Hierarchical main path analysis to identify decompositional multi-knowledge trajectories

Sejun Yoon, Changbae Mun, Nagarajan Raghavan, Dongwook Hwang, Sohee Kim and Hyunseok Park

Abstract

Purpose – *The purpose of this paper is to propose a quantitative method for identifying multiple and hierarchical knowledge trajectories within a specific technological domain (TD).*

Design/methodology/approach – *The proposed method as a patent-based data-driven approach is basically based on patent classification systems and patent citation information. Specifically, the method first analyzes hierarchical structure under a specific TD based on patent co-classification and hierarchical relationships between patent classifications. Then, main paths for each sub-TD and overall-TD are generated by knowledge persistence-based main path approach. The all generated main paths at different level are integrated into the hierarchical main paths.*

Findings – *This paper conducted an empirical analysis by using Genome sequencing technology. The results show that the proposed method automatically identifies three sub-TDs, which are major functionalities in the TD, and generates the hierarchical main paths. The generated main paths show knowledge flows across different sub-TDs and the changing trends in dominant sub-TD over time.*

Originality/value – *To the best of the authors' knowledge, the proposed method is the first attempt to automatically generate multiple hierarchical main paths using patent data. The generated main paths objectively show not only knowledge trajectories for each sub-TD but also interactive knowledge flows among sub-TDs. Therefore, the method is definitely helpful to reduce manual work for TD decomposition and useful to understand major trajectories for TD.*

Keywords *Technological trajectories, Technology decomposition, Knowledge persistence, Citation network, Knowledge network, Technological trends*

Paper type *Research paper*

Sejun Yoon and Changbae Mun are both based at the Department of Information Systems, Hanyang University, Seoul, Republic of Korea. Nagarajan Raghavan is based at the Department of Engineering Product Development, Singapore University of Technology and Design, Singapore, Singapore. Dongwook Hwang is based at SUTD-MIT International Design Center (IDC), Singapore University of Technology and Design, Singapore, Singapore. Sohee Kim and Hyunseok Park are both based at Department of Information Systems, Hanyang University, Seoul, Republic of Korea.

Received 13 January 2020
Revised 20 April 2020
Accepted 15 May 2020

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (No. 2019S1A5A8036427) and supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (No. 2017R1A2B4012431).

1. Introduction

Main path analysis has been widely used for many studies on technological innovation (Verspagen, 2007; Lu *et al.*, 2016; Lu and Liu, 2016; Park and Magee, 2017; Fontana *et al.*, 2009; Kuhn, 1962; Mina *et al.*, 2007). The basic concept of a main path analysis is to identify the most important paths or flows within a knowledge network based on the topological or diffusion features in the given network (Park and Magee, 2017). As patents are up-to-date and reliable technical documents (Daim *et al.*, 2006; Park *et al.*, 2013; Mun *et al.*, 2019b), most main path approaches for understanding the technological change and innovation have widely used a patent citation-based knowledge network; the citation relationship between cited and citing patents denotes knowledge flows (Park and Magee, 2019; Park and Magee, 2017; Verspagen, 2007; Petruzzelli *et al.*, 2015).

The fundamental role of main path analysis is to reduce a network complexity for visually identifying the most important knowledge flows from the complex and large network. So, the

identified main paths should be as small as possible, but include all significant technological knowledge representing the developmental trajectories. In addition, given that:

- The underlying mechanism for new technology developments is knowledge recombination (Nelson and Winter, 1982; Weitzman, 1998; Fleming, 2001; Schilling and Green, 2011; Appio *et al.*, 2017; Nakamura *et al.*, 2015; Fallatah, 2018); and
- A technological domain (TD) usually consists of several sub-TDs hierarchically structured under the TD (Benson and Magee, 2015; Mun *et al.*, 2019a), a main path analysis should identify multiple paths, not single path, and knowledge flows occurred between sub-TDs (Park and Magee, 2017).

Most of conventional main path approaches have been mainly based on a network search path algorithm, suggested by Hummon and Doreian (1989). However, a search path-based approach has some critical drawbacks for analyzing TDs. First, a search path-based main path approaches often omit some important patents from main paths (Park and Magee, 2017). Second, given that the hierarchical structure under a TD, it is reasonable that TD includes the multiple number of developmental paths. But most search path approaches identify only single path (Lu and Liu, 2016). Third, in a similar vein, most search path approaches cannot show knowledge combinations occurred within a TD.

Recently, Park and Magee (2017) introduced a new main path analysis using the concept of knowledge persistence (KP) to resolve the limitations of conventional approaches. It identifies multiple paths with relatively small network scale without missing key patents in a TD. Park and Magee (2017) approach, however, cannot separately find hierarchical structure of sub-TDs. This may require experts' additional effort to understand the TD structure, such as the number of sub-TDs under the TD or the inclusive relationships of patents on the paths in each sub-TD.

As an effort to address the aforementioned limitations of Park and Magee (2017), this paper proposes a hierarchical main path to identify multiple main paths in a specific TD. The method first identifies the hierarchical structure of a TD based on patent co-classifications and hierarchical relationships between patent classifications. Then, it generates main paths for each sub-TD and whole TD separately and integrates them to the interconnected hierarchical main paths. The patents on the main paths belong to, at least, one of sub-TDs and so various insightful knowledge flows, such as knowledge flows across sub-TDs, knowledge combination of different sub-TDs, or dominantly important sub-TD over time, can be objectively identified. In addition, the method can show each sub-TD's developmental trajectories which are essential to better understand a TD's structure from the component or functional perspective, but cannot be found without automated TD decomposition.

This paper conducted an empirical analysis using Genome sequencing technology: we selected Genome sequencing because it is one of the most important breakthrough technologies for human beings, and the TD is clearly divided into functional categories and involves many complex knowledge flows among sub-TDs. The results show that the proposed hierarchical main path analysis decomposed the given TD into three sub-TDs and then generated hierarchical main paths by integrating main paths for whole TD and each sub-TD. The identified main paths include some meaningful knowledge flows across different sub-TDs and identifies relatively dominant sub-TD over time. Therefore, the method not only reduces the required manual works for TD decompositions, but also provides rich information for understanding on technological changes and developmental trajectories in a TD.

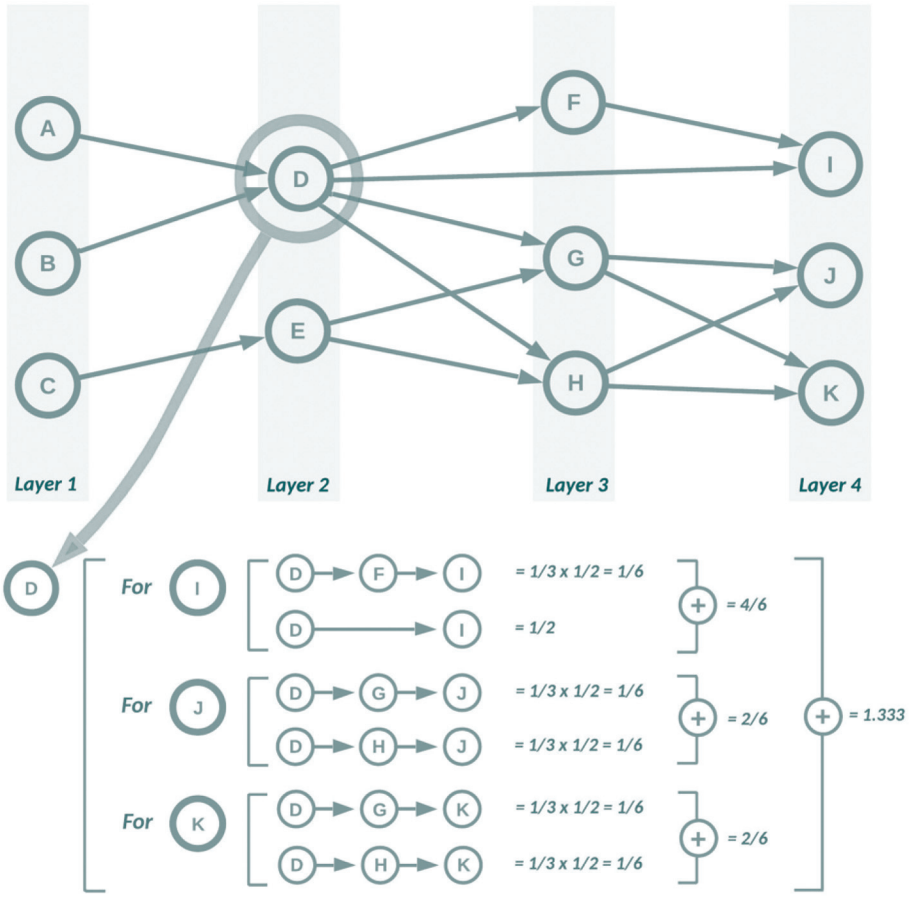
The rest of this paper is structured as follows. Section 2 reviews the related literatures. Section 3 describes the proposed method. Section 4 presents the empirical analysis and discussion of the results, and finally conclusions are drawn in Section 5.

