to each of the firm's technology strategies. Most prior studies merely identify the human resources that match a system user's needs and fail to consider their research or invention fields. By contrast, we identified the invention field of each inventor and generated performance-career portfolio maps for each invention field. This allows TBFs to select young and talented human resources for each targeted technology field. Second, our approach can screen candidate inventors that firms should consider when scouting human resources. Evaluating a large number of human resources is essential but requires a great deal of qualitative effort from HRM experts. Our approach can reduce the number of human resources the firm needs to investigate by identifying prospective or alternative inventors, saving much qualitative effort. Consequently, this approach is a tool to support human resource assessment and monitoring, so firms can use this approach to support decision-making during human resource scouting.

6. Conclusion

This study proposed an inventor profile mining approach to identify prospective human resources. Prospective inventors, the human resources identified through this approach, have the technical knowledge needed by TBFs and can immediately become active in the firm's technology strategy. The prospective inventors are young and talented human resources who can grow into well-established inventors and contribute to the sustainable growth of the firm. We analyzed patent data to evaluate the invention capabilities of human resources: invention field, invention performance, and career. The invention field was derived from LDA-based topic modeling, while invention performance and career were measured by the AW-index and first patent filing year, respectively. Performance-career portfolio maps were then generated for each invention field. The prospective inventors were identified in the performance career maps as inventors whose invention performance was greater than the second quartile and whose invention career was shorter than the third quartile. The application of this approach was demonstrated using a target firm, *Suprema*, a biometric-security SME. We identified prospective and alternative inventors in three technology fields: 1) image formation, feature extraction, and object recognition; 2) wireless communication and video/image data processing; and 3) object and attribute information identification in print, image, and fingerprint data. We then statistically verified that the prospective inventors identified by our approach were more likely to become well-established than other inventors with similar invention careers. Finally, we discussed the academic contributions and managerial implications of this study.

Despite the value of this study, it has several limitations, which provide future research possibilities. First, for each inventor, the technology that showed the greatest contribution in the inventor-technology matrix was assigned to the invention field. However, each inventor may have technical capabilities for multiple invention fields. Therefore, future studies should set a threshold for contribution values, so that each inventor can be assigned to a plurality of technology fields. Second, in identifying young inventors, we replaced inventors' youngness with their invention career. However, a short invention career may not always imply youngness. Also, we believe there will be indicators that can better represent the youngness of inventors than the first filing year used in this study. Thus, future studies should develop indicators that could replace the inventor age more accurately. Third, PatentsView provides improved inventor identification information, but there are still issues related to inventor disambiguation. Although this case has been found to have no such problem, the inventor disambiguation problem may affect the reliability of our approach. However, in the near future, if the USPTO's efforts to complete the inventor disambiguation problem come to fruition, the credibility of the proposed approach in this study is expected to increase. Fourth, there are many profile information of inventors, such as LinkedIn profile data, inventors' curriculum vitae, and publications, etc. Like patent data used in this study, this information can also represent the inventors' technical capabilities and experiences. Therefore, we think that an approach to comprehensively evaluate the competence of human resources can be proposed if various human resource data are collected and detailed inventor identification work is performed in the future studies. Finally, although the approach presented in this study screened candidate inventors, recruiting human resources still requires experts' qualitative efforts. Therefore, further studies should develop a method to prioritize identified prospective inventors. The prioritization method will not only minimize the qualitative effort required for human resource scouting, and we can more reasonably verify the performance of the method by comparing it with other state-of-the-art techniques such as the latest recommendation algorithms or multi-criteria decision making techniques.

Author contributions

Jaemin Chung: Conceived and designed the analysis, Performed the analysis, Wrote the paper.
Namuk Ko: Contributed data or analysis tools, Wrote the paper.
Hyeonsu Kim: Collected the data, Contributed data or analysis tools.

Janghyeok Yoon: Conceived and designed the analysis, Wrote the paper.

Acknowledgement

This paper was supported by Konkuk University in 2018.

