IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 16, 2017, accepted October 11, 2017, date of publication October 26, 2017, date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2764478

Monitoring Newly Adopted Technologies Using Keyword Based Analysis of Cited Patents

SUNGHYUN NAM AND KWANGSOO KIM
Industrial and Management Engineering Department, Pohang University of Science and Technology, Pohang 37673, South Korea

Corresponding author: Kwangsoo Kim (kskim@postech.ac.kr)

This work was supported by the National Research Foundation of Korea through the Ministry of Science, ICT and Future Planning under Grant NRF-2016R1A2B4008381.

ABSTRACT This paper proposes a method that can reliably monitor the adoption of existing technology by term frequency-inverse document frequency (TF-IDF) and K-means clustering using cited patents. TF-IDF and K-means clustering can extract patent information when the number of patents is sufficiently large. When the number of patents is too small for TF-IDF and K-means clustering to be reliable, the method considers patents that were cited by the originally set of patents. The mixed set of citing patents and cited patents is the new subject of analysis. As a case study, we have focused in agricultural tractor in which new technologies were adopted to achieve automated driving. TF-IDF and K-means clustering alone failed to monitor the adoption of new technology but the proposed method successfully monitored it. We anticipate that our method can ensure the reliability of patent monitoring even when the number of patents is small.

INDEX TERMS Technology planning, technology transfer, patents, technology monitoring, keyword-based model.

I. INTRODUCTION

FIRMS invest aggressively in new technologies to increase the number and quality of products that are provided to customers [1]. This is a reasonable investment strategy because technology is one of the most common factors that affects the success of new product development, directly or indirectly [2]–[4]. New products produced by new technologies for new market have resulted in between 40% and 90% of the increase in national wealth in most countries [5]. However, most companies that are undergoing successful growth are facing pressure to develop new technology and products because they are changing rapidly [6]. Therefore, tracking the emergence of new technology or adoption of new technology has an important challenge for firms.

A patent is a reliable technical document that includes important contemporary technological information [7]. One can use this information to monitor competitors, to assess technology, to manage R&D portfolio, and to identify and assess potential sources for the external generation of technological knowledge [8].

Tasks performed by exploiting patent information can be grouped into three classes: patent search, patent analysis, and patent monitoring [9]. Curran and Leker [10] have monitored convergence of nutraceuticals and functional foods and telecommunications disciplines using the number of patents in which main and secondary IPCs could be co-classified, and by quantifying the share of co-classified patents according to IPCs. Yoon and Kim [11] tried to identify rapidly evolving technological trend by patent network based on SAO (subject-action- object)-based semantic patent analysis. Lee *et al.* [12] monitored trends of technological change by Formal Concept Analysis (FCA)-based dynamic patent lattice.

Many patent analyses adopt K-means clustering [13]–[15], the problem of data separation remains. The efficiency of the K-means clustering algorithm increases as the separation between clusters increases. However, representing documents by their word tokens cannot guarantee adequate separation of documents to cluster the document set under certain conditions: e.g., when the documents share most of their tokens, or distinctive tokens are not separable due to noise. For example, when a small number of patent documents share the same 4-digit or 5-digit IPC code, K-means clustering is not a suitable analysis tool. Instead, similarities between documents can be calculated based on similarities in the keywords found in each document [16], or by adopting FCA [12] while considering patent context [17].

However, when the documents contain technical terms and the number of documents is small, keyword similarities are difficult to calculate, and the patent context can be difficult to understand. Word hierarchy structure such as WordNet can solve this problem, but the use of specialized technical terms complicates the analysis. Therefore, the primary purpose of this study is to propose a method that uses patents analysis to monitor adoption of new technologies reliably when the number of patents is small. We believe that cited patents can provide additional power to TF-IDF and K-means clustering methods without distorting the information carried by the patents. There exist more sophisticate tools for mining information from text but we chose commonly used tools to ensure that our proposed method is valid even with less sophisticate tools.

The rest of this paper is organized as follows. Section II briefly provides the theoretical background of the proposed method. Section III describes the proposed method. Section IV describes a case study of the steering mechanism of agricultural machines or implements. Section V provides conclusions and suggests future work.

