Technological Forecasting & Social Change 127 (2018) 291–303

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Early identification of emerging technologies: A machine learning approach using multiple patent indicators

Changyong Lee[a,*], Ohjin Kwon[b], Myeongjung Kim[a], Daeil Kwon[c]

[a] School of Management Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea
[b] Korea Institute of Science and Technology Information, 66 Hoegi-ro, Dongdaemun-gu, Seoul 130-741, Republic of Korea
[c] School of Mechanical, Aerospace, and Nuclear Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea

ARTICLE INFO

Keywords:
Technology forecasting
Emerging technologies
Early identification
Machine learning models
Multiple patent indicators

ABSTRACT

Patent citation analysis is considered a useful tool for identifying emerging technologies. However, the outcomes of previous methods are likely to reveal no more than current key technologies, since they can only be performed at later stages of technology development due to the time required for patents to be cited (or fail to be cited). This study proposes a machine learning approach to identifying emerging technologies at early stages using multiple patent indicators that can be defined immediately after the relevant patents are issued. For this, first, a total of 18 input and 3 output indicators are extracted from the United States Patent and Trademark Office database. Second, a feed-forward multilayer neural network is employed to capture the complex nonlinear relationships between input and output indicators in a time period of interest. Finally, two quantitative indicators are developed to identify trends of a technology's *emergingness* over time. Based on this, we also provide the practical guidelines for implementation of the proposed approach. The case of pharmaceutical technology shows that our approach can facilitate responsive technology forecasting and planning.

1. Introduction

Emerging technologies are of great interest to a wide range of stakeholders in both industry and government who aim to set up investment-related strategies (Rotolo et al., 2015). The existing literature has shown that patent citation information is useful for measuring the economic value of a technology (Lerner, 1994; Narin et al., 1987). In this respect, many methods – such as cluster analysis, association rule mining, and conjoint analysis – have been employed to identify emerging technologies using patent citation information. However, the outcomes of previous studies are not forward-looking because most have been limited to *ex post* evaluation which measures past performance, impacts, or consequences (Lee et al., 2016). The value of predictive analysis for identifying emerging technologies has seldom been addressed.

Arguably, the most scientific approaches to identifying emerging technologies use curve fitting techniques (Daim et al., 2006; Shin et al., 2013) and stochastic models (Jang et al., 2017; Lee et al., 2011; Lee et al., 2012; Lee et al., 2016; Lee et al., 2017) to show future-projected trends of a technology by estimating the future citation counts of the relevant patents as a quantitative proxy. Curve fitting techniques using least squares estimation or least absolute deviation fit growth curves to time-series patent citation data and extrapolate those curves beyond the range of the data, whereas stochastic models estimate probability distributions of patent citations in the future by analysing fluctuations observed in historical data. However, the outcomes of these methods are likely to reveal no more than current key technologies, since they can only be performed at later stages of technology development due to the time required for patents to be cited (or fail to be cited) (Haupt et al., 2007). It should be noted that the time lag between citing and cited patents is found to be between 4 and 5 years on average (Verspagen and De Loo, 1999), and the latest patents have naturally less chance to be cited by other patents (Karki, 1997). Moreover, these methods have been criticised due to their reliance on making assumptions about pre-determined growth curves and probability distributions (Jang et al., 2017; Lee et al., 2011; Lee et al., 2012; Lee et al., 2016; Lee et al., 2017; Shin et al., 2013), which are difficult to identify at early stages of technology development and are heterogeneous across technologies. Hence, curve fitting techniques and stochastic models are of little practical assistance in identifying emerging technologies, especially when a technology is at its early stages and there is no historical data (Jang et al., 2017).

As a remedy, we propose a machine learning approach to identifying emerging technologies at early stages using multiple patent

* Corresponding author.
E-mail addresses: changyong@unist.ac.kr (C. Lee), dbajin@kisti.re.kr (O. Kwon), audwnd0215@unist.ac.kr (M. Kim), dkwon@unist.ac.kr (D. Kwon).

https://doi.org/10.1016/j.techfore.2017.10.002
Received 8 April 2017; Received in revised form 10 August 2017; Accepted 2 October 2017
Available online 20 November 2017
0040-1625/

indicators that can be defined immediately after the relevant patents are issued. Economic and innovation literature has presented a wide range of patent indicators – such as patent family and originality – that may be indicative of the future citation count of patents and that further the relevant technology's economic value (Lerner, 1994; Narin et al., 1987). The tenet of this research is that analysis of those patent indicators can provide evidence for a patent's value and further the relevant technology's value in the future. For this, first, a total of 18 input and 3 output indicators are extracted from the United States Patent and Trademark Office (USPTO) database. Second, a feed-forward multilayer neural network – that is a supervised machine learning technique inspired by attempts to model the neuro-physical structure of the human brain – is employed to capture the complex nonlinear relationships between input and output indicators in a time period of interest. The primary advantage of this method for identifying emerging technologies is its ability to infer a function from observations (Buscema et al., 2017). It should be noted that there is no theoretical understanding of the relationships between those patent indicators, and moreover, the complexity and nonlinearity associated with innovation processes makes the design of a certain function impractical (Chen et al., 2012). Finally, two quantitative indicators are developed to identify trends of a technology's *emergingness* over time. Based on this, we also provide the practical guidelines for the implementation of our approach in terms of the choice of machine learning models and model update.

We applied the proposed approach to support Korean small and medium-sized high tech companies in technology forecasting at the request of the Korea Institute of Science and Technology Information (KISTI). We adopted the USPTO database for this research, since it contains the most representative data for analysing international technology (Lee et al., 2013). Our experience showed that the proposed approach can find emerging technologies at early stages, using the limited patent indicators that can be defined and extracted immediately after the relevant patents are issued. Our method also enabled us to perform systematic and continuous monitoring of emerging technologies, yielding high potential benefits at relatively low cost. Moreover, the results of our case study enabled us to identify a way to improve the proposed approach, which we expect to be a useful complementary tool to support experts' decision making in emerging technologies, especially for small and medium-sized high-tech companies. We believe that the systematic process and quantitative outcomes our approach offers can facilitate responsive technology forecasting and planning.

This paper is organised as follows. Section 2 presents the background to our research and Section 3 explains the research framework and methodology, which are then illustrate by a case study on pharmaceutical technology in Section 4. Section 5 provides the guidelines for implementation of our approach. Finally, Section 6 offers our conclusions.

2. Background

2.1. Definitions and characteristics of emerging technologies

Although emerging technologies have been the subject of many previous studies, there is no consensus as to what qualifies a technology to be emergent (Rotolo et al., 2015). As Table 1 reports, the definitions and concepts of emerging technologies presented by a number of studies overlap, but at the same time, point to different characteristics. For instance, Day and Schoemaker (2000) defined an emerging technology as a science-based innovation that has the potential to create a new industry or to transform existing ones. Porter et al. (2002) referred to an emerging technology as a technology that could exert much enhanced economic influence in the coming 15-year horizon. Considering that the economic influence of emerging technologies should exert not just on a specific domain, but also on the entire socio-economic system, Martin (1995) introduced the concept of general emerging technologies, and emphasised the wide scope and convergence of technological fields as

Table 1
Definition and characteristics of emerging technologies.

Authors	Definition	Characteristics						
		Prominent impact	Scope and coverage	Uncertainty and ambiguity	Development effort and developers' capabilities	Novelty	Science-intensity	Coherence	Growth speed
Day and Schoemaker (2000)	Science-based innovation that has the potential to create a new industry.	✓	✓	✓	✓	✓	✓	✓	
Porter et al. (2002)	Technology that could much enhanced economic influence in the coming 15-year horizon	✓		✓	✓				
Martin (1995)	Technology that exploits a wide scope and coverage of technology fields	✓	✓		✓		✓		
Cozzens et al. (2010)	Technology that shows high potential but has not yet demonstrated its value	✓		✓			✓		✓
Small et al. (2014)	Technology that has universal agreement on its novelty and growth					✓			✓

key characteristics. Taking the uncertainty and ambiguity into account, Cozzens et al. (2010) conceptualised an emerging technology as a technology that shows high potential but has not demonstrated its value or settled down into any kind of consensus. Put those things together, Rotolo et al. (2015) summarised five key characteristics of emerging technologies, which are (i) novelty, (ii) fast growth, (iii) coherence, (iv) prominent impact, and (v) uncertainty and ambiguity.

2.2. Patent-based approach to identifying emerging technologies

Given such multi-facet characteristics of emerging technologies, modelling and analysing emerging technologies is a task beset with hazards, such as uncertainty, data unreliability, and the complexity of real world feedback. As such, in practice, identifying emerging technologies has relied largely on expert-centric approaches such as Delphi and large-scale survey methods (Jeong and Kim, 1997). However, experts' judgements are mostly subjective; they often have difficulty in defending their judgements rationally (Lee et al., 2016). Moreover, these expert-centric approaches have become time-consuming and labour-intensive as technologies proliferated unceasingly and innovation cycles become shorter (Lee et al., 2012). As a consequence, industrial practitioners demand quantitative methods based on objective information as a complement to experts' judgements.

