International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Technology Predicting the Future using Data Mining Techniques

Junhwan Moon, Jinhwa Kim

Abstract: ***The technologies of the future society can make a super connective society where everything is connected and organically interacted, and it promotes "Hyper-Connected", "Hyper-Intelligent", "Hyper-Real" as it is possible to become intellectual by techniques such as artificial intelligence and big data analysis. To find out the key triggers by extracting the technologies with these characteristics and identifying what factors and forms are used allows us to grasp changes in the industry due to the rapidly changing domestic environment and the influence of innovation technologies. According to the analysis, the key triggers of major technologies in information society are keywords such as security, technology, service, virtual, and finance which have a great chance to converge with other technologies. The implications of this study are as follows. First, key technologies of domestic information society were identified based on a huge amount of reliable news data. In addition, Network Analysis based on Data, which has increased the frequency of studies all over the field including computer science, sociology and business administration, is applied to predicting the future. Based on this, the methodology has been proposed that can be used to predict the pattern of technology development or notable technologies in the future. This study wasn't conducted only on one detailed technology, but comprehensively on major technologies extracted in the era of 4th Industrial Revolution and presents Black Swan Prediction to enhance competitiveness of domestic industry in related technologies and create new added value.***

Keywords**: Technology Prediction, Big Data, Social Network Analysis.**

I. INTRODUCTION

As the Digital Transformation accelerates under the 4th Industrial Revolution, the boundaries of physical, social, industrial, and economic realms are broken down, and new technologies and content with forms, attributes, and structures different from the past are created and consumed. This makes changes in the overall industrial ecosystem. In particular, the production of new content and the development of various industries are in progress as it has developed that people, things, products and services are hyper-connected due to technologies such as the Internet of Things (IoT), Big Data, Artificial Intelligence (AI), Cloud Computing, and Virtual Reality (VR). The 4th Industrial Revolution has led to remarkable qualitative changes in all ecosystems of industries, companies and business, and is mobbing towards realizing an unpredictable and innovative future. In these times, technology forecasting has a reasonable outlook on the speed and direction of technology development in consideration of changes in customer needs and competitive environment, and importance of establishing future technology strategies is increasing (Coates, V. et al., 2001). Major countries and research institutes around the world are reflecting on the rapidly changing environment through regular environmental scanning and reflecting them in future technology forecasts. However, most future technology forecasting methods have been developed for large-scale projects, which does not help the technology forecasting activities of commercial companies. Therefore, technology forecasting methods are needed to overcome the limitations of cost and time and to identify rapidly changing business environmental changes. The main purpose of this study is to analyze the industrial changes caused by the rapidly changing domestic business environment and the influence of innovative technologies, to grasp the crisis response, and to forecast the economic future more quickly. Thus, in order to respond to the crisis and take the first-mover advantage according to the change of the new environment, this study will examine what areas the emerging innovative technologies are trying to pursue and how they are related, and predict the directions in which the future technology environment will be combined and converged.

Manuscript published on January 30, 2020.
* Correspondence Author
Junhwan Moon*, Senior Research Engineer of Department of Business Research, Sogang University, Seoul. Email: mjhpioneer@gmail.com
Jinhwa Kim*, Professor of Department of Management Information Systems, Sogang University, Seoul. Email: jinhwakim@sogang.ac.kr

II. RELATED WORKS

A. The Status of Future Technology in Information Society

In 2007, Apple introduced the iPhone, a combination of phone, computer, and the Internet, where electronics dominated the lives of consumers. These technological advances are changing our lives. OECD (2016) selected 10 future technologies that will have a significant social and economic impact over the next 10-15 years. Publishing 'OECD Science, Technology and Innovation Outlook 2016', OECD explains that these technologies can provide clues to dealing with the challenges facing the world, including aging populations, climate change, and resource depletion.

In addition, the World Economic Forum (WEF, 2016) pointed out that Artificial Intelligence (AI), Mechatronics, Internet of Things (IoT), 3D Printing, Nanotechnology, Biotechnology, New Materials Technology, and Energy Storage Technology are the innovation that leads the 4th industrial revolution.

Klaus Martin Schwab classified major innovations from three perspectives in his book 'the 4th industrial revolution': Physical technology, Digital technology, and Biotechnology. Gartner announces the attention to 4D printing, Augmented Reality (AR), Virtual Reality (VR), and Connected Homes as digital business innovation is accelerating and technology is increasingly human-centered in 2016 with the release of a new technology hype cycle. Unprecedented advances in computing power will enable companies to exert their influence through machine learning, virtual personal assistants, smart robots, and autonomous vehicles. Also, the advent of the platform revolution will serve as a bridge between humans and technology. Key technologies that facilitate building such platforms include Neural Hardware, Quantum Computing, Block chain, and IoT Platforms (Gartner, 2017).