II. BACKGROUNDS

A. INTERNATIONAL PATENT CLASSIFICATION

Patent information can be used in three important areas of technology management. (1) To obtain relevant information about the competitor's R&D strategies and to assess the competitive potential of technologies. (2) To identify and assess options for the external generation of technological knowledge. (3) To store relevant knowledge as a core element of knowledge management [8].

The retrieval of patent documents is crucial to patent-issuing authorities, potential inventors, research and development units, and other users who are concerned with the application or development of technology [18]. To facilitate the process of retrieval, patent include bibliographic information such as applicants, inventors, registration dates, citation, and patents classification codes. To exploit such information, various technology classifications have been used by different institutions [19]; existing systems include the United States Patent Classification (USPC), European Classification (ECLA), Cooperative Patent Classification (CPC), International Patent Classification (IPC).

IPC classifies patents hierarchically. IPC is a code that consists of sections, classes, sub-classes, main group, and sub-groups. Many studies utilized this hierarchy system in their research under different levels, depending on the purpose of the research, and in the method being used. Leydesdorff *et al.* [20] developed basemaps and overlays using 3-digit and 4-digit levels of the IPC. Engelsman and van Raan [21] proposed a co-citation map that use 2-digit IPC codes. Park and Yoon [22] have conducted co-classification analysis under 4-digit IPC code.

B. TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY

Term frequency – inverse document frequency (TF-IDF) [23] is used when individual documents are needed to be distinguished from collection of documents [24]. TF is the number of times a certain term occurs in a document, and IDF calculates a logarithmically-scaled inverse number of documents in which that term has appeared: i.e.,

$$\mathrm{w_{i,j}} = tf_{i,\mathrm{j}} * \log(\frac{\mathrm{N}}{df_{\mathrm{i}}}) \tag{1}$$

where $\mathrm{w_{i,j}}$ is the TF-IDF value of a term i in document j, and $tf_{i,\mathrm{j}}$ is the frequency of the term i in document j, and N is the total number of documents, and df_{i} is the number of documents which contains the term i [25]. High TF value suggests that the term is widely used, and high IDF value indicates that the term is uncommonly used. Therefore, we can conclude that terms with high TF-IDF value can distinguish or represent a document aside from others [26]. In this paper, we use the basic TF-IDF formula among many variations.

C. K-MEANS CLUSTERING

K-means clustering finds a partition such that the squared error between empirical mean of a cluster and the squared error between the empirical mean of a cluster and the points in the cluster is minimized [27]. It is also known as Lloyd's algorithm [28] and is expressed as

$$J(S) = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{2}$$

where μ_i is the mean of cluster S_i, and K is the number of the cluster. The goal is to find the value K that minimizes $J(S)$. K-means clustering starts with a random initial partition and keeps reassigning the patterns to clusters based on the similarity between the pattern and the cluster centers until a convergence criterion is met [29].

Text documents can be analyzed effectively and efficiently by clustering [30]. To analyze text documents by clustering, text must be transformed to numeric form; one popular way to do this is to use a vector space model [31]. In this model, unique content-bearing words are extracted from documents as features, and individual documents are represented as vectors in this feature space [32].

III. METHOD

A. DATA SELECTION AND TF-IDF CALCULATION

This method begins with collection of patents' raw data. Issued patents in the target IPC can be downloaded from the USTPO website. When the number of patents is too small for TF-IDF and K-means clustering to be reliable, the method then searches for patents; that have been cited by the originally-collected patents.The mixed set of citing patents and cited patents is the new subject of analysis. The title, abstract, patent number, and date of patent are extracted to construct a database.

The next phase is to tokenize title and abstract to assemble a relevant vocabulary. For each vocabulary, the document frequency is counted and tokens are excluded if their document frequency is remarkably high or low [33].

Remarkably high document frequency means that the token appears in almost document in the set, and therefore cannot distinguish among documents. Remarkably low document frequency means thatthe token appears in only a few documents in the set, and therefore may excessively separate these documents from the others. An expert can set the threshold values to eliminate these two classes of terms.