2. Theoretical background

2.1 Knowledge persistence-based main path analysis

Park and Magee (2017) developed a KP-based main path analysis. KP is a metric to measure how much knowledge in a patent is inherited to the recent developments in a knowledge network (see Figure 1). KP can quantify patent's value from the global citation perspective and so has been recognized as a good metric to identify technologically important patents in a TD (Martinelli and Nomaler, 2014; Park and Magee, 2017, 2019, Mun *et al.*, 2019b). Park and Magee (2019) found that global citation metrics outperformed a local citation metrics in technological discontinuity identification. Even though a local citation metric can underestimate some critical patents for consecutive knowledge flows if they do not have many direct forward citations, a global citation metric, e.g. KP, usually assesses them as important patents (Park and Magee, 2019). Therefore, KP-based main path analysis can minimize the possibility to omit the significant patents from main paths. In addition, KP-based main path analysis searches backward and forward paths from each

Figure 1 Knowledge persistence calculation



Notes: Layer denotes the topological structure of knowledge flows in a TD. The number of layers in the TD can be calculated based on the longest sequence of citation flow from endpoints to startpoints. Once the longest knowledge flow is identified, each patent can be assigned to one of the layers. The inheritance proportion is 1/n, and n is the number of backward citations of the patent in the next layer

important patent, and so it can identify multiple main paths for any TD. Park and Magee (2017) empirically compared the performance KP-based main path analysis with the representative search path-based approach by applying them to Desalination and Solar photovoltaic technologies. They found that the KP-based approach dramatically reduces network complexity, about 10 times smaller than the baseline approach, and contains about 20% more of the dominant patent in the TDs (Park and Magee, 2017).

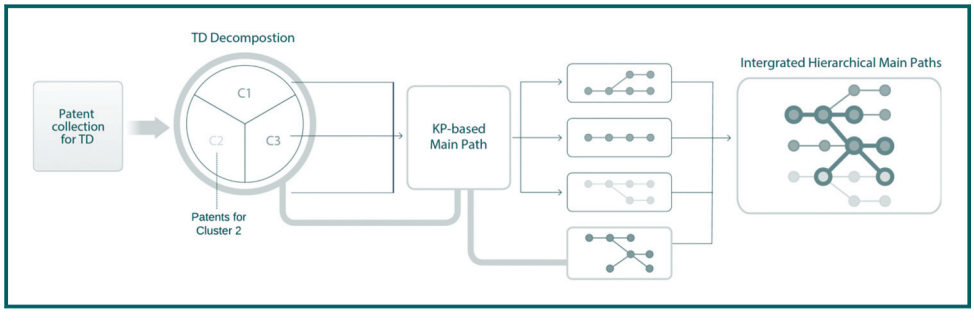
2.2 Hierarchical structure in technological domain

Fundamental mechanism for creating new knowledge is a combinatorial process of existing knowledge (Nelson and Winter, 1982; Weitzman, 1998; Fleming, 2001; Schilling and Green, 2011; Appio *et al.*, 2017; Nakamura *et al.*, 2015). Every technology is also developed based on other prior technologies, and so a technology, or TD, can be considered as either the TD itself, which consists of other technologies, or a sub-TD of other technologies (Benson and Magee, 2015). The hierarchical structure in a TD can be analyzed from functional perspective and/or compositional perspective (Keuneke, 1991). For example, telecommunication, or telephone, TD can be simply decomposed to be functionalities, e.g. *transmitter, receiver and notify*, or components, e.g. *microphone, speaker, ringer and repeating coil*. The decomposition perspective can be generally decided by the characteristics of TD. For some TDs, two decomposition perspectives show almost similar hierarchical structures, such as a method or process related TD, e.g. Genome sequencing. But, for example, software technology can be decomposed only from functional perspective. The hierarchical decomposition of a TD is an essential task for any TD-level analysis. In particular, since each sub-TD, as a major component of TD, directly affects to and is affected by other sub-TDs and their interrelationships are one of major drivers for overall progress of the TD, the TD decomposition is a significant initial task for better understanding of the developmental trajectories. The decomposition task has been highly relied on domain expert's knowledge. But, the number of sub-TDs and their hierarchical depth in the TD can directly affect to the further analysis. Therefore, there has been high demand for a way to objectively identify sub-TDs under a TD.

3. Method

The overall concept of the proposed method is that the focal TD is first decomposed into multiple sub-TDs, and then main paths for each sub-TD and whole TD are generated using the KP-based main path algorithm. The identified main paths at sub-TD level do not have any connecting points among them, but most patents on main paths at the whole TD level are included in, at least, one of sub-TDs and so all main paths at sub-TD level can be integrated into one hierarchical main paths. The process for the method consists of the following four steps (Figure 2). First, patent data for a specific TD is collected. Second, the

Figure 2 Process for hierarchical main path analysis



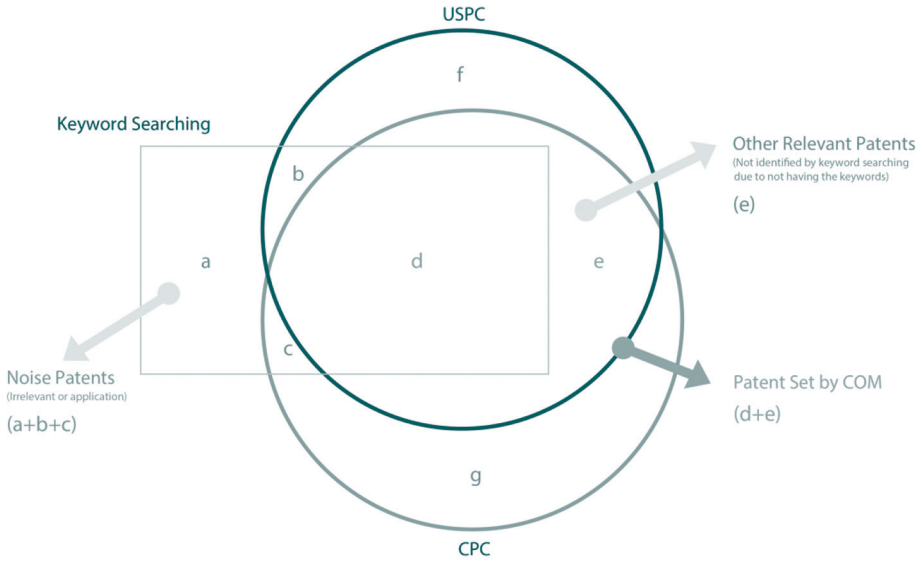
TD is decomposed based on the characteristics of patent co-classifications and hierarchical relationships between patent classifications. Third, main paths for each sub-TD and whole TD are identified. Lastly, all identified main paths are merged to generate hierarchical main paths.

3.1 Data set construction

Data collection is a fundamentally critical step for any data-driven analysis. Even though most previous research has adopted a keyword-based patent search, it has some critical limitations. First, even though patents having specific keywords related to the focal TD, some of them are not directly related to the TD and should not be involved in the patent set (noise patents in Figure 3). Second, some patents, which are clearly related to the TD but do not have any keywords used for keyword searching, cannot be identified by a keyword searching. One possible case is an emerging TD. Based on the innovation theory on knowledge creation (Nelson and Winter, 1982; Weitzman, 1998; Fleming, 2001; Schilling and Green, 2011; Appio *et al.*, 2017; Nakamura *et al.*, 2015), new knowledge is created based on existing knowledge. In a similar vein, an emerging TD is also based on previous technological knowledge, but emerging TDs are recognized as specific TDs after receiving specific names, e.g. magnetic resonance imaging TD. So, relevant but relatively old patents are usually omitted by a keyword searching.

To avoid the potential problems, this paper adopted the classification overlap method (COM), developed by (Benson and Magee, 2013, 2015), to collect patents for a specific TD.

Figure 3 Comparison between keyword searching vs COM



Notes: *a*. patents only identified through keyword searching but not directly related to the focal TD; *b*. patents identified through overlap between keyword searching and USPC and related to applications using the focal TD; *c*. patents identified through overlap between keyword searching and CPC and related to applications using the focal TD; *d*. patents identified through overlap among keyword searching, USPC and CPC and directly related to the focal TD; *e*. patents identified through overlap between USPC and CPC and directly related to the focal TD, but they do not include relevant keywords; *f*. patents identified only through USPC but not directly related to the focal TD; *g*. patents identified only through CPC but not directly related to the focal TD

COM makes a search query by overlaps of the different patent classification systems, such as cooperative patent classification (CPC) or United States patent classification (UPC) . Benson and Magee (2015) found that a technological space identified by overlaps of two different patent classification can well represent a specific TD, and the data relevancy is on average 86% for 28 TDs (Park and Magee, 2019).

The COM process is as follows. First, an initial patent set is collected by using simple keywords related to the TD. Second, major CPC and UPC codes are identified by calculating Mean-Precision-Recall (MPR) whose formulation is as follows:

$$MPR = \frac{(precision + recall)}{2}$$

$\text{MPR} = \frac{(\text{precision} + \text{recall})}{2}$,where *precision*, defined as the fraction of relevant instances among the retrieved instances, is $\frac{\#\text{ of patents in the initial patent set within the patent class}}{\#\text{ of patents in the patent class}}$ and *recall*, defined as the fraction of the total amount of relevant instances that were actually retrieved, is $\frac{\#\text{ of patents in the initial patent set within the patent class}}{\#\text{ of the collected patents in the initial patent set}}$. Third, the combinations of CPCs and UPCs which provide the highest MPR are selected as a patent search query for TD.

