Invention property analysis of patents: a case of augmented reality technology

Hayoung Choi
Department of Industrial Management
Konkuk University, Seoul, Korea
Email: mammam@konkuk.ac.kr

Janghyeok Yoon †
Department of Industrial Engineering
Konkuk University, Seoul, Korea
Email: janghyoon@konkuk.ac.kr

Abstract. Understanding innovation concepts of a system is a key factor to determine technology development directions in R&D planning. Patents are an up-to-date and reliable resource for technical advancement, and thus innovation directions of a system can be described by invention properties portrayed in patents. A property indicates what a system is or has, or a specific characteristic of a system or its subsystems, so it has the capability to represent innovation concepts targeted by various inventions. This study defines such innovation concepts as invention properties and aims to analyze technology trends by mining invention properties from patents. Our approach for invention property analysis developing patent-invention property vectors by extracting adjectives with technical meaning from the text of massive patents in a given technology, followed by deriving the relationships among invention properties and constructing of the invention property portfolio map to identify the technology trends in light of innovation concepts. Our approach providing a different type of view in patent analysis is illustrated using patents of augmented reality technology. This approach contributes to determining evolving trends of innovation concepts within a given technology system.

Keywords: Invention property; Patent mining; Innovation concept; Technology trend; Augmented reality

1. INTRODUCTION

With a diversified and rapidly changing technology environment, continuous innovation is a key to successful businesses for firms; firms perform R&D planning and make an effort to have a technological competitiveness for accomplishing successful innovations. In this regard, it is important to understand the innovation concept or what a technology aims to perform, and can help determine technology development directions in R&D planning.

As the result of innovation and R&D process, companies get patents in order to protect their invention knowledge (Park, Kim et al. 2013). A patent is an up-to-date data and also provides information about related technology for the invention. For this reason, patents function effective resources for technology analysis, including monitoring technology trends, identifying technology opportunities and forecasting new technology (Mogee 1991; Liu and Shyu 1997). Based on patent analysis, much research has conducted to support R&D planning. Prior studies conducted analysis to forecast emerging technology by integrating the use of bibliometrics and patent analysis(Daim, Rueda et al. 2006). In addition, much research conducted text mining based patent analysis to development networks to identify trends and generate patent maps to discover new technology opportunities (Yoon and Park 2004; Kim, Suh et al. 2008; Lee and Jeong 2008; Yoon and Kim 2011; Yoon and Kim 2011). The prior studies are based on analyzing quantitative information and keywords related to technology in patents, because it is reliable resource and also it is simple and easy to identify them. Despite the contributions of the prior studies, they have still some difficulty in figuring out innovation concepts that are key factors of technology development.

In this paper, we propose a method to analyze patents in terms of innovation concepts. Innovation concepts represent what a technology aims to perform and solve and they thus describe the features of a technology. Innovation concepts can be described by invention properties in patents. A property

indicates "what a system is or has" and a specific characteristics of a system; prior studies considered properties to be expressed as an adjective in patents (Dewulf 2006; Verhaegen, D'hondt et al. 2009; Yoon, Choi et al. 2010). Therefore, invention properties in patents have the capability to represent innovation concepts targeted by various inventions.

Therefore, this research defines invention properties as innovation concepts. The proposed procedure consists of collecting patents, extracting invention properties, structuring patents in terms of invention properties, constructing invention property networks and positioning the invention properties.

Our approach providing a different type of view in patent analysis is illustrated using patents of augmented reality technology. This approach contributes to determining evolving trends of innovation concept within a given technology system. In addition, this approach can assist decision makers in R&D planning to formulate strategies for technology development.

2. Theoretical Background

2.1 Properties and functions of systems

The concept of property was proposed to generate connections between products, processes and systems based on their properties and functions by Dewulf (2006). A property means a specific characteristic of a system, what a system is or has; a function means an action to realize properties of a system, what a system does or undergoes. And Dewulf suggests that with analysis of about 16,000 patents from the US patent and Trademark Office, properties are described by adjectives and functions are described by verbs(Dewulf 2006). Therefore, properties and functions could be extracted in sentence using grammatical analysis (Verhaegen, D'hondt et al. 2009).