In academia, highlighting possible avenues for methodological adaptation, there have been certain shifts in the direction of research on the identification of emerging technologies, from expert-based qualitative approaches to engineering-centric quantitative approaches. We can summarise the major studies' results as follows: Cho and Shih (2011) presented structural holes indicators to identify rapidly developing technologies using patent citation networks. Breitzman and Thomas (2015) developed the emerging clusters model to identify emerging technologies using patent citation information from multiple systems. Arora et al. (2013) proposed an updated search approach for identifying emerging fields of technology by using inclusion and exclusion terms. Lee et al. (2015) proposed integration of text mining techniques and the local outlier factor to identify novel patents. Gerken and Moehrle (2012) introduced a semantic patent analysis to measure the distance between patents to identify high novelty invention. Yoon and Kim (2011) proposed a subject-action-object (SAO)-based semantic patent analysis to identify rapidly evolving technological trends. Joung and Kim (2017) presented a technical keyword-based patent analysis to monitor emerging technologies. Ju and Sohn (2015) proposed a quality function deployment (QFD) framework to facilitate R&D planning for emerging technologies that reflect customers' future needs.

Focusing more on the dynamics of emerging technologies, Shin et al. (2013) employed curve fitting techniques to calculate the expected number of patent citations and its variance, as proxy for future returns and risks of molecular amplification technologies. Lee et al. (2012) and Jang et al. (2017) proposed a stochastic patent citation analysis to assess the future impacts of image superposition technologies and molecular amplification technologies in a time period of interest by employing the future citation count as a proxy. Lee et al. (2016) and Lee et al. (2017) developed a stochastic technology life cycle analysis to examine and forecast a technology's progression through its life cycle using patent indicators, and conducted case studies of molecular amplification technologies and lithography technologies, respectively.

However, while all these previous studies have proved valuable in using quantitative data and scientific methods, and providing insights into emerging technologies, they are not effective when a technology is at the early stages of technology development, and cannot incorporate the multi-facet characteristics of emerging technology into analyses. These drawbacks provide our underlying motivation and are fully addressed in this study. Table 2 summarises the difference between previous research and the current paper.

3. Methodology

Fig. 1 shows the overall process of the proposed approach. Given the complexities involved, the proposed approach is designed to be executed in four discrete steps: data collection and pre-processing; defining and extracting patent indicators; assessing the value of patents; and identifying trends of a technology's *emergingness*.

3.1. Data collection and pre-processing

Once a technology field of interest is chosen, the relevant patents (Set 1) are collected based on certain search conditions. The patent documents collected at this stage are a mixture of both structured and unstructured data in either HTML or XML formats. The documents are thus parsed according to the types of information (e.g. patent number, assignee, and class) and stored in a structured patent database. In addition to this, the patents that are published by the assignees of Set 1 patents (Set 2) are collected and pre-processed to analyse the capabilities and efforts of technology developers. Moreover, the patents that cite the Set 1 and Set 2 patents (Set 3) are collected and pre-processed to analyse the patent citations made by later patents. The integrated patent database thereby includes information on both citing and cited patents.

3.2. Defining and extracting patent indicators

This study employs multiple patent indicators that may be indicative of the value of patents and further the value of the relevant technologies, instead of using a single indicator such as patent applications (Von Wartburg et al., 2005) and citations (Jang et al., 2017; Lee et al., 2012; Shin et al., 2013). Previous studies have identified the key characteristics of emerging technologies, but not all of them can be considered in the early stages of technology development. It should be noted that such characteristics as *fast growth* and *coherence* need continuous monitoring and surveillance to be defined and measured. For this reason, considering *what* patent indicators measure and *when* they can be extracted from the database, we define a total of 21 patent indicators, which are composed of two broad categories: (1) potential impact as output indicators and (2) technological characteristics as input indicators. Table 3 summarises the patent indicators employed in this study.

3.2.1. Potential impact as output indicators

There is a nearly universal agreement that emerging technologies have significant potential impact (Martin, 1995). Many empirical studies have found that there is a significant positive relationship between the number of forward citations of patents and technological impact or economic value of the patents (Lerner, 1994; Narin et al., 1987). Following this convention, we employ the number of forward citations of a patent as a proxy for the potential impact of the patent, which is used as a target variable of our model. Specifically, we measure the number of forward citations of a patent over the next three, five, and ten years after the patent is issued so as to aid decision making in short-, mid-, and long-term technology planning.

3.2.2. Technological characteristics as input indicators

Economic and innovation literature has presented a variety of patent indicators to assess the characteristics of patented technologies. This study employs a total of 18 indicators to capture the key characteristics of emerging technologies, which are divided into five subcategories: (1) novelty, (2) science-intensity, (3) growth speed, (4) scope and coverage, and (5) development effort and capabilities. It should be noted that these indicators can be extracted from patent databases immediately after the relevant patents are issued. Moreover, some indicators are defined at different levels (e.g. class and mainline sub-class level or core area, peripheral area, and total level) which

Table 2
Comparisons of previous research and the current paper.

Factor	Previous research	Current paper
Approach	Probabilistic approach	Machine learning approach
Data	Mainly patent citation information	Multiple patent indicators
Method	Curve fitting techniques and stochastic models	Supervised models (e.g. artificial neural networks, random forest, and support vector machine)
Results and implications	Current key technologies	Emerging technologies
Time available	Later stages of technology development	Early stages of technology development – immediately after the relevant patents are issued

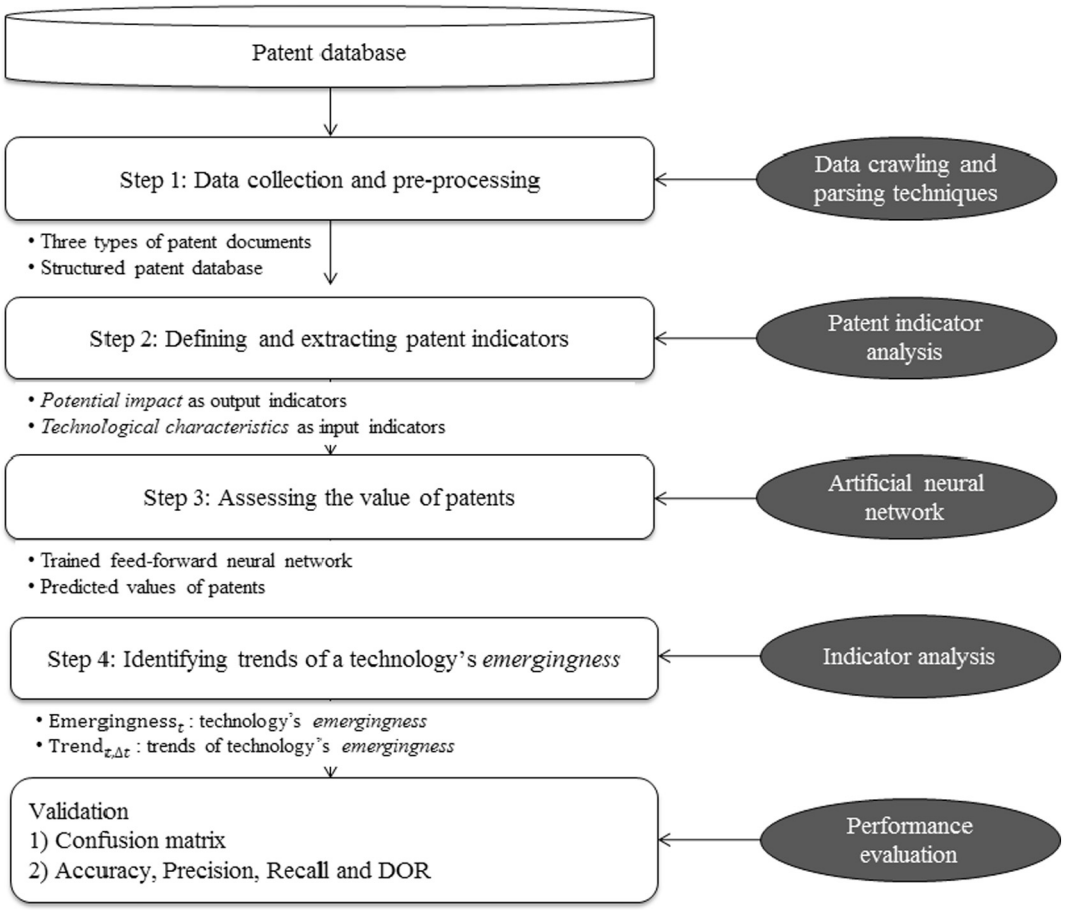

Fig. 1. Overall process of the proposed approach.

enable a fine-grained characterisation of technological novelty and firms' development effort and capabilities (Aharonson and Schilling, 2016; Fleming, 2001). These indicators are used as input variables of our model, respectively.