3.2 Decomposition of technological domain into sub-technological domains

To identify sub-TDs under the TD, this paper adopts the concept of minimum overlap classification (MOC) (Mun *et al.*, 2019a). MOC denotes the smallest technology space which can be generated by combination of the deepest classes in two different patent classifications. The group of technological similar MOCs can represents a sub-TD under the TD, and so the hierarchical clustering based on MOC distances can identify hierarchical structure of the TD.

To calculate MOC distances, the method adopted two metrics (Mun *et al.*, 2019a): Patent overlap-based distance (*PODist*) and Class hierarchy-based distance (*CHDist*). Patent overlap-based distance is calculated based on the patent co-classifications. Since patents usually classified to multiple patent classes, MOCs including same patents can be considered as similar MOCs. Therefore, the patent overlap-based distance between the MOCs is calculated by the cosine similarity for MOC vectors; each dimension for MOC vector is the patents:

$$\text{CosDist}(\text{MO}C_i,\ \text{MO}C_j) = 1 - \frac{\text{MO}C_i \cdot \text{MO}C_j}{\|\text{MO}C_i\|\|\ \text{MO}C_j\|},$$

where MOC_i is a vector representation of the *i-th* MOC, MOC_i ssMOC_j is dot product of two MOC vectors, $\|\text{MO}C_i\|$ is Euclidean length of MOC_i, the range of the distance is [0, 1]. However, high dimensionality can produce the indistinguishable similarity values and so it should be reduced. For this, the method applies a logistic function to weight clear difference in MOC distances. The final formulation for Patent overlap-based distance is as follows:

$$PODist(MOC_i, MOC_j) = \frac{1}{1 + e^{-10\left(CosDist\left(MOC_i, MOC_j\right) - 0.5\right)}}.$$

Class hierarchy-based distance is calculated based on the relationship between upper and lower classes in a patent classification system. The patent classification system is structured as a hierarchical tree network, so a link length-based semantic similarity can be applied to calculate technological distance between two classes (Mun *et al.*, 2019a). The formulation for link length-based distance (*LLDist*) between two patent classes is as follows:

$$LLDist(Class_i,\ Class_j) = 1\ - \frac{2\ \cdot d(LCS(Class_i,\ Class_j))}{d(Class_i) + d(Class_j)},$$

where $Class_i$ is the specific patent class in a patent classification system, e.g. IPC, CPC, or UPC, $d(Class_i)$ is the number of links from the root class in the given patent system to $Class_i$, and $LCS(Class_i, Class_j)$ is the least common subsumer of $Class_i$, and $Class_j$ under the hierarchical structure of the given patent classification system. The range of *LLDist* is [0,1]. Since a MOC is the combination of two different patent classification systems, the class hierarchy-based distance between MOCs should consider similarities between CPCs and UPCs. Based on this, the formulation of *CHDist* is as follows:

$$CHDist(MOC_i,\ MOC_j) = \frac{LLDist(CPC_i,\ CPC_j) + LLDist(UPC_i,\ UPC_j)}{2}.$$

Based on the patent overlap-based distance and class hierarchy-based distance, the distance between MOCs is calculated by multiplication of them, and the formulation is as follows:

$$MOCDist = PODist \times CHDist\ .$$

As the next step, similar MOCs are clustered using a hierarchical agglomerative clustering algorithm. Because the agglomerative clustering groups similar entities as a cluster and the identified clusters are merged to whole TD: this process is the reverse direction of TD decomposition, but makes the same structure with the expected decomposition (Mun *et al.*, 2019a). The TD decomposition, or MOC clustering, identifies many sub-TDs in different hierarchies; a sub-TD can consist of lower-level sub-TDs. The purpose of this research is to generate hierarchical main paths for a specific TD, and so sub-TDs in the first hierarchy are selected as sub-TDs. The technological definition for each sub-TD is qualitatively determined based on the descriptive definition of related CPC and UPC codes for MOCs and the top 20 highly cited patents in the sub-TD.

3.3 Identification of knowledge persistence-based main paths

The hierarchical main paths are integrated paths for all sub-TDs and whole TD. Therefore, all main paths should be first generated using KP-based main path analysis (Park and Magee, 2017). KP-based main path analysis first identifies the dominantly significant patents, i.e. high KP patents, by calculating KP of each patent in a citation network. The procedure for KP calculation is as follows. First, all *endpoint* patents, which do not have forward citations, and *startpoint* patents, which do not have backward citations, are identified. Second, the longest citation link between *startpoint* and *endpoint* patents is identified. This flow is recognized as the number of layers of the TD. Third, all patents are rearranged by layer. Fourth, KP, how much knowledge is inherited by *endpoint* patents in the layer-based citation network, is calculated. The inheritance proportion is 1/the number of backward citations of the patent in the next layer (see example in Figure 1), and specifically, KP of a patent can be calculated as follows (Park and Magee, 2017):

$$KP(P_A) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\prod_{k=1}^{l_j-1}\frac{1}{BWDCit(P_ijk)},$$

where P_A means the patent A, P_{ijk} is the *k-th* patent on the *j-th* backward path from P_i to P_A; $BWDCit(P_{ijk})$ is the number of backward citations of P_{ijk}, without considering backward citations by patents included in between the first layer and layer *t-1*, when P_A belongs to layer *t*; l_j is the number of patents on the *j-th* backward path from P_i to P_A; m_i is all possible backward paths from P_i to P_A; *i* is the number of patents in the last layer, which are indirectly connected to P_A.

The dominantly significant patents are determined by normalizing KP values from the global perspective (global persistence: GP) or local perspective (layer persistence: LP); the global persistence (GP) is calculated by dividing by the maximum KP in the TD; the layer persistence (LP) is calculated by dividing by the maximum KP in the layer. Since KP as one of citation-based metrics has a time-effect (Park and Magee, 2019), GP generally cannot identify relatively recent patents as significant ones and so LP is essential metric to solve the time effect problem. Based on the Park and Magee (2017), this paper considers patents whose GP $\geq$ 0.3 or LP $\geq$ 0.8 as dominantly significant patents. Then, the backward and forward searching from the identified high KP patents identifies main paths (Figure 4). Since the mechanism of the backward and forward searching is to select patents having the highest value of global persistence among the directly linked patents on the citation network, main paths from *starting* patents to *endpoint* patents can be identified.

3.4 Generation of hierarchical main paths

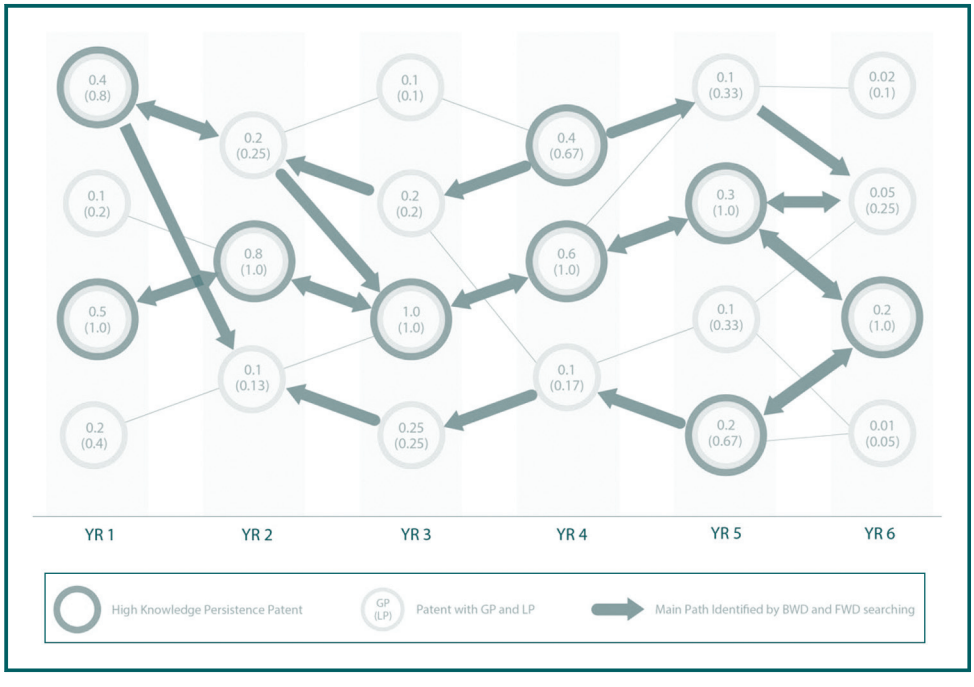
This step generates the hierarchical main paths by integrating all identified main paths for sub-TDs and whole TD. Basically, KP-based main paths is a layer-based citation network, and so it is difficult to be merged with other KP-based main paths due to the different layer scales. Therefore, all identified main paths are rearranged by year and then connected by patents duplicated in more than two main paths. Even though all sub-TDs are separately generated and so do not have any citation relation among them, main paths for whole TD is mapped on all across the sub-TDs. So the integrated hierarchical main paths have knowledge flows across the sub-TDs.

