Computers & Industrial Engineering 98 (2016) 289–299

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Generating patent development maps for technology monitoring using semantic patent-topic analysis

Mujin Kim, Youngjin Park, Janghyeok Yoon *

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

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form 16 February 2016
Accepted 5 June 2016
Available online 9 June 2016

Keywords:
Patent development map
Technology monitoring
Patent map
Bibliometrics
Latent Dirichlet allocation
Topic analysis
3D printing technology

ABSTRACT

Patent development maps (PDMs) are a useful visual and monitoring tool for technology-trend identification, and therefore proper technology planning, because they provide an overall understanding of a technology's historical development and current stage. The rapid increase in technical data, however, has made it costly and time-consuming to monitor the technology development progress manually. Although some studies have suggested how to identify development paths among patents, little attention has been paid to synthetic consideration of the two core factors for PDMs: (1) the succession relationship among patents in terms of technological content and (2) the technological taxonomies of individual patents. Therefore, this paper suggests a semantic patent topic analysis-based bibliometric method for PDM generation.

The method consists of (1) collecting and preprocessing patents, (2) structuring each patent into a term vector, (3) identifying the technological taxonomies of patents by applying latent Dirichlet allocation, and (4) visualizing the development paths among patents through sensitivity analyses based on semantic patent similarities and citations. This method is illustrated using patents related to 3D printing technology. This method contributes to quantifying PDM generation and, in particular, will become a useful monitoring tool for effective understanding of the technologies including massive patents.

1. Introduction

With the move towards a globalized technological environment, firms are competing for new technologies and securing intellectual property rights to assist in their technological competitiveness. R&D in such environments regards patent analysis before a new project as an essential prerequisite. It is reported that up to 30% of all R&D expenditures is wasted on redeveloping existing inventions (European Commission (EC) 2007). In particular, patents, as the most prolific and up-to-date technology source, contain up to 80% of recent technical information worldwide, because most patent applications are published within 18 months after their first filing, irrespective of their country of origin. These statistics suggest that monitoring current and historical patent advancements is important when developing new technologies.

Patent development maps (PDMs) are one useful tool for monitoring technology. They describe the development relationships among patents within a given technology domain over time. PDMs, as the output of prior art searches, have been widely used in industry, and they include two typical core components (Fig. 1): (1) Technological succession relationships among patents over time and (2) the technological taxonomies of the patents (Yoon & Choi, 2012). First, the technological succession relationship indicates the knowledge link between a former patent and its succeeding patents; thus, it shows the development paths among them. Second, the technological taxonomies in PDMs indicate the subtopics constituting a given technology, and each patent is assigned to one of the taxonomies. Therefore, PDMs with these two core components can provide R&D planners and researchers with an appropriate understanding of the current stages and historical progress of a technology, and thereby assist effective R&D planning from an evolving technological perspective (Choi & Park, 2009).

Customary approaches for PDM development rely on experts creating them manually. However, the rapid increase in the number of global patents has made it difficult to construct PDMs in this manner (Yoon, Park, & Kim, 2013). This problem, in particular, grows more serious in the case of rapidly evolving technologies, such as emerging and high technologies (Yoon, Park, Kim, Lee, & Lee, 2014). Thus, some studies have suggested how to construct patent maps in the forms of network and positioning maps; they are largely grouped into citation-based and content-based approaches.

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

http://dx.doi.org/10.1016/j.cie.2016.06.006
0360-8352/

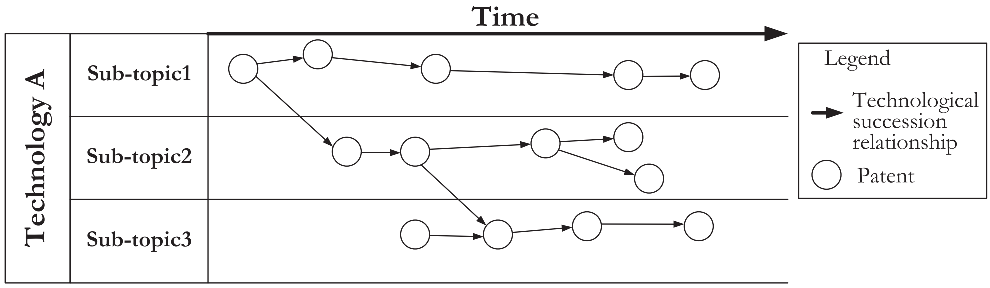

Fig. 1. PDM schematic.

Citation-based studies consider the citations between two patents as knowledge flows (Gress, 2010); therefore, they visualize the knowledge flows (Choi & Park, 2009; Hung & Wang, 2010). Despite their simplicity and ease of use, however, the studies are limited in their ability to identify substantive succession relations among patents in PDMs because they neglect the patent contents; rather, citation-based approaches are a good aid to monitoring the overall trend of widely ranging technologies using large-scale patents.

On the other hand, content-based studies employ text-mining techniques to measure the content similarity between pairs of patents. By exploiting such similarity information, the studies have suggested patent networks (Chang, Wu, & Leu, 2010; Yoon & Kim, 2011; Yoon & Park, 2004) or two-dimensional positioning maps (Bergmann et al., 2008; Yoon & Kim, 2012). However, they have paid little attention to PDM generation that combines the substantive succession relationship among patents with technological taxonomy identification.

As noted above, prior studies used citations or content similarities to generate patent maps in the forms of networks and positioning maps, but, despite their usefulness, have not sufficiently addressed the quantified development of PDMs. Therefore, this study suggests a method of generating PDMs by combining patent citations and semantic patent-topic analysis. The method consists of (1) collecting and preprocessing the patent data of a given technology, (2) structuring each patent into a term vector, (3) identifying technological taxonomies through semantic patent-topic analysis based on latent Dirichlet allocation (LDA), and (4) visualizing the semantic succession relationships among the patents assigned to one of the taxonomies. This method is illustrated using 3D printing technology patents. This method contributes to quantifying the PDM generation process. In addition, as a monitoring tool for a technology's current stages and historical paths, the method will help technology experts understand high technologies; in particular, those which contains massive patents.

The organization of this paper is as follows: First, we present a brief overview of the groundwork, followed by the proposed method. Then, we use 3D printing technology as an illustrative example, and conclude with a discussion and further research.

2. Background

This paper describes how to construct PDMs based on semantic patent topic analysis; therefore, this section presents a brief overview of the technological development path identification, followed by LDA-based topic analysis studies.

2.1. Related work on technological development paths

Visualizing technological development paths is an effective method of providing an overall understanding of the historical stages of a specific technology over time. According to the authors' best knowledge, only a few methodological studies have been conducted to identify and visualize technological development paths. One citation-based study proposed a novel measure called the forward citation node pair by multiplying the number of forward citations of the two linked patents, and then visualizing the relationships among flash memory system patents into a network (Choi & Park, 2009). The study was an initial attempt to identify technological development paths, but its limitation lay in its lack of consideration of the technological content. Another limitation was that each former patent could have a succession relationship with only one of the later patents that cited the former patent.

In light of technological content, a semantic keyword network was suggested to visualize major technology topics and their relationships over time (Kim, Suh, & Park, 2008). The semantic network approach was interesting, but it focused on the chronological technology keywords over time. It did not deal with large-scale patents, their succession relationships, or their technological taxonomies.

Some commercial patent services, including PIAS (www.kipris.or.kr) and WINTELIPS (www.wintelips.com), currently provide patent-based technology development matrices. PIAS and WINTELIPS use international patent classification codes as technological taxonomies and simply allocate each patent into a matrix cell by classification code and year. WINTELIPS also visualizes a network that is composed of patents citing or cited by a patent, but its service logic does not consider the patents' technological content.