References

Afzal, M. T., & Maurer, H. A. (2011). Expertise recommender system for scientific community. *Journal of Universal Computer Science*, *17*(11), 1529–1549.
Agiakloglou, C., Drivas, K., & Karamanis, D. (2016). Individual inventors and market potentials: Evidence from US patents. *Science and Public Policy*, *43*(2), 147–156.
Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289.
Andersson, H., & Berggren, C. (2007). Individual inventors in the R&D factory. *Creativity and Innovation Management*, *16*(4), 437–446.
Becker, B., & Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *The Academy of Management Journal*, *39*(4), 779–801.
Bercovitz, J. E., & Feldman, M. P. (2007). Fishing upstream: Firm innovation strategy and university research alliances. *Research Policy*, *36*(7), 930–948.
Blei, D. M., & Lafferty, J. D. (2009). Topic models. In *Text mining*. pp. 101–124. Chapman and Hall/CRC.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(January), 993–1022.
Bornmann, L., & Marx, W. (2011). The h index as a research performance indicator. *European Science Editing*, *37*(3), 77–80.
Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H.-D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, *5*(3), 346–359.
Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine*.
Buller, P. F., & McEvoy, G. M. (2012). Strategy, human resource management and performance: Sharpening line of sight. *Human Resource Management Review*, *22*(1), 43–56.
Charan, R., Drotter, S., & Noel, J. (2010). . *The leadership pipeline: How to build the leadership powered company* (Vol. 391) John Wiley & Sons.
Cifariello, P., Ferragina, P., & Ponza, M. (2019). Wiser: A semantic approach for expert finding in academia based on entity linking. *Information Systems*, *82*, 1–16.
Cucchetti, A., Mazzotti, F., Pellegrini, S., Cescon, M., Maroni, L., Ercolani, G., et al. (2013). The use of the Hirsch index in benchmarking hepatic surgery research. *The American Journal of Surgery*, *206*(4), 560–566.
Daud, A., Abbasi, R., & Muhammad, F. (2013). Finding rising stars in social networks. *International Conference on Database Systems for Advanced Applications*.
Daud, A., Ahmad, M., Malik, M., & Che, D. (2015). Using machine learning techniques for rising star prediction in co-author network. *Scientometrics*, *102*(2), 1687–1711.
Daud, A., Aljohani, N. R., Abbasi, R. A., Rafique, Z., Amjad, T., Dawood, H., et al. (2017). Finding rising stars in co-author networks via weighted mutual influence. *Proceedings of the 26th International Conference on World Wide Web Companion*.
Davoodi, E., Kianmehr, K., & Afsharchi, M. (2013). A semantic social network-based expert recommender system. *Applied Intelligence*, *39*(1), 1–13.
Ernst, H., Leptien, C., & Vitt, J. (2000). Inventors are not alike: The distribution of patenting output among industrial R&D personnel. *IEEE Transactions on Engineering Management*, *47*(2), 184–199.
Ferrucci, E., & Lissoni, F. (2019). Foreign inventors in Europe and the United States: Diversity and patent quality. *Research Policy*, *48*(9), p. 103774.
Franceschini, F., & Maisano, D. (2012). Publication and patent analysis of European researchers in the field of production technology and manufacturing systems. *Scientometrics*, *93*(1), 89–100.
Göktepe-Hultén, D. (2008). Academic inventors and research groups: Entrepreneurial cultures at universities. *Science and Public Policy*, *35*(9), 657–667.
Guan, J. C., & Gao, X. (2009). Exploring the h-index at patent level. *Journal of the American Society for Information Science and Technology*, *60*(1), 35–40.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, *102*(46), 16569–16572.
Jeong, B., & Yoon, J. (2017). Competitive intelligence analysis of augmented reality technology using patent information. *Sustainability*, *9*(4), 497.
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. *International Journal of Information Management*, *48*, 280–290.
Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R-and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.
Khey, D. N., Jennings, W. G., Higgins, G. E., Schoepfer, A., & Langton, L. (2011). Re-ranking the top female academic "stars" in criminology and criminal justice using an alternative method: A research note. *Journal of Criminal Justice Education*, *22*(1), 118–129.
Kiehne, D.-O., & Krill, B. (2017). The influence of the amount of inventors on patent quality. *Management*, *47*, 300–308.
Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, *117*, 228–237.
Ko, N., Jeong, B., Choi, S., & Yoon, J. (2017). Identifying product opportunities using social media mining: Application of topic modeling and chance discovery theory. *IEEE Access*, *6*, 1680–1693.
Lepak, D. P., Takeuchi, R., & Snell, S. A. (2003). Employment flexibility and firm performance: Examining the interaction effects of employment mode, environmental dynamism, and technological intensity. *Journal of Management*, *29*(5), 681–703.
Li, X.-L., Foo, C. S., Tew, K. L., & Ng, S.-K. (2009). Searching for rising stars in bibliography networks. *International Conference on Database Systems for Advanced Applications*.
Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., et al. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Research Policy*, *43*(6), 941–955.
Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management*, *41*(6), 1462–1480.
Long, H., Boggess, L. N., & Jennings, W. G. (2011). Re-assessing publication productivity among academic "stars" in criminology and criminal justice. *Journal of Criminal Justice Education*, *22*(1), 102–117.
Mairesse, J., & Sassenou, M. (1991). *R&D productivity: A survey of econometric studies at the firm level*. National Bureau of Economic Research.
Mariani, M., & Romanelli, M. (2007). "Stacking" and "picking" inventions: The patenting behavior of European inventors. *Research Policy*, *36*(8), 1128–1142.
Moehrle, M. G., Walter, L., Geritz, A., & Müller, S. (2005). Patent-based inventor profiles as a basis for human resource decisions in research and development. *R&D Management*, *35*(5), 513–524.
Moreira, C., & Wichert, A. (2013). Finding academic experts on a multisensor approach using Shannon's entropy. *Expert Systems with Applications*, *40*(14), 5740–5754.
Naeem, M., Khan, M. B., & Afzal, M. T. (2013). Expert discovery: A web mining approach. *Journal of AI and Data Mining*, *1*(1), 35–47.
Narin, F., & Breitzman, A. (1995). Inventive productivity. *Research Policy*, *24*(4), 507–519.
Nie, Y., Zhu, Y., Lin, Q., Zhang, S., Shi, P., & Niu, Z. (2019). Academic rising star prediction via scholar's evaluation model and machine learning techniques. *Scientometrics*, *120*(2), 461–476.
No, H. J., & Lim, H. (2009). Exploration of nanobiotechnologies using patent data. *The Journal of Intellectual Property*, *4*(3), 109–129.