4. Empirical test

4.1 Data

The patents for Genome sequencing technology were collected by using COM. The detailed information on data is shown in Table 1 and the descriptive definitions for patent classifications used for COM query are shown in Table 2.

Figure 4 Backward and forward searching for main path identification



4.2 Results

We first decomposed Genome sequencing TD by using MOC similarities and hierarchical clustering, and three sub-TDs were identified. The specific definition for each sub-TD was qualitatively defined based on the title of top 20 highly cited patents in each sub-TD (Appendix Table A1 shows the list of top 20 patents): sub-TD 1 was defined as performance enhancement methods for sequencing and analysis, sub-TD 2 was defined as genome-based applied approach, experimental tool and applications, and sub-TD 3 was defined as sample preparation process. Main paths for three sub-TDs and whole TD were generated respectively by using KP-based main path analysis. The detailed information on the main paths by TD decomposition is shown in Table 3. The hierarchical main paths by integrating all main paths are shown in Figure 5.

The proposed hierarchical main paths provide three types of breakthrough insight. First, the hierarchical main paths show relative dominance of each sub-TD over time (Figure 6). This information is useful to understand the overall knowledge trends in a TD from the sub-TD perspective. The overall trends of Genome sequencing can be described by qualitatively analyzing Figure 6 and whole TD-level main paths:

- Sanger sequencing (Sanger *et al.*, 1977) was introduced in the early 1980s and is related to the sub-TD 3. The techniques related to sample preparation such as DNA polymerization and chain termination attracted attention. However, this

Table 1 Data overview

Search query	*# of patents*	*Data range*
C12Q and (435/6.11 or 435/6.12 or 536/24.3)	16,468	US-granted patents from 1971.01.01 to 2018.12.31 (Application date)

Table 2 Definition for genome sequencing related patent classifications

patent classification	*Descriptive definition*
UPC 435/6.11	Nucleic acid based assay involving a hybridization step with a nucleic acid probe, involving a single nucleotide polymorphism (SNP), involving pharmacogenetics, involving genotyping, involving haplotyping, or involving detection of DNA methylation gene expression
UPC 435/6.12	With significant amplification step (e.g., polymerase chain reaction (PCR), etc.)
UPC 536/24.3	Probes for detection of specific nucleotide sequences or primers for the synthesis of DNA or RNA
CPC C12Q	Measuring or testing processes involving enzymes, nucleic acids or micro-organisms immunoassay; compositions or test papers therefor; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes

Table 3 Summary of main paths (see list of patents in appendix Table A2)

Sub-TD # (Color/ Highlight)	*Name of sub-TD*	*# of patents*	*# of nodes on citation network*	*# of edges on citation network*	*# of nodes on main paths*	*# of edges on main paths*
1 (Gray)	Performance enhancement methods for sequencing and analysis	6,135	3,916	16,425	34	35
2 (Red)	Genome-based applied approach, experimental tool and applications	7,207	6,725	35,213	56	68
3 (Blue)	Sample preparation process	8,886	4,659	26,286	35	54
4 (Bold circle)	Whole TD	16,491	12,705	101,391	46	67

Figure 5 Hierarchical main paths for Genome sequencing technology

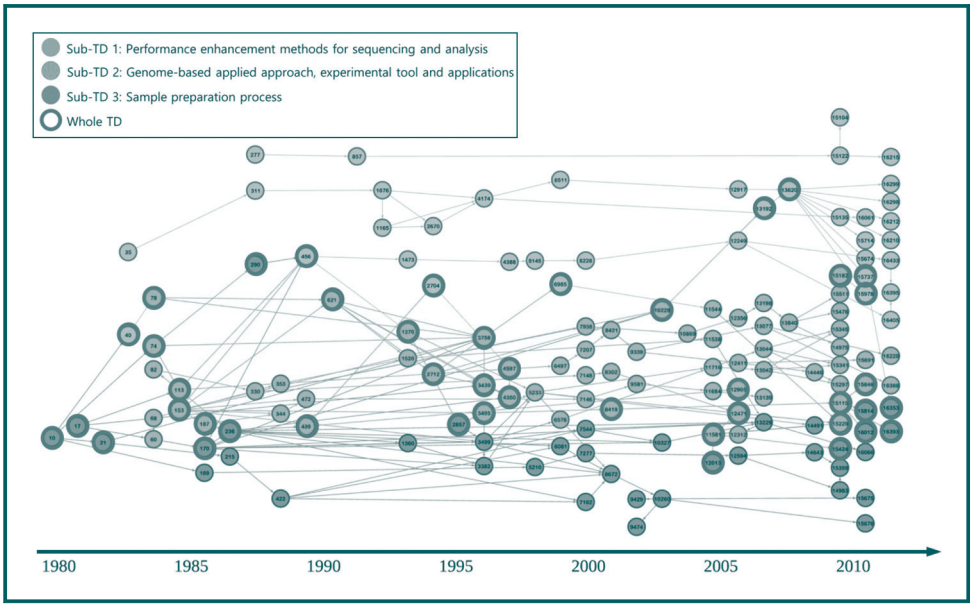
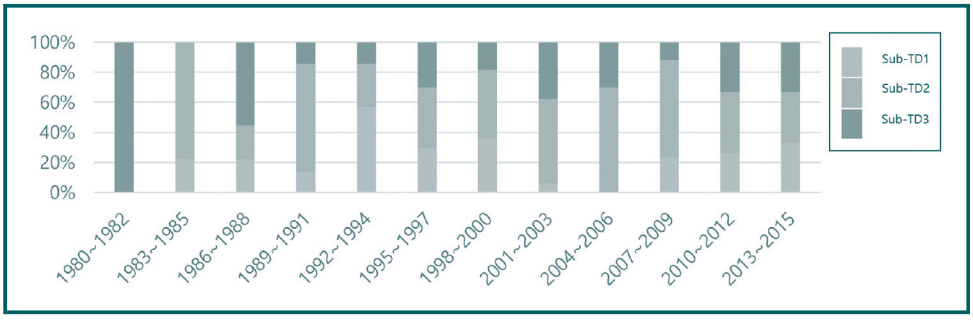


Figure 6 Proportion of patents for each sub-TD in specific periods



sequencing process was difficult to apply to experiments on human genomes, because the process was too complicated and time consuming (Heather and Chain, 2016).

- In the 1990s, sub-TD 1 and sub-TD 2 dominantly contributed for the TD (Figure 6). Automated sequencing became popular in terms of genome-based applied approaches. This technique was a mass automation technique that has been widely developed for capillary electrophoresis and fluorescent labeling. Polymerase chain reaction (PCR) was developed as a new sequencing method (sub-TD 1) and had been developed for denaturation and annealing. In particular, since 1995, genome-based applied devices (sub-TD 2) had received increasing attentions. This movement was related to the generalization of PCR equipment and the commercialization of mass analysis equipment (e.g. ABI PRISM 3700: Sequencer of Applied Biosystems (Marziali and Akeson, 2001)). Moreover, many techniques for parallelization analysis and reaction using automated sequencing devices were actively developed from the late 1990s and so the importance of sub-TD 3 had increased again.

- Next generation sequencing (NGS) techniques were introduced and had actively evolved during the 2000s. In particular, clonal amplification, massively parallel, and base calling techniques were introduced for NGS. NGS technique simplifies the entire sequencing process by eliminating the cloning process and library building process. In particular, the whole human genome project was completed, and all required technologies, i.e. all sub-TDs, were enough matured, so real applications using NGS were available at that time. Because NGS sequencing is fast, accurate and inexpensive, personal genetic sequencing services had been increased. About the experimental tools and applications, which are related to sub-TD 2, a variety of commercial products, such as SOLiD 5500 by Life Technology (ThermoFisher, 2019) or 454 GS FLX by Roche's Roche (Heather and Chain, 2016), were developed.
- The next generation NGS devices were introduced in the 2010s, and Pacific Bioscience's SMRT technology became particularly important in sub-TD 2. In sub-TD 3, techniques for diversifying the reaction types for base search, such as fluorescence pulse, potential difference, or hydrogen ion change, have been developed.

Second, the hierarchical main paths show knowledge flows across different sub-TDs. Specifically, this information provides the dynamic knowledge recombination of different knowledge from different sub-TDs. For example:

- The knowledge flows by the node 153, 170 and 456 are about knowledge combination by different sub-TDs. It shows that the sequencing method (sub-TD 1) was developed based on the knowledge from both genome-based applied approach (sub-TD 2) and sample preparation processes (sub-TD 3). Specifically, the node 153 (US 4683202) in the sub-TD 2 is about a kit that can be used for detection by amplifying only a specific position. The node 170 (US 4683195) in the sub-TD 3 is a technique for preparing nucleic acid from fragments having smaller nucleotides than synthesized fragments. The node 456 (US 5210015) in the sub-TD 1 is about a technique for detecting target nucleic acids that can be used for analysis by PCR amplification. Therefore, the basic knowledge for the detecting method in the node 456 came from different sub-TDs in the given TD.
- The paths drawn by the node 17, 170, 236, 439, 2704 and 4597 show the technological developments across different sub-TDs. The first three nodes (17, 170 and 236) are about a sample preparation process in the sub-TD 3. The two later nodes (429 and 4597) are about sequencers or devices for bioinformatics, which are related to the sub-TD 2, and the node (2704) is about the sequencer method. The paths show the important technological advancements in Genome sequencing. The node 17 (US 4395486) is about the basic methodology related to Sanger sequencing and the node 236 (US 4965188) and 170 (US 4683195) are about the techniques for cloning and amplification. Based on the sample preparation technique from the sub-TD 3, NGS techniques such as the node 439 (US 5547839) in the sub-TD 2 were developed. This node is about a device that performs parallel processing for massively parallel sequencing by using fluorescent labels. Since new applications or devices require better detection methods and vice versa, new devices lead to further advancements of methods and new methods also impact on new devices. The node 2704 (US 5741462) in the sub-TD 1 is about the method for effective detecting process by using Microarray technology. The node in the end of the paths is 4597 (US 6190857) in the sub-TD 2 and was about a novel kit for testing Messenger RNA (mRNA) to diagnose prostate cancer based on the new detection method.

