Advanced Science and Technology Letters
Vol.85 (Information Technology and Computer Science 2015), pp.25-28
http://dx.doi.org/10.14257/astl.2015.85.06

Patent Big Data Analysis using Graph Theory

Sunghae Jun

Department of Statistics, Cheongju University, Chungbuk 360-764, Korea
shjun@cju.ac.kr

Abstract. Patent big data analysis is to analyze patent document data by big data technologies. A patent document contains diverse and complete information of researched and developed technology, such as patent title and abstract, patent number, issued date, claim, drawing, etc. The velocity of increasing the number of applied patents is so fast. So, patent data have characteristic of big data, 3V (volume, variety, and velocity). In this paper, we propose a graph model for patent analysis because visualization by graph theory is representative method for big data. We can apply the results of patent analysis to R&D planning and technological innovation for company and nation.

Keywords: Big data, graph theory, patent document, visualization, technology analysis.

1 Introduction

Big data can be defined as various descriptions by academic and industrial fields. In statistics, big data is focused on analysis of large data set. Also big data is about collecting, storing, and preprocessing of huge and heterogeneous data in computer science. On the other side, big data is managed for making and implementing good decisions. Generally big data has three features, volume, variety, and velocity [1]. Patent documents are considered as big data, because patent data also have three characteristics of big data. So in this paper, we propose patent analysis using big data approach. We use graph theory of data structure for big data visualization of patent data, and find relationships between technologies. To show how our proposed method could be applied to real problem, we perform a case study for target technology.

2 Related Works

2.1 Graph Theory and Big Data Visualization

Data structure is data collection for organizing data in computer memory [2]. General data structure includes array, list, stack, queue, tree, and graph. Using data structure efficiently, we can manipulate data such as storing and analysis. In this paper, we use

ISSN: 2287-1233 ASTL

graph for big data analysis. Graph is a data structure representing relationships between connected objects. A graph (G) consists of vertex (V) and edge (E) as follow; G=(V, G). The V means object with diverse characteristics, and the E is relationship between vertices. In big data, graph data structure is a useful tool for finding dependencies among data objects. So we can visualize the relationships between things in big data, where the things represent all objects such as keywords, people, companies, documents, etc.

2.2 Patent and Technology Analysis

Patent is an intellectual property (IP), and inventor's exclusive right granted by nation or government for a limited period of time [3]. A patent document contains various results of developed technology such as title, abstract, International Patent Classification (IPC) codes, applied date, claims, figure, drawing, etc. [4]. So, patent data are useful for technology analysis. We can perform R&D planning by technology analysis using patent documents.

3 Patent Big Data Analysis using Graph Theory

Big data can be defined as various descriptions by academic and industrial fields. In statistics, big data is focused on analysis of large data set. Also big data is about collecting, storing, and preprocessing of huge and heterogeneous data in computer science. On the other side, big data is managed for making and implementing good decisions. Generally big data has three features, volume, variety, and velocity [1]. Patent documents are representative big data because patent includes heterogeneous data types such as text, number, and figure. Also massive amounts of patents have been filed and registered to patent offices around the world. Therefore we use the concept of big data for patent analysis in this paper. Fig. 1 shows our procedure of patent big data analysis.

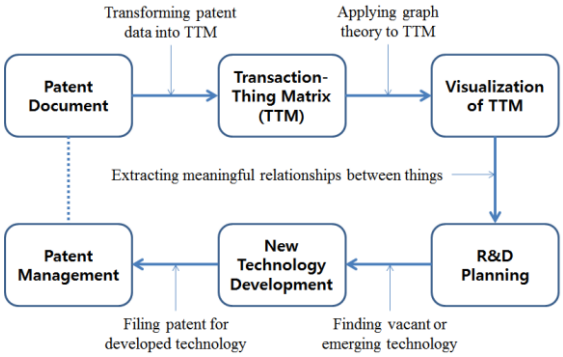
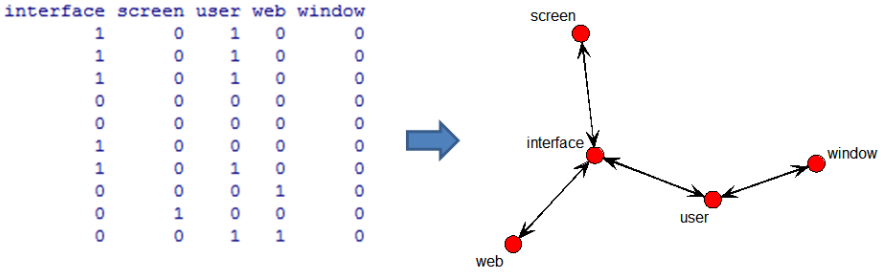


Fig. 1. Procedure of patent big data analysis.