3.2.2.1. Novelty. This sub-category has two types of patent indicators to represent the novelty of patents. The first indicator is technological originality which has been a widely accepted characteristic of emerging technologies (Jaffe and Trajtenberg, 2002). Given that inventions represent a combination of existing ideas, patents based on a wider set of ideas may indicate more valuable knowledge (Fernández-Ribas, 2010). Previous studies have found the significant relationships between the technological originality of a patent and the value of the relevant technology (Bessen, 2008). This indicator uses the diversity of prior arts for a patent and captures how much the patented invention draws from different sets of technologies. We measure the technological originality of patents at two different levels. The class-level technological originality is calculated from classes of cited patents, whereas the mainline subclass-level technological originality is calculated from mainline subclasses of cited patents, as shown below:

$$1 - \sum_{j \in S_B} B_j^2 \quad (1)$$

where B_j is the ratio of the number of cited patents that belong to class (or mainline subclass) j to the total number of cited patents and S_B is the set of classes of cited patents. The second indicator, prior knowledge, represents the relationship between the target patent and prior arts. According to US patent law, inventors must cite all related prior publications and patents in their patent applications (Harhoff et al., 1999), and a patent examiner is responsible for insuring that all appropriate patents have been cited (Alcacer et al., 2009; Criscuolo and Verspagen, 2008; Lanjouw and Schankerman, 2001). Prior literature has identified that patents with large numbers of backward citations have a relatively low novelty and low monetary value (Harhoff et al., 2003). This indicator has also been used as a key variable for detecting a technology's stage transitions in its life cycle (Haupt et al., 2007). For this reason, we measure the number of backward citations of a patent as a proxy for prior knowledge of the patent.

Table 3
Summary of patent indicators employed in this research.

Category	Sub-category	Patent indicator	Operational definition
Potential impact as output indicators	Potential impact	Forward citations over three years (FC_3)	Number of forward citations over the next three years after a patent is issued
		Forward citation over five years (FC_5)	Number of forward citations over the next five years after a patent is issued
		Forward citation over ten years (FC_{10})	Number of forward citations over the next ten years after a patent is issued
Technological characteristics as input indicators	Novelty	Class-level technological originality (CTO)	Herfindahl index on classes of cited patents
		Mainline subclass-level technological originality (STO)	Herfindahl index on mainline subclasses of cited patents
		Prior knowledge (PK)	Number of backward citations
	Science-intensity	Scientific knowledge (SK)	Number of non-patent literature references
	Growth speed	Technology cycle time (TCT)	Median age of cited patents
	Scope and coverage	Main field (MF)	Main class to which a patent belongs
		Technological scope (TS)	Number of classes to which a patent belongs
		Commercial scope (CS)	Number of patents registered in multiple countries with the coverage of the same invention
		Protection coverage described in independent claims (PCID)	Number of independent claims
		Protection coverage described in dependent claims (PCD)	Number of dependent claims
	Development effort and capabilities	Collaboration (COL)	1 if a patent has more than one assignee, otherwise 0
		Inventors (INV)	Number of inventors
		Total know-how (TKH)	Number of patents issued by an assignee
		Core area know-how (CKH)	Number of patents in a technology field of interest issued by an assignee
		Peripheral area know-how (PKH)	Number of patents in other technology fields issued by an assignee
		Total technological strength (TTS)	Number of forward citations of patents issued by an assignee
		Core area technological strength (CTS)	Number of forward citations of patents in a technology field of interest issued by an assignee
		Peripheral area technological strength (PTS)	Number of forward citations of patents in other technology fields issued by an assignee

3.2.2.2. Science-intensity. This sub-category has one indicator to represent the science-intensity of patents. The basic assumption of this indicator is that more scientific knowledge contained in the patented invention may lead to the development of a more innovative and influential technology (Cozzens et al., 2010; Day and Schoemaker, 2000). The closeness to the scientific knowledge can be measured with the number of non-patent literature (NPL) references (Trajtenberg, 1990), and this provides a notion of the science-intensity of the patent. NPL references consist not only of scientific articles but other types of publications, such as conference proceedings and books (Callaert et al., 2012). It has been verified that there exists a close relationships between the number of NPL references of a patent and the value of the relevant technology (Callaert et al., 2006). For this reason, we measure the number of NPL references as a proxy for science-intensity invention.

3.2.2.3. Growth speed. This sub-category has one indicator to represent the growth speed of technology development. As stated earlier, growth speed cannot be fully observed at early stages of technology development without continuous monitoring and surveillance. Instead, this study employs technology cycle time that captures the degree of newness of prior knowledge or the pace of technology progress (Bierly and Chakrabarti, 1996; Kayal and Waters, 1999). This indicator is measured by the median age of patents cited by the target patent, and thus shorter technology cycle time reveals a fast technological progress from old to new technologies.

3.2.2.4. Scope and coverage. This sub-category has four types of patent indicators to represent the scope and coverage of patents. The first indicator is the main field to capture the major technological field of a patent. Patents are assigned to one or more classes to delineate the technology field they cover. While a patent belongs to multiple classes in many cases, the main class best represents the field where the patented technology can be applied (Lee et al., 2009). In this respect, we use the main class of a patent as a proxy for the main field of the patent, which is useful for modelling different patenting behaviours across technological fields (Chen and Chen, 2011). The second indicator is the technological scope that represents the scope of the technological fields of a patent. We use the number of classes to which a patent belongs as a proxy for scope of technological field of the patent. The third indicator represents the commercial scope which is measured by the number of patents registered in multiple countries with the coverage of the same invention (i.e. patent family). Patent law is territorial in nature (OuYang and Weng, 2011), and therefore only important patents are applied to many national or regional patent offices. Several studies have found the positive relationships between the size of patent family and the economic value of a patent (Guellec and de la Potterie, 2000). The last indicator is the protection coverage which indicates the scope of the legal protection conferred by a patent claims (Lanjouw and Schankerman, 2001). A patent's claims reflect the coverage of its protection, and are positively correlated with the scope and usefulness of the patent, and thus its value (Ernst, 2003). The number of claims has been used as a convenient proxy of a patent's potential value by patent-holders who are considering trying to renew their patent (Jeong et al., 2016). Given that patent claims can be divided into two parts – independent claims describing the essential features of a patent and dependent claims covering additional details, we utilise the number of dependent and independent claims as a proxy for protection coverage.

3.2.2.5. Development effort and capabilities. This sub-category has four types of patent indicators to capture the development effort and capabilities. The first indicator is the collaboration, which captures collaboration on the assignee side of patents. Previous studies have found that there is a significant positive relationship between co-assignee and the value of a patent (Ma and Lee, 2008; Meyer, 2006).

This indicator is 1 if a patent has more than one assignee, otherwise 0. Similarly, previous studies have shown that patents by multiple inventors are more significant than those by a single inventor (Ernst, 2003; Ma and Lee, 2008). For this reason, we measure the number of inventors as a proxy for human resources. The third and fourth indicators are concerned with developers' capabilities. The third indicator, know-how, measures the level of a firm's knowledge stock, and can be interpreted as technological and commercial interest in that field (Meyer, 2006). Specifically, we measure this indicator at three different levels. The total know-how is the total number of patents issued by an assignee, the core area know-how is the number of patents in a technology field of interest issued by an assignee, and the peripheral area know-how is the number of patents in other technology fields issued by an assignee. The fourth indicator, technological strengths, measures the value of a firm's prior patents and can be interpreted as the firm's technological strength or competitive position in a technological field (Ernst, 2003). Similar to know-how, we measure this indicator at three different levels. The total strength is the total number of forward citations of patents issued by an assignee, the core area strength is the number of forward citations of patents in a technology field of interest issued by an assignee, and the peripheral area strength is the number of forward citations of patents in other technology fields issued by an assignee.

3.3. Assessing the value of patents

We employ a feed-forward multilayer neural network to capture the complex nonlinear relationships between input and output indicators in a time period of interest. This method – that is inspired from attempts to model the neuro-physical structure of the human brain (Dayhoff and DeLeo, 2001) – is considered useful for assessing the value of patents for the following reasons. First, unlike previous curve fitting techniques and stochastic models, it does not require any assumptions about predetermined growth curves and probability distribution, since neural networks develop nonlinear mathematical models through empirical, inductive, and repeated learning processes (Bode, 1998). Second, it learns both the input and output patterns in the associative memory so that any minor variation in its input would be unlikely to affect the prediction quality of the output (Buscema et al., 2017). In many cases, neural networks have been found to outperform traditional statistical methods (i.e. linear regression, logistic regression, and discriminant analysis) in terms of accuracy and robustness.

The basic structure of feed-forward neural networks is shown in Fig. 2. The network consists of multiple layers of computational units and each node in a layer has directed connections to the nodes of the subsequent layer. In a feed-forward network, the information flows in a forward direction and the outputs of each intermediate layer are the inputs to the following layer. Specifically, a node in the input layer transmits the value it receives to the network. A node in the hidden and output layers receives an activation signal, which subtracts a bias factor from a weighted sum on its inputs. It then transforms the signal by applying activation functions such as linear, sigmoid, and ReLU functions, as shown in Eq. (2), where g is the activation function, w_{ij} is the weight of the link between nodes i and j, and β_{hj} is the bias factor for node j in the layer h.

$$N_j = g\left(\sum_i w_{ij}x_i + \beta_{hj}\right) \tag{2}$$

The operation of neural networks involves two steps: training and testing. In the training step, the network is taught to solve certain problems or identify specific patterns based on given information. Specifically, the network uses a training dataset to formulate connection weights for each node in the network. One of the most widely used training methods is the back-propagation algorithm. This algorithm modifies the input connection weights and bias factors so that the difference between output and target is minimised. The objective function E is shown in Eq. (3), where Y_i and $f(w_{ij}, x_i)$ indicate the output of the network and the actual target value.

$$\mathrm{E} = \sum_i \frac{1}{2}(Y_i - f(w_{ij}, x_i))^2 \tag{3}$$

The weights and bias factors are first modified for the nodes in the output layer, and the error is then propagated backward to the nodes that point to them, until the nodes in the input layer are reached, as shown in Eqs. (4) and (5), where l is the learning rate or decay parameter ranging between 0 and 1.

$$\beta_i^{new} = \beta_i^{old} + l(err_i) \tag{4}$$

$$\mathrm{w}_{ij}^{new} = w_{ij}^{old} + l(err_{ij}) \tag{5}$$

In the testing step, the performance of the trained neural network is measured by using a test dataset. Once the network is developed with an acceptable level of the error, it can be applied to new data.

