Asia Pacific Journal of Information Systems
Vol. 29 No. 4 (December 2019), 817-837

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2019.29.4.817

Special Issue: Data Analytics in Artificial Intelligence Era

Detecting Knowledge structures in Artificial Intelligence and Medical Healthcare with text mining

Hyun-A Lim[a], Pham Duong Thuy Vy[b], Jaewon Choi[c,*]

[a] M.S. Student Department of Business Administration, Global Business School, Soonchunhyang University, Korea
[b] M.S. Student Department of Business Administration, Global Business School, Soonchunhyang University, Korea
[c] Assistant Professor, Department of Business Administration, Global Business School, Soonchunhyang University, Korea

A B S T R A C T

The medical industry is rapidly evolving into a combination of artificial intelligence (AI) and ICT technology, such as mobile health, wireless medical, telemedicine and precision medical care. Medical artificial intelligence can be diagnosed and treated, and autonomous surgical robots can be operated. For smart medical services, data such as medical information and personal medical information are needed. AI is being developed to integrate with companies such as Google, Facebook, IBM and others in the health care field. Telemedicine services are also becoming available. However, security issues of medical information for smart medical industry are becoming important. It can have a devastating impact on life through hacking of medical devices through vulnerable areas. Research on medical information is proceeding on the necessity of privacy and privacy protection. However, there is a lack of research on the practical measures for protecting medical information and the seriousness of security threats. Therefore, in this study, we want to confirm the research trend by collecting data related to medical information in recent 5 years. In this study, smart medical related papers from 2014 to 2018 were collected using smart medical topics, and the medical information papers were rearranged based on this. Research trend analysis uses topic modeling technique for topic information. The result constructs topic network based on relation of topics and grasps main trend through topic.

Keywords: Smart Medicine, Medical Information, Healthcare, Topic Modeling, Research Trend Analysis

Ⅰ. Introduction

Artificial intelligence (AI) technology can be used widely in medicine, finance, manufacturing, service, education (Swaraja, 2019). The medical industry is rapidly developing into telemedicine, mobile health, EHR·EMR·PHR, wireless medical field and precision medical care through the fusion of artificial in-

*Corresponding Author. E-mail: jaewonchoi@sch.ac.kr Tel: 82415301240

telligence technology and ICT (Zawadzki and Krzysztof, 2016). The combination of medical and healthcare with artificial intelligence improves the accuracy of diagnosis. This combination utilizes it as a new medical technology or medical assistant (Yu and Kim, 2019). Recently, As the environment changes due to elderly society and economic industry convergence, the demand for smart medical services is increasing. Smart medical healthcare has become possible to diagnose and treat through AI robot, and autonomous surgical robot has appeared. Also, it is possible to predict drug interactions, prescribe and dispense. In other words, smart medical means intelligent service based on medical information in the form of combining medical, artificial intelligence and ICT technology. Smart medical provide healthcare services such as preventive and state-of-the-disease diagnosis, diagnosis, treatment, and health care anytime and anywhere. Smart medical care is divided into technologies for measuring, analyzing and managing medical information, including devices and platforms for providing services. Major technologies of Smart medical include Big Data and Artificial Intelligence, Internet of Things (IoT), Virtual Reality and Augmented Reality.

Research on smart medical healthcare has been actively conducted on data, such as providing personal health record data and digital health solutions (Wissam et al., 2018). Google, Facebook, and IBM have developed a learning and reasoning-based artificial intelligence platform that integrates video, voice, and text into the healthcare field. Babylon Healthcare utilizes an artificial intelligence chatbot to collect personal health record data from users and provide customized counseling services to help patients cope with emergencies and connect doctors who meet the symptoms. However, as smart medical information can be collected and shared anytime and anywhere, problems related to medical information security are occurring. Information protection is very important because medical information contains personal information.

The medical and healthcare industry provide smart services through intelligent medical platforms, but there are healthcare hacking issues through vulnerable areas. In smart medical healthcare, malignant code found on a MRI CD in a hospital can be found, and malignant code in software embedded in medical equipment, such as pacemakers, can be life threatening (Chen et al., 2017). The Identity Theft Resource Center, a study of US personal information leak cases, reported that data breaches in healthcare and medical sector accounted for about 34.5 percent of the total in 2018. The number of healthcare data thefts and outbreaks has increased by about 400% since 2005. Hacking of medical device has a life-threatening effect on an individual's life and causes personal medical information security problems (Lee and Wu, 2017). The security of smart medical healthcare threats occur such as devices, networks, platform services, and healthcare information (Kim and Han, 2018). Therefore, it is necessary to study security technology to build security system quickly in smart medical.

Smart medical is focusing on the implementation of artificial intelligence robotic medical devices or medical applications of intelligent robotic platforms (Silverman et al., 2015). Medical information regulatory and information policy trends and the privacy of personal medical information (Zahid et al., 2019). However, in the field of medical information, only research on the necessity of privacy and personal information protection is under way (Price and Cohen, 2019), it is difficult to find research on the protection of medical information. Analysis of medical information protection trends will provide opportunities for creating new value by utilizing the tech-

nologies that are emerging at present. It can also be used as a resource for solving social issues in medical information hacking and forecasting changes in future society.

In this study, we collected medical information papers and examined the research trends related to medical information protection. The purpose of this study is to confirm the research on information security in the field of smart medical information and to suggest the direction of academic research by comparing research topics on information security. Medical information related papers were collected from 2014 to 2018 using topics related to medical information. The research trend analysis used text mining technique for topic information of theses, constructed a topic network based on the relation of topics. It is possible to identify the main trends through topics (Benites-Lazaro et al., 2018).

Ⅱ. Literature Review

2.1. Smart Medical Healthcare

As population aging rapidly progresses worldwide, the incidence of chronic and serious diseases increases with the increasing variety of serious diseases that increase the patient's burden on treatment expenses. In addition, quality improvement and expectations for treatment methods are expected to increase steadily in terms of medical demand. Recently, along with the development of the 4th industrial revolution, artificial intelligence (AI), the combination and application of information technology to medical equipment, robotics are expected to be more innovative and effective in population aging solving problem (Tawalbeh et al., 2016). That also being expected as a solution. Especially, it focuses on AI and robot technology, and shows a trend that is being utilized variously in medical field. Recently, in domestic enterprises, universities and research institutes, research and development of robotics medical equipment combine with AI technology has been actively carried out by paying attention to the advantages of AI and robotic technology (Dhukaram and Baber, 2013). One of the remarkable achievements of the application of robotic and AI is virtual nursing assistant, can save nurses from unnecessarily having to visit patients and can keep patients from unnecessarily visiting the hospital (Yu and Kim, 2019).