Oldroyd, J. B., & Morris, S. S. (2012). Catching falling stars: A human resource response to social capital's detrimental effect of information overload on star employees. *Academy of Management Review*, *37*(3), 396–418.
Panagopoulos, G., Tsatsaronis, G., & Varlamis, I. (2017). Detecting rising stars in dynamic collaborative networks. *Journal of Informetrics*, *11*(1), 198–222.
Park, H., Yoon, J., & Kim, K. (2012). Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics*, *90*(2), 515–529.
Patel, V. M., Ashrafian, H., Bornmann, L., Mutz, R., Makanjuola, J., Skapinakis, P., et al. (2013). Enhancing the h index for the objective assessment of healthcare researcher performance and impact. *Journal of the Royal Society of Medicine*, *106*(1), 19–29.
Rousseau, R., & Leuven, K. (2008). Reflections on recent developments of the h-index and h-type indices. *COLLNET Journal of Scientometrics and Information Management*, *2*(1), 1–8.
Schettino, F., Sterlacchini, A., & Venturini, F. (2013). Inventive productivity and patent quality: Evidence from Italian inventors. *Journal of Policy Modeling*, *35*(6), 1043–1056.
Sun, J., Xu, W., Ma, J., & Sun, J. (2015). Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework. *International Journal of Production Economics*, *165*, 185–193.
Velmurugan, C., & Radhakrishnan, N. (2015). A scientometric analysis of research papers published on pharmacognosy as reflected in the web of science. *Advances in Pharmacognosy and Phytomedicine*, *1*(1), 27–40.
Wang, B., Liu, S., Ding, K., Liu, Z., & Xu, J. (2014). Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: A case study in LTE technology. *Scientometrics*, *101*(1), 685–704.
Wang, Q., Ma, J., Liao, X., & Du, W. (2017). A context-aware researcher recommendation system for university-industry collaboration on R&D projects. *Decision Support Systems*, *103*, 46–57.
Xu, Y., Guo, X., Hao, J., Ma, J., Lau, R. Y., & Xu, W. (2012). Combining social network and semantic concept analysis for personalized academic researcher recommendation. *Decision Support Systems*, *54*(1), 564–573.
Yang, M.-s., Kang, N.-k., Kim, Y.-j., Kim, J.-s., Choi, K.-n., & Kim, Y.-k. (2013). Expert recommendation system based on analyzing expertise and networks of human resources in National Science & Technology Information Service. *Journal of Central South University*, *20*(8), 2212–2218.
Yang, G.-C., Li, G., Li, C.-Y., Zhao, Y.-H., Zhang, J., Liu, T., et al. (2015). Using the comprehensive patent citation network (CPC) to evaluate patent value. *Scientometrics*, *105*(3), 1319–1346.
Zhang, J., Ning, Z., Bai, X., Wang, W., Yu, S., & Xia, F. (2016). Who are the rising stars in academia? In *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*.
Zhang, J., Xia, F., Wang, W., Bai, X., Yu, S., Bekele, T. M., et al. (2016). Cocarank: A collaboration caliber-based method for finding academic rising stars. *Proceedings of the 25th International Conference Companion on World Wide Web*.
Zhu, L., Zhu, D., Wang, X., Cunningham, S. W., & Wang, Z. (2019). An integrated solution for detecting rising technology stars in co-inventor networks. *Scientometrics*, *121*(1), 137–172.