The value of a patent cannot be ascertained fully: what we can predict is the future citation count of a patent that is found to be indicative of the value of the patent. However, predicting the exact future citation count of a patent is not the focus of our analysis. For example, can we say that a patent with 50 citations is more valuable than another patent with 45 citations? Considering this issue, we use a feed-forward neural network to classify patents according to their value measured by the expected future citation counts of the patents.

3.4. Identifying trends of a technology's emergingness

This step identifies emerging technologies by monitoring the development of valuable patents for a technology of interest. Here, the way of defining a technology may be different across the objectives of analysis. For instance, a technology can be defined as a set of patents that have common keywords or that belong to the same class. We define two indicators describing the trends of a technology's *emergingness*. Specifically, $\mathrm{Emergingness}_t$ is the weighted sum of the numbers of patents in value categories issued in time t, as shown in Eq. (6), whereas $\mathrm{Trend}_{t,\Delta t}$ measures the increasing and decreasing ratio of a technology's *emergingness*, as shown in Eq. (7). Here, how to define the weights depends on technological fields and analysis contexts (i.e., how much a specific value category is more important than others to identify the trends of a technology's *emergingness* in the field of interest). These indicators can be defined and calculated over the next three, five, and ten years given the issued year of the relevant patents, facilitating decision making in short-, mid- and long-term technology planning.

$$\mathrm{Emergingness}_t = \sum_{i=0}^{C} w_i N(PC_i) \tag{6}$$

$$\mathrm{Trend}_{t,\Delta t} = \frac{\mathrm{Emergingness}_{t+\Delta t} - \mathrm{Emergingness}_t}{\Delta t} \tag{7}$$

These two indicators can provide insight into a technology's *emergingness* in the present and in the future. For example, if the values of these two indicators for a subject technology are greater than those of another, the subject technology is emerging in the present and this trend is expected to deepen in the future. If the value of *emergingness* is greater than that of another but the value of *trend* is less than that of another, the subject technology is emerging in the present but this trend is expected to be slackened in the future.

4. Empirical analysis and results

4.1. Overview

We conducted a case study of pharmaceutical technology for three reasons. First, a patent normally equals a product in the pharmaceutical industry so that the technological value of a patent is directly related to

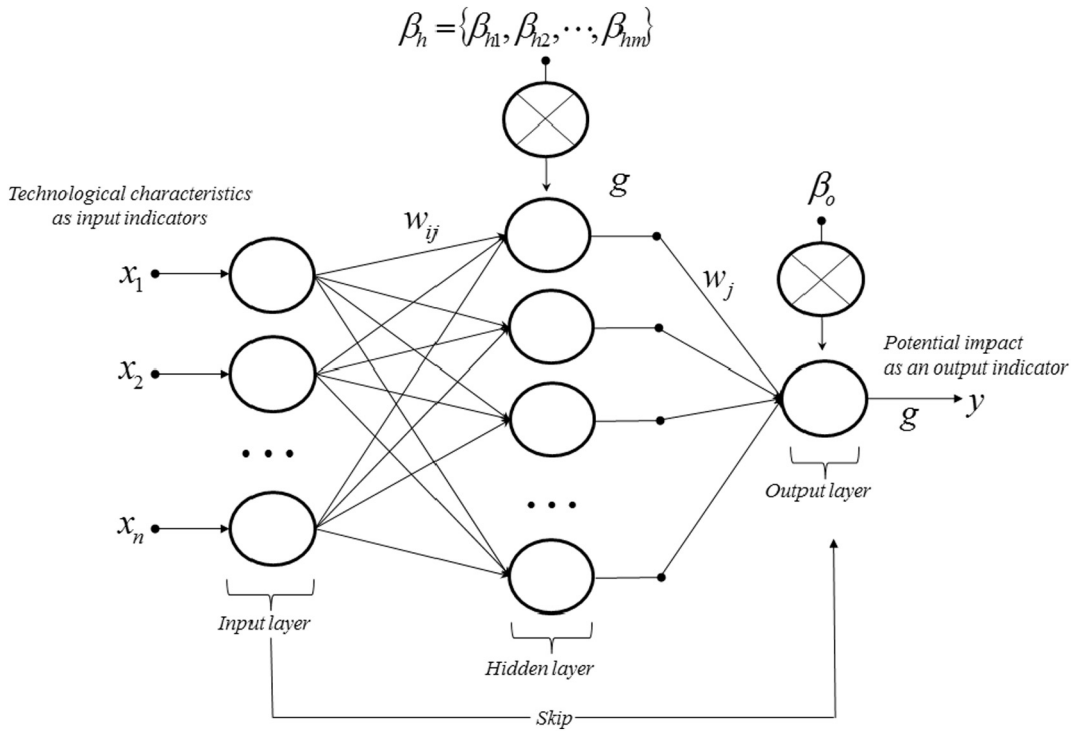

Fig. 2. Structure of feed-forward neural networks.

its commercial value (Chen and Chang, 2010). Second, patent management activities such as valuation and protection is especially important in this industry compared to those of other industries since the manufacturing process is relatively easy to replicate and can be copied with a fraction of the investment (Chaudhuri, 2005). Finally, industrial practitioners demand for objective information based on scientific methods to assess the future prospects of pharmaceutical technology development since this technology involves large scale investment and high risks for R&D (Chen and Chang, 2010). Failure in investment-related decision making may lead to huge losses. It is therefore worthwhile analysing the extensive information that patents pertaining to the pharmaceutical technology provide, so as to identify the emerging technologies and their future prospects.

4.2. Neural network approach to identifying emerging technologies

4.2.1. Data collection and pre-processing

The USPTO (http://www.uspto.gov) serves as our data collection source, since the US is the world's largest patent market – the majority of patents submitted to the USPTO are also submitted in other countries – and so is considered appropriate for analysing international technologies (Lee et al., 2013). The database is also well-organised and holds historical information back to 1976 (Lee et al., 2017).

A total of 35,356 patents (Set 1) that belongs to 424 class (entitled “drug, bio-affecting and body treating compositions”) were collected over the reference period, 2000 to 2009. Here, a Java-based web mining program was developed to download patents in HTML formats automatically, since the number was sufficiently large that we could not collect them all manually. These documents were then parsed based on their structures, distinguishing each document by its content, and details about patent numbers, assignees, citations, claims, classes, and other information were stored in a patent database that we constructed using Microsoft Office Access. In addition to this, the patents that are published by the assignees of Set 1 patents (Set 2) and the patents that cite the Set 1 and Set 2 patents (Set 3) were collected and stored in the same manner.

4.2.2. Defining and extracting patent indicators

We extracted a total of 18 input indicators and 3 output indicators from the database. As previously noted, predicting the exact future citation count of a patent is not the focus of our analysis: we employ a neural network to classify patents according to their value measured by the expected future citation counts of the patents. As such, we grouped patents into four categories according to the (expected) number of forward citations of patents in three, five, and ten years, as summarised in Table 4. Here, L1 patents are the most valuable while L4 patents are the least valuable. The resulting matrix was a 35,357 by 22 patent indicator matrix, which was used to train and test the neural network model. The matrix is not reported here in its entirety owing to lack of space, but part of the matrix is shown as Table 5. In the table, the first column represents the patent number followed by 18 input indicators, and the last three columns represent the categories of patents. For the 10,269 patents published after 2006, they do not have the value of FC_{10} since the number of forward citations in 10 years cannot be defined. Finally, the input indicators were normalised to make training faster and to prevent our neural network from getting stuck in local optimal values.

4.2.3. Assessing the value of patents

To begin with, the number of layers and the number nodes in a layer for a feed-forward neural network should be carefully determined. This task rather depends on the size and the nature of the data set (Chen et al., 2012). Although deep learning approaches using multiple hidden layers (i.e., convolutional neural networks and recurrent neural networks) have been found to be useful in such tasks as computer vision

Table 4
Four categories of patents according to the (expected) number of forward citations.

Category	(Expected) number of forward citations
L1	Above 20
L2	10–19
L3	2–9
L4	0 and 1

Table 5
Parts of the patent indicator matrix.