After the vocabulary assembly is complete, the TF-IDF value is calculated for each token. The output of this process is a matrix in which the columns represent patents, and rows represent tokens; each cell contains the TF-IDF value of a token in a document.

B. CLUSTERING AND CLUSTER REPRESENTATION

In this stage, patents are clustered using the K-means clustering algorithm and some keywords are selected to represent each cluster. Using the result of 3.1., each document can be represented as a vector in the vector space. In this space, documents could be clustered using K-means clustering algorithm. An expert uses heuristic methods to determine the number of clusters. Ideally, the number of clusters should be set equal to the number of technology types in the IPC considered. However, patents may not be issued in every technology type. Also, when the adoption of new technology is monitored by considering keywords, patents with new technology tend to form a new cluster. The following three steps would help to find the appropriate number of clusters.

First, an initial estimate (n, x, y) set is set arbitrarily, where n is the number of clusters, x is the maximum document frequency and y is the minimum document frequency. Only tokens with document frequency between x and y are retained in the vocabulary. Then the K-means algorithm is run and six keywords that are closest to the cluster centroid are chosen. Then, (x, y) set are reset and whether the selected keywords representing a single technology is evaluated. Third, n is changed according to the result of previous step, the second step is repeated until an appropriate (n, x, y) set is found. The ideal (n, x, y) set should have minimum values of n and y, and the maximum value of x.

C. IDENTIFYING NEWLY ADOPTED TECHNOLOGY

Using the optimized (n, x, y) set from 3.2, the TF-IDF values are counted and patents are clustered to find newly-adopted technology. New technology can be identified by distinctive keywords, which might form a new cluster, or appear in an existing cluster. Even though a patent contains new technology, TF-IDF and the K-means clustering algorithm might include the patent in an existing cluster under certain conditions. When the new technology is described by a small number of distinctive terms, patent with the new technology might be positioned near patents that use conventional technology. Patents with new technology will form new cluster only if it is described with a sufficient number of new terms. Therefore, adoption of new technology should be recognized not by the appearance of a new cluster, but by the use of a new keyword.

When a new keyword is monitored, an expert can investigate the patent that uses the keyword, to determine whether the patent presents new technology. If the patent within a cluster has an IPC of interest, expert can analyze the patent in detail because it means that the patent is our original subject, which is citing patent. If the patent is not under IPC of interest, find citing patent and expert can analyze the citing patent in detail.

IV. CASE STUDY

A case study of patents with IPC A01B69 is presented to illustrate the suggested method. In IPC 2017.01 version, A01B69 corresponds to ''Steering of agricultural machines or implements; Guiding agricultural machines or implements on a desired track'' and has four subgroups. This IPC is suitable example for two reason. First, from 2006.01 to 2016.06, at least five patents were issued per year. Also without our proposed method, TF-IDF and K-means clustering algorithm failed to identify new technology. Second, self-driving or driverless tractors have become an important research topic in agricultural vehicles since 2010. This feature requires new foreign-domain technology such as GPS or computer vision analysis. Monitoring this IPC to identify such technology adoption would prove that this proposed method is valid.

A. DATA COLLECTION

We have downloaded every issued U.S. patent from January 2006 to June 2016. By selecting patents for which the main IPC was A01B69, 137 patents were found. These patents cited 570 different patents. After eliminating duplicated patents, 696 were selected. From these the title, abstract, patent number, date of patent were extracted.

B. CHECK NEWLY ADOPTED TECHNOLOGY FROM ORIGINAL PATENTS

To ensure the validity of our proposed method, we have analyzed 137 patents that had IPC A01B69 to check whether TF-IDF and K-means clustering algorithm can identify the new technology. We tested all combination of $n \in (2,3,4,5)$, $x \in (0.6,0.8)$, $y \in (0.01,0.1)$ when n is cluster number, x is maximum document frequency, and y is minimum document frequency. This approach did not find any newly adopted technology when it used any of these (n, x, y) settings.

C. TF-IDF CALCULATION

From the 137 citing patents, we extracted 12,699 tokens from the title and abstract; stopwords were excluded. To include compound words, we also used 2-grams. From the total of 696 patents, 17,127 tokens were found. Using TF-IDF, we can place 696 patents in a 17,127-dimensional vector space.