The technology of the future information society we will meet is a hyper-connected era where everything is connected and interacts in common. Intelligence through technologies such as artificial intelligence and big data will not only enable on-demand services, but also ultimately produce and develop new technologies and content with surreal forms, attributes, and structures.

This study predicts that as modern technologies becomes Hyper-Connected, Hyper-Intelligent, and Hyper-Real, which direction the technologies will evolve and what factors and forms are the technologies emerging and used in the future. For this purpose, this study examines what factors are developed as trigger focusing on IoT and Fin-tech technologies enable a hyperconnected society, Artificial Intelligence and Big Data to promote super-intelligence, and Virtual Reality and 3D Printing Technology which can lead to surrealization.

B. Social Network Analysis

Social network theory is based on the view that the structural characteristics of social relations affect behavior and uses social network analysis techniques. Social network analysis refers to methods for abstracting the networks between individuals and groups into nodes and links and then identifying their structural characteristics and social phenomena (Newman and Girvan, 2004; Wasserman and Faust, 1994). This technique considers individuals as nodes and individual social relationships as links to obtain social networks, and is used to find out how it spreads through analyzing the structure and strength of social networks, who is an influencer that acts as a hub or center to word of mouth on the network (Freeman 2017; Kwahk, 2017). In particular, social network analysis is used as one of the big data analysis techniques, which is easy to understand the semantics of the network by grasping the frequency of appearance of key words. The most commonly used indicator among social network analysis indicators is centrality, which allows one actor to identify which nodes play a major role in the entire network (Park et al., 2015). Indicators for centrality analysis include degree centrality and closeness centrality based on the connectivity between objects, betweenness centrality based on intermediary between entities, and eigenvector centrality based on the weight of the relative connectivity of individuals (Scott, 2017; Wasserman & Frust, 1994).

III. METHODOLOGY

The purpose of this study is to derive the main technologies of the information society and to attempt their future impact and prediction. Since the data collected in this study are large and unstructured types, it is not appropriate to apply the traditional statistical analysis method assuming that the data is structured (Oh et al., 2009). Therefore, it was determined that text mining analysis, which aims to extract and process useful information from unstructured or semi-structured text data based on natural language processing techniques, is useful for extracting concepts represented in text and identifying relationships among them (Paranyushkin, 2011). In this study, the frequency of occurrence was calculated by extracting key words from the news articles collected using Python and R programs, big data analysis tools, and refining them into a form suitable for normalization and analysis. Using the NodeXL program, this study identified the linkage structure between keywords, analyzed the centrality of linkage degree, and clustered highly interconnected keywords for network visualization. The overall research process is like <table 1>.

Table 1 Research process

Phase	Research procedure	Detail
Phase 1	Extraction keywords	• Main keywords of Information society (Internet of Things, Fintech, Artificial Intelligence, Big Data, Virtual Reality, 3D printing
Phase 2	Data collection	• Web crawling and data filtering for 242,645 news
Phase 3	Data Cleaning	• Extraction keywords through preprocessing and Natural Language Processing
Phase 4	Data Analysis and visualization	• Keyword frequency analysis, social network analysis
Phase 5	Conclusion	• Result and Implications

A. DATA COLLECT

In this study, news articles related to major technologies of information society were collected using news homepage (https://media.daum.net/). News is a representative medium for understanding issues that have been noticed by the public in a certain era (Dos Rieis et al., 2015).

Reliable data collection is essential for forecasting future technologies. As GIGO, Garbage In, Garbage Out, says "If the data are not appropriate, the conclusion is unreliable," choosing the right data before using technology prediction is indispensable, too (Batini, C. et al., 2011; Bizer, C. et al, 2012).

In order to collect news articles using the Web crawling method, we need to know the address of the web documents to be collected. You can find the URL by entering 'keyword' and 'search period' on the DAUM News homepage (https://media.daum.net/). Based on a theoretical background, the search period is two years based on an event in Google Trend where interests in each keyword reaches its peak. The most things we want to know about the future are the weather, health, and the economy and so on. The future can be predicted to a certain extent if a phenomenon from some point in the past to the present is understood in advance. The status of data collection by main keyword selected for this study is like <table 2>.