Third, the hierarchical main paths show developmental trajectories of each sub-TD as well. Even though sub-TDs as major knowledge pillars for the whole TD should be analyzed for profound understanding of TD, most previous main path approaches do not consider

hierarchical structure under a TD. The hierarchical main paths can include all technological trajectories of sub-TDs and whole TD by the TD decomposition. For example:

- The knowledge paths by the node 1076, 1165, 2670, 4174, 6511, 12917 and 13620 are related to the sub-TD 1 (Figure 7) and about one significant screening method for diagnosis in genome sequencing methods, which are an essential information for understanding the whole TD. In overall, the first node 1076 (US 5434049) is about the basic method for screening by using fluorescent labeling and capillary electrophoresis, the next nodes 1165 (US 5605662), 2670 (US 5849486) and 4174 (US 6051380) are related to the major breakthrough in this trajectory, and the last three nodes 6511 (US 6355431), 12917 (US 7292742) and 13620 (US 8153375) complement and improve this knowledge stream. Specifically, the first node 1076 is about the probe screening method for diagnostics. This technique reduces the inconvenience of genetic testing by analyzing target polynucleotides at once. As a similar technological advance, the node 1165 is a technique for observing and controlling the reactions to various molecules. Thus, this method can be the basis for clinical diagnostic analysis. The node 2670 is a technique for antibody reactions and clinical diagnostics, which can be used for diagnosis by analyzing fluorescence signals using a microelectronic system. The node 4174 uses DNA hybridization reactions to more precisely control sequencing and improve detection capabilities. The node 6511 is a technique for improving sequencing accuracy by amplifying only the targets, and the node 12917 uses a zero-mode waveguide for increasing efficiency. The last node 13620 in the trajectory enables the fast analysis of diagnostic sequencing applications. Thus, the entire development in this trajectory shows the clinical diagnostic applications and devices.
- The knowledge paths by the node 10869, 12356, 13198 and 13840, related to the sub-TD 2 (Figure 8), is an important knowledge stream to prevent diagnostic mistakes which might occur by careless use of target samples. The paths developed from the basic techniques of sequencing (the first node 10869) to the techniques for precise identification (later three nodes), such as forensics, gene profiles, and unique tagging methods for each investigator. This path begins with 10869 (US 7238486) which is a technique for detecting a target nucleic acid using labeled oligonucleotides. The next two nodes 12356 (US 7501253) and 13198 (US 7635565) are about a technique to determine the length of the nucleotide target for identification and forensic medicine by using the Branch Migration Assay method. In particular, the node 13198 can link the gene profile with the patient's molecular fingerprint by analyzing the STRs (Short Tandem Repeats) and repeated sequence elements. The last node 13840 (US 8021842) is an independent tag-based technique for multiple test subjects.

Figure 7 Hierarchical main paths for sub-TD 1

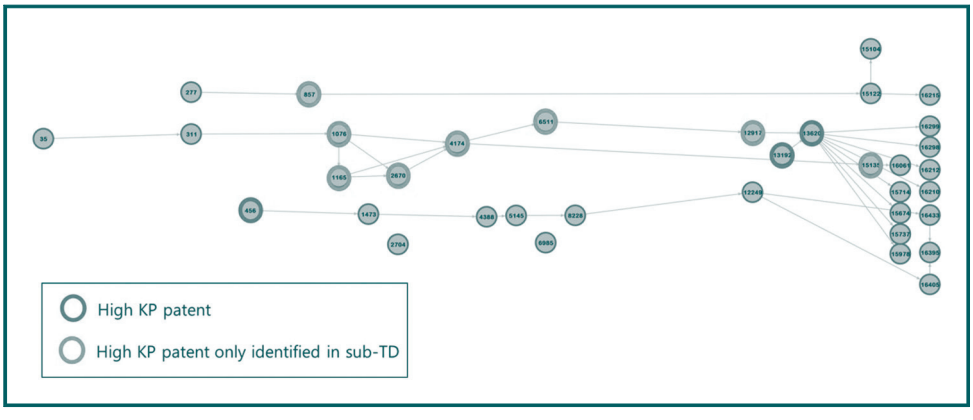
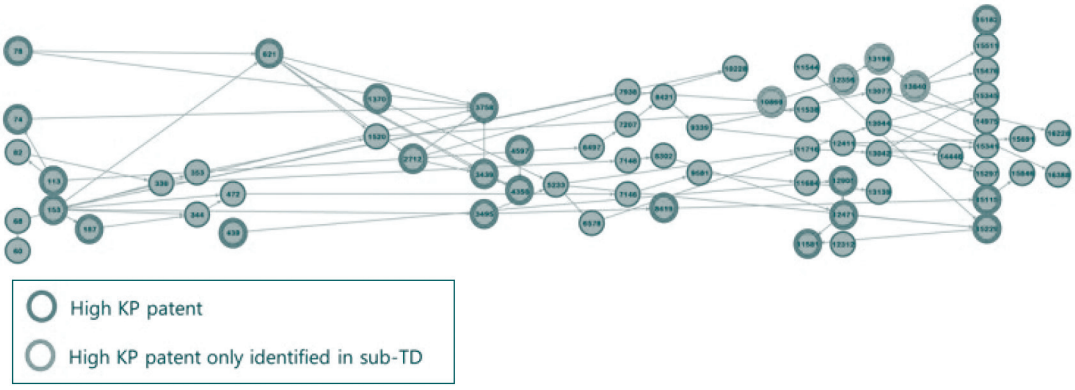


Figure 8 Hierarchical main paths for sub-TD 2



5. Conclusion

This paper proposes a hierarchical main path analysis. The proposed method first analyzes hierarchical structure under a TD based on patent co-classification and hierarchical relationships between patent classifications. Then, main paths for each sub-TD and whole-TD are generated by knowledge persistence-based main path approach. The all generated main paths at different level are integrated into the hierarchical main paths. In particular, as the TD decomposition step objectively identifies sub-TDs in the TD, hierarchical main paths can have clear benefits:

- dominantly important sub-TDs in each period can be identified;
- specific main paths for each sub-TD can be generated; and
- complex knowledge flows among sub-TDs can be clearly identified.

This paper conducted an empirical analysis using Genome sequencing technology. The results show that the proposed method automatically identifies three sub-TDs which are major functionalities in the TD and generates the hierarchical main paths. The generated main paths show knowledge flows across different sub-TDs and the changing trends in dominant sub-TD over time. The information seems to be helpful to reduce manual works for TD decomposition and useful to understand hierarchical trajectories for TD.

However, some issues should be considered for further research. First, even though the method automatically identifies sub-TDs, it still requires qualitative analysis to characterize the technologies of the identified sub-TDs. Natural Language Processing can be an applicable technique to objectively and automatically define the technologies. For example, a keyword extraction technique and comparing occurrence frequencies of keywords by sub-TD can provide meaningful textual information for defining sub-TDs. Second, as KP-based main paths usually generate multiple nodes in the last layer, the further developmental directions seem to be unclear. Therefore, further works should develop the way to identify very small number of nodes in the last layer, and one possible strategy is to adopt Radicalness (Shane, 2001) index to identify the patents having high possibility of radical innovation and then to minimize the number of last nodes. Third, given that each citation does not have same amount of knowledge inheritance, KP algorithm, which gives the same weight for every citation, needs to be improved to consider the relative difference of citation weights.

References

Appio, F.P., Martini, A. and Fantoni, G. (2017), "The light and shade of knowledge recombination: insights from a general-purpose technology", *Technological Forecasting and Social Change*, Vol. 125, pp. 154-165.

Benson, C.L. and Magee, C.L. (2013), "A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field", *Scientometrics*, Vol. 96 No. 1, pp. 69-82.

Benson, C.L. and Magee, C.L. (2015), "Technology structural implications from the extension of a patent search method", *Scientometrics*, Vol. 102 No. 3, pp. 1965-1985.

Daim, T.U., Rueda, G., Martin, H. and Gerdsri, P. (2006), "Forecasting emerging technologies: use of bibliometrics and patent analysis", *Technological Forecasting and Social Change*, Vol. 73 No. 8, pp. 981-1012.