Visualizations by prior studies and commercial services are limited in providing the core PDM components. Therefore, our quantified PDM generation method has advantages in that it can identify semantic technological succession relations among patents and extract sub-topics within a given technology.

2.2. LDA in patent analysis

Latent Dirichlet allocation (LDA) is a generative topic model which finds latent topics in a text corpus, based on the assumption that authors generally write documents with respect to specific topics (Blei, Ng, & Jordan, 2003). Using the LDA process, a document is represented as a mixture of topics that produce words with certain probabilities (Fig. 2). Unlike latent semantic analysis (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990), the

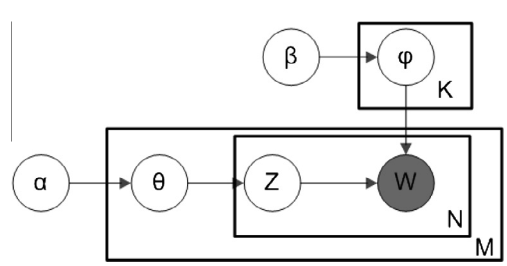

Fig. 2. LDA concept (Blei et al., 2003).

topics coming from LDA are easier to interpret, because they are represented by combinations of words with contribution probabilities for each topic (Wang & Blei, 2011). In addition, LDA is known to outperform other dimension-reduction techniques when dealing with a large corpus and interpreting the identified latent dimensions (Blei et al., 2003).

LDA assumes the following generative process for a corpus D consisting of M documents, each of length N_i:

1. Choose $\theta_i \sim Dir(\alpha)$, where $i \in \{1, \ldots, M\}$
2. Choose $\varphi_k \sim Dir(\beta)$, where $k \in \{1, \ldots, K\}$
3. For each word position i, j, where $j \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, M\}$
   A. Choose a topic $z_{ij} \sim Multinomial(\theta_i)$.
   B. Choose a word $w_{ij} \sim Multinomial(\varphi_{z_{ij}})$,

where α is the parameter of the Dirichlet prior on the per-document topic distribution, β is the parameter of the Dirichlet prior on the per-topic word distribution, θ_i is the topic distribution for document i (the sum of θ_i is 1.0), φ_k is the word distribution for topic k, z_{ij} is the topic for the jth word in document i, and w_{ij} is the specific word.

Many studies have used LDA for web spam-filtering applications (Bíró, Szabó, & Benczúr, 2008), fraud detection (Xing & Girolami, 2007), human action recognition (Wang, Sabzmeydani, & Mori, 2007), and scientific article and website recommendation (Das, Datar, Garg, & Rajaram, 2007; Jin, Zhou, & Mobasher, 2005; Krestel, Fankhauser, & Nejdl, 2009; Wang & Blei, 2011m). Regarding patent-based analysis, studies have applied LDA to the technological trend identification of greenhouse gas reduction technology (Kim, Park, & Jang, 2014), knowledge organization system development (Hu, Fang, & Liang, 2014), and firms' technological concentration trends on patent subjects (Wang, Liu, Ding, Liu, & Xu, 2014).

LDA is a distinguished tool for latent topic distribution for a large corpus. Therefore, it has the ability to identify sub-topics for a technology area composed of many patents, and represent each of the patents in an array of topic distributions. In this paper, we apply LDA to identify the patent technological taxonomies and measure the semantic patent similarities for technological succession relationships.

3. Proposed method

This section proposes a method for PDM construction. The proposed method is composed of four steps: (1) Collecting and preprocessing patent data, (2) structuring each patent into a term vector, (3) identifying technological taxonomies by applying LDA, and (4) visualizing the semantic relation among the patents assigned to one of the taxonomies (Fig. 3). The following sections will describe each step in detail.

3.1. Collecting and preprocessing patent data

The prerequisite for our method is to collect a set of valid patents related to a technology for PDM construction. For this, a patent retrieval query should be prepared. Technical keywords are chosen through a literature study, and these keywords are properly combined to define a patent retrieval query. The search query is then input to obtain patents from patent databases, such as the United States Patent and Trademark Office (USPTO). The searched patents can be stored in the form of an electronic file, such as a Microsoft Excel file, and used as the basis for the next steps.

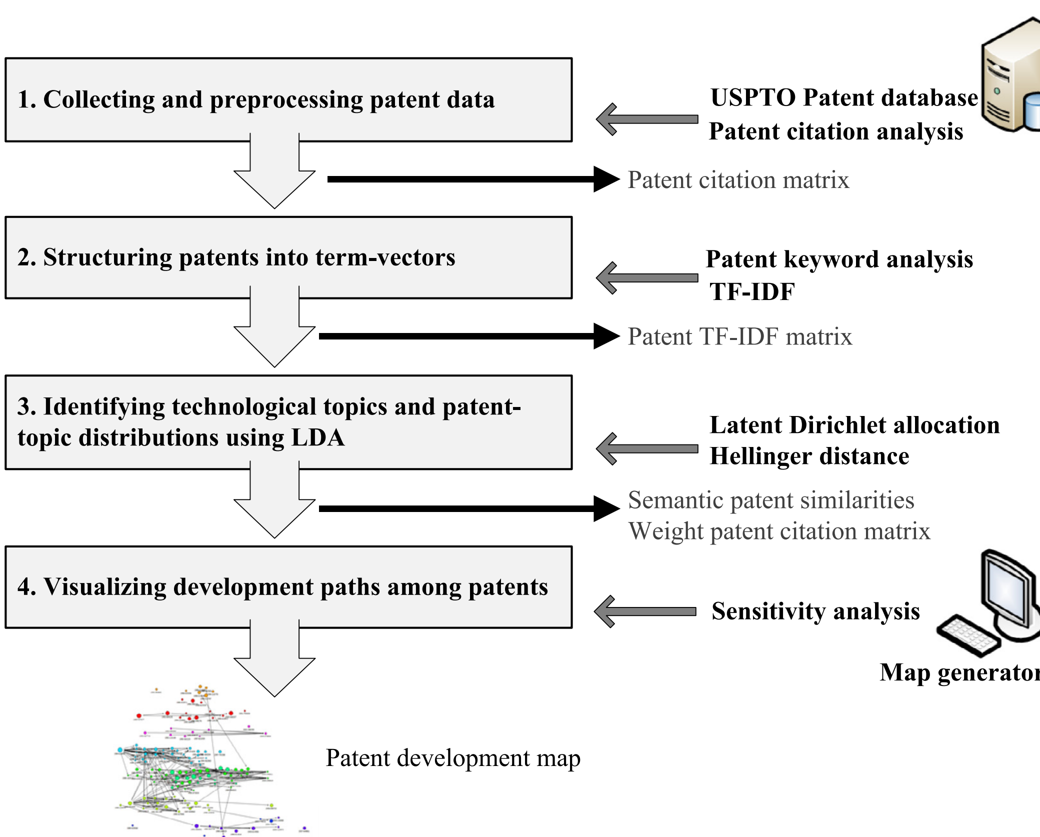

Fig. 3. Steps of the proposed method.

Given a set of patents, this step analyzes the citation-based relationships among the patents (Fig. 4). Each patent has backward citations for its prior inventions, and these citations can be extracted from the bibliographic information of the stored patents. By gathering backward citations, this step generates an $N \times N$ patent-citation network for N patents. This network will be incorporated into a semantic patent-similarity matrix to generate a weighted patent-citation matrix.

3.2. Structuring patents into term vectors

This step includes two major processes to construct a term frequency (TF)-inverse document frequency (IDF) matrix of patents: (1) Extracting the terms and (2) structuring the patents into term vectors. First, this step extracts word-level terms from the patents' textual fields, including abstracts and claims. This step includes the abstract and claim sections because they are believed to contain the most essential information about the invention (Miller, 2005; Park, Kim, Choi, & Yoon, 2013). After excluding irrelevant terms, such as "this," "method," "invention," and "part," the terms are identified using lexical databases, such as WordNet3.0, and a term list is obtained.

Second, each patent is represented into a term vector, or array of term frequencies, and finally a term frequency (TF)-inverse document frequency (IDF) matrix is generated by incorporating the patent term vectors (Fig. 5). TF-IDF is an index for weighting terms in text mining, and the TF-IDF value of a term indicates its importance in a document. The TF-IDF value of a term t is calculated as:

$$\begin{aligned} TF-IDF(t,d,D) &= TF(t,d) \times IDF(t,D) \\ &= TF(t,d) \times \log\left(1+\frac{|D|}{|\{d \in D : t \in d\}|}\right), \end{aligned} \tag{1}$$

where $TF(t,d)$ is the number of appearances of term t in document d, $|D|$ is the total number of documents, and $|\{d \in D : t \in d\}|$ is the number of documents that include term t.

3.3. Identifying technological topics and patent-topic distributions using LDA

This step identifies the taxonomical topics for a technology using a topic-modeling technique, followed by representing patents as patent-topic distributions. To this end, the TF-IDF matrix obtained in the previous step is used for the LDA process. By applying LDA with a proper number of topics, this step outputs technological topics, which are represented as an array of term-contribution probabilities. In addition, this step structures each patent as a patent-topic distribution vector, or an array of topic distribution probabilities with respect to the patent. The topic that contributes the most to a patent becomes its technological category.

Given patent-topic distribution vectors, this step constructs an $N \times N$ semantic patent-similarity matrix for N patents by measuring the similarities between pairs of patent-topic distribution vectors (Fig. 6). We adopt the Hellinger distance, which quantifies the distance between two probability distributions in probability and statistics (Hellinger, 1909). Given two discrete probability distributions $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_k)$, the Hellinger distance, $0 \leqslant H(P,Q) \leqslant 1$, is defined as:

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}(\sqrt{p_i}-\sqrt{q_i})^2}. \tag{2}$$

If two probability distributions are identical, the Hellinger distance is 0. Then, a similarity measure $S(P,Q)$ for the two probability distributions can be defined as:

$$S(P,Q) = 1 - H(P,Q). \tag{3}$$

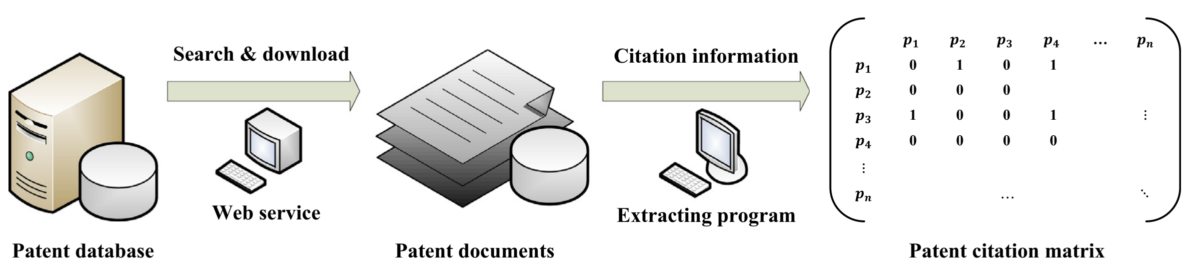

Fig. 4. Concepts of collecting and preprocessing patent data.

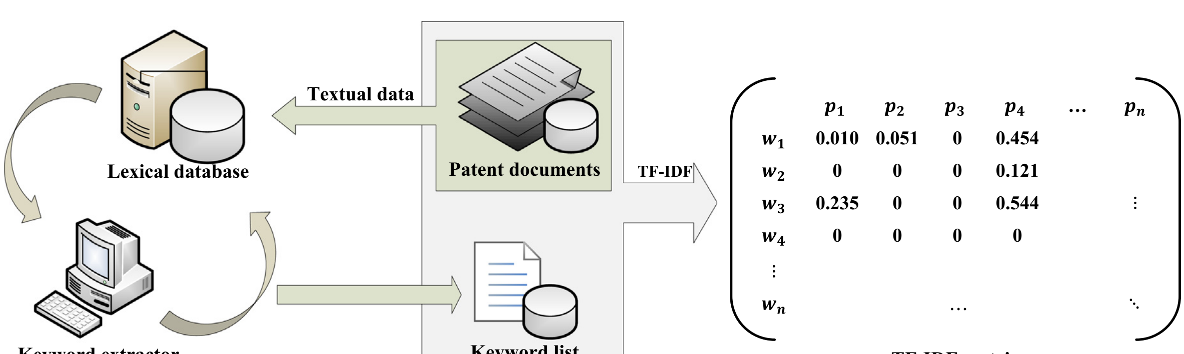

Fig. 5. Concepts of structuring patents into term vectors.

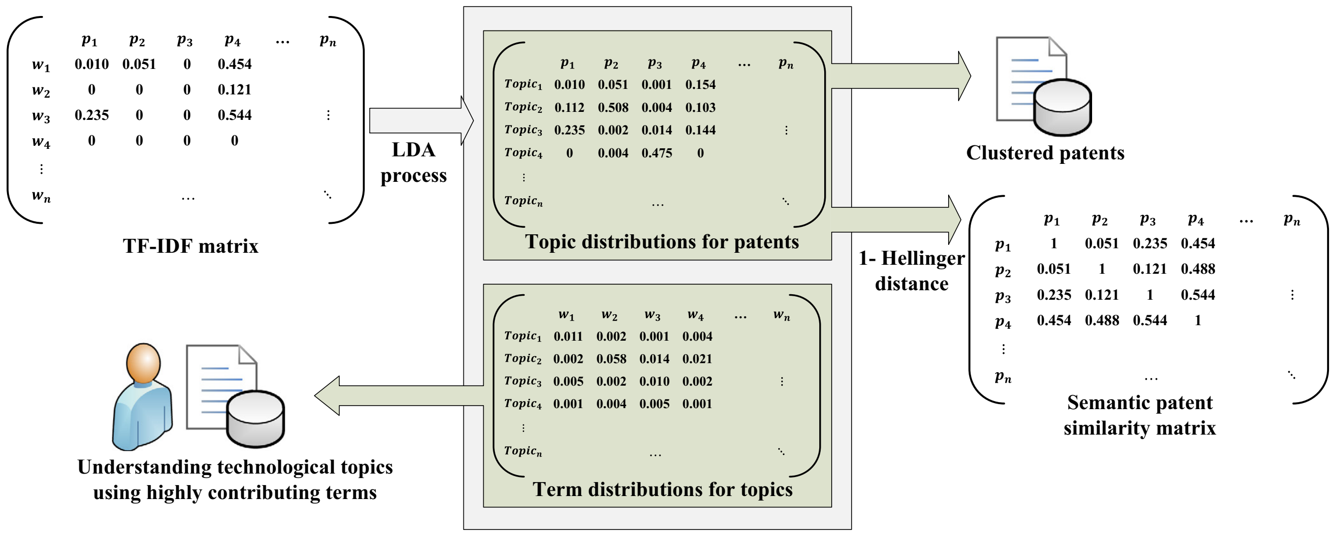

Fig. 6. Concepts of identifying technological taxonomies and measuring semantic patent similarities.

3.4. Visualizing the development paths among patents

This step constructs a PDM by sensitivity analysis of the current impact index (CII) and patent similarity values. First, by incorporating the semantic patent-similarity matrix and the patent-citation matrix, this step generates a weighted patent-citation matrix that contains the strength of the knowledge succession among patents. Applying threshold value ρ_f to pairs of patents in the weighted patent-citation matrix eliminates the weakly related knowledge succession relationships. Second, major patents are chosen using CIIs, which indicate the number of forward citations of a patent in the last five years (Hirschey & Richardson, 2001). If a patent has a higher CII value, it is considered to be more core and marketable at the time of analysis (Albert, Avery, Narin, & McAllister, 1991). This step uses threshold value ρ_c to selectively show the major patents in PDMs.

By filtering major patents and succession relationships, our PDM generator visualizes PDMs (Fig. 7). In a PDM, nodes are patents, node sizes indicate the patents' CII values, and patents are arranged in time order from left to right using patent application dates. Layers, or technological taxonomies, classify the patents using different colors, and directed links between nodes indicate the technological succession relationships between patents.

4. Illustrative example: 3D printing technology

In this study, we apply the proposed PDM generation method to 3D printing technology. 3D printing or additive manufacturing, which has a history of 30 years, is any of various processes used to make three-dimensional objects (Gibson, Rosen, & Stucker, 2010), and has recently received much attention as a future manufacturing alternative, due to its advantages in various industrial applications, such as distributed manufacturing, mass customization, and rapid prototyping. 3D printing is a high technology area where patents are increasingly produced. In addition, the number of patents related to 3D printing technology was appropriate to show the working process of the proposed method. Therefore, we used it as an example.

To construct a PDM, the 3D printing patents were prepared. To this end, three researchers and a patent attorney, who well understood 3D printing technology, identified keywords in the literature of this technology, followed by formulating a patent retrieval query

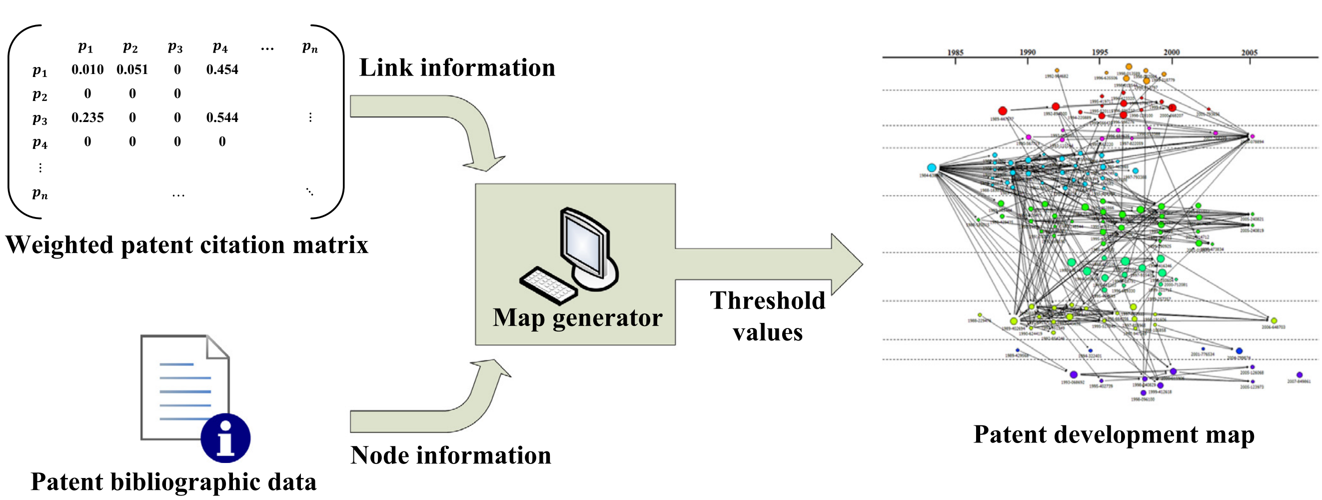

Fig. 7. Concepts of PDM visualization.

Table 1
Patent retrieval query for 3D printing technology.

Patent retrieval query	Initial patent set	Valid patent set
((Additive near2 (Manufactur∗ or Mfg∗ or Fab∗)) or ((3D or 3-D or (three-dimension∗) or (3 adj dimension∗) or 3-dimension∗) near2 print∗) or (Rapid near2 Proto∗) or (((Laminated near2 Object near2 Manufactur∗)) or ((Electron near2 Beam near2 Freeform near2 Fab∗)) or ((Direct near2 Metal near2 Deposit∗) or (Direct near2 Metal near2 Tool∗)) or ((Electron near2 Beam near2 Melt∗)) or ((Selective near2 Laser near2 Melt∗)) or ((Selective near2 Heat near2 Sinter∗)) or ((Digital near2 Light near2 Process∗)) or (Polyjet or (Poly near2 jet∗)) or ((Multi near2 Jet∗ near2 Model∗) or (Multi near2 Jet∗)) or (3D near Print∗) or ((Fused near2 Deposition near2 Model∗) or (Fused near2 Filament near2 Fab∗)) or ((Direct near2 Metal near2 Laser near2 Sinter∗)) or ((Selective near2 Laser near2 Sinter∗)) or ((Stereo near Lithography∗) or stereolithography∗) or (Laser near2 Engineered near2 Net near Shap∗)))	5416	837

with the keywords (Table 1). We obtained 3D printing patents granted until 2013 in the USPTO database through WINTELIPS, a commercial patent database service. We located the initial set of 5416 patents from the database. Then, after excluding irrelevant patents from the initial patent set, we obtained 837 valid patents related to 3D printing technology.

The first patent for 3D printing was granted in 1984, and the yearly numbers of patent applications increased overall until 2010, but have rapidly decreased since 2011 (Fig. 8). Patent applications between 1995 and 2011 formed 86.7% of the total patent applications (727 of 837). The number of patent applications between 2012 and 2013 appears to be small, because many of the patent applications for these two years are still in the examination process (average difference between application dates and granted dates = 2.94 years).

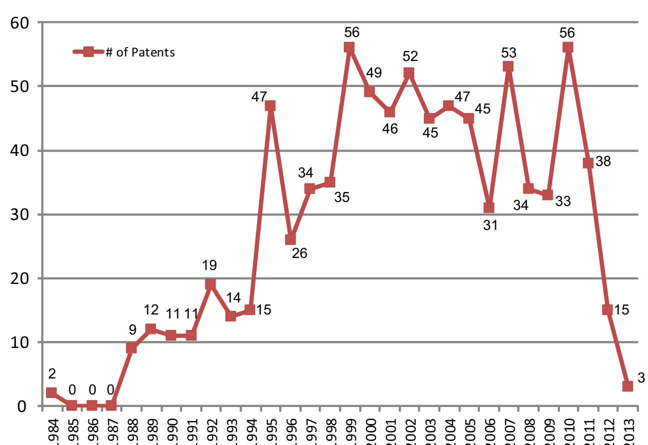

Fig. 8. Patent registration trends in 3D printing technology.

By analyzing the backward citations of 837 patents, we were able to obtain a patent-citation matrix that contains citation relationships between all pairs of patents (Table 2). The patent-citation matrix includes the bibliographical-succession relationships among the patents. Cited patents are considered to be topically related to the later patents citing them; inventors and patent examiners can cite former patents with objectives similar to their patents, although the pairs of patents do not have technological succession relationships (Park, Ko, & Yoon, 2015).

Then, we needed to identify the substantive-succession relationships using semantic patent-topic analysis. We structured each patent using word-level terms. To this end, we extracted an initial set of terms from the patent abstract and claim sections, and excluded stop words or irrelevant terms. As a result, 3413 total terms were selected. Using these terms, we structured each patent into an array of TF-IDF term values, and finally obtained a term-based TF-IDF matrix for the 837 patents (Table 3). The TF-IDF matrix showed the TF-IDF term values for all patents; terms with a high TF-IDF value were regarded as important within the patents containing them. This TF-IDF matrix was the input for our LDA processing.

In order to identify technological topics for 3D printing technology and measure semantic patent topic similarities, we applied LDA techniques to the TF-IDF matrix obtained in the previous step. In LDA, the number of topics indicates the number of latent dimensional factors, and it can be determined properly through a test that measures the average patent similarities by the number of topics (Wang et al., 2014). We found the relationship between the number of topics and the average semantic patent similarities; the numbers were set from two to 12 (Fig. 9). From the first 12 topics, we chose the number with the lowest average similarity.

Table 2
Part of the patent-citation matrix.

Application number	2011-166513	2012-365028	2008-972485	2011-016189	2008-035743	2006-439805	2004-848831	2006-335282	2003-650086	…
2005-240819	1	1	0	0	0	0	0	0	0	…
2004-781304	0	0	1	1	1	1	1	0	0	…
2003-606881	0	0	1	1	1	1	1	0	0	…
1996-771009	0	0	1	1	1	1	1	1	1	…
1996-722326	1	1	1	1	1	1	1	0	1	…
1997-822059	0	0	1	1	1	1	1	1	0	…
1996-722326	1	1	1	1	1	1	1	0	1	…
1997-790005	1	1	1	1	1	1	1	0	0	…
1997-793388	0	0	0	0	0	0	0	1	0	…
1993-009249	1	1	1	1	1	1	1	0	1	…
1996-766956	1	1	0	0	0	0	0	0	0	…
1995-480670	1	1	1	1	1	1	1	0	0	…
1997-876695	1	1	0	0	0	0	0	0	0	…
1996-611914	0	0	1	1	1	1	1	0	1	…
1984-638905	1	1	1	1	1	1	1	1	1	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Table 3
Part of the keyword-based TF-IDF matrix to structure patents.

Patent	Powder	Laser	Polymer	Stereo	Fabric	Selected	Lithography	Energy	Layer	...
2005-295008	8.728	3.426	3.765	0.000	0.000	0.000	0.000	0.910	5.177	...
2010-830452	0.000	0.685	2.259	0.000	2.624	0.607	0.000	0.000	0.000	...
2005-212711	0.000	1.370	3.012	1.116	0.000	3.036	1.248	11.835	1.479	...
2010-862546	0.000	0.000	2.259	1.116	1.312	0.000	1.248	0.910	0.986	...
2005-579783	2.182	0.000	0.753	0.000	0.656	0.607	0.000	0.000	0.000	...
2006-503628	2.182	1.370	3.012	0.000	0.656	0.000	0.000	0.000	0.000	...
2006-561191	0.000	6.167	5.271	0.000	3.936	2.429	0.000	0.910	0.000	...
2004-903379	0.000	1.370	9.789	1.673	0.656	0.000	1.873	0.000	0.000	...
2004-831052	0.727	0.685	0.000	0.000	2.624	1.215	0.000	2.731	0.986	...
2002-127019	0.000	0.685	5.271	2.231	1.968	0.607	3.12	1.820	1.232	...
1995-463203	1.455	0.685	5.271	0.558	0.656	0.000	0.624	0.000	0.493	...
1996-597805	5.819	1.370	1.506	0.000	0.656	0.607	0.000	0.000	1.232	...
1993-044971	5.819	1.370	1.506	0.000	0.656	1.215	0.000	1.820	1.479	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

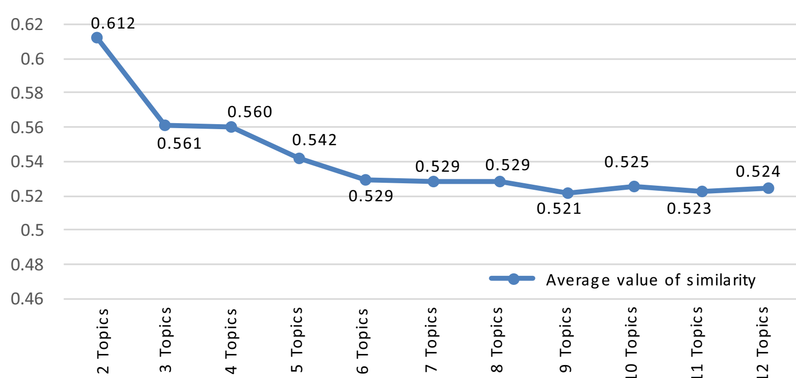

Fig. 9. Average patent similarity by number of topics.

Setting the number of topics to nine returned the lowest average similarity value of 0.521; therefore, we selected nine topics as the optimal topic model.

We obtained LDA results for the TF-IDF matrix using nine topics. Each topic was regarded as a technology category, or a mixture of different terms and their contributions to the topic (Table 4). We considered each topic as a technological taxonomy constituting 3D printing technology by examining the main contributing terms of each topic.

Topic 1 contained 81 patents related to bonding adhesive particulates on a surface or plate for 3D printing. For example, it included patent 1989-447677, "three-dimensional printing techniques." This patent discloses a process for making a 3D object by depositing layers of a fluid porous material, including adhesive powders (Sachs, Haggerty, Cima, & Williams, 1993).

Topic 2 had 95 patents related to methods of depositing metal materials for 3D printing, and Topic 3 included 121 patents mainly related to powder or particle sintering technology, such as selective laser sintering (SLS). Topic 4 had 112 patents related to supplementary applications of 3D printing, including selected region recoating and the addition of 3D objects to an existing 3D object. Topic 5 covered 91 patents about 3D printing-based dental techniques, such as artificial tooth implants and tooth molding.

Topic 6 had 120 patents that presented basic methods and principles of 3D printing technology; they were mostly related to the use of laser beams. Topic 7 contained 73 patents related to 3D printing methods or devices customized for various materials. Topic 8 with 59 patents was found to be laser or light-based sintering technologies that specifically used photopolymer materials, including plastics and resins. Topic 9 included 85 patents related to the conversion algorithms of 3D CAD images into laminates and their device applications.

As another LDA result, we were able to generate a patent-topic distribution matrix (Table 5). To simplify the assignment process of

Table 4
Part of the technology topics identified by LDA.

Topics	# of patents	Descriptions	Main contributing keywords
Topic 1	81	Adhesive particulate bonding	Plane, particulate, bond, surface, adhesive, cool, slice, melting, plate, lane, ...
Topic 2	95	Metal materials depositing	Metal, substrate, molten, droplet, liquid, deposit, filament, composite, ...
Topic 3	121	Powder or particle sintering	Powder, ceramic, bind, particle, solvent, organic, temperature, soluble, sintered, water, ...
Topic 4	112	Supplementary applications	Build, surface, layer, coat, movable, region, section, lamina, synergistic, intermediate, ...
Topic 5	91	Dental applications	Model, dental, patient, custom, implant, formation, bone, rapid, medical, ...
Topic 6	120	Basic methods and principles	Fluid, beam, laser, signal, scan, subject, target, simulation, control, elective, ...
Topic 7	73	Devices customized to materials	Radiation, liquid, formula, epoxy, acrylate, polymer, vinyl, aliphatic, ...
Topic 8	59	Laser or light-based sintering	Light, resin, plastic, thermoplastic, photo, hard, stereo, ...
Topic 9	85	3D CAD image conversion and its application device	Mold, mask, optical, complementary, rough, block, model, design, exterior, rapid, ...

Table 5
Part of the patent-topic distribution matrix.

	Topic-1	Topic-2	Topic-3	Topic-4	Topic-5	Topic-6	Topic-7	Topic-8	Topic-9
1995-419711	0.869	0.006	0.004	0.011	0.019	0.009	0.005	0.004	0.006
2010-787075	0.011	0.872	0.013	0.009	0.006	0.006	0.005	0.008	0.007
2005-579783	0.007	0.008	0.767	0.008	0.013	0.015	0.038	0.026	0.006
2007-876153	0.009	0.016	0.010	0.782	0.012	0.011	0.006	0.007	0.038
2009-650169	0.005	0.005	0.005	0.005	0.889	0.005	0.005	0.004	0.005
1997-866600	0.008	0.011	0.008	0.009	0.021	0.751	0.019	0.007	0.032
1997-901303	0.005	0.005	0.008	0.005	0.005	0.005	0.866	0.013	0.004
2010-764234	0.010	0.017	0.007	0.013	0.008	0.010	0.009	0.853	0.010
2006-335282	0.034	0.032	0.066	0.115	0.043	0.423	0.018	0.016	0.039
2005-068487	0.136	0.013	0.260	0.023	0.013	0.028	0.232	0.015	0.052
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

each patent to a technology topic, the topic that contributed the most to a patent was regarded as the patent's main technological category; for example, patent 2010-787075 was assigned to Topic 2 because Topic 2 had a contribution value of 0.872 to the patent. In fact, this patent on "systems and methods for fabricating a direct metal deposition structure having fully forged structural qualities" (Newkirk, Liou, & Francis, 2013) was strongly related to metal deposition methods.

By using the patent-topic distribution matrix and Eq. (3), we obtained semantic similarity values between all patent pairs (Table 6). Then, by combining this semantic patent-similarity matrix and the patent-citation matrix (Table 2), we finally obtained a weighted patent-citation matrix, which contains substantive technological succession relations between pairs of patents (Table 7).

For the 837 patents constituting 3D printing technology, our PDMs are intended to show the development paths among major patents by sensitivity analysis. To this end, highly cited patents in 3D printing technology were chosen using a threshold (ρ_c) for CII values, and weak-succession relationships were cut off using a threshold (ρ_f) for succession relationship.

Our PDM generator imports CII patent values, patent bibliographic information, defined technological topics, and the weighted patent-citation matrix; it then visualizes the PDMs by adjusting the two threshold values. Through trial and error, we found that threshold values $\rho_f = 0.70$ and $\rho_c = 20$ well displayed the major patents and their development paths for 3D printing technology. Following are the highly cited patents (Table 8) and the generated PDMs (Fig. 10). On the PDM, the patents are spread by application year and colored by their technological topics; their

Table 6
Part of the semantic patent-similarity matrix.

Patents	2011-230270	2000-711128	2007-593970	2011-032283	2012-541811	2012-355400	2011-166513	2013-874948	…
2011-230270	1.000	0.505	0.652	0.582	0.582	0.531	0.536	0.500	…
2000-711128	0.505	1.000	0.496	0.736	0.399	0.560	0.350	0.363	…
2007-593970	0.652	0.496	1.000	0.557	0.482	0.646	0.427	0.684	…
2011-032283	0.582	0.736	0.557	1.000	0.395	0.628	0.334	0.432	…
2012-541811	0.582	0.399	0.482	0.395	1.000	0.492	0.912	0.361	…
2012-355400	0.531	0.560	0.646	0.628	0.492	1.000	0.433	0.443	…
2011-166513	0.536	0.350	0.427	0.334	0.912	0.433	1.000	0.301	…
2013-874948	0.500	0.363	0.684	0.432	0.361	0.443	0.301	1.000	…
2009-490685	0.714	0.443	0.571	0.565	0.368	0.426	0.310	0.548	…
2007-652876	0.494	0.320	0.701	0.396	0.286	0.384	0.235	0.616	…
2010-979484	0.369	0.370	0.487	0.404	0.387	0.688	0.326	0.328	…
2011-205526	0.639	0.512	0.588	0.552	0.677	0.570	0.598	0.492	…
2006-581633	0.362	0.352	0.453	0.382	0.347	0.578	0.275	0.355	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Table 7
Part of the weighted patent-citation matrix.

Patents	2011-166513	2007-652876	2012-365028	2010-916818	2009-646632	2007-725925	2011-019729	2011-176190	…
2005-123973	0.517	0.000	0.540	0.000	0.000	0.000	0.000	0.000	…
2001-924608	0.457	0.495	0.496	0.000	0.000	0.000	0.000	0.000	…
2000-615906	0.429	0.000	0.457	0.738	0.000	0.000	0.000	0.632	…
1998-040829	0.000	0.000	0.000	0.742	0.000	0.000	0.000	0.607	…
1999-286213	0.421	0.000	0.473	0.591	0.000	0.000	0.000	0.531	…
1997-920428	0.320	0.000	0.372	0.000	0.357	0.000	0.000	0.000	…
1997-790005	0.451	0.000	0.497	0.648	0.000	0.478	0.537	0.601	…
1995-405812	0.468	0.000	0.518	0.000	0.000	0.444	0.502	0.000	…
1995-469284	0.379	0.000	0.412	0.515	0.000	0.000	0.494	0.521	…
1994-322401	0.000	0.372	0.000	0.000	0.000	0.000	0.563	0.000	…
1989-447677	0.000	0.450	0.000	0.000	0.000	0.584	0.799	0.000	…
1989-365444	0.365	0.387	0.409	0.585	0.566	0.000	0.566	0.541	…
1984-638905	0.365	0.372	0.394	0.000	0.764	0.000	0.480	0.000	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Table 8
Part of the highly cited patent list (137 patents; $\rho_c = 20$).

#	Application #	# of citation	CII(2009–2013)
1	1996-722155	233	176
2	1989-447677	383	145
3	1984-638905	542	137
4	1993-138345	233	133
5	1994-200636	210	114
6	1999-350604	133	92
7	2000-568207	186	85
8	1992-894100	214	76
9	1999-416346	101	71
⋮	⋮	⋮	⋮
135	1995-468288	35	20
136	1988-268907	63	20
137	1988-183015	66	20

succession relationships are represented as directed links and the technological topics are reorganized for effective PDM visualization.

PDM visualization provided an overall understanding about the historical development and current stage of 3D printing technology. We observed several patterns. First, Topics 2 and 7 tended to be non-successive. In particular, patents in Topic 2 (depositing metal materials) and Topic 7 (methods or devices customized for various materials) tended to be independent of each other; patents in the same topic did not have succession relationships. According to our examination, this was because the patents used different types of materials, and therefore provided depositing methods customized to the materials.

Second, Topics 1, 5, 8, and 9 tended to be internally successive. Some former patents in the four topics were advanced by their follow-up patents. For example, patents in Topic 1 (bonding adhesive particulates on a surface or plate) showed technological succession relationships among themselves, but only several patents with a high CII value had technological succession relationships. For example, patent 1989-447677 is directly succeeded by its later patent 1992-894100; they share a process composed of depositing a layer of a powder material in a confined region, applying a further material to one or more selected powder material regions, and depositing a selected number of successive layers. Then, both patents are advanced by patent 1999-410943, which includes an improved process of depositing multiple layers of reinforcement and adhesive compositions for forming a net-shaped reinforcement pre-form.