Patent no.	CTO	STO	PK	SK	TCT	TS	CS	…	INV	TKH	CKH	TTS	CTS	Category (FC3)	Category (FC5)	Category (FC10)
6010679	0.625	0.625	4	0	1994	3	36	…	2	–	–	–	–	L4	L3	L3
6010680	0.442	0.442	22	11	1992.5	1	144	…	4	81	62	1846	1419	L4	L4	L1
6010681	–	–	0	5	–	2	30	…	5	2	2	57	57	L3	L3	L3
6010682	0.423	0.474	28	9	1990	3	25	…	2	15	13	1128	1043	L4	L4	L3
6010683	0.692	0.86	39	0	1986	2	14	…	1	32	21	769	695	L3	L3	L3
6010684	–	–	0	0	–	2	15	…	1	1	1	25	25	L4	L4	L4
6010685	0	0	1	0	1984	2	19	…	3	2	1	11	5	L4	L4	L3
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
6495608	0.74	0.74	10	0	1992	3	1	…	1	18	7	280	110	L4	L4	L4
6497858	0	0	2	0	1994.5	2	8	…	3	4	4	31	31	L4	L4	L4
6497859	0	0	1	0	1994	2	3	…	2	7	7	89	89	L4	L4	L3
6497860	0	0.444	3	0	1988	2	17	…	2	12	11	58	52	L4	L4	L3
6497861	0.625	0.722	12	0	1998	2	35	…	5	43	41	615	601	L3	L2	L1
6497862	0.5	0.5	2	1	1992	3	11	…	4	1	1	6	6	L3	L3	L3
6497863	0.5	0.5	2	0	1987	2	18	…	4	4	3	6	11	L4	L4	L3
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
7638143	0.43	0.628	11	4	1999	2	35	…	2	4	4	53	53	L4	L2	–
7638144	0.695	0.836	16	3	1998	5	42	…	1	3	2	37	25	L4	L4	–
7638145	0.5	0.5	2	27	2005	2	32	…	2	77	49	884	462	L4	L4	–
7638146	–	–	0	4	–	3	7	…	1	–	–	–	–	L4	L4	–
7638147	–	–	0	0	–	1	9	…	1	1	1	0	0	L4	L4	–
7638148	0.694	0.694	7	14	1998	2	44	…	1	1	1	1	1	L4	L4	–
7638149	0.5	0.5	4	8	2001.5	3	21	…	3	8	7	16	16	L4	L4	–

Table 6
Results of assessment of the value of patents.

Patent no.	CTO	STO	PK	SK	TCT	TS	CS	…	INV	TKH	CKH	TTS	CTS	Category (3-year forecast)	Category (5-year forecast)	Category (10-year forecast)
6010679	0.625	0.625	4	0	1994	3	144	…	2	–	–	–	–	L4	L4	L3
6010680	0.442	0.442	22	11	1992.5	1	30	…	4	81	62	1846	1419	L4	L4	L1
6010681	–	–	0	5	–	2	25	…	5	2	2	57	57	L4	L4	L4
6010682	0.423	0.474	28	9	1990	3	14	…	2	15	13	1128	1043	L3	L3	L1
6066324	–	–	0	9	–	3	45	…	4	6	6	75	75	L4	L4	L4
6066325	0.743	0.847	48	8	1994	2	75	…	5	2	1	475	328	L4	L2	L1
6066326	0	0	2	3	1996.5	3	8	…	2	740	616	6853	5920	L4	L4	L3
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
6066339	0.108	0.108	35	0	1993	1	20	…	3	14	11	761	745	L4	L3	L1
6066340	0.531	0.599	18	13	1991.5	2	18	…	3	4	3	69	67	L4	L3	L2
6066341	0.611	0.778	6	0	1991.5	2	1	…	1	1	1	15	15	L4	L4	L3
6248323	0.667	0.667	3	1	1982	5	15	…	2	3	3	33	33	L4	L4	L3
6248324	0	0	6	0	1999.5	4	121	…	2	–	–	–	–	L2	L4	L4
6548078	0.32	0.32	5	6	1998	1	18	…	2	4	4	459	459	L4	L2	L1
6569463	0.24	0.24	15	0	1996	3	94	…	2	14	13	2921	2908	L1	L1	L1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
6994843	0.724	0.855	87	82	1995	4	480	…	2	67	65	2139	2122	L1	L1	L1
6994849	0	0	2	0	2000.5	5	15	…	1	1	1	41	41	L4	L3	L1
6994850	0.72	0.858	34	40	2002	1	36	…	1	36	25	201	101	L4	L4	L4
6994851	0.681	0.836	79	41	1991	3	146	…	2	10	7	173	156	L4	L3	L3
7622139	0.741	0.849	37	16	1994	2	43	…	3	119	87	1256	854	L4	L3	L1
7632489	0.544	0.825	26	0	2000	6	22	…	4	449	352	8179	6814	L4	L3	L3

and speech recognition that require extensive feature engineering work, increasing the number of hidden layers may lead to over-fitting problems and result in poor out-of-sample forecasting performance (Zhang, 2003). Previous studies have presented that one hidden layer is sufficient to approximate any continuous functions when there are limited numbers of input variables (Bode, 1998; Chen et al., 1990). Moreover, we observed that using multiple hidden layers did not result in significant additional performance in our case study, but did increase the run time considerably. As such, a parsimonious model with one hidden layer that performs sufficiently well was employed in this study. Similarly, the use of many nodes in a layer may cause the same problems. Considering these issues, following Kastra and Boyd (1996), we employed one input layer with fifteen nodes, one hidden layer with four nodes (i.e. square root of $n \times m$, where n and m denote the number of inputs and outputs), and one output layer with one node.

We used the *nnet* package provided by R to develop a feed-forward neural network with a back-propagation algorithm (Ripley et al., 2016). A logistic function was employed as an activation function since the use of sigmoid-type functions (i.e., hyperbolic tangent and arctangent function), in particular the use of losticistic functions, is the most popular (Dayhoff and DeLeo, 2001). Here, while this study employed a logistic function as an activation function, other globally diffrentiable functions can be utilised as an activation function of feed-forward multilayer neural networks trained by backpropagation algorithms according to the performance of analysis results (Chen and Chang, 1996). Moreover, the network was set to allow to skip the hidden layer (*skip*), if necessary, so that the optimal solution can be derived by connecting the input and output layer directly. The maximum number of iterations (*maxit*) for the training of the model was set to 1000, *entropy* was set to *TRUE* to fit maximum conditional likelihood and provide pobabilistic outputs, and *decay* was set to 0.001 after several experiments using cross-validation to find an optimal value.

Under this circumstance, a neural network model was developed to classify patents according to the predicted number of forward citations of a patent over the next three, five, and ten years immediately after the patent was issued. We used a 5-fold stratified sampling technique since the number of patents for each category is imbalanced. In other words, after patents are divided into homogeneous strata according to their citation counts of patents (i.e. L1, L2, L3, and L4), the random sampling was applied within each stratum with the sampling fraction that is proportional to the population of a stratum. A total of 80% of the patents were used as a training dataset, while the remaining 20% of the patents were employed as a test dataset.

The result is not reported here in its entirety owing to lack of space, but parts of the results are given in Table 6. The last three columns represent the predicted value categories when the corresponding patents were employed as a test dataset. Patents show different patterns of dynamics, although most of them are classified as L4 patents. Considering that a small number of patents receive multiple citations while the vast majority are not cited in their lifetime, it is reasonable that most patents are classified as L4 patents. For instance, patent 6,994,843 (delivery of stimulants through an inhalation route) – that has high values for originality, commercial scope, developers' technological strength – is classified as L1 patents for three different time periods (i.e. three-, five-, and ten-year forecasts). Patent 6,010,681 (biodegradable blood-pool contrast agents) – that has low values for almost all input variables – belong to L4 patents for all time periods. More interestingly, patents 6,066,325 (fragmented polymeric compositions and methods for their use) and 6,548,078 (method for treating and/or preventing retinal diseases with sustained release corticosteroids) – both of which have high values for originality and commercial scope – are considered the least valuable for three-year forecasts, but the most valuable for ten-year forecasts. They only have minimal impacts in the beginning of their life cycles, but are recognised as valuable patents after verification process in the market. The reasons for such temporary stagnation stem from (1) high technological uncertainty, (2) low applicability of the technology, (3) low customer acceptance, and (4) expensive cost (Abernathy and Utterback, 1978; Callon, 1980).

4.2.4. Identifying emerging technologies

This step identifies the trends of emerging technologies by monitoring the development of valuable patents at the mainline sub-class level. The *emergingness* and *trend* indicators were calculated with the patents for pharmaceutical technology and their predicted value categories, after w_1, w_2, w_3, and w_4 were set to 10, 5, 3, and 1, respectively. Here, the weights were defined by experts' judgments, considering how much a specific value category is more important than others to identify a technology's *emergingness* in the field of pharmaceutical industry. They were then rescaled to make comparisons easier, ranging from 0 to 1.

Table 7 shows the *emergingness* and *trend* indicators calculated by using a five-year forecast. As stated earlier, *emergingness* represents the development of valuable patents for a technology of interest, whereas *trend* measures the increasing and decreasing ratio of a technology's *emergingness*. While most of mainline sub-classes are stagnant, showing up-and-down fluctuations of *emergingness*, some mainline sub-classes show clear increasing or decreasing patterns of *emergingness*. For instance, mainline sub-classes 424/178.1 (conjugate or complex of monoclonal or polyclonal antibody, immunoglobulin, or fragment thereof with nonimmunoglobulin material) and 424/1.11 (radionuclide or intended radionuclide containing; adjuvant or carrier compositions; intermediate or preparatory compositions) have few of valuable patents for a whole period and their *emergingness* fluctuates up-and-down. On the other hand, mainline sub-class 424/400 (preparations characterised by special physical form) has steadily been a dominant technology and its *emergingness* has been decreasing since 2006. Mainline subclasses 424/184.1 (antigen, epitope, or other immunospecific immunoeffector) and 424/130.1 (immunoglobulin, antiserum, antibody, or antibody fragment, except conjugate or complex of the same with non-immunoglobulin material) show increasing *emergingness* patterns but have different *trends*. The *trend* of 424/130.1 has been much greater than that of 424/184.1 since 2012.

