

# TECHNOLOGY ANALYSIS FOR INTERNET OF THINGS USING BIG DATA LEARNING

Sunghae Jun<sup>1</sup>

<sup>1</sup>Professor, Department of Statistics, Cheongju University, Chungbuk, Korea

## Abstract

The internet of things (IoT) is an internet among things through advanced communication protocols without human's operation. The main idea of IoT is to reach and improve the common goal by their intelligent networking. The IoT is an integrated technology of several sub technologies, such as wireless sensor or semantic. The technology of IoT has been evolved according to the environment based on information communication technology and social infrastructure. So we need to know the technological evolution of IoT in the future. In this paper, we analyze the IoT technology to find its technological relationship. We use patents and papers related to the IoT, and consider big data learning as tool for the IoT technology analysis. To verify the performance of our proposed methodology, we perform and show our case study using collected the patent and paper documents.

**Keywords:** Big data learning, Internet of things, Technology analysis, and Patent analysis

\*\*\*

## 1. INTRODUCTION

Since the ubiquitous computing was introduced in the late 1980s, the technologies of communication between humans and things, or between things and things have been researched and developed rapidly [1]. The technologies of internet of things (IoT) that the things can communicate their information each other without the help of human have been spread at the beginning of the 2000s [2]. Many researches of IoT have been performed in diverse academic and industrial fields, and the developed results have been published as papers and applied as patents [3]. The papers of IoT have been published for a long time. By comparison, the patents of IoT have been applied and registered relatively recently. So, we can get the most recent information of IoT technology from the published journal and conference papers. But in general, the paper has the limitation of developed technology because of its characteristics. On the contrary to this, a patent have detailed and complete information of developed technology because of its exclusive right for the technology in applied and registered patent. In this paper, we consider both paper and patent data for technology analysis of IoT. The analytical result of IoT paper data gives the newest trend about IoT technology. Also, we get more detailed result from IoT patent data analysis. To analyze the paper and patent data, we use big data learning (BDL). The BDL includes statistics and machine learning algorithms. Next section shows the related works which are the introductory IoT and BDL. In section 3, we propose our methodology for IoT technology analysis using BDL, we also show the experimental result of proposed methodology in this section. Our conclusion is shown in section 4.

## 2. RELATED WORKS

The research backgrounds related to our study are IoT and BDL. First we show the description of IoT.

### 2.1 Internet of Things

We are interested in how the things with intelligence are connected each other and acquire new knowledge, finally provide novel value to us. This is IoT which is a system to make what we want to by networking between intelligent things on the internet without human's intervention [2]. In addition, the IoT is a networking of thing space based on sensor, network, and information over distributed environment [3]. We should develop the technologies of sensor, network, and interface for the IoT. In this paper, we analyze the technologies related to IoT for its technological evolution. So, we use patent and paper data for technology analysis of IoT.

### 2.2 Big Data Learning

We are in the age of big data [4],[5]. Big data has three typical characteristics which are volume, variety, and velocity [6]. That is, the size of big data is extremely large, the data types of big data are diverse such as number, text, and figure. In addition, the data processing of big data is so rapidly. So we should consider these characteristics when we analyze the big data. The BDL is a technique of learning from big data [7]. BDL is distinguished from traditional data analysis such as statistics. We have to transform the big data into a structured data. This is suitable to data analysis using statistics and machine learning algorithm. And then we analyze the structured data by learning tools which are statistics and machine learning. Finally, we apply the results to real domain for optimal decision. In this paper, we consider paper and patent to big data because the paper and patent data include number, text, and figure, also they are so large and the velocity of their publication and application is rapidly. The Figure 1 shows the process of BDL.