Fallatah, M.I. (2018), "Does value matter? An examination of the impact of knowledge value on firm performance and the moderating role of knowledge breadth", *Journal of Knowledge Management*, Vol. 22 No. 3, pp. 678-695.

Fleming, L. (2001), "Recombinant uncertainty in technological search", *Management Science*, Vol. 47 No. 1, pp. 117-132.

Fontana, R., Nuvolari, A. and Verspagen, B. (2009), "Mapping technological trajectories as patent citation networks. An application to data communication standards", *Economics of Innovation and New Technology*, Vol. 18 No. 4, pp. 311-336.

Heather, J.M. and Chain, B. (2016), "The sequence of sequencers: the history of sequencing DNA", *Genomics*, Vol. 107 No. 1, pp. 1-8.

Hummon, N.P. and Doreian, P. (1989), "Connectivity in a citation network: the development of DNA theory", *Social Networks*, Vol. 11 No. 1, pp. 39-63.

Keuneke, A.M. (1991), "Device representation-the significance of functional knowledge", *IEEE Expert*, Vol. 6 No. 2, pp. 22-25.

Kuhn, T.S. (1962), "The structure of scientific revolutions",

Lu, L.Y. and Liu, J.S. (2016), "A novel approach to identify the major research themes and development trajectory: the case of patenting research", *Technological Forecasting and Social Change*, Vol. 103, pp. 71-82.

Lu, L.Y., Hsieh, C.-H. and Liu, J.S. (2016), "Development trajectory and research themes of foresight", *Technological Forecasting and Social Change*, Vol. 112, pp. 347-356.

Martinelli, A. and Nomaler, Ö. (2014), "Measuring knowledge persistence: a genetic approach to patent citation networks", *Journal of Evolutionary Economics*, Vol. 24 No. 3, pp. 623-652.

Marziali, A. and Akeson, M. (2001), "New DNA sequencing methods", *Annual Review of Biomedical Engineering*, Vol. 3 No. 1, pp. 195-223.

Mina, A., Ramlogan, R., Tampubolon, G. and Metcalfe, J.S. (2007), "Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge", *Research Policy*, Vol. 36 No. 5, pp. 789-806.

Mun, C., Yoon, S. and Park, H. (2019a), "Structural decomposition of technological domain using patent co-classification and classification hierarchy", *Scientometrics*, Vol. 121 No. 2, pp. 633-652.

Mun, C., Yoon, S., Kim, Y., Raghavan, N. and Park, H. (2019b), "Quantitative identification of technological paradigm changes using knowledge persistence", *PloS One*, Vol. 14 No. 8, pp. e0220819.

Nakamura, H., Suzuki, S., Sakata, I. and Kajikawa, Y. (2015), "Knowledge combination modeling: the measurement of knowledge similarity between different technological domains", *Technological Forecasting and Social Change*, Vol. 94, pp. 187-201.

Nelson, R.R. and Winter, S.G. (1982), *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA.

Park, H. and Magee, C.L. (2017), "Tracing technological development trajectories: a genetic knowledge persistence-based main path approach", *PloS One*, Vol. 12 No. 1, p. e0170895.

Park, H. and Magee, C.L. (2019), "Quantitative identification of technological discontinuities", *IEEE Access*, Vol. 7, pp. 8135-8150.

Park, H., Ree, J.J. and Kim, K. (2013), "Identification of promising patents for technology transfers using TRIZ evolution trends", *Expert Systems with Applications*, Vol. 40 No. 2, pp. 736-743.

Petruzzelli, A.M., Rotolo, D. and Albino, V. (2015), "Determinants of patent citations in biotechnology: an analysis of patent influence across the industrial and organizational boundaries", *Technological Forecasting and Social Change*, Vol. 91, pp. 208-221.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of the National Academy of Sciences*, Vol. 74 No. 12, pp. 5463-5467.

Schilling, M.A. and Green, E. (2011), "Recombinant search and breakthrough idea generation: an analysis of high impact papers in the social sciences", *Research Policy*, Vol. 40 No. 10, pp. 1321-1331.

Shane, S. (2001), "Technological opportunities and new firm creation", *Management Science*, Vol. 47 No. 2, pp. 205-220.

ThermoFisher (2019), available at: www.thermofisher.com/order/catalog/product/4460730, (accessed 25 October 2019).

Verspagen, B. (2007), "Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research", *Advances in Complex Systems*, Vol. 10 No. 01, pp. 93-115.

Weitzman, M.L. (1998), "Recombinant growth", *The Quarterly Journal of Economics*, Vol. 113 No. 2, pp. 331-360.

Corresponding author

Hyunseok Park can be contacted at: hp@hanyang.ac.kr

Appendix

Table A1 Top 20 highly cited patents in each sub-TD

Cluster #	*Label*	*Patent#*	*Title*	*# citations*
1	170	US4683195	Process for amplifying, detecting, and/or-cloning nucleic acid sequences	475
1	857	US5445934	Array of oligonucleotides on a solid substrate	206
1	456	US5210015	Homogeneous assay system using the nuclease activity of a nucleic acid polymerase	126
1	775	US5270163	Methods for identifying nucleic acid ligands	123
1	267	US5002867	Nucleic acid sequence determination by multiple mixed oligonucleotide probes	120
1	2706	US5653939	Optical and electrical methods and apparatus for molecule detection	106
1	547	US5302509	Method for sequencing polynucleotides	102
1	1165	US5605662	Active programmable electronic devices for molecular biological analysis and diagnostics	100
1	171	US4868103	Analyte detection by means of energy transfer	96
1	872	US5492806	Method of determining an ordered sequence of subfragments of a nucleic acid fragment by hybridization of oligonucleotide probes	88
1	344	US5124246	Nucleic acid multimers and amplified nucleic acid hybridization assays using same	87
1	439	US5547839	Sequencing of surface immobilized polymers utilizing microflourescence detection	84
1	907	US5573906	Detection of nucleic acids using a hairpin forming oligonucleotide primer and an energy transfer detection system	81
1	1813	US6040138	Expression monitoring by hybridization to high density oligonucleotide arrays	79
1	270	US4988617	Method of detecting a nucleotide change in nucleic acids	77
1	1677	US5547835	DNA sequencing by mass spectrometry	77
1	2670	US5849486	Methods for hybridization analysis utilizing electrically controlled hybridization	72
1	2406	US5605798	DNA diagnostic based on mass spectrometry	71
1	1583	US5512439	Oligonucleotide-linked magnetic particles and uses thereof	69
1	599	US5278043	Ruthenium-lumazine energy transfer systems	68
2	153	US4683202	Process for amplifying nucleic acid sequences	832
2	187	US4800159	Process for amplifying, detecting, and/or cloning nucleic acid sequences	271
2	456	US5210015	Homogeneous assay system using the nuclease activity of a nucleic acid polymerase	182
2	2279	US5925517	Detectably labeled dual conformation oligonucleotide probes, assays and kits	169
2	621	US5202231	Method of sequencing of genomes by hybridization of oligonucleotide probes	168
2	2712	US5744305	Arrays of materials attached to a substrate	148
2	344	US5124246	Nucleic acid multimers and amplified nucleic acid hybridization assays using same	143
2	1370	US5538848	Method for detecting nucleic acid amplification using self-quenching fluorescence probe	139
2	267	US5002867	Nucleic acid sequence determination by multiple mixed oligonucleotide probes	136
2	442	US4996143	Fluorescent stokes shift probes for polynucleotide hybridization	134
2	547	US5302509	Method for sequencing polynucleotides	134
2	1878	US5854033	Rolling circle replication reporter systems	133
2	727	US5399491	Nucleic acid sequence amplification methods	124
2	2823	US5800992	Method of detecting nucleic acids	111
2	3499	US5866336	Nucleic acid amplification oligonucleotides with molecular energy transfer labels and methods based thereon	109

(*continued*)

Table A1

Cluster #	*Label*	*Patent#*	*Title*	*# citations*
2	113	US4683194	Method for detection of polymorphic restriction sites and nucleic acid sequences	108
2	432	US4994373	Method and structures employing chemically-labelled polynucleotide probes	108
2	238	US5030557	Means and method for enhancing nucleic acid hybridization	106
2	364	US5130238	Enhanced nucleic acid amplification process	106
2	270	US4988617	Method of detecting a nucleotide change in nucleic acids	105
3	170	US4683195	Process for amplifying, detecting, and/or-cloning nucleic acid sequences	661
3	236	US4965188	Process for amplifying, detecting, and/or cloning nucleic acid sequences using a thermostable enzyme	323
3	442	US4996143	Fluorescent stokes shift probes for polynucleotide hybridization	144
3	1878	US5854033	Rolling circle replication reporter systems	143
3	10	US4358535	Specific DNA probes in diagnostic microbiology	131
3	2823	US5800992	Method of detecting nucleic acids	125
3	3499	US5866336	Nucleic acid amplification oligonucleotides with molecular energy transfer labels and methods based thereon	117
3	1671	US5700637	Apparatus and method for analyzing polynucleotide sequences and method of generating oligonucleotide arrays	113
3	364	US5130238	Enhanced nucleic acid amplification process	110
3	344	US5124246	Nucleic acid multimers and amplified nucleic acid hybridization assays using same	107
3	439	US5547839	Sequencing of surface immobilized polymers utilizing microflourescence detection	98
3	1972	US5837832	Arrays of nucleic acid probes on biological chips	95
3	3999	US6174670	Monitoring amplification of DNA during PCR	92
3	872	US5492806	Method of determining an ordered sequence of subfragments of a nucleic acid fragment by hybridization of oligonucleotide probes	87
3	2774	US6143495	Unimolecular segment amplification and sequencing	85
3	5128	US6210891	Method of sequencing DNA	84
3	1316	US5508169	Indexing linkers	81
3	2728	US5770367	Tag reagent and assay method	81
3	1980	US5871697	Method and apparatus for identifying, classifying, or quantifying DNA sequences in a sample without sequencing	81
3	2732	US5786146	Method of detection of methylated nucleic acid using agents which modify unmethylated cytosine and distinguishing modified methylated and non-methylated nucleic acids	74