4.3. Validation

The distinct characteristics of the proposed approach stem from the use of neural networks for a multi-class classification problem; assessing the value of patents is equivalent to classifying all the patents into four classes according to the expected number of forward citations of patents. Several metrics using 5-fold cross validation techniques[1] were examined to assess the performance of our approach after a confusion matrix was constructed, as summarised in Tables 8 and 9.

Firstly, we measured the accuracy per class and the average accuracy of the proposed approach, as defined in Eqs. (8) and (9).

$$Accuracy_i = \frac{tp_i + tn_i}{fp_i + fn_i + tp_i + tn_i} \tag{8}$$

$$Average\ accuracy = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{fp_i + fn_i + tp_i + tn_i}}{l} \tag{9}$$

Here, true positive (tp_i), true negative (tn_i), false positive (fp_i), and false negative (fn_i) for class i represent the number of positive examples correctly predicted, the number of negative examples correctly predicted, the number of negative examples wrongly predicted as positive, and the number of positive examples wrongly predicted as negative; l is the number of classes. Although there are differences in the degree of the accuracy across different time periods of forecasts, the proposed approach is found to be effective in assessing the value of patents immediately after the relevant patents are issued. Moreover, it is noteworthy that the proposed approach provides more accurate and significant results for the most and the least valuable patents among the four categories.

Secondly, although the accuracy indicator is a basic metric, it is not reliable in our case study since it yields misleading results when the data set is unbalanced (Kim et al., 2017). For this reason, we calculated the precision (positive predictive value), recall (true positive rate or sensitivity), and diagnostic odds ratio (DOR) to evaluate the performance of the proposed approach. Precision is the number of true positive results divided by the number of all positive results, whereas recall is the number of true positive results divided by the number of positive results that should have been returned. DOR is a measure of the overall effectiveness of a classifier, and is defined as the ratio of the probability odds of the classification being positive if the subject is actually positive to the probability odds of the classification being positive if the subject is actually negative, as shown in Eq. (10). Here, sensitivity is equivalent to the recall, whereas specificity is the proportion of negatives that are correctly identified. This indicator is independent of prevalence or balanced sets, and ranges from zero to infinity. DOR of exactly one means the test is equally likely to predict a positive outcome whatever the true condition, and thus gives no information, whereas higher DOR indicates better performance.

$$DOR_i = \frac{sensitivity \times specificity}{(1 - sensitivity) \times (1 - specificity)} \tag{10}$$

DOR shows that the proposed approach provides reliable and significant performance, although there are differences in the degree of

[1] k-Fold cross-validation is a statistical technique for assessing how the results of analysis will generalise to an independent data set and how accurately a predictive model will perform in practice. This technique partitions data into k nearly equally sized folds. Subsequently k iterations of training and validation are performed such that, in each iteration, a different fold of the data is held-out for validation while the remaining k-1 folds are used for learning a model. Upon completion, k samples of the performance metric are available and they are combined to derive a more accurate estimate of model performance.

Table 7
Emergingness and *trend* indicators for pharmaceutical technologies.

(a) *Emergingness*

Patent sub-class	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
424/1.11	0.071	0.062	0.041	0.048	0.036	0.028	0.049	0.035	0.034	0.036
424/130.1	0.099	0.095	0.104	0.087	0.102	0.076	0.123	0.164	0.183	0.236
424/178.1	0.016	0.012	0.014	0.013	0.015	0.008	0.020	0.009	0.016	0.014
424/184.1	0.217	0.235	0.220	0.232	0.199	0.164	0.218	0.215	0.230	0.252
424/400	0.829	1	0.983	0.964	0.587	0.464	0.491	0.399	0.335	0.251
…	…	…	…	…	…	…	…	…	…	…
424/520	0.016	0.026	0.032	0.027	0.019	0.031	0.019	0.017	0.014	0.015
424/59	0.046	0.044	0.054	0.031	0.021	0.021	0.020	0.027	0.022	0.007
424/600	0.051	0.053	0.055	0.055	0.042	0.042	0.054	0.031	0.036	0.027
424/63	0.011	0.011	0.013	0.011	0.007	0.002	0.003	0.003	0.003	0.001
424/65	0.026	0.017	0.035	0.025	0.017	0.008	0.009	0.013	0.007	0.001
…	…	…	…	…	…	…	…	…	…	…
424/78.08	0.026	0.030	0.043	0.039	0.019	0.018	0.023	0.010	0.008	0.012
424/84	0.005	0.007	0.007	0.005	0.005	0.006	0.003	0.002	0.002	0.002
424/85.1	0.049	0.044	0.054	0.046	0.036	0.039	0.037	0.039	0.038	0.028
424/9.1	0.094	0.089	0.063	0.074	0.055	0.041	0.066	0.050	0.044	0.054
424/93.1	0.101	0.112	0.113	0.110	0.091	0.066	0.089	0.088	0.071	0.070
424/94.1	0.070	0.057	0.051	0.043	0.036	0.048	0.038	0.033	0.021	0.026

(b) *Trend*

Patent sub-class	2006	2007	2008	2009	2010	2011	2012	2013	2014
424/1.11	− 0.009	− 0.020	0.006	− 0.011	− 0.008	0.020	− 0.013	− 0.002	0.002
424/130.1	− 0.004	0.010	− 0.018	0.015	− 0.026	0.047	0.041	0.019	0.053
424/178.1	− 0.003	0.002	− 0.002	0.002	− 0.007	0.011	− 0.011	0.007	− 0.002
424/184.1	0.018	− 0.015	0.012	− 0.033	− 0.035	0.053	− 0.003	0.014	0.023
424/400	0.171	− 0.017	− 0.019	− 0.377	− 0.123	0.027	− 0.092	− 0.065	− 0.083
…	…	…	…	…	…	…	…	…	…
424/520	0.010	0.006	− 0.005	− 0.008	0.012	− 0.012	− 0.002	− 0.003	0.001
424/59	− 0.001	0.009	− 0.023	− 0.010	0	− 0.001	0.007	− 0.005	− 0.015
424/600	0.002	0.002	− 0.001	− 0.012	0	0.0114	− 0.023	0.006	− 0.009
424/63	0.001	0.002	− 0.002	− 0.004	− 0.005	0.001	0.001	0	− 0.002
424/65	− 0.009	0.018	− 0.010	− 0.009	− 0.008	0.001	0.005	− 0.006	− 0.006
…	…	…	…	…	…	…	…	…	…
424/78.08	0.004	0.013	− 0.005	− 0.020	− 0.001	0.005	− 0.013	− 0.002	0.004
424/84	0.003	0	− 0.002	0	0.001	− 0.003	− 0.001	0	0
424/85.1	− 0.005	0.010	− 0.008	− 0.009	0.003	− 0.002	0.002	− 0.001	− 0.010
424/9.1	− 0.005	− 0.026	0.012	− 0.020	− 0.013	0.025	− 0.016	− 0.006	0.010
424/93.1	0.011	0.001	− 0.003	− 0.019	− 0.025	0.0228	− 0.001	− 0.018	− 0.001
424/94.1	− 0.013	− 0.006	− 0.008	− 0.008	0.012	− 0.010	− 0.005	− 0.011	0.005

Table 8
Confusion matrix.

	Classified as L1	Classified as L2	Classified as L3	Classified as L4
(a) Three-year forecasts				
L1	42	0	18	37
L2	1	1	77	251
L3	6	0	416	5618
L4	4	0	369	28516
(b) Five-year forecasts				
L1	71	5	259	102
L2	19	4	686	481
L3	29	6	3399	7625
L4	6	1	2315	20348
(c) Ten-year forecasts				
L1	693	64	426	863
L2	315	106	964	1498
L3	218	120	2628	6318
L4	35	13	1081	5245

Table 9
Performance metrics.

Measure	L1	L2	L3	L4	Overall
(a) Three-year forecasts					
Accuracy	0.998	0.991	0.828	0.822	0.910
Precision	0.793	1	0.473	0.828	0.773
Recall	0.433	0.003	0.069	0.987	0.373
DOR	2446.969	–	4.600	7.262	819.61
(b) Five-year forecasts					
Accuracy	0.988	0.966	0.691	0.702	0.837
Precision	0.568	0.250	0.510	0.713	0.510
Recall	0.163	0.003	0.307	0.898	0.343
DOR	125.249	9.599	2.863	4.781	35.623
(c) Ten-year forecasts					
Accuracy	0.907	0.856	0.557	0.524	0.711
Precision	0.550	0.350	0.515	0.377	0.448
Recall	0.339	0.037	0.283	0.823	0.370
DOR	16.207	3.392	1.411	2.962	5.993

effectiveness of classification across different classes. It also confirms that our method is especially effective in identifying the most valuable patents in three and five years. Given that the precision values are greater than recall measure, it is also noted that our method provides conservative results. Taking these results together, the proposed approach was proved to be accurate and significant, supporting our contention that it assesses the value of patents at early stages of technology development.