Table 2 Summary of keywords

keyword	Google Trend (vertex)	period
Virtual Reality	Dec. 2016.	2014.12.01 –2016.12.31
Big Data	Nov. 2017.	2015.11.01 - 2017.11.31
Internet of Things	Dec. 2017.	2015.12.01 –2017.12.31
Artificial Intelligence	July. 2017.	2015.07.01 –2017.07.31
Fintech	April. 2015.	2013.04.01 - 2015.05.31
3D printing	March. 2015.	2013.03.01 –2015.04.31

Resulting in searching for news related to the main keyword on DAUM News, a total of 242,645 news articles from 64 media outlets were searched. In this study, the news articles are crawled based on Python's Beautifulsoup4 library. An open source package for web crawling is also provided by the R program, but this study used Python, not R. It is the reason that python is better at processing web crawls. Thus, the researchers used Python code written by themselves to collect news articles by sing the URL of the web document found at the DAUM News homepage.

B. DATA PREPROCESSING

In the data preprocessing step, words and phrases that are unnecessary for analysis are deleted from the collected text data, and a normalization operation of the same and similar but differently expressed words is repeatedly performed. In this study, various special symbols (e.g., special characters, etc.) and stop-words that do not need have meanings are removed from the text data collected through the web crawling method. After that, the same meanings but various abbreviations are unified with the same word. In the normalization work, the researcher wrote the R program code and performed the preprocessing work.

IV. ANALYSIS OF OVERALL RESULTS

A. Analysis of Network Structure

The calculated values through the NodeXL program are shown in <Table 3>. A total of 96 nodes were gathered in the network, which is connected by 124 links, of which 16 are redundantly connected. The density, which means the degree of connectivity between individuals in the network, was found to be 0.014473684. The larger the value of the density index of the network, the more active the information exchange between individuals in the network, which can be interpreted as the active exchange of information between entities (Burt 1992; Rowley, 1997). Network density values were derived from 0 to 1, and the analysis showed that the values were significantly lower. This shows that the exchanges are concentrated around few specific key words, rather than organically connected and exchanged between words about the future technology of information society.

Table 3 Network statistic value

Graph Metric	Value
Vertices	96
Unique Edges	124
Edges With Duplicates	16
Total Edges	140
Self-Loops	0
Reciprocated Vertex Pair Ratio	0.023255814
Reciprocated Edge Ratio	0.045454545
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	96
Maximum Edges in a Connected Component	140
Maximum Geodesic Distance (Diameter)	8
Average Geodesic Distance	4.392144
Graph Density	0.014473684
Modularity	0.611429
NodeXL Version	1.0.1.409

※ NodeXL pro

B. Analysis of Powerful Words in the Network

The centrality (degree, betweenness, closeness, power) analysis was conducted to identify the entities that have the greatest influence in the network and those who act as mediators (gatekeepers, opinion leaders) in the network. The centrality (degree, betweenness, closeness, power) analysis was conducted to identify the entities that have the greatest influence in the network and those who act as mediators (gatekeepers, opinion leaders) in the network. <Table 4> shows the high results based on the degree centrality of future technologies related to information society. <Table 4> shows the top 20 words with high degree centrality values. In the result, representative words related to technology are derived from various field such as technology, automobile, service, virtual, transaction, finance, education and security.

The high degree centrality means a word that is frequently used together with other representative words in a main technology with high word correlation, which can be regarded as a key word with high possibility of convergence with other technologies.

As a result of analyzing the betweenness centrality, it was seen that security, technology, service, prospect, reality, convergence, and finance have high values. On the other hand, words such as development, car and trading were relatively low. Compared to the results of degree centrality, the words of betweenness centrality in the top 20 have relatively high values such as prospect (1353.494), reality (1142.244), convergence (1136.659), analysis (1051.899), and autonomy (952.270). This means that it can serve as a linkage to integrate with existing core keywords. From this point of view, words such as security (1952.721), technology (1506.158), and service (1483.697) can be interpreted as mediators like gatekeepers or opinion leaders who exchange and interchange various information within the network. Predominance is an extension of degree centrality. It takes into account the centrality of directly connected entities as well as indirectly linked entities, so that words with overall influence in the network can be identified. Assuming that words with high centrality values affect other words (Park et al., 2015), a user with a high centrality can be regarded as the most influential core user. Words such as hypothetical (0.039), trading (0.037), technology (0.035), future (0.030), finance (0.030), and hacking (0.030) can be interpreted as having high effects. Lastly, closeness centrality refers to a word with high influence to access other words. The closeness centrality of words such as technology, security, service, education, and future have similar values to degree centrality. The closeness centrality of words such as research (0.0051), prospect (0.0048), fusion (0.0047), analysis (0.0042), and hack (0.0039) showed high values.

Table 4 Result of centrality (Top 20)

No	Core Keyword	Deg.	B.C.	C.C.	E.C.
1	Technology	7	1506.158	0.0054	0.035
2	Automobile	7	814.510	0.0027	0.027
3	Service	7	1483.697	0.0049	0.018
4	Virtual	6	940.760	0.0028	0.039
5	Deal	6	541.413	0.0033	0.037
6	Finance	6	1052.427	0.0037	0.030
7	Education	6	1017.316	0.0046	0.023
8	Security	6	1952.721	0.0052	0.020
9	Development	6	875.137	0.0027	0.016
10	Future	5	937.428	0.0044	0.030
11	Innovation	5	799.829	0.0034	0.023
12	Reality	5	1142.244	0.0027	0.023
13	Automatic	5	952.270	0.0028	0.021
14	View	5	1353.494	0.0048	0.020
15	News	5	853.157	0.0027	0.018
16	Fusion	5	1136.659	0.0047	0.016
17	Experience	5	481.829	0.0026	0.015
18	Broadcast	5	647.791	0.0024	0.012
19	Safety	5	701.881	0.0025	0.009
20	Hacking	4	775.170	0.0039	0.030

※ Deg.: Degree
B.C.: Betweenness Centrality
C.C.: Closeness Centrality
E.C.: Eigenvector Centrality

Cluster analysis was performed to identify more significant relationships in the network analysis results. Cluster analysis is a method of exploring meaningful structures and patterns without prior knowledge of the contents of the data. It is suitable for visually indicating which words belong to the same group, so it is easy to understand the structural relationship between clusters. Looking at the clustering results, it was confirmed that a total of 10 clusters are formed. Nine clusters created by applying the Parallel Nearest Neighbor Clustering (PNNC) algorithm were applied to the network map as shown in <Fig 1>, and classified into the color and shape of the nodes included in each cluster. Each cluster contains as few as four and as many as fifteen key words, with an average of nine, and the results of each cluster are shown in <Table 5>.

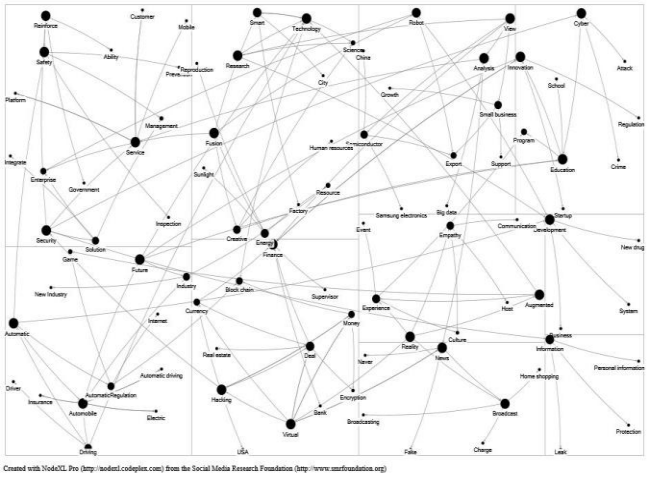

Figure 1 Clustering network

Table 5 Clustering node and list

Cluster	No of nodes	Keyword category
C1	15	Security, Service, Reinforce, Government, Ability, Solution, Integrate, Safety, Management, Enterprise, Mobile, Platform, Customer, Prevention, Inspection
C2	13	Future, Automobile, Automatic, Game, Self-regulation, Internet, Driving, Driver, Automatic driving, Insurance, Electric, Industry, New industry
C3	13	Fusion, Technology, Smart, Factory, City, Energy, Sunlight, Reproduction, Creative, Human resources, Resource, Science, Research
C4	12	Finance, Virtual, Bank, Supervisor, Hacking, Currency, USA, Deal, Real estate, Encryption, Money, Block chain
C5	11	View, Robot, Export, Analysis, Big data, Semiconductor, China, Samsung electronics, Small business, Growth, Support
C6	9	Education, Innovation, Cyber, Crime, Attack, Program, Startup, Regulation, School
C7	8	Reality, Augmented, Empathy, Host, Communication, Culture, Experience, Event
C8	7	News, Broadcast, Home shopping, Charge, Broadcasting, Naver, Fake
C9	4	Development, System, New drug, Business
C10	4	Information, Personal information, Protection, Leak

V. CONCLUSION

This study has identified the areas and the factors and patterns that the innovative technologies are aiming for based on the major technologies that have recently become issues in the information society, and predicted Black Swan in the future technology environment based on these results. As a result of social network analysis, the most influential words in the network are words such as 'technology', 'car', 'service', 'virtual', 'transaction', 'financial', 'education', and 'security', which show a high degree of Degree Centrality. Technologies which are connected with these words can be regarded as keywords with a high possibility of fusion with other technologies. As a result of performing cluster analysis to understand the structure and pattern of the network in detail, it was confirmed that a total of 10 clusters were formed. This shows that there is a high interest in security technology, and it can be inferred that security-related matters are important for future major technologies of the information society. Considering the results of this analysis, it is recognized that the interest and importance of security technology is increased, and related services and solutions need to be strengthened to manage the protection and leakage of personal information, and energy convergence and smart technology convergence are major factors along with the importance of security. In addition, as the future technology, it was confirmed that the innovative education and industry development through starting a business or loosening of financial regulations are needed along with the technologies such as autonomous driving and others related to automobiles. With the rapid pace of technological evolution around the world, high value-added technologies such as autonomous vehicles and artificial intelligence are commercialized. Related industries are where manufacturing, service, and information and communication industries can all be tied together, and they have a great influence on related industries such as insurance, finance, and repair. Therefore, it can be expected that the automobile industry and related fields will play a major role in the 4th industrial revolution. Big data analysis is still important, and it is prospected that SME-led growth will be required along with semiconductor and robot technologies. Global IT companies are restructuring and empowering their organizations to focus on data analytics to dominate the market in advance and lead the big data market. Also, they are preparing for the big data market by developing a large data processing technology for collecting big data information. In this study, it is possible to guarantee objectivity because this study analyzes news data that excludes subjectivity of researchers, compared to surveys, scenario techniques, and interviews. So, this study can be used as a basis for predicting future technology and establishing strategies and policies of related industries.

REFERENCES

1. C. Batini, C. Cappiello, C. Francalanci, A. Maurino, and G. Viscusi. (2011, August). "A capacity and value based model for data architectures adopting integration technologies," In AMCIS.
2. C. Bizer, P. Boncz, M. L. Brodie, and O. Erling. (2012). "The meaningful use of big data: four perspectives—four challenges," ACM Sigmod Record, 40(4), 56-60.
3. D. H. Oh, S. M. E. Rahman, J. M. Kim, and G. J. Bahk. (2009). "The statistics probability analysis of pork-cutting processing conditions for microbial risk assessment," Journal of Food Hygiene and Safety, 24(1), 63-68.
4. D. Paranyushkin. (2011). "Identifying the pathways for meaning circulation using text network analysis," Nodus Labs, 26.
5. Gartner (2017). "2016 Annual Report".
6. J. C. J. Dos Rieis, F. B. de Souza, P. O. S. V. de Melo, R. O. Prates, H. Kwak, and J. An. (2015, April). "Breaking the news: First impressions matter on online news," In Ninth International AAAI conference on web and social media.
7. K. Y. Kwahk. (2017). Social Network Analysis, 2nd ed., Cheongram, Seoul.
8. L. C. Freeman. (2017). Research methods in social network analysis, Routledge.
9. M. E. Newman, and M. Girvan. (2004). "Finding and evaluating community structure in networks," Physical review E, 69(2), 026113.
10. M. G. Burt. (1992). "The justification for applying the effective-mass approximation to microstructures," Journal of Physics: Condensed Matter, 4(32), 6651.
11. OECD Science (2016). "Technology and Innovation Outlook 2016".
12. Scott, R. A. (2017). The making of blind men, Routledge.
13. S. J. Park, Y. S. Lim, and H. W. Park. (2015). "Comparing Twitter and YouTube networks in information diffusion: The case of the "Occupy Wall Street" movement", Technological Forecasting and Social Change, 95, 208-217.
14. Rowley, T. J. (1997). "Moving beyond dyadic ties: A network theory of stakeholder influences," Academy of management Review, 22(4), 887-910.
15. V. Coates, M. Farooque, R. Klavans, K. Lapid, H. A. Linstone, C. Pistorius, and A. L. Porte. (2001). "On the future of technological forecasting," Technological forecasting and social change, 67(1), 1-17.
16. Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8), Cambridge university press.
17. WEF. (2016). "The Future of Jobs".

AUTHORS PROFILE

Dr. Jun Hwan Moon, is working as Senior Research Engineer in Sogang University. He has received Ph.D. degree in Management Information System from Sogang University, Korea. He holds a M.S. in Global Service Management from Sogang University, Korea. His research interests include data science, data mining, knowledge management, business model and consumer behavior in online.

Jin Hwa Kim, is Full Professor of Management Information Systems, Sogang University, Korea. He received Ph.D. (MIS) from University of Wisconsin-Madison (2001). His research interests include bigdata, CRM, Simulation of Human Learning.