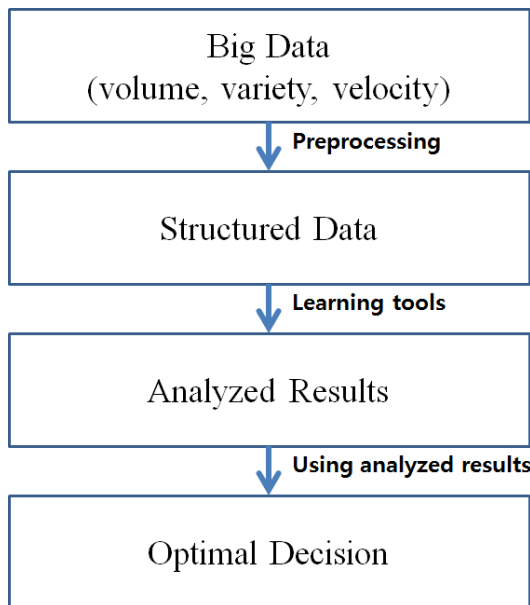


Fig -1: BDL process

Using BDL process, we can do optimal decision from big data. Next we propose a methodology for IoT technology analysis using BDL

### 3. IoT TECHNOLOGY ANALYSIS USING BIG DATA LEARNING

In this paper, we use the patent and paper data for IoT technology analysis. We also perform the BDL based on statistics and machine learning algorithm to analyze the data. Figure 2 shows the process of IoT technology analysis using BDL.

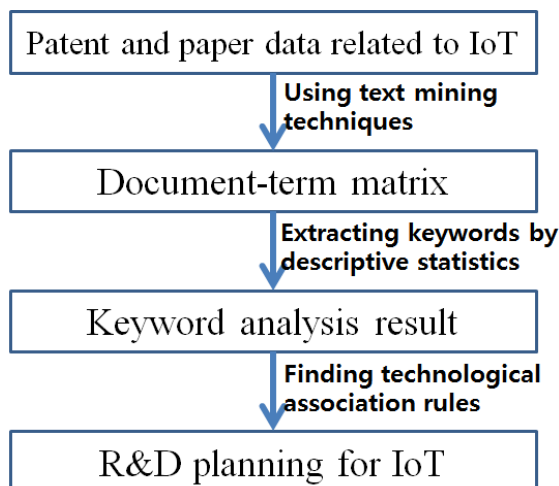


Fig -2: IoT technology analysis by BDL

First we search patent and paper documents related to IoT from patent and paper databases such as KIPRIS, USPTO or IEEE xplore [8],[9],[10]. These data are big data because they include text, number, and others. So we can apply the BDL to analyze patent and paper data. In general, patent and paper consist of texts mainly, and these are not suitable to statistics and machine learning. So we transform the text

data into structured data which is called as document-term matrix [11]. This matrix consists of document and term as row and column respectively. Each cell is the occurred frequency of a term in a document. For this transformation, we use text mining techniques and social analysis by “tm” and “sna” packages of the R project [11],[12]. Next we extract high ranked keywords by descriptive statistics for keyword analysis. In our research, we can have efficient R&D planning for IoT using the result of keyword analysis.

For our case study, we collected the patents and papers related to IoT from KIPRIS and IEEE respectively [8],[9]. First of all, we retrieved total 1,517 patent documents from the patent database which were consisted of the nations of the United States (19 patents), Europe (6 patents), the Patent Cooperation Treaty (PCT) (46 patents), Japan (1 patent), China (1441 patents), the United Kingdom (2 patents), and Taiwan (2 patents). The number of applied patents by year is shown as follow.