5. Guidelines for implementation of our approach

Newly developed methods should be carefully deployed in practice. There are many issues to be considered for practical implementation. First, we employed *classification* models to assess the value of patents since predicting the exact future citation count of a patent is not the focus of our analysis. However, the value of patents can also be assessed by using *regression* models with such performance metrics as mean absolute error (MAE) and mean absolute percentage error (MAPE). Second, although a feed-forward neural network was employed in this study, many other machine learning models can be employed to identify emerging technologies using patent indicators. For this reason, we conducted additional analyses using two other models: random forests and support vector machines. Random forest is an ensemble method which operates by constructing a multitude of decision trees, to improve accuracy and avoid the over-fitting problem seen when a single decision tree is used. These improvements are achieved via ensemble learning techniques that uses different subset of training data with a single learning model, different parameters with a single training method, and/or different learning methods. The support vector machine has its roots in statistical learning theory and uses the concept of maximum-margin hyperplanes that has the largest distance to the nearest training data point of any class. The concept of maximum-margin hyperplanes can be employed not only in linear classification tasks but also in nonlinear classification tasks based on kernel trick which implicitly transforms the data in raw representation into user-specified feature vector representation. As shown in Table 10, although all the models show acceptable performance, there are slight differences in performance between models (i.e. neural network, random forest, support vector machine) for different time periods. The artificial neural network shows better performance for short-term forecasting while random forest shows better performance for long-term forecasting.[2] Hence, industrial practitioners need to select and customise appropriate methods according to the analysis context (e.g. the type of technologies and time periods).

Third, our method can serve as a monitoring tool for identifying emerging technologies, as the data and throughput are reusable, and new data can be added and analysed easily. If new patents are issued, extracting input indicators for the patents are the only additional requirements. Finally, the model developed should be reviewed and updated. For this, the systematic processes for adjusting and updating the model need to be defined, although they may differ across technological contexts.

6. Conclusions

This study has proposed a machine learning approach for identifying emerging technologies at early stages using multiple patent indicators that can be defined immediately after the relevant patents are issued. The central tenet of the proposed approach is that patent indicators – such as patent family and originality – can provide evidence for a patent's value and further the relevant technology's value in the future. To this end, a total of 18 input and 3 output indicators were extracted from the United States Patent and Trademark Office database and a feed-forward multilayer neural network was employed to capture the complex nonlinear relationships between those patent indicators and the future citation count of a patent in a time period of interest. Finally, two quantitative indicators were developed to identify trends of a technology's *emergingness* over time. The specific case of pharmaceutical technology verified that the proposed method can facilitate responsive technology forecasting and planning.

The contributions of this research are two-fold. First, from an academic perspective, this study contributes to technology forecasting research by developing a systematic method to identify emerging technologies at early stages of technology development. This is made possible by the use of machine learning models and patent indicators that can be defined immediately after the relevant patents are issued. Moreover, unlike previous methods such as curve fitting and stochastic models that employ single indicators (e.g. patent citations and applications), the proposed method incorporates the multi-facet characteristics of emerging technologies into analysis, thereby enhancing the reliability of the proposed approach. Second, from a practical standpoint, the results provide information about the future prospects of a technology's *emergingness*, enabling the quick analysis of wide ranging technologies and supporting decision making, at acceptable levels of time and cost. Moreover, the proposed approach is of more practical use than previous methods, in that it does not require any assumptions about pre-determined curves and probability distributions, which are difficult to identify at early stages of technology development. We expect the proposed approach and software system could be useful as a complementary tool to support expert decision making in emerging technologies, especially for small and medium-sized high-tech companies.

Despite its usefulness, this study is subject to certain limitations, which should be complemented by future research. First, the number of forward citations of a patent is employed as a proxy for the patent's value and further the relevant technology's value. Although previous studies have found significant positive relationships between them, the value of patents cannot be ascertained fully by this indicator. Some complementary indicators such as licencing and transaction price need to be incorporated into analyses. Second, we employed the limited number of patent indicators to assess the multi-facet characteristics of emerging technologies. The use of other types of patent indicators (i.e., examination time, novelty measured from patent landscapes, and novelty measured from patent descriptions) and more advanced analysis techniques (i.e., text mining) could be helpful for assessment of the value of patents and identification of emerging technologies. Moreover, the approach suggested here could be further enhanced by including other types of databases such as journal publications and newspapers. Third, it is difficult to understand the role of input indicators and the detailed relationships between indicators (i.e., which input indicators influence on the value of patents and how much) from the proposed approach using feed-forward multilayer neural networks. Our approach needs to be integrated with such methods as perturb, profile, and stepwise method to understand the importance of input indicators (Gevrey et al., 2003; Olden and Joy, 2004). Moreover, the use of Bayesian networks (Jensen, 2001) and random forests (Genuer et al., 2010) could also be helpful for investigating the relationships between indicators. Finally, our case study has been limited to the pharmaceutical technology where the technological value of a patent is directly related to its commercial value. The external validity of our approach needs to be established through further testing on diverse technologies

Table 10
Performance comparisons of other machine learning models.

Measure	Accuracy	Precision	Recall	DOR
(a) Three-year forecasts				
Artificial neural network	0.910	0.773	0.373	819.61
Random forest	0.910	0.594	0.344	292.666
Support vector machine	0.908	0.565	0.377	297.797
(b) Five-year forecasts				
Artificial neural network	0.837	0.510	0.343	35.623
Random forest	0.836	0.556	0.299	29.071
Support vector machine	0.830	0.590	0.337	52.497
(c) Ten-year forecasts				
Artificial neural network	0.711	0.448	0.370	5.993
Random forest	0.773	0.607	0.386	11.898
Support vector machine	0.747	0.529	0.313	7.029

[2] For more detailed results of performance evaluation, see Appendix A.

where patenting behaviours may differ. Nevertheless, we argue that the analytical power our approach offers makes a substantial contribution, both to current research and to future practice.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (No. 2017R1C1B2011434) and supported by the Future Strategic Fund (No. 1.140010.01) of Ulsan National Institute of Science and Technology (UNIST).

Appendix A. Results of performance evaluation

Measure	Average accuracy				Precision				Recall				DOR			
	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4
(a) Three-year prediction																
Artificial neural network	0.998	0.991	0.828	0.822	0.792	1	0.473	0.828	0.433	0.003	0.069	0.987	2446.969	–	4.599	7.262
Random forest	0.998	0.990	0.830	0.821	0.655	0.375	0.527	0.821	0.350	0.016	0.009	0.999	1061.816	61.465	5.462	41.922
Support vector machine	0.998	0.990	0.826	0.820	0.642	0.366	0.424	0.828	0.417	0.041	0.065	0.984	1121.046	60.574	3.757	5.812
(b) Five-year prediction																
Artificial neural network	0.988	0.966	0.691	0.702	0.568	0.250	0.510	0.713	0.162	0.003	0.307	0.898	125.429	9.599	2.863	4.781
Random forest	0.987	0.966	0.695	0.695	0.486	0.519	0.526	0.691	0.038	0.011	0.201	0.947	76.894	30.842	2.822	5.727
Support vector machine	0.988	0.966	0.688	0.678	0.659	0.506	0.501	0.692	0.126	0.063	0.261	0.897	172.900	30.944	2.642	3.503
(c) Ten-year prediction																
Artificial neural network	0.907	0.856	0.557	0.524	0.550	0.350	0.515	0.377	0.339	0.037	0.283	0.823	16.207	3.392	1.411	2.962
Random forest	0.916	0.860	0.559	0.759	0.705	0.475	0.506	0.742	0.259	0.031	0.910	0.345	28.911	5.716	3.803	9.161
Support vector machine	0.904	0.857	0.507	0.722	0.646	0.361	0.475	0.635	0.071	0.031	0.902	0.249	17.684	3.565	2.077	4.791