In recent years, with the advent of the smartization era, smart healthcare, which is an intelligent healthcare service customized for users (patients, healthcare people, etc.) has been emerging (Alma et al., 2017). Besides the continuous development of artificial intelligence, the applications of AI on robot technology have also developed, in which typically must be mentioned are: surgical assistant, rehabilitation robots, telepresence robot, etc. u-Healthcare service is an evolving model of mobile medical service (Goozee and Rhianna, 2013). It eliminates spatial and temporal constraints and provides centralized remote medical service system based on biometric information and environmental information collected from various medical sensors and devices in living space (Yiwen et al., 2018).

New technologies have influenced many parts of our daily life. Today's healthcare system has also recognized the advantages of using Information and Communication Technology (ICT) to improve the quality of healthcare, turning traditional into smart healthcare. Smart Medicine is defined by the technology that leads to better diagnostic tools, better treatment for patients, and devices that improve the quality of life for anyone and everyone (Karim et al., 2018). The key concept of smart medical healthcare

includes eHealth and mHealth services, electronic record management, smart home services and intelligent and connected medical devices. Smart medical healthcare provides solutions for personal healthcare, such as personal health devices and personal health applications (Umakanth and Estefania, 2018).

The area of health in recent years has been rapidly integrating technology in the monitoring, diagnosis and treatment of patients remotely. Thus, achieving to improve the quality of life of patients and greater traceability of information from them. Recently study discusses the concept of networked healthcare and its enablement through the mobile cloud computing and big data analytics technologies (Umakanth and Estefania, 2018). The motivation and development of networked healthcare applications and systems is presented along with the adoption of cloud computing in healthcare. A cloudlet – based mobile cloud computing infrastructure to be used for healthcare big data applications is described (Alma et al., 2017).

The EU's MobiHealth (Mobile Healthcare) project is aimed at high-risk pregnant women, patients with chronic illnesses, and those with heart disease, and provides platforms and services to provide services such as disease diagnosis and prediction and emergency response through continuous patient monitoring in daily life (Yu and Kim, 2019). Recent developments of wireless and mobile technology have laid a foundation for the new generation of wireless intelligent sensors (Alma et al., 2017). This same technology has made the implementation of intelligent medical monitors that can provide real-time feedback to the patient feasible, either as a warning of impending medical emergency or as a monitoring aid during exercise. Intelligent medical monitors can significantly decrease the number of hospitalizations and nursing visits by acting as a personal "guardian angel" that can warn the user of a medical emergency or contact a specialized medical response service (Goozee and Rhianna, 2013).

2.2. Information of Smart Medical Healthcare

In recent years, medical institutions are establishing medical information systems, and they are digitizing all processes and records such as patient's reception, examination, prescription, treatment, hospitalization, discharge, and test results (Chen et al., 2018). The medical information system is composed of homepage, OCS (Order Communication System), EMR (Electronic Medical Record), PACS (Picture Archiving Communication System), PHR (Personal Health Record). The homepage provides services such as patient's online check-up appointment and hospital guide. OCS is a prescription delivery system that delivers the prescription results created by the medical staff after the patient's treatment to the respective departments (medicines, raw materials, etc.). EMR is an electronic medical record that records patient's medical record information in electronic form rather than paper. PCAS is a system for transmitting and storing patient imaging information such as X-ray, CT, PHR is a personal health record system that stores personal health information in electronic form (Goldschmidt and Peter, 2005).

Medical information systems are connected to devices in various hospitals, mainly medical devices and supported devices (Hong and Kim, 2017). Medical devices are products classified by the Medical Law and are products intended for the diagnosis and treatment of diseases and ascending. This is partly due to the development of IT technology, which is classified as a Smart Medical Device, which is connected to systems in the hospital via Wi-Fi and Bluetooth communication (Dimitrov, 2016). Supported devices are medical service support devices that are

provided to improve patient information and convenience, such as electronic signatures, issuance of medical waiting lists, and hospital information (Javanov et al., 2003).

In smart medical healthcare, medical information plays an important role. It is possible to provide various services and develop medical robots based on medical information. However, there are problems such as institutional problems of hospitals that provide medical information and infringement of personal privacy (Alejandra et al., 2019; Appelbaum, 2000; Ross, 1996; Smith and Eloff, 1999). Research is needed to protect medical information beyond technology and smart medical healthcare services to utilize medical information.

2.3. Information Security for Smart Medicial Healthcare

EMR, which has been managed in its own form in the hospital and clinic, has recently developed into the concept of EHR, the personal electronic health record of life (Wang et al., 2018). Sharing of electronic records and prescription information between hospitals, reflecting the need for ownership and management of patient's own mandatory data, and HER intelligent and EHRS for high-quality medical services are required (Kim and Chang, 2018) In addition, EHR and EHRS are currently conducting standardization of data sharing, medical terminology unification, messaging method and interface between medical institutions to support interoperability and connection. In addition, there are some issues and researches on security threat factors that prevent activation of EHR market such as safe keeping and exchange of EHR data and protecting personal privacy of patients (Gollakota, 2011).

The establishment of advanced medical information system poses many risks in terms of protection personal information despite advantages for improving public health. Since medical information is professional and public information related to the people, special protection is required among personal information, and digital information is easy to reproduce (Lee and Wu, 2017). In a certain aspect, medical information theft could be much more dangerous than credit card data theft. Because credit card data theft can be easily recognized and reported to banks, which can act immediately. While medical information theft, on the other hand, which is usually difficult to be recognized by the subjectivity of the victim.