Table A2 List of patents on hierarchical main paths for genome sequencing. *Note*: patents having no persistence value are located on the last layer

Label	*Patent #*	*Year*	*Sub-TD #*	*Persistence*	*GP*	*LP*	*Layer*	*Title*
10	US4358535	1980	3	752.4502	0.753	1	1	Specific DNA probes in diagnostic microbiology
17	US4395486	1981	3	687.6806	0.68819	0.91392	1	Method for the direct analysis of sickle cell anemia
21	US4486539	1982	3	315.4531	0.31568	0.31568	2	Detection of microbial nucleic acids by a one-step sandwich hybridization test
35	US4689295	1983	1	34.0488	0.12105	0.12105	1	Test for Salmonella
40	US4533628	1983	1	94.4713	0.094541	0.094541	2	Colony hybridization method
60	US4710465	1984	2	102.578	0.2323	0.37369	2	Junction-fragment DNA probes and probe clusters
68	US4883750	1984	2	171.8616	0.3892	0.79216	3	Detection of specific sequences in nucleic acids
74	US4820630	1984	2	311.2989	0.31153	0.71369	4	Assay for nucleic acid sequences, particularly genetic lesions, using interactive labels
78	US4613566	1984	2	327.8361	0.32808	0.50592	3	Hybridization assay and kit therefor
82	US4775619	1984	2	89.0169	0.20159	0.41031	3	Polynucleotide determination with selectable cleavage sites
113	US4683194	1985	2	648.005	0.64848	1	3	Method for detection of polymorphic restriction sites and nucleic acid sequences
153	US4683202	1985	2	542.5699	0.54297	0.72107	1	Process for amplifying nucleic acid sequences
169	US4925785	1986	3	72.3015	0.12922	0.51148	4	Nucleic acid hybridization assays
170	US4683195	1986	3	999.2665	1	1	2	Process for amplifying, detecting, and/or-cloning nucleic acid sequences
187	US4800159	1986	2	330.6274	0.33087	0.51022	3	Process for amplifying, detecting, and/or cloning nucleic acid sequences
215	US4795699	1987	3	130.2775	0.23284	0.4073	3	T7 DNA polymerase
236	US4965188	1987	3	383.9338	0.38422	0.88022	4	Process for amplifying, detecting, and/or cloning nucleic acid sequences using a thermostable enzyme
277	US4888278	1988	1	82.1907	0.29221	0.29221	1	In-situ hybridization to detect nucleic acid sequences in morphologically intact cells
290	US5011769	1988	3	378.8748	0.37915	1	5	Methods for detecting nucleic acid sequences
311	US5089387	1988	1	68.7369	0.24438	0.61564	2	DNA probe diffraction assay and reagents
330	US5118605	1988	2	65.5528	0.14845	0.48053	4	Polynucleotide determination with selectable cleavage sites
344	US5124246	1989	2	113.1087	0.25614	0.82913	4	Nucleic acid multimers and amplified nucleic acid hybridization assays using same
353	US5137806	1989	2	120.3427	0.27253	0.88216	4	Methods and compositions for the detection of sequences in selected DNA molecules
422	US5001050	1989	3	141.3586	0.25264	1	4	PH phi 29 DNA polymerase
439	US5547839	1990	2	120.7423	0.12083	0.55561	7	Sequencing of surface immobilized polymers utilizing microflourescence detection
456	US5210015	1990	1	408.0464	0.40835	1	6	Homogeneous assay system using the nuclease activity of a nucleic acid polymerase
472	US5200314	1990	2	72.261	0.16364	1	5	Polynucleotide capture assay employing in vitro amplification
621	US5202231	1991	2	436.1794	0.4365	1	4	Method of sequencing of genomes by hybridization of oligonucleotide probes
857	US5445934	1992	1	111.6518	0.39696	1	2	Array of oligonucleotides on a solid substrate

(*continued*)

Table A2

Label	Patent #	Year	Sub-TD #	Persistence	GP	LP	Layer	Title
1076	US5434049	1993	1	113.1982	0.40245	1	3	Separation of polynucleotides using supports having a plurality of electrode-containing cells
1165	US5605662	1993	1	146.3339	0.52026	1	4	Active programmable electronic devices for molecular biological analysis and diagnostics
1360	US5567583	1994	3	88.4461	0.15808	1	5	Methods for reducing non-specific priming in DNA detection
1370	US5538848	1994	2	217.3143	0.21747	1	7	Method for detecting nucleic acid amplification using self-quenching fluorescence probe
1473	US5491063	1994	1	53.2511	0.18932	0.47042	3	Methods for in-solution quenching of fluorescently labeled oligonucleotide probes
1520	US5503980	1994	2	72.0319	0.16312	0.99683	5	Positional sequencing by hybridization
2670	US5849486	1995	1	105.3812	0.37466	1	5	Methods for hybridization analysis utilizing electrically controlled hybridization
2704	US5741462	1995	1	205.6949	0.20585	1	8	Remotely programmable matrices with memories
2712	US5744305	1995	2	178.8204	0.17895	0.82287	7	Arrays of materials attached to a substrate
2857	US5691145	1996	3	146.0151	0.14612	0.70986	8	Detection of nucleic acids using G-quartets
3382	US6090552	1997	3	34.0879	0.060924	1	7	Nucleic acid amplification oligonucleotides with molecular energy transfer labels and methods based thereon
3439	US6054270	1997	2	352.0861	0.35234	1	11	Analyzing polynucleotide sequences
3495	US6117635	1997	2	90.3718	0.090438	0.27466	10	Nucleic acid amplification oligonucleotides with molecular energy transfer labels and methods based thereon
3499	US5866336	1997	3	230.8857	0.23106	0.81223	9	Nucleic acid amplification oligonucleotides with molecular energy transfer labels and methods based thereon
3758	US6344316	1997	2	118.1325	0.11822	1	12	Nucleic acid analysis techniques
4174	US6051380	1997	1	71.1227	0.25286	1	6	Methods and procedures for molecular biological analysis and diagnostics
4350	US5925525	1998	2	329.0326	0.32927	1	10	Method of identifying nucleotide differences
4388	US6030787	1998	1	36.5866	0.13008	0.34718	5	Hybridization assay using self-quenching fluorescence probe
4597	US6190857	1998	2	284.2622	0.28447	1	9	Diagnosis of disease state using MRNA profiles in peripheral leukocytes
5145	US6258593	1999	1	20.75	0.073773	0.29175	6	Apparatus for conducting chemical or biochemical reactions on a solid surface within an enclosed chamber
5210	US6569647	1999	3	13.8364	0.024729	0.67323	8	Nucleic acid amplification method: ramification-extension amplification method (RAM)
5233	US6277607	1999	2	22.5384	0.05104	1	8	High specificity primers, amplification methods and kits
6081	US6344329	2000	3	30.5152	0.054539	0.89519	7	Rolling circle replication reporter systems
6497	US6251639	2000	2	19.4249	0.043989	0.40571	7	Methods and compositions for linear isothermal amplification of polynucleotide sequences, using a RNA-DNA composite primer
6511	US6355431	2000	1	25.1028	0.089248	1	7	Detection of nucleic acid amplification reactions using bead arrays
6578	US6380377	2000	2	14.1667	0.032082	0.97306	9	Nucleic acid hairpin probes and uses thereof
6985	US6582908	2000	1	83.1128	0.083174	1	13	Oligonucleotides

(*continued*)