D. CLUSTERING

Before clustering, we set two thresholds to filter tokens with remarkable high or low document frequency. Under the same (n, x, y) condition as the experiment done in 4.2., we analyzed the 696 patents. Six keywords were obtained from each cluster. Every (n, x, y) included a new technology that was represented by words 'image', 'sensor' or 'detect'. These keywords were not shown when the same method was applied to the 137 patents that had IPC A01B69.

Judging by the representativeness of keywords, we conclude that (3, 0.8, 0.01) achieved the best clusters of the patents. Under that condition, 67 patents form a single cluster.Variety of 3-digit IPCs were found in the cluster (see Fig. 1). IPC B60Q had the most patents (14 patents). B60Q corresponds to "arrangement of signaling or lighting devices, the mounting or supporting thereof circuits therefor, for vehicles in general". This technology is not new technology because a tractor is a subset of vehicle. G02B and G06K are meaningful IPCs. G02B corresponds to "optical elements, system, or apparatus" and G06K corresponds to "recognition of data; presentation of data; record carriers; handling record carriers". The keywords 'image', 'light', 'display', 'mirror' properly represent such technology.

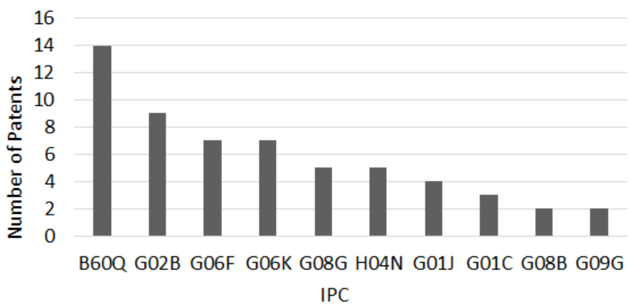

FIGURE 1. Number of patents found in top 10 IPC within the cluster.

The 67 patents found in the cluster were cited by 9 patents under A01B69 IPC (see Table 1). We can classify parking assistance, driving assistance, and ground condition recognition as 'self-driving technology', which is new type of technology in this domain. In case of tractor, checking the ground condition is important in self-driving. Parking assistant and rear collision technology uses camera to achieve its goal. They are useful in self-driving, but it can be used without self-driving. They are not directly applicable to, but important elements to self-driving.Therefore, we can say that this cluster is actually newly-adopted technology for self-driving, or driverless tractors,and the first patent directly about self-driving appeared in 2013

E. IDENTIFYING NEWLY ADOPTED TECHNOLOGY

In 4.4., we showed that patents can demonstrate newly-adopted technology. In this step, we applied the same method on yearly bases. The (n, x, y) setting remained at (3, 0.8, 0.01). For each year from 2006 to 2016, we applied the proposed method to get keywords. As the first patent related to self-driving appeared in 2013, the yearly keywords must show that the keywords related to this newly-adopted technology first appeared in 2013.

TABLE 1. ipc a01b69 patents within the cluster.

A01b69 Patent and Its Content	Year	Number of Cited Patents in the Cluster
7848865 (Material delivery)	2010	1
8170752 (Parking assist)	2012	3
8433479 (Front lighting)	2013	9
8392065 (Driving assist)	2013	1
8437913 (Driving assist)	2013	1
8700272 (Crest detection)	2014	1
8874317 (Parking assist)	2014	25
9205864 (Driving assist)	2015	1
9327693 (Rear collision)	2016	44

TABLE 2. Yearly keywords from every ipc a01b69 patents.

Year	Cluster 1 Keywords	Cluster 2 Keywords	Cluster 3 Keywords
2012	steering, wheel, angle, steering wheel, control, steering angle	steering, torque, power, electric, motor, control	vehicle, control, method, position, signal, work
2013	steering, control, wheel, angle, torque, motor	vehicle, control, method, sensor, operate, information	antennas, position, path, track, used, received

The use of the keywords changed over time. Meaningful keywords were detected in cluster 3 in 2013 (seeTable 2). The keyword 'antennas' had never appeared before. By analyzing the patent that used this term, we learned that the patent describes how to guide or navigate the vehicle using GNSS(global navigation satellite system). The cluster was composed of six patents (see Table 3). They use the term 'antennas' in their title or abstract, and five of them were not under IPC A01B69. If patents under A01B69 in 2013 are clustered, patent 8442722 is represented by keywords 'steering', 'control', 'wheel', 'vehicle', 'torque' and 'angle', which are not representatives of newly-adopted technology. By adding cited patents to our database, we could identify patent 8442722.

After tracking the cited patents in the cluster with 'antennas', we found three candidates that might include new technology (Table 4). Patent number 8583326 is limited to

TABLE 3. Patents within the cluster that has the keyword 'antenna'.

Patent number	IPC	Title
7027918	G01C21	Satellite navigation system using multiple antennas
7271766	G01S3	Satellite and local system position determination
7317977	G06F7	Dynamic stabilization and control of an earthmoving machine
7885745	G06F7	GNSS control system and method
8140223	G01C21	Multiple-antenna GNSS control system and method
8442722	A01B69	Corner unit guidance control system using two antennas

showing the safe path using GNSS contour but it is stated in the patent that this technology can be used for self-driving. Patent number 8386129 and 8392065 are directly related to self-driving technology and are clearly stated in patent that this technology is for auto-steering. Therefore, we can conclude that our methodology has successfully identified newly adopted technology.

V. CONCLUSION

This paper presents a method to monitor newly-adopted technology in patents by combining TF-IDF with K-means clustering, even when the number of patents is small. The idea behind TF-IDF and K-means clustering is that the technology is described using distinctive keywords. Cited patents can add distinctive keywords into the original keywords pool to increase the notability of distinctive keywords. This process increases the reliability with which newly-adopted technology can be monitored. Case study shows that this method is valid.

This method uses some assumptions that should be validated before it can be applied to other technology domains. First, every type of technology must have set of keywords that is different from other types of technology. Second, every patent must cite patents when it uses technology from a foreign domain. Third, a patent's title and abstract must describe its content accordingly.

The first assumption may not be always true. The case study showed a very reliable result because technology about steering of tractors uses a very different set of words than doestechnology about GNSS and digital data processing. However, when the new technology shares a common principle, physical component or process, the keyword pool may share most of the terms. In that case, a simple keyword-based monitoring method may fail to identify the new technology. To monitor the new technology, information related to existing technology should be organized in a specific and structured way. Solving the limitation of simple keyword-based monitoring method remains as our future research.

REFERENCES

[1] C. M. Christensen, *Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA, USA: Harvard Business Review Press, 2013, p. 13.
[2] K. Atuahene-Gima, "Differential potency of factors affecting innovation performance in manufacturing and services firms in Australia," *J. Product Innov. Manage.*, vol. 13, no. 1, pp. 35–52, 1996.
[3] R. Balachandra and J. H. Friar, "Factors for success in R&D projects and new product innovation: A contextual framework," *IEEE Trans. Eng. Manag.*, vol. 44, no. 3, pp. 276–287, Aug. 1997.
[4] M. Bastic, "Success factors in transition countries," *Eur. J. Innov. Manage.*, vol. 7, no. 1, pp. 65–79, 2004.
[5] R. S. Campbell, "Patent trends as a technological forecasting tool," *World Patent Inf.*, vol. 5, no. 3, pp. 137–143, 1983.
[6] J. Cho and J. Lee, "Development of a new technology product evaluation model for assessing commercialization opportunities using Delphi method and fuzzy AHP approach," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5314–5330, 2013.
[7] W. Seo, N. Kim, and S. Choi, "Big data framework for analyzing patents to support strategic R&D planning," in *Proc. IEEE 14th Int. Conf. Dependable, Autonomic Secure Comput., 14th Int. Conf. Pervasive Intell. Comput., 2nd Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2016, pp. 746–753.
[8] H. Ernst, "Patent information for strategic technology management," *World Patent Inf.*, vol. 25, no. 3, pp. 233–242, 2003.
[9] D. Bonino, A. Ciaramella, and F. Corno, "Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics," *World Patent Inf.*, vol. 32, no. 1, pp. 30–38, 2010.
[10] C.-S. Curran and J. Leker, "Patent indicators for monitoring convergence—Examples from NFF and ICT," *Technol. Forecasting Social Change*, vol. 78, no. 2, pp. 256–273, 2011.
[11] J. Yoon and K. Kim, "Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks," *Scientometrics*, vol. 88, no. 1, pp. 213–228, 2011.
[12] C. Lee, J. Jeon, and Y. Park, "Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach," *Technol. Forecasting Social Change*, vol. 78, no. 4, pp. 690–702, 2011.
[13] G. Kim and J. Bae, "A novel approach to forecast promising technology through patent analysis," *Technol. Forecasting Social Change*, vol. 117, pp. 228–237, Apr. 2017.
[14] Y. G. Kim, J. H. Suh, and S. C. Park, "Visualization of patent analysis for emerging technology," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1804–1812, 2008.
[15] J. Yoon and K. Kim, "Detecting signals of new technological opportunities using semantic patent analysis and outlier detection," *Scientometrics*, vol. 90, no. 2, pp. 445–461, 2012.
[16] J. Joung and K. Kim, "Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data," *Technol. Forecasting Social Change*, vol. 114, pp. 281–292, Jan. 2017.
[17] R. Wille, "Restructuring lattice theory: An approach based on hierarchies of concepts," in *Ordered Sets*. Dordrecht, The Netherlands: Springer, 1982, pp. 445–470.
[18] C. J. Fall, A. Törcsvári, K. Benzineb, and G. Karetka, "Automated categorization in the international patent classification," *ACM SIGIR Forum*, vol. 37, no. 1, pp. 10–25, 2003.
[19] U. Schmoch, *Concept of a Technology Classification for Country Comparisons*, document IPC/CE/41/5, World Intellectual Property Org., Geneva, Switzerland, 2008.
[20] L. Leydesdorff, D. Kushnir, and I. Rafols, "Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC)," *Scientometrics*, vol. 98, no. 3, pp. 1583–1599, 2014.

[21] E. C. Engelsman and A. F. van Raan, "A patent-based cartography of technology," *Res. Policy*, vol. 23, no. 1, pp. 1–26, 1994.

[22] H. Park and J. Yoon, "Assessing coreness and intermediarity of technology sectors using patent co-classification analysis: The case of Korean national R&D," *Scientometrics*, vol. 98, no. 2, pp. 853–890, 2014.

[23] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *J. Documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[24] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.*, vol. 24, no. 5, pp. 513–523, 1988.

[25] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.

[26] A. Aizawa, "An information-theoretic perspective of tf–idf measures," *Inf. Process. Manage.*, vol. 39, no. 1, pp. 45–65, 2003.

[27] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[28] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[29] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[30] M. Iwayama and T. Tokunaga, "Cluster-based text categorization: A comparison of category search strategies," in *Proc. 18th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 1995, pp. 273–280.

[31] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, 1986.

[32] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 269–274.

[33] B. C. M. Fung, K. Wang, and M. Ester, "Hierarchical document clustering using frequent itemsets," in *Proc. SIAM Int. Conf. Data Mining*, 2003, pp. 59–70.

SUNGHYUN NAM received the B.E. degree in industrial and management engineering (IME) from the Pohang University of Science and Technology (POSTECH) in 2014, where he is currently pursuing the Ph.D. degree in IME from the POSTECH Graduate School. His current interests include patent mining, patent analysis, and technology planning.

KWANGSOO KIM received the B.E. and M.S. degrees in industrial engineering from Seoul National University, South Korea, in 1977 and 1979, respectively, and the Ph.D. degree from the University of Central Florida, Orlando, FL, USA, in 1985. He has been a Professor with the Department of Industrial Management Engineering, Pohang University of Science and Technology (POSTECH), for 30 years and has been planning and operating POSTECH Creative Entrepreneurship Omphalos (PCEO) providing gifted education for inventions as the Director of PCEO for eight years. He is the author of over 70 international journal papers and holds five patents. His research interests include patent mining, technology planning, business planning, future planning, and creative invention education.