The collected patent documents should be transformed into transaction-thing matrix (TTM) for big data analysis, because most analytical tools of big data such as statistics as well as graph theory need to structured data type consisted of row and column [5]. The row and column of our TTM are transaction and thing respectively. Using the TTM, we get visualization of TTM by graph theory. This result is used for R&D planning, and we can find vacant or emerging areas of target technology. Finally we develop new technology, and file patent for the developed technology. This process is important to patent management.

To verify the performance of our study, we did a case study for graphic user interface (GUI) technology. We searched 100 patents related to GUI from KIPRIS (Korea Intellectual Property Rights Information Service) [6]. Fig. 2 shows TTM of GUI patents and its visualization by graph theory.



(a) TTM of GUI patents (b) Visualization from TTM

Fig. 2. Transaction-thing matrix (TTM) and its visualization for patent big data analysis.

In (a) of Fig. 2, the TTM shows 10 transactions (patents) from 100 patent documents. Also the TTM contains five vertices (things), which are “web”, “screen”, “interface”, “window”, and “user”. The (b) in Fig. 2 shows graph structure for finding relationships between things. We used correlation coefficient matrix as a measure for constructing visualization of things. Fig. 3 represents correlation coefficient matrix of TTM.

	interface	screen	user	web	window
interface	1.00000000	0.12311830	0.62770106	0.03964318	-0.03429642
screen	0.12311830	1.00000000	-0.03342169	0.03254217	-0.02271667
user	0.62770106	-0.03342169	1.00000000	0.02436706	-0.03703352
web	0.03964318	0.03254217	0.02436706	1.00000000	-0.03357213
window	-0.03429642	-0.02271667	-0.03703352	-0.03357213	1.00000000

Fig. 3. Correlation coefficient matrix of five vertices (things).

We used R data language for statistical computing and visualization in this paper [7]. The R provides various useful functions and packages for big data analysis. From the result of Fig. 2, we conclude that the technology of “window” affects development of “interface” technology, and the technologies of “user” and “web” influence to development of “interface” technology directly. Using the result of visualization of GUI patents, we can perform R&D planning for new technology development related

to GUI. In this case study, we considered only five keywords as vertices of graph structure for technology analysis of GUI. But, we can use more keywords for patent analysis and visualization, and get more diverse results for R&D planning in patent management.

4 Conclusions

In this paper, we proposed patent big data analysis using graph theory in data structure. For given target technology, we visualized relationships between keywords of collected patent documents. The keyword used as vertex of graph theory. To get more efficient visualization, we made a TTM consisting of patent (transaction) and keyword (vertex or thing). Using the TTM, we can get diverse visualization results. Our research contributes to plan R&D policy of nation or company. In addition, we can find vacant and emerging technology areas from the visualization result by this study. But, increasing vertices of patent data will be burden for understanding visualization result. This is limitation of our research. So in our future work, we will study advanced big data analysis to settle this limitation, also develop more efficient methodology for patent visualization systematically.

References

1. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data–The next frontier for innovation, competition, and productivity, McKinsey Global Institute (2010)
2. Weiss, M. A.: Data Structure and Algorithm Analysis in C++, 3rd edition, Prentice Hall (2006)
3. Hunt, D., Nguyen, L., Rodgers, M.: Patent Searching Tools & Techniques, Wiley (2007)
4. Roper, A. T., Cunningham, S. W., Porter, A. L., Mason, T. W., Rossini F. A. Banks J.: Forecasting and Management of Technology, John Wiley & Sons (2011)
5. Jun, S.: Patent and Big Data, What's the Connection? Proceedings of KIIS Fall Conference, vol. 24, no. 2, 43-44. (2014)
6. KIPRIS, Korea Intellectual Property Rights Information Service, www.kipris.or.kr
7. R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