The sharing as well as distribution of patient information enables productive medical research, proper treatment of patients, and improvement in health care quality. On the other hand, electronic patient records pose a challenge to maintain information confidentiality, integrity, and availability (Halperin et al., 2008). The computerized record infers that the requester has the same hardware and software communication protocol and thus enables easy access to data (Sweeney, 2002). With the advent of electronic patient records, patients' data can easily be shared among physicians, health care providers, nurses, supporting staff, medical research, and public health care services (Williams, 2008). Basically, most health care information including patients' data is not generated solely within a physician and patient relationship, but is generated from the diverse sources (Jin et al., 2013). This may open doors to the unauthorized parties who may unscrupulously steal patients' data for personal benefits, alter patients' records, and expose patients' medical history. In other words, the high accessibility of patients' data has made it easier for perpetrators to invade patients' information confidentiality, integrity, and availability and commit

health care fraud (William, 2008).

Medical identity theft can cause difficulties like lost insurance coverage, mixed up records, higher premiums, medical harm and false diagnosis. But the reason healthcare data hacking exists is that there's a lucrative market for the data, and a working supply-and-demand model with both buyers and sellers. Centralized storage make the central node exiting a high risk. Once the system is invaded, it can lead to disclosure of enormous patient's privacy information and the theft or illegal use of electronic medical records (Wang et al., 2018). Therefore, protecting patient's medical records should be every hospital's and physician's office concern. But with many issues in the healthcare industry vying for attention, security may fall through the cracks.

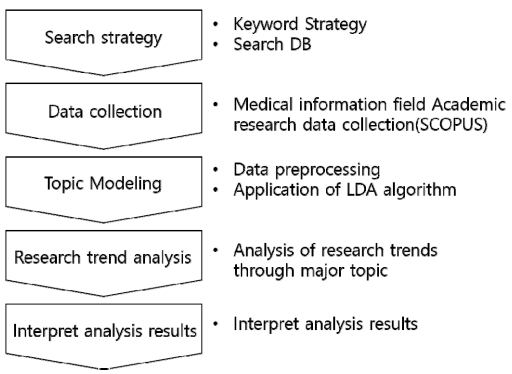
It will be much easier to maintain legal system if technological security for medical information protection is enabled and abuse of medical information can be reduced. In terms of technology, medical information should be equipped with systems such as encryption level, processing method, and processing method of medical information according to access control and use authority of medical records through user identification. In order to protect medical information from this point of view, joint research among various disciplines should be preceded (Huang et al., 2009).

The development of a security policy remains a key step in the process of securing the system. Unfortunately, existing policies are unable to cover all the security needs of medical information systems. Moreover, in practice, the absence of a security policy in the majority of current health information systems (passwords shared by several users, arbitrary use and without prior study of security mechanisms, etc). The security requirements of our system must therefore be defined in a very precise way and new methods must be devised to apprehend them (Lee and Wu, 2017). Information-access control: Current research on information access has primarily focused on technological solutions. There are very few economic studies that offer deeper insights on managing information-access control in a cost-effective manner. Healthcare organisations must invest in many information security measures, such as access control systems, intrusion detection systems, policies and personnel (Gollakota et al., 2011). The iOS and Android App Stores offer a wide selection of mHealth apps. Analysis of rating counts indicates, however, that less than a quarter of available apps are in more or less widespread use. An issue impeding app dissemination might be users' information security and privacy concerns (Kai et al., 2018).

Ⅲ. Research Methodology

3.1. Data Collection

This study was conducted based on the following analysis procedure <Figure 1>. We selected smart medical healthcare keywords to collect objective data suitable for analysis purposes. The SPRINGER website collected data based on smart medical health-



<Figure 1> Research Framework

care-related keywords. We analyzed the basic trends of smart medical healthcare research trends, conducted data preprocessing for topic modeling analysis, and applied the LDA model. Finally, we analyzed the trends of smart medical healthcare research and suggested the implications of the results.

The analysis data for this study was collected on the Springer website. We chose the data collection period from 2014 to 2018. Smart medical healthcare keywords are 'smart medicine', 'smart health care', 'telemedicine', 'medical robot' and 'AI medical care'. We collected all articles with keywords in the title, author keywords, and abstracts. From 2014 to 2018, 1,517 cases, 1,611 cases, 2,014 cases, 2,189 cases and 3,297 cases were collected. The paper with matching author names and keywords were considered the same paper. Finally, 10,562 articles were collected by removing duplicate articles. The collected data is organized by title, author, date, keywords, journal, and date.

3.2. Topic Modeling

Topic Modeling is an advanced technique of text mining, which is a technique for identifying topics that are latent in text. Topic modeling is an analytic technique that extracts information from the text of a document and creates a model to search for meaning that could not be found using existing keyword analysis alone (Blei, 2012). Topic modeling refers to the process of finding a topic through words in documents and documents using algorithms to find themes in a large unstructured collection of documents. Topic modeling, a technique used in text mining, allows you to find patterns in a corpus and search for clusters of words in topics in many documents. Topic modeling mainly uses two models. The first is a latent semantic analysis (LSA), which assigns vector values to segmented words and reduces the size of the words to close them. Second, a latent dirichlet allocation (LDA) can be used to estimate the probability that a word exists on a certain topic and the probability that a certain topic will exist based on the probability (Blei, 2012). In recent years, the LDA method, which complements the disadvantages of LSI and pLSA, has been widely used in research to find the topic of the document. In this study, topic modeling was performed using the LDA model.

The LDA algorithm is a generation model that finds hidden topics in the literature (Blei, 2003). The creation model is a process of creating actual documents. In order to create a report document, it is necessary to model which topics are to be included in each document, and which words are to be selected and placed in each topic (Blei, 2012). Therefore, it is aimed to infer hidden variables such as the ratio of the subject and the assignment of the subject by the word through observed variables such as documents and words, and as a result, Rate, and distribution of words to be included in each topic. Topic modeling is mainly used in research to extract new information from unstructured text documents, and it is proposed as an effective means to identify topics and trends in specific fields (Cho et al., 2017).

In order to identify the words (tokens) in the document to be analyzed as frequency-based, we aggregated the entire article data and recognized them as vectors for computing words in individual documents (Blei, 2006). Based on this, a Term-Document Matrix was created that calculates the frequency of words appearing in each document. In the analysis result, the complicated text form is refined in order to reduce the noise generation. In general, it is performed by applying a predetermined code to the tasks such as unification in small letters, elimination

of insolence words, extraction of roots, elimination of numbers, removal of spaces and removal of punctuation marks. The application of the LDA algorithm begins by constructing a term-matrix that constructs a matrix based on the collected documents, eliminating word order, sentence, and grammar structure and finally based on the frequency of words and words (Benites et al., 2018; Blei, 2012). The final stages of topic modeling are the definition and classification of derived topics. It means interpretation of n words with high probability of appearing for each subject derived through topic modeling and make definition of the subject. Finally, it is possible to apply the finally derived topics to experts and field workers through validation It is necessary to carry out the process of refining.