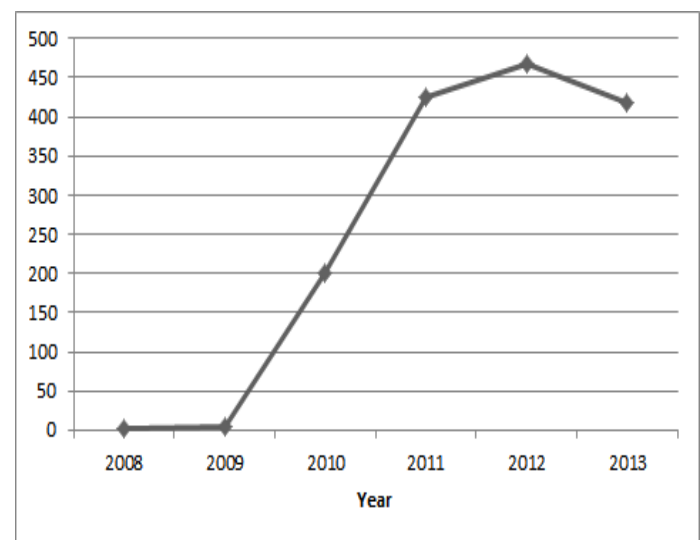


Chart -1: Number of patents by year

We found most technologies related to IoT were developed recently. Also, from the year 2010, the number of applied and registered patents has been increased rapidly. To analyze the patent documents, we constructed patent-term matrix. Using text mining preprocessing, we made this matrix, and the dimension of this matrix was 1517 by 1833. That is, the number of occurred terms was 1833. We extracted top ranked keywords from the patent-term matrix as follow.

Table -1: Top Ranked Keywords: Patent

Rank	Frequent Term
1 <sup>st</sup>	Control, data, device, information, intelligent, management, monitoring, system, technology
2 <sup>nd</sup>	Terminal
3 <sup>rd</sup>	Gateway, equipment
4 <sup>th</sup>	Communication, platform, service
5 <sup>th</sup>	Access, network, wireless

From this result, we found that the technologies based on the IoT patents were focused on the actual usage and service, because the first ranked terms contained “control”, “device”, “system”, and “management”. Also the IoT patents showed the importance of the technologies with “intelligent”, “wireless”, “platform”, and “terminal”. Next we considered another technology analysis using papers related to IoT. We retrieved them from IEEE journal site. We selected 100 papers and make paper-term matrix. This matrix consisted of 3375 terms. Like the case of patent data, we extracted top ranked keywords from the matrix as follow.

**Table -2:** Top Ranked Keywords: Paper

Rank	Frequent Term
1 <sup>st</sup>	Application, information, network
2 <sup>nd</sup>	Data, system, technology, smart
3 <sup>rd</sup>	Management, architecture, communication
4 <sup>th</sup>	Model, framework, social, use, device, object, security
5 <sup>th</sup>	Service, cloud, analysis, design, global, semantic, integration, sensor

The IoT papers had the technologies based on “application”, “information”, and “network”. These terms were the first ranked keywords for IoT technology. This shows more scientific point of view than the IoT technology based on the patents. In addition, the IoT papers contain the technologies of “model”, “social”, “security”, “cloud”, “analysis”, “design”, “global”, “semantic”, “integration”, and “sensor”.

Next we combined the keyword results of patent and paper cases for finding keyword distribution of IoT technology. Table 3 shows the keyword distribution of IoT patents and papers.

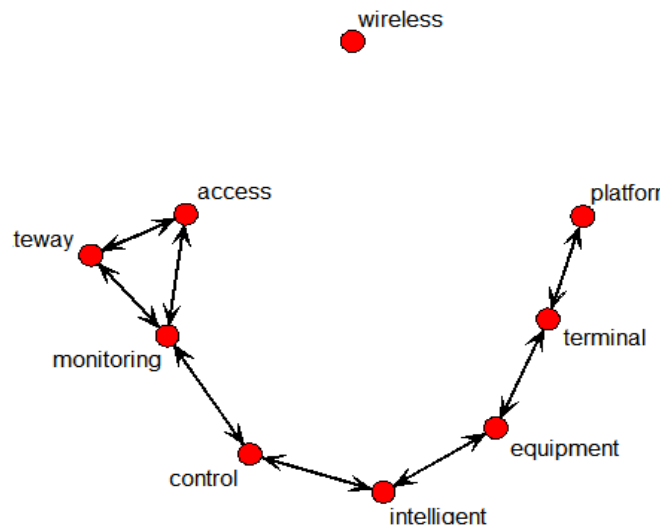
**Table -3:** Keyword Distribution

Data	Frequent Term
Patent & Paper	Data, device, information, management, system, technology, communication, network, service
Patent	Control, intelligent, monitoring, terminal, gateway, equipment, platform, access, wireless
Paper	Application, smart, architecture, model, framework, social, use, object, security, cloud, analysis, design, global, semantic, integration, sensor

From the frequent terms included in patent and paper at the same time, we can know the following technologies are researched and developed in academic and industrial fields concurrently; “data and information”, “device and system”, “management and service”, and “communication and network”. Also we can find the industrial trend of IoT technology from the frequent terms of patent keyword distribution. Using the frequent terms of IoT patents, we decided the technologies of “intelligent platform for controlling and monitoring wireless access” was popular issues for industrial domain of IoT. Especially the monitoring technology is important in the wireless

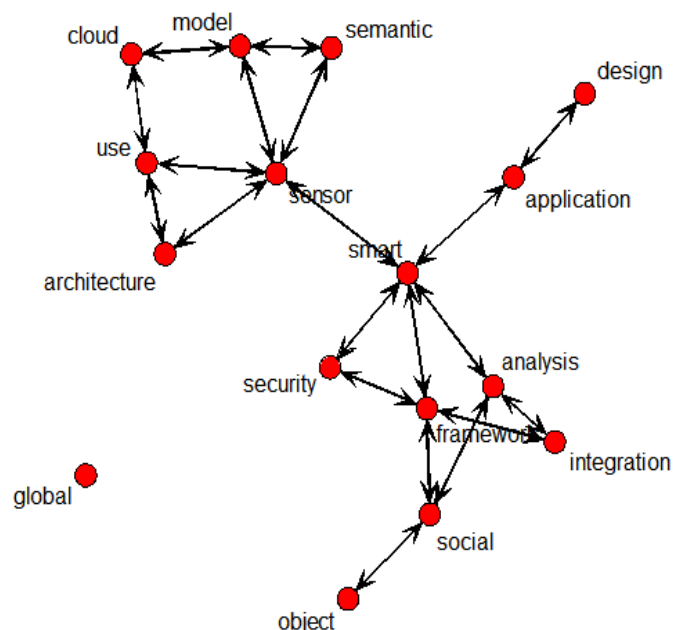
technology of IoT [13]. The technologies of “analytical model for social and security frameworks, and smart and semantic architecture for sensor and application” were researched heavily in academic field.

To know more detailed relationship between the keywords of IoT in patent document data, we structured a SNA graph using correlation coefficients of the keywords. Figure 3 shows the SNA result of IoT patent data.



**Fig -3:** SNA graph based on correlation of IoT patents

The technologies of “gateway”, “access”, and “monitoring” are developed collaboratively. We found the technology of “wireless” is located isolated. So, we knew this technology is developed independently. The other technologies are linearly related. Also the technology of “intelligent” is intermediate technology of IoT, because it is in center of SNA graph. Next figure shows SNA graph of IoT paper using the correlation matrix of its keywords.



**Fig -4:** SNA graph based on correlation of IoT papers

We found that this SNA graph had three components. The first component consists of “cloud”, “model”, “semantic”, “use”, “architecture”, and “sensor”. These represent the technology of system and architecture of IoT. The terms of “smart”, “design”, “application”, “security”, “framework”, “analysis”, “integration”, social”, “object” are assigned to the second component. This component shows the technology of IoT design and application. The last component is “global”, but we cannot add a technological meaning to this component. Also the first and second components were connected by the terms of “sensor”, and “smart”. Next two figures show other SNA plots by keyword matrix for IoT technology analysis. Figure 5 is a SNA graph using IoT keyword matrix.

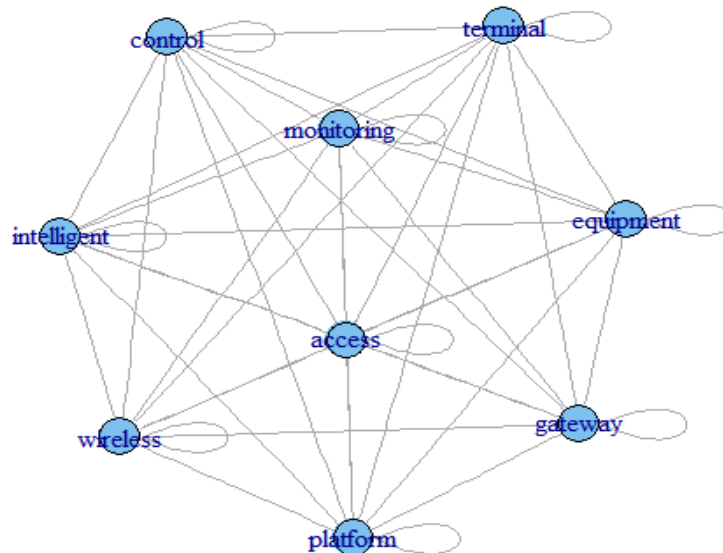


Fig -5: SNA graph based on keyword matrix of IoT patents

Using the result of above matrix, we decided that the terms of “monitoring”, and “access” were located in central areas on the IoT. So, in the patent aspect of IoT, the technology of monitoring and accessing can be considered as meaningful technology. Figure 6 shows another SNA graph of IoT keyword matrix in paper data.

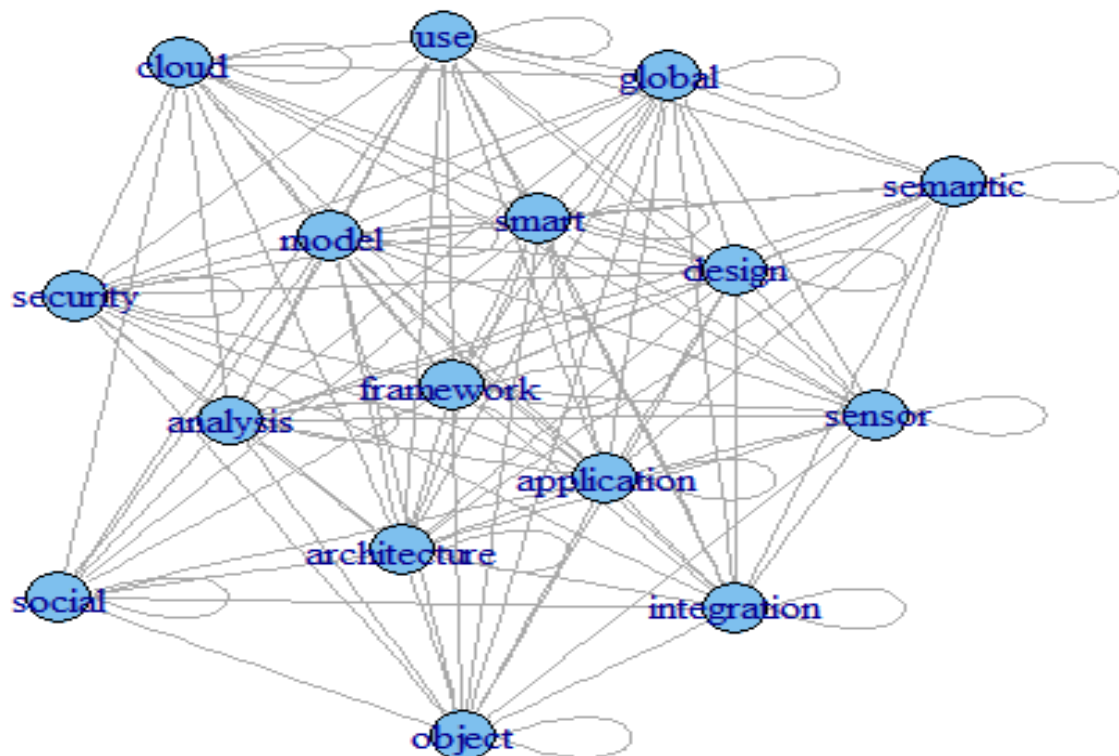


Fig -6: SNA graph based on keyword matrix of IoT papers

The terms of “smart”, “model”, “analysis”, “framework”, “design”, “architecture”, and “application” were regarded as central meanings from the point of view on IoT papers. In this experiment, we found that there were shown more diverse studies in paper than patent.

#### 4. CONCLUSION

In this paper, we analyzed the patents and papers related to IoT technology. We collected the IoT patent and paper data from KIPRIS and IEEE. Using the text mining preprocessing, we made structured data for descriptive statistics, and SNA. Also the results of statistics and SNA graph provided the understanding the technological relationship in IoT technology. From our study, we knew the main researched and developed areas on IoT technology were about the following two aspects. The first is the technology of system and architecture for IoT design. The second is the technology of monitoring and accessing between objects in the IoT environment.

There are so many methods for technology analysis based on big data learning. In our future studies, we will apply them to the technology analysis works as well as IoT technology analysis.

#### REFERENCES

- [1]. D. Bandyopadhyay, and J. Sen, (2011) “Internet of Things: Applications and Challengers in Technology and Standardization,” *Wireless Pers Commum Springer*, Vol. 58, pp. 49-69.
- [2]. L. Atzori, A. Iera, and G. Morabito, (2010) “The Internet of Things: A survey,” *Computer Networks*, Vol. 54, Iss. 15, pp. 2787-2805.
- [3]. L. Atzori, A. Iera, G. Morabito, and M. Nitti, (2012) “The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization,” *Computer Networks*, Vol. 56, Iss. 16, pp. 3594-3608.
- [4]. S. Lohr, (2012) *The Age of Big Data*, The New York Times, February 11, 2012.
- [5]. N. R. Shikalgar, and D. Badgular, (2013) “Online Review Mining for Forecasting Sales,” *International Journal of Research in Engineering and Technology*, Vol. 2, Iss. 12, pp. 53-55.
- [6]. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, (2011) *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.
- [7]. S. Jun, and J. Choi, (2014) “Big Data and Statistics, What’s the Connection?” *Proceedings on KIIS Spring Conference*, Vol. 24, No. 1, pp. 183-184.
- [8]. KIPRIS (2014) *Korea Intellectual Property Rights Information Service*, <http://www.kipris.or.kr>.
- [9]. IEEE (2014) *IEEE xplore*, <http://ieeexplore.ieee.org>.
- [10]. USPTO (2014) *The United States Patent and Trademark Office*, <http://www.uspto.gov>.
- [11]. I. Feinerer, K. Hornik, and D. Meyer, (2008) “Text Mining Infrastructure in R,” *Journal of Statistical Software*, Vol. 25, Iss. 5, pp. 1-54.
- [12]. C. T. Butts, (2008) “Social Network Analysis with sna,” *Journal of Statistical Software*, Vol. 24, Iss. 6, pp. 1-51.
- [13]. P. S. Kumar, and V. A. Kumar, (2014) “An Efficient Monitoring System for Sports Person Using Wi-Fi Communication,” *International Journal of Research in Engineering and Technology*, Vol. 3, Iss. 11, pp. 20-23.

#### BIOGRAPHIES



Sunghae Jun is a professor of Statistics at Cheongju University, Korea. He received the BS, MS, and PhD degrees in department of Statistics, Inha University, Korea, in 1993, 1996, and 2001. Also, he received PhD degree in department of Computer Science, Sogang University, Korea in 2007. Lastly he got PhD degree in department of Information Management & Security, Korea University, Korea in 2013. From 2009 to 2010, he was a visiting scholar in department of Statistics, Oklahoma State University, the United States. He has researched big data and management of technology.