A system constituted by properties and functions that could show innovation concepts of a system (Dewulf 2006). Thus, properties could be related to methods and materials of a system, and functions could be related to uses or objective of a system. This means properties and functions could be key concepts of an invention (Yoon, Choi et al. 2010).

Related study with the property and function, Dewulf(2006) suggests the concept of "product DNA" that describes the set of properties of a system and identified innovation concepts related to products. Verhaegen et al(2009) extracted properties and functions automatically and related them to evolutionary trends of technology to predict improvement directions of a product (Verhaegen, D'hondt et al. 2009). Yoon proposed an approach that extracts properties and functions from patents and generates invention property-function networks based on co-word analysis and social network analysis (Yoon, Choi et al. 2010).

Based on the concept that a property represents an attribute and characteristics of a system, this research defines invention properties as innovation concepts of a technology.

2.2 Vector space model in patent analysis

The vector space model is an algebraic model that represents text documents as vectors of identifiers(Salton, Wong et al. 1975). Thus it is used for disambiguating entities across documents(Bagga and Baldwin 1998).

In the vector space model, each document is expressed as a vector of weights. A document space consists of documents D_{ij}, and each document is identified by one or more index terms T_j. The terms may be weighted according to their importance. Each document D_i is represented by a t-dimensional vector $D_i=(d_{i1}, d_{i2}, \ldots, d_{il})$, d_{ij} representing the weighting of the jth term(Salton, Wong et al. 1975).

There are several different ways of computing the weighting (value). One of the best known schemes is term frequency-inverse document frequency model(TF-IDF). TF-IDF is expressed by two statistics, term frequency (TF) and inverse document frequency (IDF). The term frequency is the number of times term T_j occurs in document D_{ij} (Luhn 1957). The document frequency is number of documents where the term T_j appears. And, the inverse document frequency is the logarithmically scaled inverse fraction of DF and thus IDF measures how much information the word provides (Sparck Jones 1972). Thus, the TF-IDF, multiplying TF by IDF, means how important a word is to a document (Ullman, Leskovec et al. 2011).

- ***Tf(t,d) : term frequency of term t occur in document d***
- ***Idf(t,D) : inverse document frequency of term t in set of document D***

$$= log \frac{N}{|\{d \in D : t \in d\}|}$$

- ***N : total number of documents in the corpus N = |D|***
- ***|d ∈ D : t ∈ d| : number of documents where the term t appears***

- ***Tf-Idf(t,d,D)= Tf(t,d) · Idf(t,D)***

Formula 1 : TF-IDF

One of the advantages of the vector space model is to allow measuring the degree of similarity between documents. Given the two documents, it is possible to compute a similarity between the documents, $s(D_i,D_j)$, using various similarity measures (Salton, Wong et al. 1975).

2.3 Network analysis

Network analysis is derived from graph theory. If a system consists of actors and interaction, a network is generated based on the interactions among actors. Thus,

network analysis allows investigation of interactive relationships among actors.

The application of network analysis is wide and diverse. Especially, it can be adopted to identify technology trends in bibliometrics, which is a study to measure and analyze texts and information (Lee and Jeong 2008; Yoon and Kim 2012). In most cases, patents are adopted for technology analysis using cited-citing relationships.

By adopting social network analysis, it is possible to compute the degree of relationship from various perspectives, such as density, degree of centrality and betweenness (Otte and Rousseau 2002).

3. Proposed approach

The procedure for invention property analysis is composed of five steps (Fig 1): (1) Patents are collected in a given technology area, (2) invention properties are extracted from the patents, (3) patents are structured in terms of properties, (4) invention property network is constructed and (5) invention properties are positioned in portfolio maps.