Table A2

Label	Patent #	Year	Sub-TD #	Persistence	GP	LP	Layer	Title
7146	US6534273	2001	2	7	0.015852	0.48081	9	Two-step hybridization and capture of a polynucleotide
7148	US6664079	2001	2	19.4945	0.044147	0.86494	8	Massive parallel method for decoding DNA and RNA
7162	US6573051	2001	3	17.7738	0.031766	1	9	Open circle probes with intramolecular stem structures
7207	US6812005	2001	2	10.975	0.024854	0.48695	8	Nucleic acid detection methods using universal priming
7277	US6797470	2001	3	13.1778	0.023552	0.74142	9	Detection of nucleic acid sequence differences using coupled ligase detection and polymerase chain reactions
7544	US6977148	2001	3	16.6878	0.029825	0.81196	8	Multiple displacement amplification
7938	US6815164	2001	2	5.3429	0.012099	0.23706	8	Methods and probes for detection and/or quantification of nucleic acid sequences
8228	US6875619	2001	1	9.75	0.034664	0.3884	7	Microfluidic devices comprising biochannels
8302	US7057026	2002	2	14.5588	0.03297	1	9	Labelled nucleotides
8419	US6977163	2002	2	6.5	0.006505	0.36167	14	Methods and systems for performing multiple reactions by interfacial mixing
8421	US6955901	2002	2	8.3	0.018796	0.5701	9	Multiplex ligatable probe amplification
8672	US6977153	2002	3	16	0.028596	1	10	Rolling circle amplification of RNA
9339	US7153658	2003	2	7	0.015852	1	10	Methods and compositions for detecting targets
9429	US7955795	2003	3	0	0	0	12	Method of whole genome amplification with reduced artifact production
9474	US8043834	2003	3	0	0	0	12	Universal reagents for rolling circle amplification and methods of use
9581	US7097979	2003	2	1	0.002265	0.14286	10	Detection of HIV-1 by nucleic acid amplification
10228	US7169560	2004	2	17.9721	0.017985	1	14	Short cycle methods for sequencing polynucleotides
10260	US7618776	2004	3	3.5	0.006255	0.72917	11	Rolling circle replication reporter systems
10327	US8158354	2004	3	4.8	0.008579	1	11	Methods for rapid purification of nucleic acids for subsequent analysis by mass spectrometry by solution capture
10869	US7238486	2005	2	4	0.009058	0.27475	9	DNA fingerprinting using a branch migration assay
11538	US7741036	2006	2	7.3333	0.016607	0.5037	9	Method for rapid detection and identification of bioagents
11544	US7459275	2006	2	4.8257	0.004829	0.45383	15	Sequencing of surface immobilized polymers utilizing microfluorescence detection
11581	US8796432	2006	2	2	0.002001	1	19	Chemically cleavable 3'-o-allyl-DNTP-allyl-fluorophore fluorescent nucleotide analogues and related methods
11684	US7425417	2006	2	1	0.002265	0.33333	11	Detection of HIV-1 by nucleic acid amplification
11716	US7442510	2006	2	7	0.015852	1	10	Method of identifying hairpin DNA probes by partial fold analysis
12013	US9169510	2006	3	0	0	0	20	Pyrosequencing methods and related compositions
12249	US8137912	2007	1	5	0.017777	0.63063	8	Methods for the diagnosis of fetal abnormalities
12312	US8802372	2007	2	0	0	0	12	Methods for rapid forensic analysis of mitochondrial DNA and characterization of mitochondrial DNA heteroplasmy

(*continued*)

Table A2

Label	Patent #	Year	Sub-TD #	Persistence	GP	LP	Layer	Title
12356	US7501253	2007	2	4	0.009058	0.57143	10	DNA fingerprinting using a branch migration assay
12411	US8198027	2007	2	2	0.004529	0.66667	11	Methods and compositions for nucleic acid amplification
12471	US7883869	2007	2	2.6667	0.002669	1	17	Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators
12594	US7833716	2007	3	4.4762	0.008	0.27976	10	Tagged oligonucleotides and their use in nucleic acid amplification methods
12905	US7713698	2007	2	3.525	0.003528	0.64091	16	Massive parallel method for decoding DNA and RNA
12917	US7292742	2007	1	7.9286	0.028188	1	8	Waveguides for performing enzymatic reactions
13042	US7858314	2008	2	1.5	0.003397	0.5	11	Probe, probe set, probe carrier, and testing method
13044	US7858315	2008	2	1.5	0.003397	0.5	11	Probe, probe set, probe carrier, and testing method
13077	US8008010	2008	2	2	0.004529	0.66667	11	Chimeric oligonucleotides for ligation-enhanced nucleic acid detection, methods and compositions therefor
13139	US7723040	2008	2	0	0	0	12	Detection of HIV-1 by nucleic acid amplification
13192	US7767400	2008	1	10.6333	0.010641	1	15	Paired-end reads in sequencing by synthesis
13198	US7635565	2008	2	3	0.006794	1	11	Analyzing blood type with identification of patient by genotyping
13229	US8407010	2008	3	1.3333	0.002383	1	12	Methods for rapid forensic analysis of mitochondrial DNA
13620	US8153375	2009	1	5.5	0.005504	1	16	Compositions and methods for nucleic acid sequencing
13840	US8021842	2009	2	3	0.006794	1	12	Nucleic acid analysis using sequence tokens
14446	US8512955	2010	2	0	0	0	12	Methods and compositions for nucleic acid amplification
14491	US8088575	2010	3	1	0.001787	0.75	12	Massive parallel method for decoding DNA and RNA
14643	US8034570	2010	3	0.5	0.000894	0.10417	11	Tagged oligonucleotides and their use in nucleic acid amplification methods
14975	US8932989	2011	2	0	0	0	13	Sieving of nucleic acid samples
14983	US8278052	2011	3	0.5	0.000894	0.10417	11	Tagged oligonucleotides and their use in nucleic acid amplification methods
15104	US9581549	2011	1	0	0	0	9	Nucleic acid target detection using a detector, a probe and an inhibitor
15115	US8323900	2011	2	2.25	0.002252	0.2116	15	Microfluidic system for amplifying and detecting polynucleotides in parallel
15122	US8535889	2011	1	2	0.007111	0.25225	8	Digital analyte analysis
15135	US8206917	2011	1	1	0.003555	1	10	Combinatorial decoding of random nucleic acid arrays
15182	US8936911	2011	2	0	0	0	18	Purified extended polymerase/template complex for sequencing
15229	US8298792	2011	2	3	0.003002	1	18	Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators
15297	US7993846	2011	2	0	0	0	12	Probe, probe set, probe carrier, and testing method
15341	US8026061	2011	2	0	0	0	12	Probe, probe set, probe carrier, and testing method

(*continued*)

Table A2

Label	Patent #	Year	Sub-TD #	Persistence	GP	LP	Layer	Title
15345	US8206900	2011	2	0	0	0	12	Probe, probe set, probe carrier, and testing method
15359	US8574847	2011	3	0	0	0	12	Use of blocker oligonucleotides in selective amplification of target sequences
15424	US9255292	2011	3	0	0	0	20	Synthesis of four-color 3'-O-allyl modified photocleavable fluorescent nucleotides and related methods
15476	US9428799	2011	2	0	0	0	13	Method for determining an allele profile of nucleic acid
15511	US8445205	2011	2	0	0	0	13	Nucleic acid analysis using sequence tokens
15674	US9279159	2012	1	0	0	0	10	Quantification of adaptive immune cell genomes in a complex mixture of cells
15675	US8551708	2012	3	0	0	0	12	Methods for localized in situ detection of mRNA
15676	US8551710	2012	3	0	0	0	12	Methods for localized in situ detection of mRNA
15691	US8642268	2012	2	0	0	0	12	Methods and compositions for nucleic acid amplification
15714	US8535886	2012	1	0	0	0	10	Methods and compositions for nucleic acid sample preparation
15737	US8455193	2012	1	1	0.001001	0.375	17	Compositions and methods for nucleic acid sequencing
15814	US9464319	2012	3	0	0	0	16	Multivolume devices, kits and related methods for quantification of nucleic acids and other analytes
15846	US9080207	2012	2	0	0	0	16	Microfluidic system for amplifying and detecting polynucleotides in parallel
15978	US8658364	2012	1	1	0.001001	0.375	17	Isolation of polymerase-nucleic acid complexes
16012	US9447461	2012	3	0	0	0	16	Analysis devices, kits, and related methods for digital quantification of nucleic acids and other analytes
16061	US8563246	2012	1	0	0	0	11	Combinatorial decoding of random nucleic acid arrays
16066	US8921047	2012	3	0	0	0	13	Secondary structure defining database and methods for determining identity and geographic origin of an unknown bioagent thereby
16210	US9267168	2013	1	0	0	0	10	Methods and compositions for isolating template nucleic acids
16212	US9394567	2013	1	0	0	0	10	Detection and quantification of sample contamination in immune repertoire analysis
16215	US9441266	2013	1	0	0	0	9	Digital analyte analysis
16228	US9193994	2013	2	0	0	0	12	Polynucleotide and use thereof
16298	US9506119	2013	1	0	0	0	10	Method of sequence determination using sequence tags
16299	US9528160	2013	1	0	0	0	10	Rare clonotypes and uses thereof
16353	US9404146	2013	3	0	0	0	18	Compositions and methods for nucleic acid sequencing
16388	US8889355	2013	2	0	0	0	12	Chimeric oligonucleotides for ligation-enhanced nucleic acid detection, methods and compositions therefor
16393	US9339812	2013	3	0	0	0	16	System and method for processing and detecting nucleic acids
16395	US9639657	2013	1	0	0	0	10	Methods for allele calling and ploidy calling

(*continued*)

Table A2

Label	*Patent #*	*Year*	*Sub-TD #*	*Persistence*	*GP*	*LP*	*Layer*	*Title*
16405	US9163282	2013	1	0.5	0.001778	0.0625	9	Methods for non-invasive prenatal ploidy calling
16433	US9424392	2013	1	0.5	0.001778	0.0625	9	System and method for cleaning noisy genetic data from target individuals using genetic data from genetically related individuals