In topic modeling, the number of topics is usually determined by the researcher (Benites et al., 2018; Blei, 2012). That is, the researcher designates the number of topics to be extracted by performing a number of preliminary tests. This is very important because it affects the interpretability of the results of the study. If you specify the number of topics, all the words in the data or document are automatically calculated in the probability of belonging to each topic and arranged in ascending order of value (Blei, 2006). Key words arranged in each topic are automatically presented, and they are used to specify the topic. Because this automated analysis of results is possible, topic modeling has the advantage that it can quickly and clearly identify semantic relationships for large amounts of data without prior knowledge of the researcher, word classification (labeling), or manual coding (Blei, 2012).

In this study, a total of 10,562 papers were used to identify trends in smart medical healthcare research. The thesis words were extracted by only nouns, and 37,374 words were removed by topic modeling by removing the etymologies of et and al. The number of topics is changed from 5 to 15 in order to determine the appropriate number of topics, and the results of the perplexity function provided by the LDA analysis algorithm, the interpretability of the topic, and the meaningfulness of the topic are determined.

3.3. Claume Newman Moore (CNM) Algorithm

The Claume Newman Moore (CNM) algorithm is based on the Girvan and Newman (GN) algorithms and is a new algorithm that increases the efficiency of GN algorithms (Clauset et al., 2004). The Girvan and Newman algorithms propose a modularity classification scheme that describes how to distinguish specific groups in a network (Girvan and Newman, 2002). It is a useful method for grasping the structure of communities in a network by quantitatively grasping the structural characteristics of links. It utilizes the Betweenness centrality to determine how many points a certain point (k) appears in the shortest distance of the network and confirms the shortest distance route. Thereafter, the frequency of the node k appearing therein is examined and numerically represented. Investigate the nodes and connections that serve as a bridge between the communities through the inter-centicity (Girvan and Newman, 2002). The higher the Betweenness centrality, the more important it is in the inter-community network. Calculate the modularity for each group that breaks one connection between the communities. Finally, the modularity point having the largest value is confirmed as having the best community detection.

The CNM algorithm is a 'greedy' algorithm that maximizes 'modularity' and is a greedy modularity-based algorithm (Han and Wang, 2010). Each node begins to compose a community and aggregates

small clusters together until they have the best fit (Newman, 2004). The Modularty algorithm finds the community structure within the network. In general, the community structure is a set of relatively densely connected nodes in the network (Han and wang, 2010). Modulartiy means the degree to which the links connected to the community of each community are connected to the links connected to the outside. All nodes are assumed to have their own community structure, and community numbers are assigned to the number of nodes. We combine two communities into one community and calculate the variation of modularity Q (Clauset et al., 2004). For all links, the above procedure is performed to construct a community by connecting the communities with the largest variation of Q. Finally, it is repeated until the community of the entire network becomes one (Clauset et al., 2004). And finds the group with the maximum modularity value calculated by the CNM algorithm (Han and wang, 2010). In this study, topics of research related to medical information security are classified through topic modeling and sub-groups are classified through CNM algorithm.

Ⅳ. Data analysis and results

4.1. The trend of Smart Medical Healthcare Research

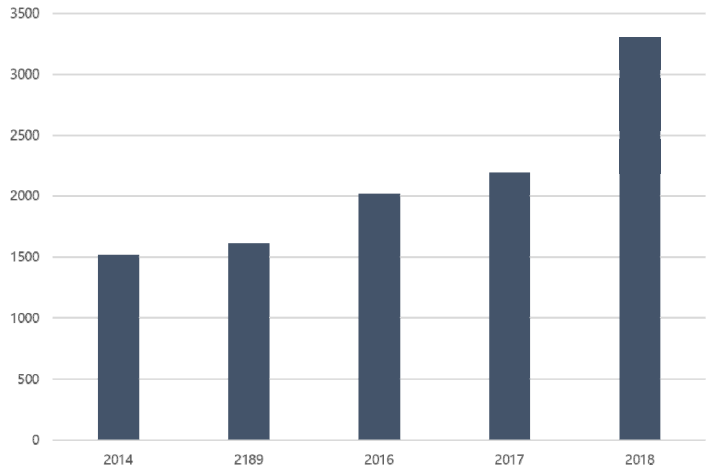
The <Figure 2> shows trends in the field of smart medical healthcare published from 2014 to 2018, and it is confirmed that it has been gradually increasing over the last five years. The number of papers collected was 1517, 1611, 2014, 2189, and 3297 from 2014 to 2018.

The Keyword Trends in smart medical healthcare

The keywords frequency was confirmed by collecting the top 10 words based on frequency. Mainly used keywords are 'datum' and so on. The most frequently used keywords were datum, health, system, information, patient, model, care, method, analysis, technology etc. <Table 1> shows the top 10 words in frequency.

Smart medicine Topic modeling

In this study, five topics were analyzed to analyze



<Figure 2> Increasing the Number of Smart Medical Healthcare Research by Year

<Table 1> Kewords Frequency

Rank	Word	Frequency
1	datum	16,174
2	health	12,107
3	system	9,824
4	information	9,281
5	patient	7,490
6	model	5,946
7	care	5,170
8	method	5,118
9	analysis	4,849
10	technology	4,294

the topic modeling results. We used title, keyword, and abstract for topic analysis. The final stage of the topic modeling analysis is to determine a topic that represents each topic group based on the derived topic. The combination of words derived through topic modeling was profiled as a representative subject of academic research in the field of smart medical healthcare as follows.

In Topic 1, words such as health, patient, datum, record, hospital, disease, and treatment were derived and confirmed that topics related to hospital records in smart medicine care were extracted. In Topic 2, words such as datum, technology, healthcare, service, application, information, data, and device were derived, and it was confirmed that topics related to health care service of smart medicine were extracted. Topic 3 was derived from the words robot, intelligence, AI, science, interaction, and person. In topic 4, words such as information, security, privacy, user, system, and problem were derived, which confirmed that topics on the security of smart medicine were derived. In Topic 5, words such as algorithm, method, model, datum, image, classification, and technique were derived, and it was confirmed that topics about smart medical healthcare technology and algorithm were derived. Therefore, Topic 1 was chosen as the hospital record, Topic 2 as health care, Topic 3 as the smart medical robot, Topic 4 as the security of smart medical healthcare, and Topic 5 as the smart medical healthcare technology (algorithm). A total of 10,562 papers were categorized as 2,646 of topic 1, 2,490 of topic 2, 1,002 of topic 3, 1,894 of topic 4, and 2,530 of topic 5.

4.2. Smart Medicine Information Research Trends

Information related topics were detected multiple in smart medical healthcare papers. Therefore, in the case of smart medicine related papers, the papers containing the words 'information' and 'data' in the title and keywords were rearranged. The number of medical information papers in the smart medicine field was 894 in 2014, 965 in 2015, 1,076 in 2016, 1,068 in 2017 and 1,347 in 2018. Medical information surveys decreased slightly in 2017 but increased sharply in 2018. Research on medical information is expected to be actively discussed. In this study, duplicate entries were deleted when author name and keyword were the same. Finally, 5,332 articles were used. Keyword frequency was verified by collecting the top 10 words <Table 3>.

<Table 2> Topic List of Smart Medicine

	Topic-1	Topic-2	Topic-3	Topic-4	Topic-5
1	health	datum	robot	information	method
	0.035	0.047	0.01	0.037	0.023
2	patient	system	intelligence	system	model
	0.031	0.031	0.005	0.024	0.022
3	care	technology	interaction	user	algorithm
	0.02	0.019	0.005	0.015	0.017
4	datum	health	person	knowledge	datum
	0.019	0.019	0.005	0.014	0.016
5	record	healthcare	theory	paper	network
	0.012	0.014	0.005	0.01	0.014
6	hospital	research	research	approach	image
	0.011	0.013	0.004	0.01	0.014
7	disease	information	AI	research	feature
	0.01	0.011	0.004	0.008	0.012
8	risk	service	behavior	privacy	learning
	0.009	0.009	0.004	0.006	0.012
9	year	development	life	domain	approach
	0.008	0.009	0.004	0.006	0.012
10	outcome	application	science	problem	classification
	0.008	0.009	0.004	0.006	0.01
11	factor	management	ethic	process	machine
	0.008	0.008	0.004	0.006	0.01
12	analysis	challenge	agent	use	performance
	0.007	0.008	0.003	0.006	0.009
13	treatment	Data	relationship	security	technique
	0.007	0.007	0.003	0.005	0.009
14	use	process	human	tool	problem
	0.007	0.007	0.003	0.005	0.009
15	information	device	action	design	paper
	0.007	0.006	0.003	0.005	0.008

<Table 3> Keywords Frequency of Medical Information

Rank	Word	Frequency
1	datum	12,151
2	health	11,274
3	information	7,329
4	patient	6,252
5	system	5,216
6	study	5,193
7	care	4,601
8	record	3,250
9	analysis	2,994
10	healthcare	2,830

Medical Information Topic modeling

In this study, a total of 5,332 papers were used to identify the research trends related to medical information. The words were extracted from nouns by topic modeling with 37,374 words. The title, keywords, and abstract were used for LDA topic modeling analysis. The LDA topic modeling analysis results of this study are as follows. In Topic 1, words such as intervention, practice, provider, physician, implementation, communication, participant, survey, review were extract and it was confirmed that topics related to information collection through doctors and participants were extracted. In Topic 2, words such as disease, risk, woman, age, child, association, and status were extracted and confirmed for patient information related to the disease. Topic 3 was derived

<Table 4> Topic List of Medical Information

	Topic-1	Topic-2	Topic-3	Topic-4	Topic-5
1	intervention	risk	hospital	image	technology
	0.008	0.015	0.025	0.014	0.011
2	practice	woman	cost	algorithm	application
	0.008	0.012	0.013	0.012	0.009
3	service	population	treatment	classification	service
	0.007	0.011	0.011	0.008	0.009
4	need	age	disease	feature	management
	0.007	0.011	0.01	0.007	0.008
5	provider	factor	cancer	mining	privacy
	0.006	0.011	0.01	0.007	0.008
6	physician	disease	diagnosis	technique	process
	0.006	0.009	0.008	0.007	0.007
7	implementation	level	drug	dataset	challenge
	0.006	0.008	0.008	0.006	0.007
8	communication	child	rate	performance	user
	0.006	0.008	0.008	0.006	0.007
9	group	diabetes	case	set	security
	0.006	0.008	0.007	0.005	0.006
10	factor	rate	outcome	disease	development
	0.005	0.007	0.007	0.005	0.006
11	participant	group	medication	Medical	framework
	0.005	0.007	0.006	0.005	0.006
12	outcome	effect	efficiency	learning	network
	0.005	0.007	0.006	0.005	0.005
13	survey	CI	database	text	access
	0.005	0.007	0.006	0.005	0.005
14	trial	association	admission	machine	device
	0.005	0.007	0.006	0.005	0.005
15	review	status	event	accuracy	solution
	0.005	0.006	0.006	0.005	0.005

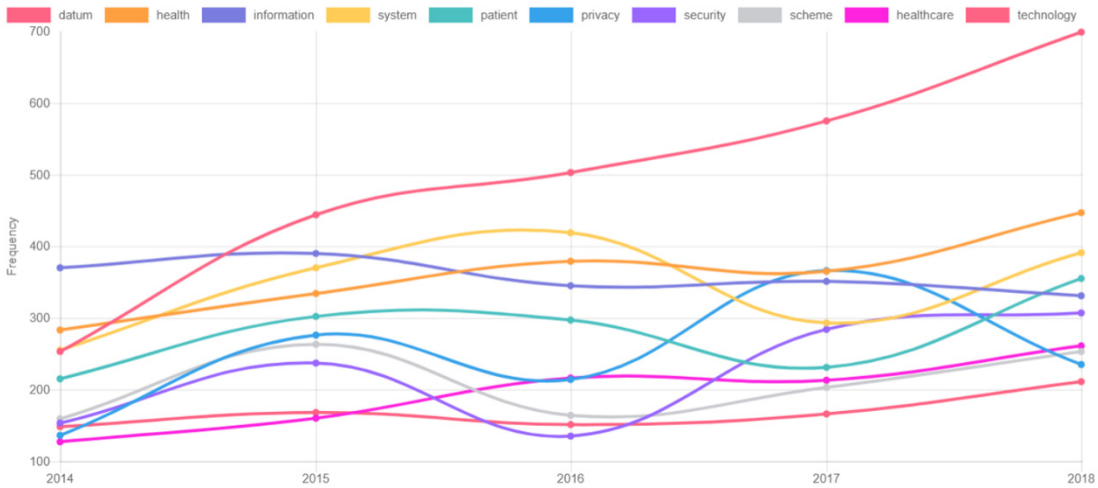
from words such as hospital, treatment, disease, cancer, diagnosis, drug, case, outcome, efficiency, database and drug, and the subject was defined as hospital treatment information. Topic 4 extracted words such as algorithm, classification, mining, technology, dataset, performance, machine, text and accuracy, and topics were defined as medical information technology. In Topic 5, words such as application, service, privacy, security, device, and access were derived and selected as a topic for health information security. Finally, the topic of topic 1 was selected as medical information collection, and a total of 1,246 documents were classified. The topic of topic 2 was selected as medical information and 849 documents were classified. The topic of topic 3 topics were selected for medical information and 805 documents were categorized. The topic of topic 4 was selected as medical information utilization technology, and 632 documents were classified. Topic 5 was selected as a topic for medical information security and 1800 documents were classified.

4.3. The Research Trend of Security for Smart Medical Healthcare

We identified keywords for medical information security targeting 949 articles from topic 5. Based on the words used in the paper on healthcare information security in the field of smart medicine, draw a word cloud in <Figure 3>. Looking at the most used words, words such as 'datum', 'security', 'health', 'technology', 'health' and 'privacy' were visually highlighted. In addition, we confirmed that the words 'application', 'device', 'network', 'care', etc. stand out. The number of papers collected was 103, 177, 199, 222, and 268 from 2014 to 2018.

In the paper on medical information security in the smart medical healthcare field, the 10 most frequently used words are as follows in <Figure 4>. In 2017, the frequency of 'system' related terms dropped, and 'privacy' and 'security' words increased rapidly. In 2018, it was confirmed that nine words except for 'information' were rising.

<Figure 3> Wordcloud for Security Issues in Smart Medical Healthcare



<Figure 4> Top Frequency Word Graph

Topic modeling for medical information security

In order to conduct research on medical information security in smart medical healthcare, it is divided into three topics. The topic results are as follows <Table 5>. Topic 1 is defined as hospital operation and insurance through extracted words such as model, management, hospital, policy, solution, implementation, and health. Topic 2 was selected as a topic related to health care service through extracted words such as scheme, authentication, network, sensor, IoT, encryption, device, security, and Internet. Topic 3 is defined as the protection of images or the digital watermarking theme extracted words such as image, watermarking, method, condition, disease, treatment, and time. Smart medical information security papers were categorized as 641 in topic 1, 182 in topic 2, and 126 in topic 3.

Applied CNM algorithm for medical information security

In this study, subcategorized by each topic were recategorized through CNM algorithm. The modularity analysis was performed during the cohesion analysis of the top 100 words extracted for each topic, and the CNM algorithm was used. The reclassified clusters for each topic were formed so that the links of the clusters were more than those of the different clusters. This means that the subject words included in the group are observed more frequently than the main group words.

As a result of the analysis, topics related to hospital information were classified into topics on patient information management and security. Topics related to healthcare were classified into topics of hospital medical information security and healthcare information using IoT and sensors. Finally, the third topic is classified as the subject of storage security technology and patient information consent.

Ⅴ. Discussion and Implications

5.1. Discussion of Findings

In this study, we used topics modeling and CNM

<Table 5> Topic List of Medical Information Security

	Topic-1		Topic-2		Topic-3	
1	model	0.006	scheme	0.051	image	0.02
2	management	0.006	authentication	0.04	disease	0.011
3	Data	0.006	attack	0.025	method	0.009
4	hospital	0.005	cloud	0.023	treatment	0.008
5	Health	0.005	protocol	0.021	family	0.006
6	risk	0.005	network	0.016	watermarking	0.006
7	policy	0.005	sensor	0.011	time	0.005
8	requirement	0.005	IoT	0.011	group	0.005
9	practice	0.005	server	0.01	cancer	0.005
10	design	0.005	encryption	0.009	condition	0.005
11	framework	0.005	device	0.009	outcome	0.005
12	quality	0.005	Security	0.009	trial	0.005
13	solution	0.005	performance	0.008	person	0.005
14	implementation	0.004	Internet	0.008	support	0.004
15	control	0.004	environment	0.007	intervention	0.004

<Table 6> CNM Result of Medical Information Security

	Topic-1		Topic-2		Topic-3	
1	EHR	Security	Ehealth	Authentication	ECG	adult
2	EMR	assessment	Encryption	RFID	Image	age
3	physician	law	IoT	Telecare	Medical	blood
4	hospital	platform	Network	biometric	Telemedicine	cancer
5	Data	policy	Security	cryptography	accuracy	case
6	professional	protection	WBAN	flaw	algorithm	change
7	provider	regulation	device	medicine	capacity	child
8	Health	solution	monitoring	password	diagnosis	consent
9	Healthcare	mHealth	sensor	protection	doctor	decision
10	Information	management	attack	Scheme	heart	decisionmaking
11	benefit	monitoring	signal	addition	hospital	disability
12	collection	case	Cloud	card	image	disease
13	cost	control	Internet	efficiency	loss	disorder
14	knowledge	database	Medical	exchange	method	drug
15	method	device	architecture	impersonation	patients	emergency

algorithms to identify trends in the security of smart medical information. we analyzed the research trends related to medical information based on academic research in ‘smart medicine’, ‘medical information’, and ‘medical information security’ applying quantitative analysis technique based on objective data. In

order to grasp the global research trend of medical information, we confirmed the trend according to the number of articles per year. As a result, the number of articles related to smart medical healthcare increased overall. The topics were identified using the words used in the title, keywords, and abstracts. In the smart medical healthcare field, many keywords related to medical information were derived. This research reorganized articles containing keywords 'information', 'data', and 'record' in title and keyword in smart medical healthcare paper and conducted topic modeling. In other words, we collected research related to smart medical healthcare paper and selected papers related to medical information. Next, topic modeling was carried out with smart medical information papers to confirm research trends on medical information security in smart medical field. The subject of topic 5 derived from medical information was confirmed to be related to security. After that, I carried out a research on medical information security again with the article classified as Topic 5. Medical information security is topic modeled and subdivided by topics using CNM algorithm.

As a result of research analysis of smart medical, it was confirmed that the frequency of words such as 'datum' and 'information' is high. Papers in smart medical healthcare were classified into 5 categories (hospital record, health care, smart robot, smart medicine security, smart medicine technology). If you check the topic classification value, you can see that the information content is high in the whole topic. Especially, it can be confirmed that the information content is high in the security domain. In general, smart medical healthcare is actively researching information. In case of medical information dissertation, information collection, patient information, hospital treatment record, medical information technology and health information security were classified. In the case of medical information, it can be confirmed that there are high issues regarding the collection of patient personal information, medical information collection, information utilization technology and security based on the hospital.

Topics in medical information security papers are divided into hospital and healthcare areas through modeling analysis results. Research on storage technologies such as images and videos and copyright are also being studied. In the detailed area, the hospital has extracted topics on patient management and security. In the case of health care, topics related to IoT, sensors, and diagnosis and treatment services have been extracted. In the third topic, we have confirmed that there are security issues related to image storage, patient information such as patient consent, personal information consent, and hereditary records.

Recently, interest in smart medical healthcare has increased rapidly, privacy and security issues for medical information are attracting attention. However, after checking the topic on medical information, it can be confirmed that the topic document about information security is the least. As a result of examining topics on medical information security, it can be confirmed that the research is divided into hospital and health care areas. It can be seen that security research on information such as image storage technology of medical information area and hospital record of patient is active. Based on the results of this study, it can be confirmed that there is little research on the solution of medical information security. This shows that research on medical information security technology is insufficient and technology and collaboration in other fields are not working smoothly.

5.2. Limitations and Future Research Directions

In order to analyze the trends of academic research related to smart medical information, scientific analysis method and objective data were used to present the results. However, there are some limitations in terms of methods and subjects of research. First, in terms of methodological aspects of the research, there is a limit that the contents of the individual research or the consideration of the research field are numerically reflected only by carrying out the quantitative analysis of the data center. In the later studies, it is expected that more qualitative results can be obtained if qualitative research on individual studies is supported. Second, it is necessary to analyze data collected by SPRINGER only and collect additional data through web sites such as Web of Science. Future research will provide more accurate results if trend analysis is conducted on the whole papers of smart medical healthcare. The most important feature of the LDA technique is to select a group of topics through the author's opinion. Future studies can be analyzed by other analytical methods to supplement these areas. The purpose of this study is to confirm research trends on medical information security in smart medical field. Therefore, we did not focus on other topics in smart medical healthcare and medical information. In future studies, each topic of medical information can be identified more intensively.

5.3. Implications for Research and Practice

This study is significant in that it conducts quantitative analysis based on academic data in order to analyze the trend of smart medical information academic research scientifically. Especially, it is significant in that the traditional frequency - based quantitative analysis method and the topic modeling method are used together for comparative analysis. It also explored the global trends in the field of smart medical information and provided meaningful information in research design and topic selection. Finally, this study was conducted to confirm the issue of information security in smart medical healthcare field. We collected papers in smart medical healthcare field and classified the documents through LDA topic modeling analysis. In order to understand the research trends on medical information security, we analyzed the topic modeling analysis network analysis to identify trends.

Ⅵ. Conclusion

In this study, using the topic modeling method looked at the subject who deal primarily in the field of smart medical healthcare. The number of study of the smart medical healthcare is increasing. As a result, we have presented five subject areas in smart medical healthcare field, five areas in smart medical healthcare information field, and six subject areas in information security of the smart medical healthcare field.

Topic modeling results of smart medical information security were divided into hospital operation and insurance (641), health care services (182) and copyright protection such as protection of images or the digital watermarking theme (126). As a result of CNM analysis of smart medical information security, Topic-1 was classified into patient information management and hospital information security, Topic-2 was classified as health care service through IoT and hospital service security, and Topic-3 was classified as security technology and personal information consent. As a result of this research, the issue of smart medical information security has been increasing, but it has been confirmed that there is

lack of research on the solution. In addition, we have confirmed that research on medical information security is focused on hospital operation and insurance. In other words, Previous medical information studies were also based on hospital records or hospital data.

In future studies, we have to expand our research scope as the smart medical healthcare industry grows. First, it is necessary to concentrate on collected medical information for the healthcare service provided through the wearable device. Second, there is also a need for research on privacy issues such as personal information consent and copyright of medical information. Finally, we need to study how to get personal information consent more easily. Because, personal medical information is required for efficient healthcare services. It is expected that these studies can be used as a means to actively exchange academic knowledge among related researchers.

Acknowledgements

This research was supported by the Soonchunhyang University Research Fund.

<References>

[1] Alejandra, T., Viviana, G., Maria, D. P. D., and Victor, L. (2019). Sensitivity analysis of longitudinal count responses: A local influence approach and application to medical data. *Journal of Applied Statistics, 46*, 1021-2042.

[2] Alma, J. A., Nicole, J. M., and Pablo, P. (2017). *Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease*. The Cochrane database of systematic reviews.

[3] Appelbaum, P. S. (2000). Threats to the confidentiality of medical records: No place to hide. *Jama-journal of the American Medical Association, 283*(6), 795-797.

[4] Benites-Lazaro, L. L., Giattia, L., and Giarollab, A. (2018). Topic modeling method for analyzing social actor discourses on climate change, energy and food security. *Energy Research & Social Science, 45*, 318-330.

[5] Blei, D. M. (2012). Probabilistic topic models. *Communication of the ACM, 55*(4), 77-84.

[6] Blei, D. M., and Lafferty, J. D. (2006). Dynamic topic models. In: *Proceedings of the 23rd International Conference on Machine Learning ACM*, 113-120.

[7] Carine, B. R., Georges, B., Amir, H. E. H., and Emmanuel, A. (2018). Medical data mining for heart diseases and the future of sequential mining in medical field. *Machine Learning Paradigms, 149*, 71-99.

[8] Chen, J., Ma, X., Mingxiao, D., and Zhuping, W. (2018). A blockchain application for medical information sharing. *2018 IEEE International Symposium*, 1-7.

[9] Chen, M., Yang, J., Hao, Y., Mao, S., and Hwang, K. (2017). A 5G cognitive system for healthcare. *Big Data and Cognitive Computing, 1*(1).

[10] Cho, K. W., Bae, S. K., and Woo, Y. W. (2017). Analysis on topic trends and topic modeling of KSHSM journal paper using text mining. *The Korean Journal of Health Service Management, 11*(4), 213-224.

[11] Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), 114-123.

[12] Dhukaram, A. V., and Baber, C. (2013). Elderly cardiac patients' medication management: Patient day-to-day needs and review of medication management system. *2013 IEEE International Conference on Healthcare Informatics*, 107-114.

[13] Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare Informatics Research, 22*(3), 156-153.

[14] Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821-7826.

[15] Goldschmidt, P. G. (2005). HIT and MIS: implications

of health information technology and medical information systems. *Communications of the ACM, 48*(10), 69-74.

[16] Gollakota, S., Hassanieh, H., Ransford, B., Katabi, D., and Fu, K. (2011). They can hear your heartbeats: Non-invasive security for implantable medical devices. *ACM SIGCOMM Computer Communication Review, 41*(4), 2-13.

[17] Goozee, R. (2013). Smart medicine: How the changing role of doctors will revolutionize health care. *Journal of Mental Health, 22*(6), 580-582.

[18] Halperin, D., Thomas, S. H., Fu, K., Tadayoshi, K., and Maisel, W. H. (2008). Security and privacy for implantable medical devices. *The Community for Technology Leaders, 17*(1).

[19] Han, H., and Wang, J. (2010). Improving CNM algorithm to detect community structures of weighted network. *Computing Technology Institute, 46*(35), 86-89.

[20] Hong, H. K., and Kim, S. H. (2017). Structured design of healthcare system based on mobile to improve the quality of life for the elderly people. *The Society of Convergence Knowledge Transactions, 5*(2),79-83.

[21] Huang, L. C., Chu, H. C., Lien, C. Y., Hsiao, C. H., and Kao, T. (2009). Privacy preservation and information security protection for patients' portable electronic health records. *In Computers in Biology and Medicine, 39*(9), 743-750.

[22] Jin, W., Zhongqi, Z., Kaijie, X., Yue, Y., and Ping, G. (2013). A research on security and privacy issues for patient related data in medical organization system. *International Journal of Security and Its Applications, 7*(4), 287-298.

[23] Jovanov, L., Amanda, O. D., Raskovic, D., Cox, P. G., Adhami, R., and Andrasik, F. (2003). Stress monitoring using a distributed wireless intelligent sensor system emil. *IEEE Engineering in Medicine & Biology Magazine, 22*(3), 49-79.

[24] Kai, F., Yanhui, R., Hui, Li., and Yintang, Y. (2018). MedBlock: Efficient and secure medical data sharing via blockchain. *Journal of Medical Systems, 42*(136).

[25] Karim, A., Abderrahim, B. H., and Hayat, K. (2018). Big healthcare data: Preserving security and privacy. *Journal of Big Data, 5*(1).

[26] Kim, Y. H., and Chang, H. B. (2018). The change of future environment and the task of healthcare security. *OSIA Se&TR Journal, 31*(2), 4-9.

[27] Lee, N. Y., and Wu, B. H. (2017). Privacy protection technology and access control mechanism for medical big data. *2017 6th IIAI International Congress on*, 424-429.

[28] Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, 70*(6).

[29] Price, W. N., and Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*, 37-43.

[30] Raman, A. (2007). Enforcing privacy through security in remote patient monitoring ecosystems. *6th International Special Topic Conference*, 298-301.

[31] Ross, J. A. (1996). A security policy model for clinical information systems. *1996 IEEE Symposium*, 30-43.

[32] Silverman, B., Hanrahan, N., Bharathy, G., Gordon, K., and Johnson, D. (2015). A systems approach to healthcare: Agent-based modeling community mental health and population wellbeing. *Artificial Intelligence in Medicine, 63*(2), 61-71.

[33] Smith, E., and Eloff, J. H. P. (1999). Security in health care information systems: Current trends. *International Journal of Medical Information, 54*(1), 39-54.

[34] Swaraja, K. (2019). Medical image region based watermarking for secured telemedicine. *Multimedia Tools and Applications, 77*, 28249-28280.

[35] Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty Fuzziness and Knowledge-based System, 10*(5), 571-588.

[36] Tawalbeh, L. A., Mehmood, R., Benkhlifa, E., and Song, H. (2016). Mobile cloud computing model and big data analysis for healthcare applications. *Institute of Electrical and Electronics Engineers, 4*, 6171-6180.

[37] Umakanth, K., and Estefania, R. (2018). Asthma management in the era of smart-medicine: Devices, gadgets, apps and telemedicine. *The Indian Journal of Pediatrics, 85*(9), 757-762.

[38] Wang, Y., Xue, C. A., and Wang, L. (2018). Medical information security in the era of artificial intelligence. *Medical Hypotheses, 115*, 58-60.

[39] Williams, P. (2008). A practical application of CMM to medical security capability. *Information Management & Computer Security, 16*(1), 58-73.

[40] Wissam, A., Ali, A., Suleiman, A. Y., and Abdallah, K. (2018). Smart Medicine Dispenser (SMD). *2018 IEEE 4th Middle East Conference on Biomedical Engineering*, 2165-4255.

[41] Yiwen, D., Jianwei, L., Zhenyu, G., and Hanwen, F. (2018). A medical information service platform based on distributed cloud and blockchain. *2018 IEEE International Conference on Smart Cloud*, 34-39.

[42] Yu, H. K., and Kim, J. G. (2019). Smart navigation with AI Engine for Li-Fi based medical indoor environment. *Artificial Intelligence in Information and Communication*(ICAIIC), 195-199.

[43] Zahid, N., Sodhro, A. H., Fawad Zafar, R., Zahid, B., