ARTICLE IN PRESS
TFS-18786; No of Pages 10
Technological Forecasting & Social Change xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Technological Forecasting & Social Change

A novel approach to forecast promising technology through patent analysis

Gabjo Kim, Jinwoo Bae *

Government Cooperation Team, Korea Intellectual Property Strategy Institute, 131, Teheran-ro, Gangnam-gu, Seoul 06133, Republic of Korea

ARTICLE INFO

Article history:
Received 8 December 2015
Received in revised form 28 November 2016
Accepted 29 November 2016
Available online xxxx

Keywords:
Promising technology
Patent analysis
Technology clusters
CPC
Patent indicator

ABSTRACT

Forecasting promising technology is a relevant opportunity for management of companies and countries. Furthermore, researchers in research and development (R&D) have recently considered that patents include detailed information on developed technologies. For these reasons, we suggest a novel approach to forecasting PT using patent analysis. The overall process of the proposed methodology consists of three steps. First, to form technology clusters, we clustered patent documents on the basis of the cooperative patent classification (CPC), which represents a more detailed technology classification system than the international patent classification (IPC). Second, regarding the process of defining technology clusters, we examined the combination of CPCs of each formed clusters. Finally, patent indicators such as forward citations, triadic patent families, and independent claims are analyzed to assess whether the technology clusters are promising. We collected patent data on the wellness care industry from the United States Patent and Trademark Office (USPTO) to verify the proposed methodology.

1. Introduction

Promising technology is a key technology that underlies the steady growth of companies and countries. Its influence on a company's investments and production and on overall national industries is significant. Additionally, a promising technology is changing quickly and unexpectedly. As such, companies and countries that focus rapidly on promising technologies to lead the industry are able to increase their competitiveness such that it becomes directly connected to survival (Jeong and Yoon, 2015).

Hence, properly forecasting promising technologies is integral for decision makers of both corporations and countries. Upon further examination, first, it was found that the establishment of an efficient research and development (R&D) strategy establishment is possible. Deriving a comprehensive notion that can satisfy the needs of a future envisioned society as well as the market, forecasting promising technology is being recognized as an essential stage in the R&D process (Albright, 2002). Specifically, a national R&D agenda budget can be influenced by the political environment during the R&D planning stage (Halal et al., 1998). This greatly increases the anxiety towards failure regarding future R&D endeavors. Accordingly, the allocation of an R&D budget should be based on objectivity and validity, as to allow for adequately selecting and focusing on a promising technology sector. In other words, by exploring a future environment that humans may face, forecasting promising technology can present a direction on the right path for the early stages of R&D.

Next, the autonomously led development of products and services becomes possible. As an output of R&D, the developed technology acts as a firm foundation for the development of future products and services (Wang et al., 2015). That is, being able to forecast products and services that can lead the future allows for heightened competitiveness compared to rival players.

Accordingly, in the past, such as with the Delphi and scenario methods, attempts have been made to forecast promising technology using expert opinions in related technology areas. However, the promising technology forecasting approach of these expert groups caused problems, including complexity and excessive time required for the procedure, social costs attributable to the mobilization of a large labor force, the absence of credibility for each expert's scope of technology, and different opinions on the interpretation (Choi and Jun, 2014).

As a solution, companies and countries have been establishing future promising technology forecasting strategies using patent analysis (Lee et al., 2009). Patents not only provide legal protection for intellectual property rights, but also include detailed information about the developed technology (Park et al., 2015a). Therefore, forecasting promising technology via patent analysis is significant for establishing management strategies. For instance, it can be used to prevent R&D investments in unnecessary technology areas (Kim et al., 2008), evade loyalty payments required by accidental technology infringements on rival companies (Kim et al., 2015), and design R&D projects to secure core patents (Ju and Sohn, 2015). In other words, from the viewpoint of R&D decision makers, this study can provide significant insights to strengthen the future competitiveness of a company and become an important means for technology management decisions (Ernst, 2013). Therefore, the purpose of this study is to forecast future promising technology through

* Corresponding author.
E-mail addresses: kkjjo@kista.re.kr (G. Kim), bjw8751@kista.re.kr (J. Bae).

http://dx.doi.org/10.1016/j.techfore.2016.11.023

patent analysis. In other words, we utilize historical patent data to objectively and quantitatively determine what the promising technologies of the future may be.

We have come to the following conclusions and summarized the following contributions through our study on the methodology of forecasting promising technology.

1. We are able to identify the specific detailed technological areas that compose different technologies and industries. For this, we utilize the systematic and accurate properties of the patent classification system, as included in the patent information. These technological areas may be convergence, existing, or even technologies that can create disruptive innovation.
2. We are able to determine and verify the evidence of promising prospects in existing technological fields. Based on the application year of the collected patent data, we postulate the training set from 2002 to 2009, and the test set from 2010 to 2013. In doing so, we test the restricted model by examining how much of an influence the current promising aspect valuation of technologies have had on the future. Here, we evaluate technology valuations using patent information.

This study is not based on technology forecasts made using the subjective judgments of experts, but upon patent data analysis, which contains detailed information about the technology used. Therefore, even those without technological expertise can produce objective promising technology forecasts. By identifying the opportunity creation of new technologies, we hope to utilize this as a tool to forecast the promising technologies that can substitute current ones.

This paper is organized as follows. Section 2 reviews the case studies on patent analysis and technology forecasting. Section 3 reveals the proposed methodology for the actual forecasting promising technology analysis. Section 4 presents the experimental results, while Section 5 provides the conclusions and implications of the study.

2. Literature review

2.1. Patent analysis

A patent is an accessible document that contains information on both the developed technology and its usage rights (Park et al., 2005). Such patents grant exclusive rights in exchange for disclosing the technology (Trappey et al., 2011). Therefore, in the case of important software or algorithms, where it is difficult to determine if a competitor has infringed on patents, companies often own these technologies as know-how instead of applying for patents. Furthermore, in some countries, which had inadequate systems for protecting intellectual property rights in the past, companies intentionally did not apply for patents, as their intellectual property rights were not properly protected compared to the high costs and time required for application.

However, presently, countries and companies are increasingly changing their perception and patenting their technologies (Manap et al., 2016) when such technologies include clear technological concepts and reverse engineering is possible. Accordingly, with the growing emphasis on the importance of intellectual property rights that assure exclusive rights over a developed technology, there is an increase in recent efforts to obtain patents (Park et al., 2015b). Furthermore, there is an emergence of non-practicing entities (NPEs) that acquire high quality patents with wide scope (Fischer and Henkel, 2012) to legally challenge the companies that practice patent infringement or make a profit by reselling the patents to other companies (Pénin, 2012). It implies that a patent can contribute significantly towards profit creation if it is utilized for offensive or defensive purposes by an organization.

Patents are also used to develop technologies that are more advanced than the existing versions around these patents (Belvard, 2000). That is, they not only have offensive and defensive functions but can also be used as an efficient technology management strategy.

Especially at a national level, they can be used to establish public policy. For instance, research has been conducted on the direction of future policy for detailed technologies, based on pinpointing leading patent applicants and countries for electrochemical energy storage technology (Mueller et al., 2015). Additionally, research of the time range of patents in the solar thermal utilization sector, technology type distribution, and technology trend analysis had been used to establish governmental energy policies (Zhao and Zhao, 2015). There has been research on patent analysis of global wind turbine companies, which provided a direction for policies for discovering new markets in Asian and European nations and for company policies about encouraging open innovation among companies (Zhou et al., 2015).

From the perspective of technology management planning, identifying the effective technology development trends of rival competitors and reviewing whether or not to introduce new technology can be done through patent analysis. Identifying the technology life cycles of patents in telematics revealed the possibility for new technology creation by connecting mobile devices to cloud platforms (Chang and Fan, 2016). The R&D tendencies and trends in the target sectors for leading companies in the field of amorphous silicon are examined (Tseng et al., 2011). By dividing companies into those with leading technology, those with technological potential, and those with technology quality orientation, based on calculated patent indicators, detailed competitiveness of these companies can be identified. Furthermore, researchers without detailed knowledge of technology can easily determine technological trends and important technologies through patent analysis (Chang et al., 2012). In other words, the study understood the relationships between patents and discovered key patents by forming a patent network based on similar terms that were used in different patents.

Therefore, this study approaches patents, which have become important for countries and corporations, from the establishing policies and technology management perspective because earlier studies have drawn technological implications from patent analysis.

2.2. Forecasting promising technology

Forecasting promising technology plays an important role in decision making for enterprises' and countries' management of technology. In the past, qualitative analysis, such as the Delphi and scenario approach methods, was based on technology forecasts.

In the Delphi analysis approach, the consensus process among experts heightens objectivity and persuasiveness. For example, Delphi analysis has been conducted to predict the future technologies in the public relations sector. Technology experts from various fields and from all over the world participated and reviewed dozens of topics that had previously been overlooked or ignored (Kent and Saffer, 2014). Experts were divided into groups to carry out Round 3 of the Delphi study, as to identify the determinants of business opportunities in the emergent bioenergy industry at both company and industry level (Pätäri, 2010). Therefore, Delphi analysis can be valuable when past data is unavailable or in cases where mathematical modeling is impossible.

The scenario analysis approach is useful as a base for establishing a strategy for a variety of uncertain factors. Such scenario analysis postulates and generalizes how various possible uncertain situations may evolve. Having filled out the technological road map, scenario planning is a prerequisite for making precise predictions possible (Saritas and Aylen, 2010). Using the current technological situations, market needs, evaluation and understanding of products and services, and technological influence analysis, a scenario analysis to satisfy the market needs considering future uncertainties was conducted (Holmes and Ferrill, 2005).

Moreover, a combination of the Delphi and scenario analyses can provide accurate predictions as well. The probability of occurrence for certain future events can be estimated in Round 3of the Delphi analysis,

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

based on the scenario analysis for predicting mobile broadband traffic (Lee et al., 2016).

However, these qualitative analyses have drawbacks. First, the Delphi draws its conclusions by exchange and development of expert opinions in the relevant area. Because the method relies entirely on subjective judgments, experts' biased judgments of a technology forecast are a disadvantage. The instances are experts' tendency to keep their existing R&D areas, personal or organizational inclinations, halo effect that relies on well-known researchers or institutions, and different criteria for selecting promising technology. Furthermore, it is very difficult to find willing participants for this method (Kent and Saffer, 2014). The scenario method also has reliability limitations because it is based on limited future aspects with a high possibility of occurrence. Additionally, the composition of the scenario can be arbitrary (Amer et al., 2013). As a result, although qualitative technology forecasts will continue to exist, it should be accompanied by data-based quantitative analysis (Bengisu and Nekhili, 2006).

To overcome these limitations, numerous studies on future promising technology forecasts analyzed patents with detailed information on the patented technologies. This is because patents provide sufficient data for drawing reliable conclusions in studies examining technological change and innovation (Chang et al., 2012).

For instance, research is being conducted to forecast future technology using a growth curve based on the number of patent applications. The logistic growth curve is applied to nano-sized ceramic powder technology patents (Cheng and Chen, 2008) and to building integrated photovoltaic technology patents (Chiu and Ying, 2012), respectively. The logistic growth curve and the Bass model are then applied to patents on information and communications technology (ICT) applications (Meade and Islam, 2015). Furthermore, the growth curve of forward citations for TFT and LCD, flash memory systems, and personal digital assistants is applied as a criterion to forecast future technology (Altuntas et al., 2015). As such, these studies forecasted promising technology by commonly examining future demand and the diffusion of patents.

There are also studies that provide technology forecasts by applying various data-mining approaches to patents. Promising technology is identified using association rule mining of international patent classifications (IPCs) for patent documents (Jun et al., 2012a). Additionally, the promising aspects of technology are determined by applying association rule mining to changes in patent indicator values over time for each IPC (Shih et al., 2010). Moreover, applying network analysis to IPCs, the relationships between technologies are visually expressed and whether the technology is promising based on the centrality between IPCs and the distance between nodes is determined (Park et al., 2015c).

Studies on forecasting future promising technology by applying a text mining technique to an unstructured patent title and abstract have been conducted as well. Text mining for Apple Inc. patents is applied for identifying the promising vacant technologies (Jun and Park, 2013). Additionally, a study was conducted to forecast vacant technologies by applying the generative probabilistic model of latent Dirichlet allocation to renewable energy technology patents (Kim et al., 2015). Another study examined vacant technology areas and its promising aspects by obtaining patent keywords and then applying generative topographic mapping as a probabilistic reformation of a self-organization map (Jeong et al., 2015).

These research experiments are significant in objectively deriving data-based promising technology. However, some limitations do exist. Reliability is absent, given that the growth curve from the accumulated number of patents does not take into account a variety of environmental factors on the relationship among complex technologies. Additionally, researchers' qualitative judgments are still included in patent analysis on the basis of the keyword-based approach (No et al., 2015a). Moreover, because the IPCs of patents rely significantly on broad and abstract technology classifications, it is difficult to understand and interpret technologies in great detail.

Although limitations exist, various approaches have been attempted in studies for promising technology forecasts and insightful messages identified, such as technology areas in which researchers commonly need to be actively committed to R&D. Therefore, this study forecasts future promising technology through new approaches to patent analysis to overcome the existing drawbacks of promising technology forecasts and allow researchers to engage new markets and strengthen technology competitiveness.

3. Methodology

The purpose of this study is to forecast promising technology by analyzing wellness care industry patent data in the United States. To do so, as mentioned in Section 3, wellness care patent data were collected from the United States Patent and Trademark Office (USPTO) (The United States Patent and Trademark Office (USPTO), 2015), and data mining and patent indicator analysis on a variety of bibliographic information from the collected patent documents were performed. Fig. 1 indicates the overall process of the proposed methods for forecasting promising technology (Fig. 1).

3.1. Selection of technology field

The interest in wellness care—the optimal health condition for the physical, mental, and social state—has received significant interest. In particular, around 52% of adults older than 50 receive complementary health care in the United States (Johnson et al., 2016) and its vast investment in health care shows great potential for growth in the wellness industry (Fujii et al., 2016).

This development in the wellness care industry is expanding from the convergence of IT and various industries. In particular, the activation of the wellness care industry integrated with IT and medicine likely makes it a suitable industry for forecasting promising technology (Kim, 2015). Thus, in this study, we adopted the wellness care industry to experimentally verify the new patent analysis approach for promising technology forecasts.

3.2. Data acquisition

The United States is a marketplace with the largest demand for technology and has the fastest technology development in the wellness care industry (Anon, 2010). Accordingly, the world's leading companies continuously publish patent applications to the USPTO to obtain protection for rights on new technologies that they develop. In this study, we forecasted promising technology by analyzing patent data from the USPTO on the wellness care field.

3.3. Technology clustering

To effectively search and manage a significant amount of patent information, a patent classification system that complies with the current state of technology and user needs is necessary. Such a patent classification system is utilized in a prior art search as introduced through various nations' classification systems, such as IPC, file index (FI), and CPC.

Subdividing technology system countries with large numbers of patent applications, including the United States, European countries, and South Korea, is done through a CPC system that is flexible and that can be quickly revised (Mueller et al., 2015). The CPC has more than 250,000 entries, which is larger than FI with approximately 180,000 entries and IPC with approximately 70,000 entries (Kapoor et al., 2015). Therefore, in this study, we cluster patent documents for a similar technology using the CPC information in each document.

A patent document has more than one CPC and can be composed of a patent-CPC matrix (PCM), which is an asymmetrical matrix as shown in Fig. 2 (a). We then constructed a patent-patent matrix (PPM) as a

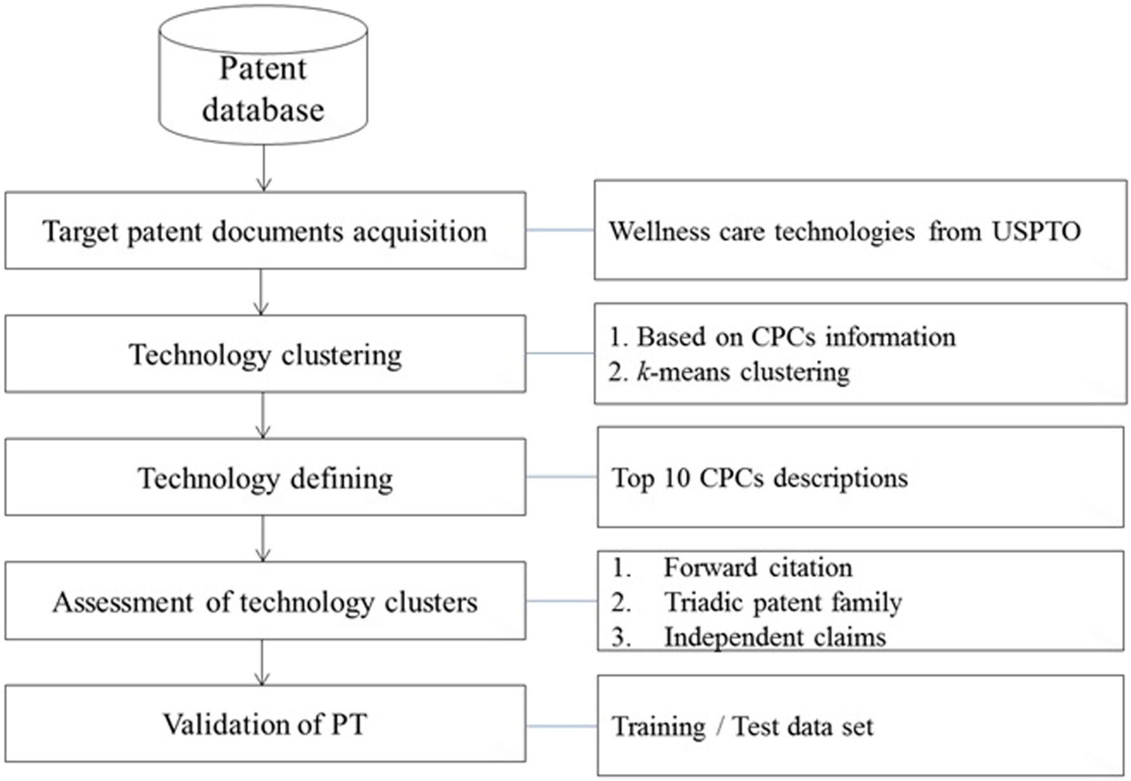

Fig. 1. Overall process of the proposed methodology.

symmetrical matrix, as shown in Fig. 2 (b). To construct the PPM, we calculated the Pearson's correlation coefficient between patent documents.

Patent documents can be mapped on a low dimensional space through multidimensional scaling (MDS). MDS is a multivariate statistical method to visually reduce data from a high dimensional space to a low dimensional space (Cox and Cox, 2001). To perform MDS, n data need to form a (dis)similarity matrix. In this study, the data are represented by a dissimilarity matrix with the range value of [0,2] by taking 1–coefficient. Next, double centering is applied as follows:

$$B = -\frac{1}{2}HD^{(2)}H \tag{1}$$

where $D^{(2)}$ is the matrix of squared proximities and is applied to the previous matrix of squared proximities. $H = I - \frac{1}{n}ee^T$, where I is an identity matrix with n observations. Next, in this study, following Eq. (2), the eigenvector decomposition of B is indicated to project patent documents onto the two dimensional space.

$$Y = V_2 \Lambda_2^{1/2} \tag{2}$$

where V_2 indicates eigenvectors e_1 and e_2, and $\Lambda_2^{1/2}$ indicates square root values of λ_1 and λ_2.

Next, we clustered patent documents with a similar technology on a two-dimensional space by adopting k-means clustering. k-means clustering is a partitioning clustering method through which the principal objective is to divide the given data into a number of k clusters with similar aspects (Hartigan and Wong, 1979).

The following briefly explains the principle of k-means clustering. First, given k seed points, the initial cluster's centroid is randomly set. Second, the set of data is assigned to the cluster that has the nearest cluster centroid. Third, the centroid of each cluster is readjusted by reassigning the average of the observations in the cluster. The repetition halts if changes no longer occur in the clusters.

To conduct the k-means clustering, the distance metrics and the number of clusters k should be selected in advance. Several methods may be used to evaluate the distance metrics, such as Euclidian distance, Manhattan distance, Chebyshev distance, and Spearman distance. In this study, we clustered the patent documents with k-means clustering using Euclidian distance.

In addition, to determine the optimal number of clusters, k, we considered the average silhouette width concept (Rousseeuw, 1987). The silhouette width is used to evaluate the validity of the clustering results,

	CPC 1	CPC 2	CPC 3	CPC 4	···	CPC m
Patent 1	1	0	1	0	···	0
Patent 2	1	0	0	0	···	0
⋮	⋮	⋮	⋮	⋮	⋱	⋮
Patent n	0	1	1	0	···	0

(a) Patent – CPCs Matrix

→

	Patent 1	Patent 2	···	Patent n
Patent 1	-	0.513	···	0.766
Patent 2	0.513	-	···	0
⋮	⋮	⋮	⋱	⋮
Patent n	0.766	0	···	-

(b) Patent – Patent Matrix

Fig. 2. Transforming PCM into PPM.

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

which can be used to determine the optimal number of clusters. The silhouette width of the ith observation is computed as follows:

$$sw_i = \frac{b_i - a_i}{max\{b_i - a_i\}} \tag{3}$$

where a_i denotes an average dissimilarity between the ith observation and the other observations within the same cluster and b_i represents an average dissimilarity between the ith observation and the other observations belonging to a neighboring cluster. sw_i has a value ranging from −1 to 1, and a value closer to 1 implies better-composed clustering. Therefore, the value of sw_i is the ith observation's silhouette width and the optimal number of clusters corresponds to the widest average silhouette width.

3.4. Technology defining

A number of previous studies used text mining or IPC on the process of defining technology after formation of technology clusters. However, the definition of technology cluster through text mining has the disadvantage of extreme subjectivity and the distribution of significant noise. This is because different researchers may produce different results if the technology clusters are defined based on the top ten (Choi and Jun, 2014) or five (Jun et al., 2012b) terms without expertise in each technology area.

On the contrary, as was previously mentioned, CPC classifies technologies in greater detail than does IPC. Thus, this study defines technology clusters based on the ten CPCs that most frequently appear in each technology cluster because this method can establish more objective definitions of technologies compared to existing technology definition methods based on terms and IPCs. There may be overlapping CPCs in the formed technology clusters since various technological factors can be amalgamated. Therefore, by also accounting for the CPC composition ratios of the formed technology clusters, we show whether the overlapping technology clusters are completely identical or are amalgamations of similar technological factors.

3.5. Assessment of technology clusters

After the formed technology clusters are defined, we calculate the patent indicators to judge the promise of the clusters. Patent citation was used as an important indicator of the technology forecast (Chang et al., 2009; Fallah et al., 2009). Additionally, patents that are entered in major markets such as the United States, Japan, and Europe—called the triadic patent family—suggest qualitatively important technologies. Moreover, the number of independent claims can be indicated as the number of inventions. This is because each patent claim denotes the legal definition of the invention (Trappey et al., 2012). Thus, the number of independent claims is proportionate to the number of technologies possessed by the patent holder. In this study, as an indicator to determine promising technology, the following three pieces of patent information are considered: (i) forward citation, (ii) triadic patent, and (iii) independent claim. These indicators are explained in the following subsection.

3.5.1. Forward citation

Forward citation in a patent states prior art documents as bibliographies. These documents are determined to be technologically closely related to the invention. That is, the fact that one patent is cited in many different patents indicates that it creates technological and economic value through important contributions to future technology development. Furthermore, one can understand knowledge flows from the information in forward and backward citations (No et al., 2015b).

The forward citation of patents has been widely applied as an important indicator to evaluate the value of technology (Lee et al., 2012). Additionally, the citation is significantly correlated with real-world patent auction prices as economic value (Fischer and Leidnger, 2014). The older and more important a patent is, the more frequently the patent tends to be cited by other patents (Yoon and Kim, 2012). We do not consider the time lag of such forward citations but rather values such as the technological influence of the patents.

In this study, the forward cites per patent about an i technology cluster is calculated as follows:

$$\text{forward cites per patent}_i = \frac{FC_i}{T_i} \tag{4}$$

where FC_i indicates the total number of forward citations of technology cluster i and T_i indicates the total number of patents of technology cluster i.

3.5.2. Triadic patent family

The triadic patent family indicates patents for which the same invention, same inventor, or applicant applied to USPTO, JPO, and EPO at same time. In 1999, the OECD used this concept to evaluate national technology competitiveness (Lee and Sohn, 2013).

Triadic patent family is now used as an important indicator when evaluating the technology level of countries or companies because the United States, Japan, and Europe are regarded as major markets (Baudery and Dumont, 2006). The right of an inventor to the invention increases through patent applications filed abroad. However, such applications take significant time and costs. Therefore, technologies with many applications in major markets have high values and strong international competitiveness. In this study, triadic patent family is calculated as follows:

$$\text{triadic patent family}_i = \frac{TP_i}{T_i} \tag{5}$$

where TP_i represents the total number of patents in technology cluster i applied simultaneously to USPTO, EPO, and JPO, and T_i represents the total number of patents of technology cluster i.

3.5.3. Independent claims

In a patent, the essential information for the configuration of the invention is written in independent claims and may be used as an indicator of the amount intellectual property rights (Petruzzelli et al., 2015). In other words, because the number of independent claims is proportional to the number of inventions, a large number of independent claims per patent may be viewed as a technology with robust rights (Lee and Lee, 2010). The independent claims per patent can be calculated as follows:

$$\text{independent claims per patent}_i = \frac{IC_i}{T_i} \tag{3}$$

where IC_i represents the total number of independent claims in technology cluster i and T_i indicates the total number of patents of technology cluster i.

4. Experimental results

We collected patent data on wellness care from 2002 to 2014 for experimental verification of the recommended methodology. However, as shown in Fig. 3, one section is not totally disclosed because all applied patents have not yet been published from 2013 to 2015 because it takes around one and a half years to disclose every issued patent. Therefore, we excluded patent data after 2013, since it is only partially disclosed.

For experimental validation, we classified patent data from 2002 to 2009 into a training set and from 2010 to 2013 into a test set. This is to empirically test whether the promising technology derived from the training set led to a large number of patents in the test set. From the entire set of 2182 patent documents collected, the training set

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

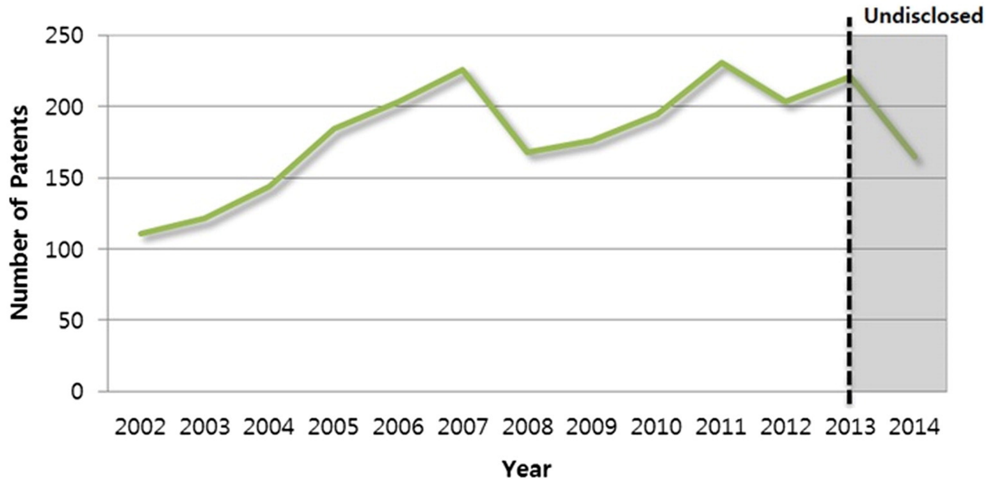

Fig. 3. Number of wellness care patents in the United States by year.

includes 1333 patent documents and the test set includes 849 patent documents. First, CPCs within patents are parsed to cluster patent documents with similar technology. The CPCs on wellness care from the collected patent data from the USPTO are distributed as indicated in Fig. 4.

After examining the results of the types of CPCs held by each patent, 1333 × 1442 PCM were formed as 1442 CPCs and a total of 1333 patents in the case of the training set and 849 × 1326 PCM were formed as 1326 CPCs and 849 patents in the case of the test set. To present each PCM as a form of a dissimilarities matrix, we calculated a value by subtracting 1 from the value of the Pearson's correlation coefficient between patent documents. The range of the matrix vector values for the training set and the test set is [0, 1.032576] and [0, 1.037801], respectively. Ranges closer to 0 indicate a dissimilarity matrix, which becomes a similar relationship. For visualization, patent documents are projected to a two-dimensional space through MDS, as shown in Fig. 6.

Next, to cluster patent documents mapped in two-dimensional space, the optimal cluster number *k* should be determined. When the number of clusters *k* of both the training set and the test set is 3, the average silhouette width value is the highest, at 0.69, as shown in Fig. 5. Therefore, the optimal number of clusters *k* from these two sets is determined to be 3. The result of performing *k*-means clustering after the value of *k* is determined is shown in Fig. 6. In the training set, cluster 1 has 483 (32.9%) of the 1333 total patent documents and cluster 2 and cluster 3 have 653 (49.0%) and 42 (18.3%) documents, respectively. Cluster 1 of the test set has 137 (16.1%) of the 849 patent documents and cluster 2 and cluster 3 of the test set has 253 (29.8%) and 459 (54.1%) documents, respectively.

We examined the top 10 CPCs included in each technology cluster to define the formed technology clusters as shown in Table 1. The Appendix A shows the descriptions corresponding to each CPC.

As a result, cluster 1 of the training set is matched to cluster 2 of the test set as a technology related to an entire wellness care business system, such as health care and patient management. Additionally, cluster 2 of the training set is matched to cluster 3 of the test set as a telemedicine technology providing clinical health care services using telecommunications and information technology at a distance. Cluster 3 of the training set is matched to cluster 1 of the test set as a data management technology, such as for patient records and medical imaging data. As a result, three of the same technology clusters from the training set and the test set are formed.

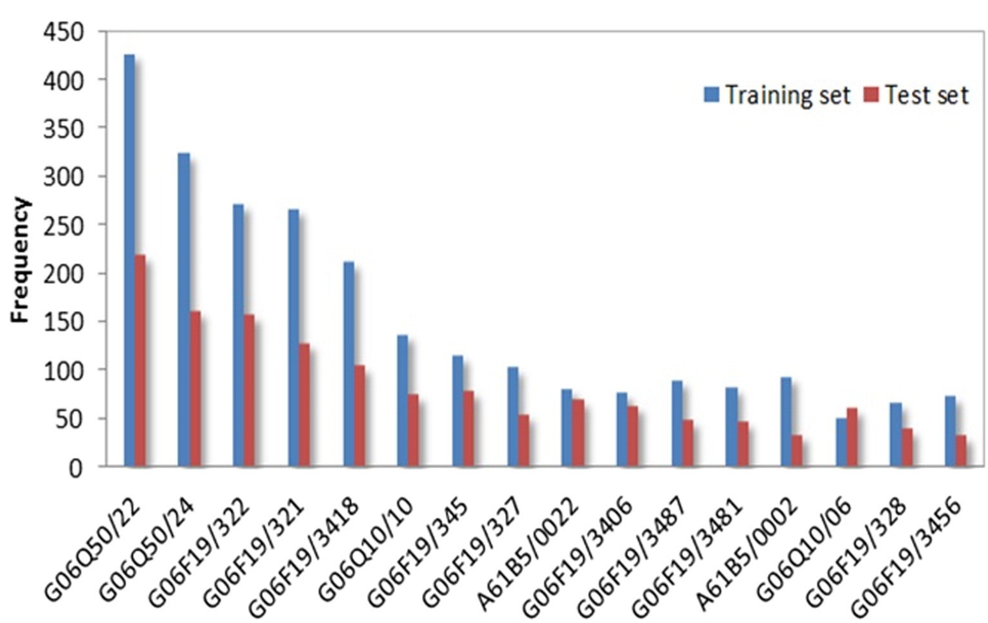

Fig. 4. CPC frequency for training and test sets on the wellness care field.

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

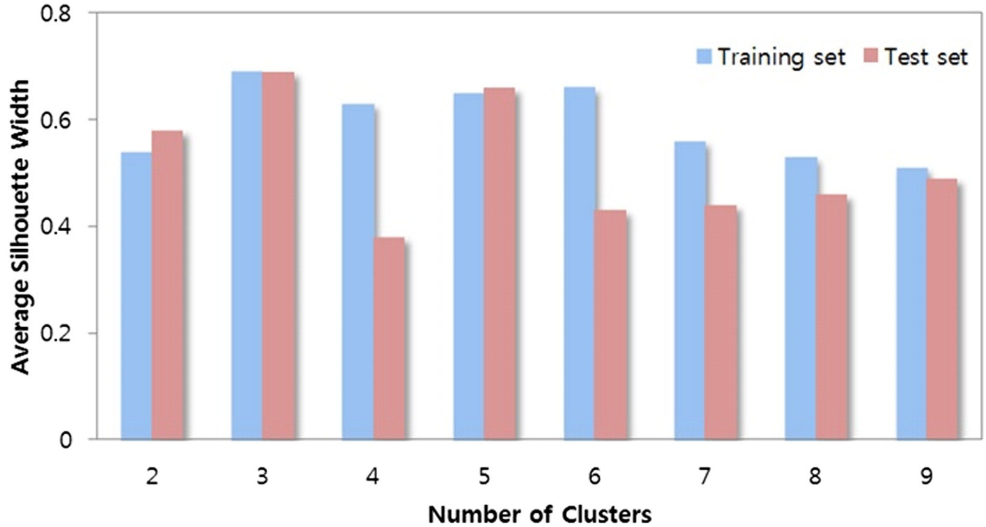

Fig. 5. Results of the average silhouette width of the training and test sets.

Here, we examined the common CPCs that appear in each formed technology cluster. There are 7 CPCs commonly contained in two technology clusters in the training set. In particular, there are more than 5 overlapping CPCs cluster 1 (wellness care business system technology) and cluster 3 (data management technology). This implies that the technology compositions of cluster 1 and cluster 3 are analogous. However, these cannot be viewed as perfectly identical technologies because, as Fig. 7 shows, the CPC composition ratios of these technology clusters are considerably different. This implies that data management technology has some mutually complementary components and exclusive relationships with data management technology (Anon, 2015). In other words, there have been integrations of similar technologies. This is also the case for cluster 1 and cluster 2 in the test set.

After the technological definition of a formed technology cluster, patent indicator analysis was performed to determine the promise of each technology cluster. As Table 2 shows, the promise for cluster 2 as determined by comparing forward citation per patent, triadic patent family, and independent claims per patent of the promising technology cluster candidates was 0.62787, 0.54242, and 4.03216, respectively. This cluster was concluded as being a promising technology cluster with the highest values among the three clusters.

Cluster 2 of the training set is the same as cluster 3 of the test set, and its portion was significantly increased from 49.0% to 54.1%. Additionally, the value of the patent indicators of cluster 1 and cluster 3 were significantly lower than that of cluster 2. Moreover, from cluster 1 and cluster 2 of the test set, these portions were decreased from 18.2% to 16.1% and 32.9% to 29.8%, respectively. This shows that, since data management and wellness care business system technology are less promising than telemedicine, R&D for the former two has been slow regarding the test set period from 2010 to 2013, while R&D for telemedicine has been active in the test set period due to its promising past achievement. Therefore, our proposed methodology accurately forecasted promising technology.

This result shows that forward citations, family patents, and the bibliographic information of independent claims can be used as indicators to evaluate the prospects of a technology. In addition, as mentioned earlier, because data management technology and business system technology have less clear technological concepts, as do algorithms and software, than telemedicine technology, that there has been relatively less vigorous patent activities in these areas.

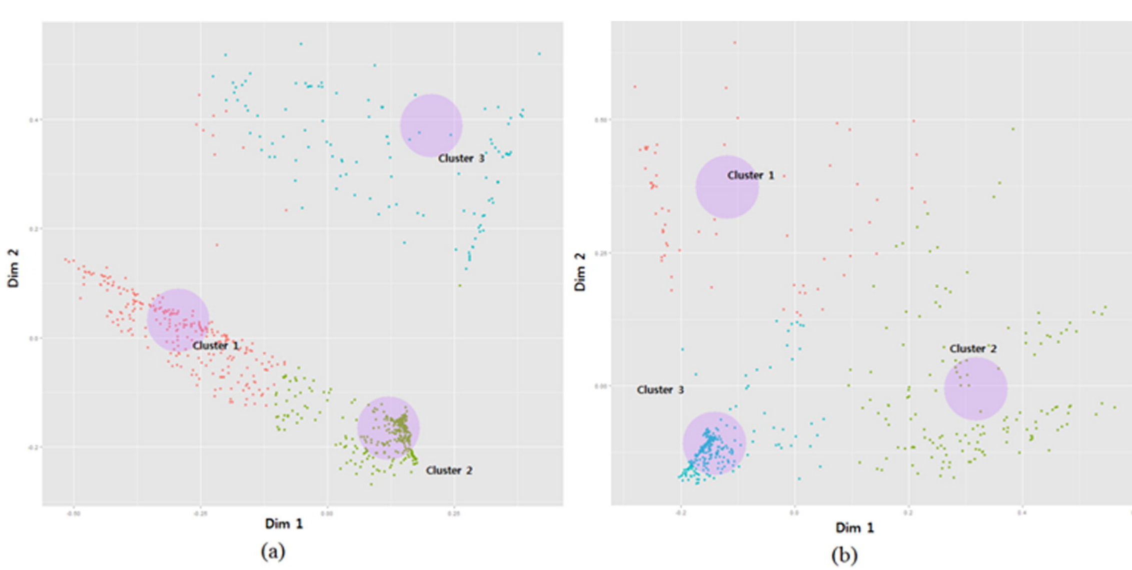

Fig. 6. Graphical results of patent projecting and clustering on the two-dimensional space of training set (a) and test set (b).

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

Table 1
The result of CPCs for each technology cluster of training and test sets.

Rank	Training set			Test set		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	g06q50/22 (359)[a]	g06f19/3418 (85)	g06f19/321 (242)	g06f19/321 (216)	g06q50/22 (206)	a61b5/0022 (60)
2	g06q50/24 (265)	a61b5/0002 (74)	g06q50/22 (59)	g06f19/322 (65)	g06q50/24 (148)	g06f19/3418 (55)
3	g06f19/322 (205)	a61b5/0022 (45)	g06q50/24 (47)	g06q50/24 (29)	g06f19/322 (82)	g06f19/345 (40)
4	g06f19/3418 (117)	g06f19/322 (37)	g06f19/322 (30)	g06q50/22 (23)	g06q10/10 (60)	g06f19/3406 (39)
5	g06q10/10 (111)	g06f19/345 (34)	g06q10/10 (17)	g06f19/3487 (15)	g06q10/06 (55)	a61b5/1118 (32)
6	g06f19/327 (84)	g06f19/3481 (34)	g06f19/3487 (15)	g06q10/10 (14)	g06f19/3418 (41)	a61b5/0205 (30)
7	g06f19/345 (78)	a61b5/14532 (33)	g06f17/30265 (13)	g06f19/327 (13)	g06f19/327 (39)	a61b5/02055 (29)
8	g06f19/3487 (64)	a61b5/0006 (31)	g06f19/3418 (10)	g06f19/3418 (12)	g06f19/328 (32)	a61b5/0002 (28)
9	g06f19/345106 (60)	a61b5/0205 (30)	g06 t7/0012 (10)	g06 t7/0012 (10)	g06f19/3487 (30)	g06f19/322 (24)
10	g06f19/328 (57)	a61b5/024 (30)	y10s707/99945 (9)	g06f19/323 (9)	g06f19/345 (29)	g06f19/3481 (24)

[a] ( ) indicates the number of CPCs.

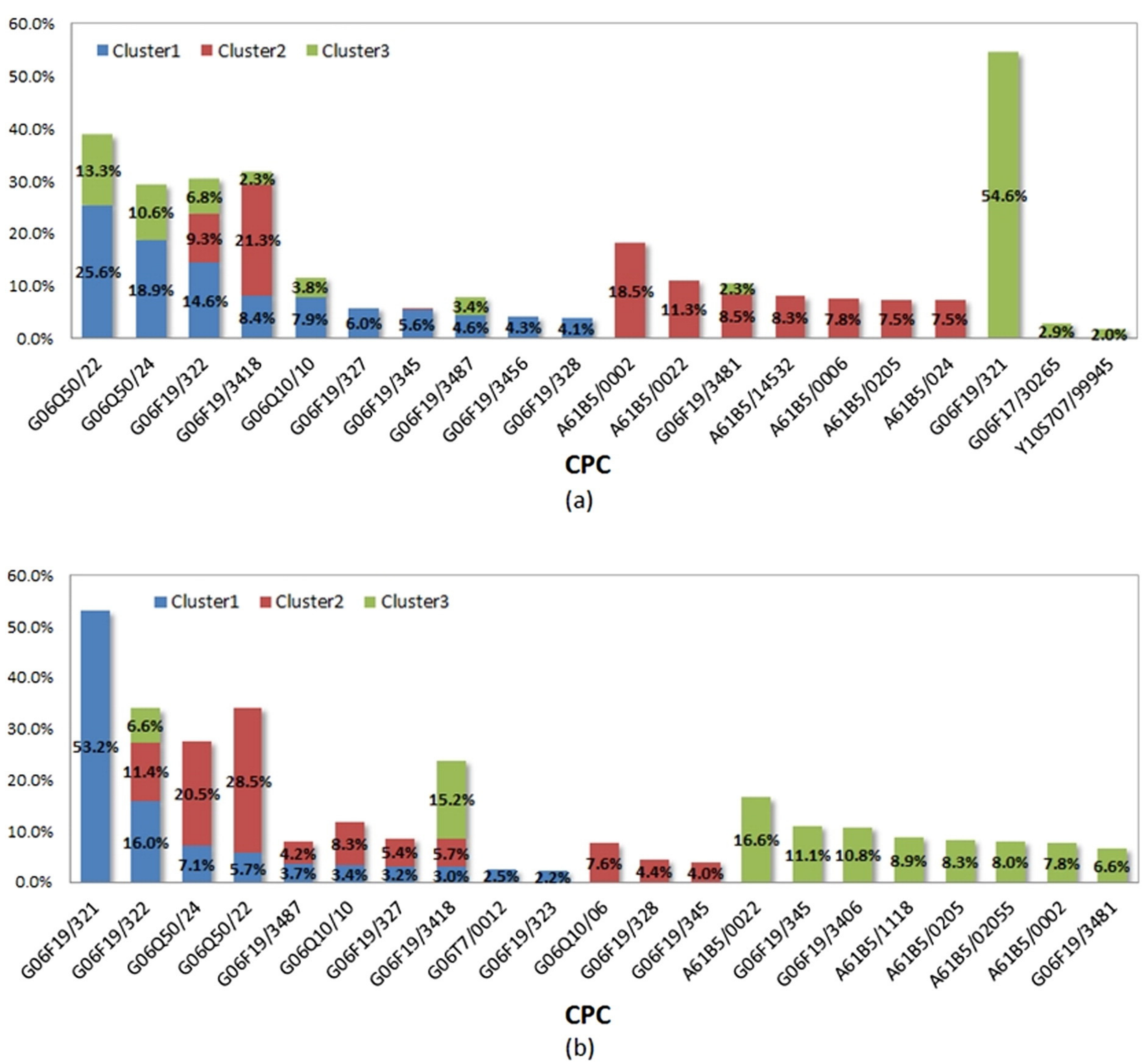

Fig. 7. CPC distribution for each formed technology cluster in the training set (a) and test set (b).

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023

Table 2
Patent indicator analysis results for the training set.

	Forward citations per patent	Triadic patent family	Independent claims per patent
Cluster 1	0.52055	0.32727	3.83333
Cluster 2	0.62787	0.54242	4.03216
Cluster 3	0.58264	0.13030	3.70661

5. Conclusion

In this paper, we forecasted the promising technology of the wellness care field using patent information from the USPTO. First, to form technology clusters with similar technologies, we considered CPCs because they include much more detailed information on technologies than do IPCs. Next, to understand the promise of each formed technology cluster, we use the patent information to calculate patent indicators, including patent family, forward citations, and independent claims. The reason is that such indicators are important pieces of information used mainly to assess the value and level of technology.

From the total data set collected for the experimental validation of the methodology, we classified the periods from 2002 to 2009 as the training set and from 2010 to 2013 as the test set. This is because it is possible to objectively confirm whether the promising technology results drawn from the promising technology forecasting model used for the first period were in fact patented in the actual future period.

Consequently, we selected telemedicine technology, which has relatively higher indicator values, as a candidate promising technology cluster of the training set. Furthermore, the portion of telemedicine technology in the promising technology cluster of the training set increased relative to other technology clusters from the test set. Moreover, the technology clusters formed in the course of forecasting promising technology were wellness care business system, data management, and telemedicine technologies. They were formed both in training set and test set with no significant difference. This indicates an incremental innovation, which means that the formed technology clusters of the wellness care industry do not show a disruptive change between the past and the future.

However, there can be a radical innovation: over time, technologies can integrate and converge with other technological factors and create a brand-new technology with an overarching and significant impact on the market. Therefore, based on this study's findings, future research could aim at forecasting a radical innovation by approaching the formation of technology clusters from a different perspective or carrying out a case study on other industries.

Previous technology forecast studies were limited because the patent analysis considered only one piece of patent information, such as abstract, citation, or number of applications. Another limitation is that the studies were too biased on the technology experts during the analytical process. However, we forecasted promising technology by considering information on CPCs, forward citations, triadic patent family, independent claims, and the number of patents, which has the advantage of providing easy access for researchers without a high level of knowledge of certain technologies.

The methodological approach proposed in this study showed that R&D decision-makers can make objective forecasts and judgments based solely on patent data. Moreover, in addition to forecasting amalgamated technologies in the long-term view, which cannot be identified, it will be possible to obtain effective results from the short-term view in judging the prospects of existing technologies. Thus, we expect that this study will be helpful for companies or countries that need fast and accurate promising technology forecasts to make wise determinations of technology management because the outcomes of technology forecasts are crucial for companies and countries in establishing technology management strategies and R&D policies.

The limitation and further direction of this study are as follows. First, this study only applied patent documents on wellness care for experimental verification. Therefore, studies that forecast promising technology through the same method for various other industries are needed. Furthermore, to evaluate the prospects of technology clusters, this study utilized patent citation information, independent claims, and family patent information.

However, there seems to be a need for future studies to take advantage of patent information with numerous technological implications, including information about patents in dispute, whether patents are standard, and technology transfer information. Additionally, we expect that more accurate future promising technology forecasts will be possible if papers or other reports on technology trends are considered together because evaluations, such as the growth prospects of the wellness market and its compatibility with other technologies and industries, may be subject to limitations if they are only based on patent analysis. Considering financial information and other types of data will enable more specific promising technology forecasts.

Appendix A. CPC descriptions

CPC	Description
a61b5/0002	Remote monitoring of patients using telemetry, e.g. transmission of vital signals via a communication network
a61b5/0006	ECG or EEG signals
a61b5/0022	Monitoring a patient using a global network, e.g. telephone networks, internet
a61b5/0205	Simultaneously evaluating both cardiovascular conditions and different types of body conditions, e.g. heart and respiratory condition
a61b5/02055	Simultaneously evaluating both cardiovascular condition and temperature
a61b5/024	Detecting, measuring or recording pulse rate or heart rate
a61b5/1118	Determining activity level
a61b5/14532	For measuring glucose, e.g. by tissue impedance measurement
g06f17/30265	Based on information manually generated or based on information not derived from the image data
g06f19/321	Related medical protocols such as digital imaging and communications in medicine protocol; Editing of medical image data, e.g. adding diagnosis information
g06f19/322	Management of patient personal data, e.g. patient records, conversion of records or privacy aspects
g06f19/323	On a portable record carrier, e.g. CD, smartcard or RFID
g06f19/327	Management of hospital data, e.g. scheduling of medical staff or operation rooms, measuring the quality or efficiency of medical staff
g06f19/328	Health insurance management, e.g. payments or protection against fraud
g06f19/3406	Dedicated hardware interfaces; Local monitoring or local control of medical devices, e.g. configuration parameters, graphical user interfaces [GUI]
g06f19/3418	Telemedicine, e.g. remote diagnosis, remote control of instruments or remote monitoring of patient carried devices
g06f19/345	Medical expert systems, neural networks or other automated diagnosis
g06f19/3456	Computer-assisted prescription or delivery of medication, e.g. prescription filling or compliance checking
g06f19/3481	Telemedicine, e.g. remote diagnosis, remote control of instruments or remote monitoring of patient carried devices
g06f19/3487	Medical report generation
g06q10/06	Resources, workflows, human or project management, e.g. organizing, planning, scheduling or allocating time, human or machine resources; Enterprise planning; Organizational models
g06q10/10	Office automation, e.g. computer aided management of

(continued on next page)

(*continued*)

CPC	Description
	electronic mail or groupware, Time management, e.g. calendars, reminders, meetings or time accounting
g06q50/22	Health care
g06q50/24	Patient record management
g06t7/0012	Biomedical image inspection
y10s707/99945	Object-oriented database structure processing

References

Albright, R.E., 2002. What can past technology forecasts tell us about the future? Technol. Forecast. Soc. Chang. 69 (5), 443–464.
Altuntas, S., Dereli, T., Kusiak, A., 2015. Forecasting technology success based on patent data. Technol. Forecast. Soc. Chang. 96, 202–214.
Amer, M., Dain, T.U., Jetter, A., 2013. A review of scenario planning. Futures 46, 23–40.
Anon, May 2010. Global Spa Summit, Spas and the Global Wellness Market: Synergies and Opportunities. SRI International.
Anon, 2015. Power to the patient: How mobile technology is transforming healthcare. Economist 1–26.
Baudery, M., Dumont, B., 2006. Comparing firms' triadic patent applications across countries: is there a gap in terms of R&D effort or a gap in terms of performances? Res. Policy 35 (2), 324–342.
Belvard, P., 2000. Defence by patent applications. Technol. Law Insur. 5 (3–4), 187–190.
Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technol. Forecast. Soc. Chang. 73 (7), 835–844.
Chang, S., Fan, C., 2016. Identification of the technology life cycle of telematics: a patent-based analytical perspective. Technol. Forecast. Soc. Chang. 105, 1–10.
Chang, S.B., Lai, K.K., Chang, S.M., 2009. Exploring technology diffusion and classification of business methods: using the patent citation network. Technol. Forecast. Soc. Chang. 76 (1), 107–117.
Chang, P.L., Wu, C.C., Leu, H.J., 2012. Investigation of technological trends in flexible display fabrication through patent analysis. Disp. 33 (2), 68–73.
Cheng, A.C., Chen, C.Y., 2008. The technology forecasting of new materials: the example of nanosized ceramic powders. Romanian J. Econ. Forecast. 9 (4), 88–110.
Chiu, Y.J., Ying, T.M., 2012. A novel method for technology forecasting and developing R&D strategy of building integrated photovoltaic technology industry. Math. Probl. Eng. 2012, 1–24.
Choi, S., Jun, S., 2014. Vacant technology forecasting using new Bayesian patent clustering. Tech. Anal. Strat. Manag. 26 (3), 241–251.
Cox, T., Cox, M., 2001. Multidimensional Scaling. second ed. Chapman Hall, Boca Raton.
Ernst, H., 2013. Patent information for strategic technology management. World Patent Inf. 25 (3), 233–242.
Fallah, M.H., Fishman, E., Reilly, R.R., 2009. Forward patent citations as predictive measures for diffusion of emerging technologies. Portland Int. Conf. Manag. Eng. Technol. 420–427.
Fischer, T., Henkel, J., 2012. Patent trolls on markets for technology – an empirical analysis of NPEs' patent acquisitions. Res. Policy 41 (9), 1519–1533.
Fischer, T., Leidnger, J., 2014. Testing patent value indicators on directly observed patent value – an empirical analysis of Ocean Tomo patent auctions. Res. Policy 43 (3), 519–529.
Fujii, H., Yoshida, K., Sugimura, K., 2016. Research and development strategy in biological technologies: a patent data analysis of Japanese manufacturing firms. Sustainability 8 (4), 351.
Halal, W.W., Kull, M.D., Leffmann, A., 1998. The George Washington University forecast of emerging technologies a continuous assessment of the technology revolution. Technol. Forecast. Soc. Chang. 59 (1), 195–204.
Hartigan, J.A., Wong, M.A., 1979. A k-means clustering algorithm. Appl. Stat. 28 (1), 100–108.
Holmes, C., Ferrill, M., 2005. The application of operation and technology roadmapping to aid Singaporean SMEs identify and select emerging technologies. Technol. Forecast. Soc. Chang. 72 (3), 349–357.
Jeong, Y., Yoon, B., 2015. Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. Technovation 39–40, 37–52.
Jeong, Y., Lee, K., Yoon, B., Phaal, R., 2015. Development of a patent roadmap through the Generative Topographic Mapping and Bass diffusion model. J. Eng. Technol. Manag. 38, 53–70.
Johnson, P.J., Jou, J., Rhee, T.G., Rochwood, T., Upchurch, D., 2016. Complementary health approaches for health and wellness in midlife and older US adults. Maturitas 89, 36–42.
Ju, Y., Sohn, S.Y., 2015. Patent-based QFD framework development for identification of emerging technologies and related business models: a case of robot technology in Korea. Technol. Forecast. Soc. Chang. 94, 44–64.
Jun, S., Park, S.S., 2013. Examining technological innovation of Apple using patent analysis. Ind. Manag. Data Syst. 113 (6), 890–907.
Jun, S., Park, S.S., Jang, D.S., 2012a. Patent management for technology forecasting: a case study of the bio-industry. J. Intellect. Prop. Rights 17 (6), 539–546.
Jun, S., Park, S.S., Jang, D.S., 2012b. Technology forecasting using matrix map and patent clustering. Ind. Manag. Data Syst. 112 (5), 786–807.
Kapoor, R., Karvonen, M., Ranaei, S., Kässi, T., 2015. Patent portfolios of European wind industry: new insights using citation categories. World Patent Inf. 41, 4–10.
Kent, M.L., Saffer, A.J., 2014. A Delphi study of the future of new technology research in public relations. Public Relat. Rev. 40 (3), 568–576.
Kim, T.I., 2015. Health technology in perspective. J. Periodontal Implant. Sci. 45 (1), 1.
Kim, Y.G., Suh, J.H., Park, S.C., 2008. Visualization of patent analysis for emerging technology. Expert Syst. Appl. 31 (3), 1804–1812.
Kim, G.J., Park, S.S., Jang, D.S., 2015. Technology forecasting using topic-based patent analysis. J. Sci. Ind. Res. 74 (5), 265–270.
Lee, Y.G., Lee, J.D., 2010. Different characteristics between auctioned and non-auctioned patents. Scientometrics 82 (1), 135–148.
Lee, W.S., Sohn, S.Y., 2013. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. Proc. Fall Conf. Korea Indust. Eng. Soc. 1153–1170.
Lee, S., Yoon, B., Lee, C., Park, J., 2009. Business planning based on technological capabilities: patent analysis for technology-driven roadmapping. Technol. Forecast. Soc. Chang. 76 (6), 769–786.
Lee, C., Cho, Y., Seol, H., Park, Y., 2012. A stochastic patent citation analysis approach to assessing future technological impacts. Technol. Forecast. Soc. Chang. 79 (1), 16–29.
Lee, S., Cho, C., Hong, E., Yoon, B., 2016. Forecasting mobile broadband traffic: application of scenario analysis and Delphi method. Expert Syst. Appl. 44, 126–137.
Manap, N.A., Shapiee, R.B., Tehrani, P.M., Shariff, A.A.M., 2016. Protecting R&D inventions through intellectual property rights. J. Intellect. Prop. Rights 21 (2), 110–116.
Meade, N., Islam, T., 2015. Forecasting in telecommunications and ICT-a review. Int. J. Forecast. 31 (4), 1105–1126.
Mueller, S.C., Sandner, P.H., Welpe, I.M., 2015. Monitoring innovation in electrochemical energy storage technologies: a patent-based approach. Appl. Energy 137 (1), 537–544.
No, H.J., An, Y., Park, Y., 2015a. A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining. Technol. Forecast. Soc. Chang. 97, 181–192.
Park, Y., Yoon, B., Lee, S., 2005. The idiosyncrasy and dynamism of technological innovation across industries: patent citation analysis. Technol. Soc. 27 (4), 471–485.
Park, I., Jeong, Y., Yoon, B., Mortara, L., 2015a. Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. Tech. Anal. Strat. Manag. 27 (7), 759–781.
Park, S., Lee, S., Jun, S., 2015b. A network analysis model for selecting sustainable technology. Sustainability 7, 13126–13141.
Park, S., Lee, S., Jun, S., 2015c. A network analysis model for selecting sustainable technology. Sustainability 7 (10), 13126–13141.
Pätäri, S., 2010. Industry- and company-level factors influencing the development of the forest energy business—insights from a Delphi study. Technol. Forecast. Soc. Chang. 77, 94–109.
Pénin, J., 2012. Strategic uses of patents in markets for technology: a story of fabless firms, brokers and trolls. J. Econ. Behav. Organ. 84 (2), 633–641.
Petruzzelli, A.M., Rotolo, D., Albino, V., 2015. Determinants of patent citations in biotechnology: an analysis of patent influence across the industrial and organizational boundaries. Technol. Forecast. Soc. Chang. 91, 208–221.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput. Appl. Math. 20, 53–65.
Saritas, O., Aylen, J., 2010. Using scenarios for roadmapping: the case of clean production. Technol. Forecast. Soc. Chang. 77 (7), 1061–1075.
Shih, M.J., Liu, D.R., Hsu, M.L., 2010. Discovering competitive intelligence by mining changes in patent trends. Expert Syst. Appl. 37 (4), 2882–2890.
The United States Patent and Trademark Office (USPTO) (2015) www.uspto.gov.
Trappey, C.V., Wu, H.Y., Dutta, F.T., Trappey, A.J.C., 2011. Using patent data for technology forecasting: China RFID patent analysis. Adv. Eng. Inform. 25 (1), 53–64.
Trappey, A.J.C., Trappey, C.V., Wu, C.-Y., Lin, C.-W., 2012. A patent quality analysis for innovative technology and product development. Adv. Eng. Inform. 26 (1), 26–34.
Tseng, F., Hsieh, C., Peng, Y., Chu, Y., 2011. Using patent data to analyze trends and the technological strategies of the amorphous silicon thin-film solar cell industry. Technol. Forecast. Soc. Chang. 78 (2), 332–345.
Wang, J., Wang, C., Wu, C., 2015. A real options framework for R&D planning in technology-based firms. J. Eng. Technol. Manag. 35, 93–114.
Yoon, J., Kim, K., 2012. An analysis of property-function based patent networks for strategic R&D planning in fast-moving industries: the case of silicon-based thin film solar cells. Expert Syst. Appl. 39 (9), 7709–7717.
Zhao, R., Zhao, L., 2015. Trends in patents for solar thermal utilization in China. Renew. Sust. Energ. Rev. 52, 852–862.
Zhou, Y., Li, X., Lema, R., Urban, F., 2015. Comparing the knowledge bases of wind turbine firms in Asia and Europe: patent trajectories, networks, and globalisation. Sci. Public Policy 1, 1–16.

Gabjo Kim received his Ph.D. in Industrial Management Engineering from Korea University in 2015. He is currently an associate research engineer on the Government Cooperation Team, Korea Intellectual Property Strategy Institute, Seoul, Korea. His research interests are technology management and forecasting through patent analysis.

Jinwoo Bae received his Ph.D. in Electronic Engineering from Kwangwoon University in 2006. He is a senior research engineer on the Government Cooperation Team, Korea Intellectual Property Strategy Institute, and an adjunct professor at the Graduate School of Management of Technology, Hanyang University, Seoul, Korea. He served as an expert adviser for the Korea Invention Promotion Association from 2006 to 2009. His research interests include technology valuation assessment, patent strategy establishment, and technology forecasting.

Please cite this article as: Kim, G., Bae, J., A novel approach to forecast promising technology through patent analysis, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.11.023