References

Abernathy, W., Utterback, J., 1978. Patterns of industrial innovation. Technol. Rev. 80, 40–47.
Aharonson, B.S., Schilling, M.A., 2016. Mapping the technological landscape: measuring technology distance, technological footprints, and technology evolution. Res. Policy 45 (1), 81–96.
Alcacer, J., Gittelman, M., Sampat, B., 2009. Applicant and examiner citations in US patents: an overview and analysis. Res. Policy 38 (2), 415–427.
Arora, S.K., Porter, A.L., Youtie, J., Shapira, P., 2013. Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs. Scientometrics 95 (1), 351–370.
Bessen, J., 2008. The value of US patents by owner and patent characteristics. Res. Policy 37 (5), 932–945.
Bierly, P., Chakrabarti, A., 1996. Determinants of technology cycle time in the US pharmaceutical industry. R&D Manag. 26 (2), 115–126.
Bode, J., 1998. Decision support with neural networks in the management of research and development: concepts and application to cost estimation. Inf. Manag. 34 (1), 33–40.
Breitzman, A., Thomas, P., 2015. The emerging clusters model: a tool for identifying emerging technologies across multiple patent systems. Res. Policy 44 (1), 195–205.
Buscema, M., Ferilli, G., Sacco, P.L., 2017. What kind of 'world order'? An artificial neural networks approach to intensive data mining. Technol. Forecast. Soc. Chang. 117, 46–56.
Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., Thijs, B., 2006. Traces of prior art: an analysis of non-patent references found in patent documents. Scientometrics 69 (1), 3–20.
Callaert, J., Grouwels, J., Van Looy, B., 2012. Delineating the scientific footprint in technology: identifying scientific publications within non-patent references. Scientometrics 91 (2), 383–398.
Callon, M., 1980. The state and technical innovation: a case study of the electrical vehicle in France. Res. Policy 9, 358–376.
Chaudhuri, S., 2005. The WTO and India's Pharmaceuticals Industry: Patent Protection, TRIPS, and Developing Countries. Oxford University Press, New Delhi.
Chen, C.T., Chang, W.D., 1996. A feedforward neural network with function shape autotuning. Neural Netw. 9 (4), 627–641.
Chen, Y.S., Chang, K.C., 2010. The relationship between a firm's patent quality and its market value—the case of US pharmaceutical industry. Technol. Forecast. Soc. Chang. 77 (1), 20–33.
Chen, Y.S., Chen, B.Y., 2011. Utilizing patent analysis to explore the cooperative competition relationship of the two LED companies: Nichia and Osram. Technol. Forecast. Soc. Chang. 78 (2), 294–302.
Chen, S., Billings, S.A., Grant, P.M., 1990. Non-linear system identification using neural networks. Int. J. Control. 51, 1191–1214.
Chen, Y.S., Chang, K.C., Chang, C.H., 2012. Nonlinear influence on R&D project performance. Technol. Forecast. Soc. Chang. 79 (8), 1537–1547.
Cho, T.S., Shih, H.Y., 2011. Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008. Scientometrics 89 (3), 795–811.
Cozzens, S., Gatchair, S., Kang, J., Kim, K.S., Lee, H.J., Ordóñez, G., Porter, A., 2010. Emerging technologies: quantitative identification and measurement. Tech. Anal. Strat. Manag. 22 (3), 361–376.
Criscuolo, P., Verspagen, B., 2008. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. Res. Policy 37 (10), 1892–1908.
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. 73 (8), 981–1012.
Day, G.S., Schoemaker, P.J., 2000. Avoiding the pitfalls of emerging technologies. Calif. Manag. Rev. 42 (2), 8–33.
Dayhoff, J.E., DeLeo, J.M., 2001. Artificial neural networks. Cancer 91, 1615–1635.
Ernst, H., 2003. Patent information for strategic technology management. World Patent Inf. 25 (3), 233–242.
Fernández-Ribas, A., 2010. International patent strategies of small and large firms: an empirical study of nanotechnology. Rev. Policy Res. 27 (4), 457–473.
Fleming, L., 2001. Recombinant uncertainty in technological search. Manag. Sci. 47 (1), 117–132.
Genuer, R., Pogi, J.M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recogn. Lett. 31 (14), 2225–2236.
Gerken, J.M., Moehrle, M.G., 2012. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. Scientometrics 91 (3), 645–670.
Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160 (3), 249–264.
Guellec, D., de la Potterie, B.V.P., 2000. Applications, grants and the value of patent. Econ. Lett. 69 (1), 109–114.
Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Rev. Econ. Stat. 81 (3), 511–515.
Harhoff, D., Scherer, F.M., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Res. Policy 32 (8), 1343–1363.

Haupt, R., Kloyer, M., Lange, M., 2007. Patent indicators for the technology life cycle development. Res. Policy 36 (3), 387–398.
Jaffe, A.B., Trajtenberg, M., 2002. Patents, Citations, and Innovations: A Window on the Knowledge Economy. MIT press.
Jang, H.J., Woo, H., Lee, C., 2017. Hawkes process-based technology impact analysis. J. Inf. Secur. 11 (2), 511–529.
Jensen, F.V., 2001. Bayesian Networks and Decision Graphs. Springer-Verlag, Inc., New York, NY.
Jeong, G.H., Kim, S.H., 1997. A qualitative cross-impact approach to find the key technology. Technol. Forecast. Soc. Chang. 55 (3), 203–214.
Jeong, Y., Park, I., Yoon, B., 2016. Forecasting technology substitution based on hazard function. Technol. Forecast. Soc. Chang. 104, 259–272.
Joung, J., Kim, K., 2017. Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. Technol. Forecast. Soc. Chang. 114, 281–292.
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.
Karki, M.M.S., 1997. Patent citation analysis: a policy analysis tool. World Patent Inf. 19 (4), 269–272.
Kastra, I., Boyd, M., 1996. Designing a neural network for forecasting economic and financial time series. Neurocomputing 10, 215–236.
Kayal, A.A., Waters, R.C., 1999. An empirical evaluation of the technology cycle time indicator as a measure of the pace of technological progress in superconductor technology. IEEE Trans. Eng. Manag. 46 (2), 127–131.
Kim, H., Hong, S., Kwon, O., Lee, C., 2017. Concentric diversification based on technological capabilities: link analysis of products and technologies. Technol. Forecast. Soc. Chang. 118, 246–257.
Lanjouw, J.O., Schankerman, M., 2001. Characteristics of patent litigation: a window on competition. RAND J. Econ. 129–151.
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, H., Lee, S., Yoon, B., 2011. Technology clustering based on evolutionary patterns: the case of information and communications technologies. Technol. Forecast. Soc. Chang. 78 (6), 953–967.
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, C., Song, B., Park, Y., 2013. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. Tech. Anal. Strat. Manag. 25 (1), 23–38.
Lee, C., Kang, B., Shin, J., 2015. Novelty-focused patent mapping for technology opportunity analysis. Technol. Forecast. Soc. Chang. 90, 355–365.
Lee, C., Kim, J., Kwon, O., Woo, H.G., 2016. Stochastic technology life cycle analysis using multiple patent indicators. Technol. Forecast. Soc. Chang. 106, 53–64.
Lee, C., Kim, J., Noh, M., Woo, H.G., Gang, K., 2017. Patterns of technology life cycles: stochastic analysis based on patent citations. Tech. Anal. Strat. Manag. 29 (1), 53–67.
Lerner, J., 1994. The importance of patent scope: an empirical analysis. RAND J. Econ. 319–333.
Ma, Z., Lee, Y., 2008. Patent application and technological collaboration in inventive activities: 1980–2005. Technovation 28 (6), 379–390.
Martin, B.R., 1995. Foresight in science and technology. Tech. Anal. Strat. Manag. 7 (2), 139–168.
Meyer, M., 2006. Are patenting scientists the better scholars?: an exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Res. Policy 35 (10), 1646–1662.
Narin, F., Noma, E., Perry, R., 1987. Patents as indicators of corporate technological strength. Res. Policy 16 (2–4), 143–155.
Olden, J.D., Joy, M.K., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178 (3), 389–397.
OuYang, K., Weng, C.S., 2011. A new comprehensive patent analysis approach for new product design in mechanical engineering. Technol. Forecast. Soc. Chang. 78 (7), 1183–1199.
Porter, A.L., Roessner, J.D., Jin, X.-Y., Newman, N.C., 2002. Measuring national 'emerging technology' capabilities. Sci. Public Policy 29 (3), 189–200.
Ripley, B., Venables, W., Ripley, M.B., 2016. Package 'nnet'. R Package Version 7–3.
Rotolo, D., Hicks, D., Martin, B.R., 2015. What is an emerging technology? Res. Policy 44 (10), 1827–1843.
Shin, J., Coh, B.Y., Lee, C., 2013. Robust future-oriented technology portfolios: black–Litterman approach. R&D Manag. 43 (5), 409–419.
Small, H., Boyack, K.W., Klavans, R., 2014. Identifying emerging topics in science and technology. Res. Policy 43 (8), 1450–1467.
Trajtenberg, M., 1990. Economic Analysis of Product Innovation: The Case of CT Scanners. Harvard University Press.
Verspagen, B., De Loo, I., 1999. Technology spillovers between sectors. Technol. Forecast. Soc. Chang. 60 (3), 215–235.
Von Wartburg, I., Teichert, T., Rost, K., 2005. Inventive progress measured by multi-stage patent citation analysis. Res. Policy 34 (10), 1591–1607.
Yoon, J., Kim, K., 2011. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. Scientometrics 88 (1), 213–228.
Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.

Changyong Lee is an associate professor of the School of Management Engineering at Ulsan National Institute of Science and Technology (UNIST). He holds a BS in computer science and industrial engineering from Korea Advanced Institute of Science and Technology, and a PhD in industrial engineering from Seoul National University. His research interests lie in the areas of applied data mining and machine learning techniques, future-oriented technology analysis, robust technology planning, intellectual property management, and service science.

Ohjin Kwon is a director of Centre for Future Information R&D at Korea Institute of Science and Technology Information. He obtained a BS and MS in computer science at Kwangwoon University, and a PhD in computer science at University of Seoul. His research areas are information systems, technology intelligence, and patent analysis.

Myeongjung Kim is a PhD student of School of Management Engineering at UNIST. He holds a BS in business administration from UNIST. His research interests include applied data mining and machine learning, technology intelligence, and intellectual property management.

Daeil Kwon is an assistant professor of system design and control engineering at UNIST. He received his PhD in mechanical engineering from the University of Maryland, and his BS in mechanical engineering from Pohang University of Science and Technology. His research interests include prognostics and health management of electronics, reliability modelling, and use condition characterisation.