Regarding Topic 8, patents 1993-068692, 1998-040829, and 2000-615906 are all stereolithographic methods using polymer precursor fluid, and patents 1998-040829 and 2000-615906 succeeded their former patent 1993-068692. Patent 1993-068692 is a method of stereolithographically producing a 3D object using a programmable photomask and a polymer precursor fluid capable of solidification. Patents 1998-040829 and 2000-615906 employ an improved radian energy source, or a spatial light modulator having an array of pixel elements that are individually controllable.

Third, Topics 3, 4, and 6 are very interactive with other topics. Most of the patents in Topic 6 were by 3D Systems which proposed

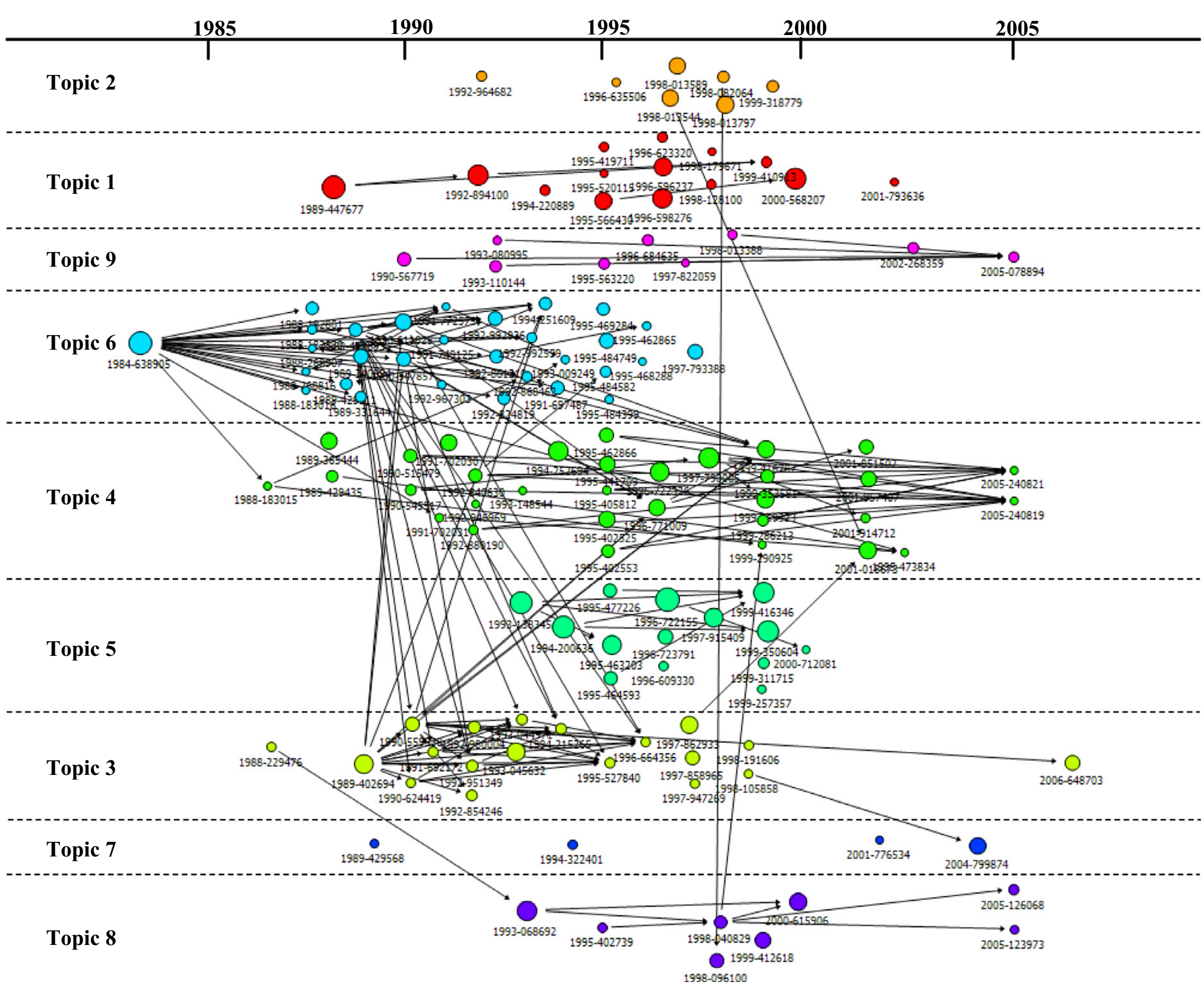

Fig. 10. Technology development map arranged by application year ($\rho_f = 0.70$, $\rho_c = 20$).

the initial concept of generating 3D objects. Especially, 3D Systems' patent 1984-638905 in Topic 6 (basic methods and principles of 3D printing technology) is the first 3D printing patent, entitled "apparatus for production of three-dimensional objects by stereolithography" (Hull, 1986); this patent had a significant impact on many later patents by 3D Systems in Topic 6, including patents 1989-340894, 1993-009249, 1995-469284, and 1995-484582.

Interestingly, there are no major patents in Topic 6 after 1997, which suggests that the basic concepts of 3D printing by stereolithography had been mostly provided by 1997. In addition, this can be explained by the many patents in Topics 4 (supplementary applications of 3D printing) and 3 (powder or particle sintering technology by SLS) which are shown to receive technological knowledge from the patents in Topic 6. For example, patent 1988-183015 in Topic 4 is a supplementary patent that incorporates specific processes to reduce the curl, stress, and other distortions of successive adjacent laminates, and many patents in Topic 3 (powder or particle sintering technology by SLS) strongly succeed the patents in Topic 6.

Another revealing insight from the PDM is that there are only a few of recent major patents, including patents 2005-078894, 2005-240821, 2006-648703 and 2005-126068 granted in August 2007, April 2009, January 2013 and September 2010, respectively. This is explained by the fact that the initial mechanisms of 3D printing have been mostly invented before the 2000s by several firms, including 3D systems and Stratasys. In fact, about 90 of the initial patents were expired since 2014 and it is expected that about 50 additional patents will be expired by 2016. The industrial focus of 3D printing is now on manufacturing low-price 3D printers using the expired core patents, instead of developing new mechanisms of 3D printing.

In this section, using patents of 3D printing technology, we illustrated our quantified method for PDM generation and then could monitor 3D printing technology by examining the nine sub-technology topics and the development paths among some major patents. In particular, we could generate the PDM of 3D printing technology with the less manual effort by the proposed method. Therefore, we believe that this approach will reduce the cost of understanding technological development in the high technologies that are newly emerging or composed of massive patents.

5. Discussion and concluding remarks

PDMs have been widely used in intellectual property-based R&D processes by technology analysts as a useful aid to support a comprehensive understanding of the historical development and current stages of a technology. Although some studies attempt to identify technology development based on patents, few or no studies have synthetically considered two core factors for PDMs: (1) the succession relationship among patents in terms of technological contents and (2) the technological taxonomies of individual patents. Therefore, generating PDMs with these two core factors would be useful for R&D planners to well understand the historical and current development stage of a given technology. For this reason, this study proposed a quantified method for PDM generation using semantic patent topic analysis.

First, we employed LDA to identify semantic technological relationships between pairs of a former patent and its successive patents. Unlike other dimension-reduction techniques, such as latent semantic analysis and principal component analysis, the results generated by the LDA process have interpretation advantages. First, topics by LDA are composed of the term contribution probabilities, so analysts can easily understand what technological category each topic describes, based on the mixture of terms. Moreover, a contribution probability vector of the topics for each patent contains the degree to which each topic contributes to the patent, so the process of categorizing patents is not necessary.

Second, our PDM generator visualized PDMs by sensitivity analysis of the CII value and the strength of the succession relationships. Using threshold values for PDMs could be controversial: setting different threshold values for CIIs and the strength values produces different PDMs. In other words, however, this can support the process of understanding a given technology's development paths with different criteria; the larger the threshold values, the stricter the PDM results. In the same vein, defining a set of keywords to structure patents could be customized by expert knowledge; experts could define a different set of keywords depending on their focus of interest. Fundamentally, our quantified method should be controlled by experts' careful knowledge; it is an expert support tool to help efficiently understand rapidly evolving or emerging technology areas.

In this study, we applied our PDM generation method to 3D printing technology. As a result and observed 137 major patents and 9 sub-topics. The number of observed development paths, in particular, was less than one of the links by patent citation relationships, because our approach considered semantic technological relatedness between patents. Therefore, this output could provide us with the enhanced understanding of the PDM for R&D planners and researchers. In addition, the PDM of 3D printing technology could be generated with less effort by the proposed quantified approach.

Our approach contributes to both academia and industry. First, this approach overcomes the methodological limitations of prior studies. The prior studies for generating PDMs for a given technology neglected the semantic relationships between a former patent and its successive patents. They assumed that a former patent would be succeeded by only one later patent, and in addition, it is hard to identify sub-technology topics for the technology. However, our approach incorporates citation information and semantic patent similarities to effectively identify the substantive technological-succession relationships among patents.

From an industrial perspective, this approach is an efficient aid for technology monitoring under the condition of insufficient time and costs for technology development analysis. For instance, patents in high technologies increasingly appear over time, so it could be costly and time-consuming to manually monitor the historical and changing paths of technology development. Our quantified approach can be flexibly controlled by an expert's examination. Therefore, we expect that this approach will reduce the cost of understanding technological development in the high technologies that are newly emerging or rapidly evolving.

Despite the contributions, challenges still remain for further research. Thus, we present limitations of the current study and conclude this paper with interesting future research topics. First, we used a threshold value of patent CIIs to selectively show major patents, which have a strong possibility of being core patents. However, recently granted patents naturally have a lower chance of being cited. Therefore, a future research topic will develop an analytical measure for the coreness of such patents, and incorporate the measure into our PDM generation method. Second, this paper illustrated the proposed method using patents related to 3D printing technology. Therefore, further research will apply this method to other high technologies, such as printed electronics or augmented reality, to verify its applicability. Finally, the proposed method necessarily involves experts' intervention. Analysts must decide the threshold values and keyword sets for PDM generation. Therefore, an interesting topic in the future will incorporate analytical measures into our approach to support experts' decision making with reduced effort.

Acknowledgement

This paper was supported by Konkuk University in 2014.

References

Albert, M. B., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*(3), 251–259.
Bergmann, I., Butzke, D., Walter, L., Fuerste, J. P., Moehrle, M. G., & Erdmann, V. A. (2008). Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips. *R&d Management, 38*(5), 550–562.
Bíró, I., Szabó, J., & Benczúr, A. (2008). Latent Dirichlet allocation in web spam filtering. In *Proceedings of the 4th international workshop on adversarial information retrieval on the web* (pp. 29–32). ACM.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.
Chang, P.-L., Wu, C.-C., & Leu, H.-J. (2010). Using patent analyses to monitor the technological trends in an emerging field of technology: A case of carbon nanotube field emission display. *Scientometrics, 82*(1), 5–19.
Choi, C., & Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting and Social Change, 76*(6), 754–768.
Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web* (pp. 271–280). ACM.
Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.
European Commission (EC) and European Patent Office (EPO) (2007). Why Researchers should care about patents. European Patent Office.
Gibson, I., Rosen, D. W., & Stucker, B. (2010). *Additive manufacturing technologies*: Springer.
Gress, B. (2010). Properties of the USPTO patent citation network: 1963–2002. *World Patent Information, 32*(1), 3–21.
Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik, 136*, 210–271.
Hirschey, M., & Richardson, V. J. (2001). Valuation effects of patent quality: A comparison for Japanese and US firms. *Pacific-Basin Finance Journal, 9*(1), 65–82.
Hu, Z., Fang, S., & Liang, T. (2014). Empirical study of constructing a knowledge organization system of patent documents using topic modeling. *Scientometrics, 100*(3), 787–799.
Hull, C. W. (1986). Apparatus for production of three-dimensional objects by stereolithography. U.S Patent and Trademark Office: No. 4,575,330.
Hung, S.-W., & Wang, A.-P. (2010). Examining the small world phenomenon in the patent citation network: A case study of the radio frequency identification (RFID) network. *Scientometrics, 82*(1), 121–134.
Jin, X., Zhou, Y., & Mobasher, B. (2005). A maximum entropy web recommendation system: Combining collaborative and content features. In *Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining* (pp. 612–617). ACM.
Kim, G., Park, S., & Jang, D. (2014). Technology analysis from patent data using Latent Dirichlet allocation. In K. M. Lee, S.-J. Park, & J.-H. Lee (Eds.), *Soft computing in big data processing* (pp. 71–80). Springer International Publishing. 271.
Kim, Y. G., Suh, J. H., & Park, S. C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications, 34*(3), 1804–1812.
Krestel, R., Fankhauser, P., & Nejdl, W. (2009). Latent Dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on recommender systems* (pp. 61–68). ACM.
Miller, J. S. (2005). Enhancing patent disclosure for faithful claim construction. *Lewis & Clark Law Review, 9*, 177.
Newkirk, J. W., Liou, F. F., & Francis, R. (2013). Systems and methods for fabricating a direct metal deposition structure having fully forged structural qualities. U.S Patent and Trademark Office: No. 8,617,661.
Park, H., Kim, K., Choi, S., & Yoon, J. (2013). A patent intelligence system for strategic technology planning. *Expert Systems with Applications, 40*(7), 2373–2390.
Park, Y., Ko, N., & Yoon, J. (2015). A patent recommendation method for technology opportunity exploration using firms own patents: The case of 3D printing. *The Journal of Intellectual Property, 10*(1), 169–200.
Sachs, E. M., Haggerty, J. S., Cima, M. J., & Williams, P. A. (1993). Three-dimensional printing techniques. U.S Patent and Trademark Office: No. 5,204,055.
Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 448–456). ACM.
Wang, B., Liu, S., Ding, K., Liu, Z., & Xu, J. (2014). Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: A case study in LTE technology. *Scientometrics, 101*(1), 685–704.
Wang, Y., Sabzmeydani, P., & Mori, G. (2007). Semi-latent Dirichlet allocation: A hierarchical model for human action recognition. In *Human motion–understanding, modeling, capture and animation* (pp. 240–254). Springer.
Xing, D., & Girolami, M. (2007). Employing Latent Dirichlet Allocation for fraud detection in telecommunications. *Pattern Recognition Letters, 28*(13), 1727–1734.
Yoon, J.-H., & Choi, S.-C. (2012). Planning future technology strategies using patent information analysis and scenario planning: The case of fuel cells. *Journal of Information Management, 43*(2), 169–197.
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.
Yoon, J., & Kim, K. (2012). Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics, 90*(2), 445–461.
Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.
Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics, 94*(1), 313–331.
Yoon, J., Park, Y., Kim, M., Lee, J., & Lee, D. (2014). Tracing evolving trends in printed electronics using patent information. *Journal of Nanoparticle Research, 16*(7), 1–15.