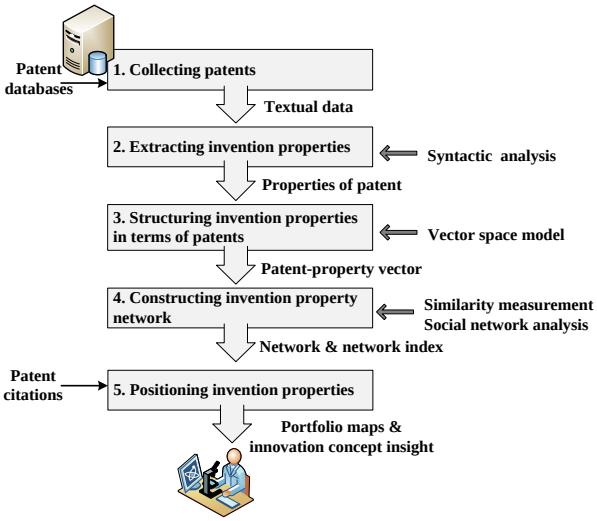

Figure 1 : The proposed procedure

3.1 Collecting patents

The first step in the proposed methodology is to collect patents in a given technology area. The patent is a result of R&D activity and describes reliable information of an invention. Patents can be collected from patent databases by using retrieval query. Then, irrelevant patents are eliminated and thus the final patent set for analysis is prepared.

The patent document has detail description about the invention using the technological terms. It has several textual sections: title, abstract, background summary, detailed description and claims. In particular, claims are considered to be the most important part in patents, because they define the boundary and complete descriptions of an invention (Fujii, Iwayama et al. 2007).

Thus, this research utilizes claims of patents to analyze invention properties.

3.2 Extracting invention properties

The second step is to extract invention properties from the claims of collected patents. The property indicates "what a system is or has"; a specific characteristic of a system (Dewulf 2006). Therefore, a property has the capability to represent an innovation concept targeted by various inventions. This study defines such innovation concepts as invention properties.

A property is expressed as an adjective to show a system attribute. Then, properties represented as adjectives in claims can be extracted by adopting syntactic analysis, however there may be some irrelevant or generic properties that cannot show innovation concepts; for example, good, bad and easy. After eliminating these irrelevant properties, the final invention property set related to a technology is prepared.

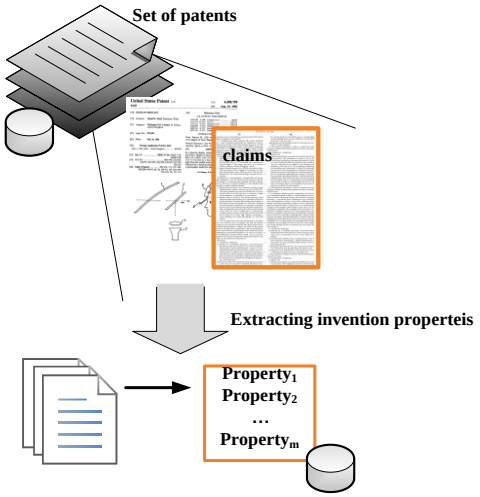

Figure 2 : Extracting invention properties from patents

3.3 Structuring patents in terms of properties

The third step is to structure patents in terms of invention properties. This study regards each patent as a set of properties, so a patent can be encoded by a vector with element values corresponding to properties. Conversely, a property vector can be constructed with element values corresponding to patents.

The occurrence of the property in each patent is assigned as the weight to construct vector space. However, using only occurrence frequencies of properties might be biased by

properties appearing frequently and undervalue some unique but important properties. Therefore, this research adopts term frequency-inverse document frequency (TF-IDF) as an effective way to reflect how important a word is in a document.

The term frequency (TF) is the number of properties in a specific patent. The document frequency (DF) is the number of patents where a specific property appears. And the inverse document frequency (IDF) is the logarithmically scaled inverse fraction of document frequency (DF). By multiplying TF by IDF, it means how important a property is to a patent. Through this process, the patent is structured in forms of the patent-property matrix. And for property vectors, the patent-property matrix is transformed to property vectors.

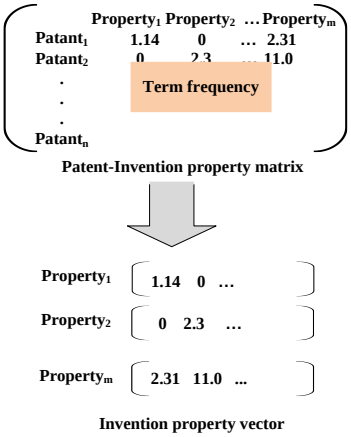

Figure 3: Structuring patents and transforming invention property vector

3.4 Constructing invention property network

The next step is to construct networks of invention properties to identify and visualize their relationships.

The relationships are identified by measuring the similarity between pairs of invention property vectors. Since the cosine similarity has the nice feature that it is 1.0 for identical vectors and 0.0 for orthogonal vectors(Singhal 2001), this research computes cosine similarity between two property vectors in the vector space model. If the similarity between a pair of properties is high, it means that two properties have relevance in a given technology area.

The relationships among properties could be constructed in the network. This network consists of properties as nodes and relatedness among properties as links. For interpreting the network, we use social network analysis that suggests meaningful indexes about relationships. Especially, the centrality index is useful to identify the importance of properties in network. Among various centrality indexes, this study adopts degree centrality, which is a measure to compute the portion of nodes that are adjacent to each node. Thus, we identify important properties in a given technology using degree centrality.

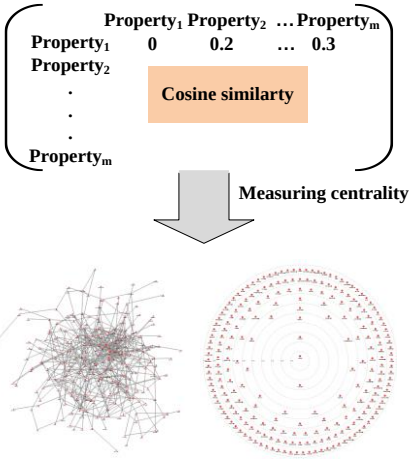

Figure 4: Constructing network of invention properties

3.5 Positioning invention properties

The final step in this research is to position invention properties in a portfolio map based on their importance and performance. This portfolio map is useful to identify technological innovation trends in terms of invention properties.

For identifying the performance of properties, we apply the citation stock of each property. Since a highly cited patent indicates a superior invention in quality to others. And, it influences later patents as the base of later invention. For this reason, the citation stock for each invention property represents the performance of the invention property. To identify the citation stock of properties, forward citation frequencies per each patent are allocated to each property in

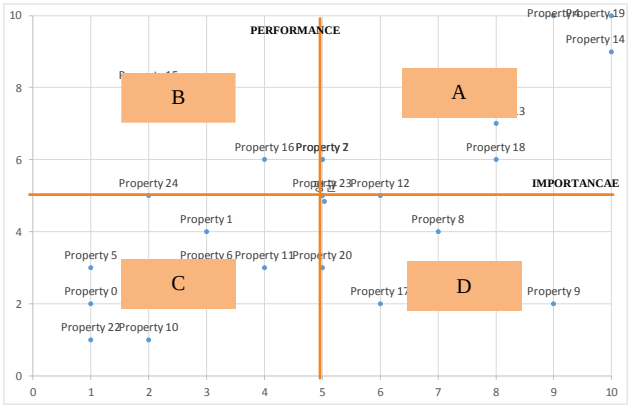

Figure 5 : Portfolio map of invention property

proportion to the rate of the property's occurrence. The total citation stock per each property represents the degree of the property's performance within a specific technology area. And to analyze the importance in specific area, we use the degree centrality index acquired from the previous step.

Through this process, the portfolio map consists of two axes for properties: importance (degree centrality) and performance (citation stock).

The portfolio map is divided into four areas according (Figure 5). Properties in area A represent that they are essential innovation concepts and also well performed or developed. Properties in area B represent that they are less essential but developed overly. And area C means that properties are less essential and less performed. The last area D means that these properties are very essential but not enough developed.

4. Illustration : the case of Augmented Reality

4.1 Augmented Reality technology

This section illustrates the proposed procedure using the case study of Augmented Reality technology. Augmented reality is a technology which enables to generate virtual images on real world environment. This technology is constituted with various components; hardware for display, sensor and input device and software. With the development modern mobile devices, Augmented Reality is getting much attentions and applied in various way, for example, military, medical, commercial and entertainment. In this regard, we conducted empirical analysis on Augmented Reality.

According to procedure of this research, first we collect patent related to Augmented Reality granted since 2015 from WIPSON that is online patent database in Korea. After eliminating irrelevant patents, 1726 patents are collected in this research.

With a selective set of patents, properties which are adjectives are extracted from claims by the syntactic analysis.

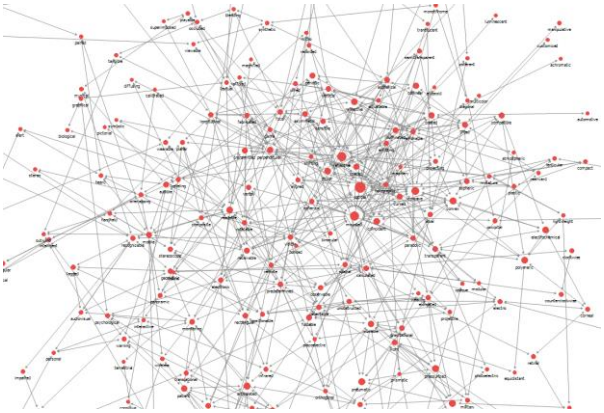

Figure 6 : The network of invention property

And stopwords and adjectives which are irrelevant with invention properties are removed. Next, we structure patents in terms of properties. The set of patents is transformed to patent-property matrix with measuring the weight of TF-IDF. And this matrix is transformed to each property vectors. Using property vectors, we construct invention property network and computing centrality of each property (Figure 6). For final step, we identify the citation rate of each property. Using the two indexes, the centrality for importance and citation rate for performance, we position each property on the portfolio map. The result of portfolio maps are shown in Figure 7.

As the result, properties such as 'optical', 'mounted', 'visible' and 'reflective' are important innovation concepts and also performed well. However, invention properties such as 'scanning', 'stereo' and 'cellular' are over performed as compared with importance. Conversely important innovation concepts such as 'acoustic', 'foldable' and 'wired' are less performed in practice. In this way, we could identify the trend of the invention property in the technology area of Augmented Reality.

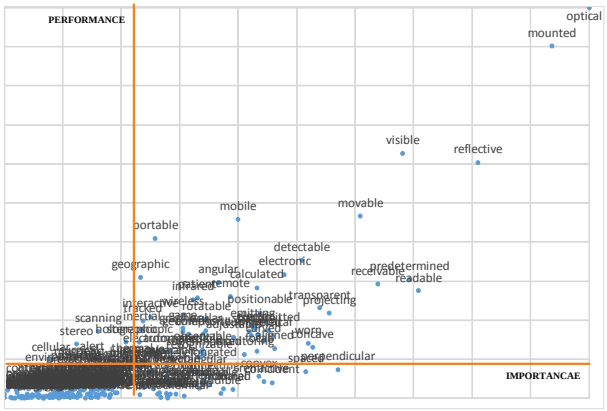

Figure 7 : The portfolio map of invention property

5. Concluding remarks

This research proposes the methodology to analyze the technology trend in the light of in invention property analysis. The proposed procedure is (1) to collect the patents, (2) to extract invention properties, (3) to structure invention property vectors, (4) to construct invention networks, and (5) to position invention properties. The proposed method was illustrates in the case study of Augmented Reality technology.

It is important to analyze the technology in term of innovation concepts that are what the technology needs to solve. Therefore, this research defines innovation concepts as invention properties that are specific characteristics in the technology and we propose the method to analyze the invention property in a specific technology. So there is no necessary to predefine the keyword or key phrase patterns. And

also because the invention property is described as adjectives in sentence, it can be automatically extracted by adopting the syntactic analysis.

This research is expected to assist decision makers to monitor the technology trends. And also the result of this study can be adopted to formulate strategy and determine the R&D planning.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(No. 2015R1A1A1A05027889)

REFERENCES

Bagga, A. and B. Baldwin (1998). Entity-based cross-document coreferencing using the vector space model. Proceedings of the 17th international conference on Computational linguistics-Volume 1, Association for Computational Linguistics.

Daim, T. U., G. Rueda, et al. (2006). "Forecasting emerging technologies: Use of bibliometrics and patent analysis." Technological Forecasting and Social Change 73(8): 981-1012.

Dewulf, S. (2006). "Directed variation: Variation of properties for new or improved function product DNA, a base for 'connect and develop'." ETRIA TRIZ Futures.

Fujii, A., M. Iwayama, et al. (2007). "Introduction to the special issue on patent processing." Information Processing & Management 43(5): 1149-1153.

Kim, Y. G., J. H. Suh, et al. (2008). "Visualization of patent analysis for emerging technology." Expert Systems with Applications 34(3): 1804-1812.

Lee, B. and Y.-I. Jeong (2008). "Mapping Korea's national R&D domain of robot technology by using the co-word analysis." Scientometrics 77(1): 3-19.

Liu, S.-J. and J. Shyu (1997). "Strategic planning for technology development with patent analysis." International Journal of Technology Management 13(5-6): 661-680.

Luhn, H. P. (1957). "A statistical approach to mechanized encoding and searching of literary information." IBM Journal of research and development 1(4): 309-317.

Mogee, M. E. (1991). "Using patent data for technology analysis and planning." Research-Technology Management 34(4): 43-49.

Otte, E. and R. Rousseau (2002). "Social network analysis: a powerful strategy, also for the information sciences." Journal of information Science 28(6): 441-453.

Park, H., K. Kim, et al. (2013). "A patent intelligence system for strategic technology planning." Expert Systems with Applications 40(7): 2373-2390.

Salton, G., A. Wong, et al. (1975). "A vector space model for automatic indexing." Communications of the ACM 18(11): 613-620.

Singhal, A. (2001). "Modern information retrieval: A brief overview." IEEE Data Eng. Bull. 24(4): 35-43.

Sparck Jones, K. (1972). "A statistical interpretation of term specificity and its application in retrieval." Journal of documentation 28(1): 11-21.

Ullman, J. D., J. Leskovec, et al. (2011). Mining of Massive Datasets, Cambridge University Press.

Verhaegen, P.-A., J. D'hondt, et al. (2009). "Relating properties and functions from patents to TRIZ trends." CIRP Journal of Manufacturing Science and Technology 1(3): 126-130.

Yoon, B. and Y. Park (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., S. Choi, et al. (2010). "Invention property-function network analysis of patents: a case of silicon-based thin film solar cells." Scientometrics 86(3): 687-703.

Yoon, J. and K. Kim (2011). "Detecting signals of new technological opportunities using semantic patent analysis and outlier detection." Scientometrics 90(2): 445-461.

Yoon, J. and K. Kim (2011). "Identifying rapidly evolving technological trends for R&D *planning using SAO-based semantic patent networks." Scientometrics 88(1): 213-228.*

Yoon, J. and K. Kim (2012). "TrendPerceptor: A property–function based technology intelligence system for identifying technology trends from patents." Expert Systems with Applications 39(3): 2927-2938.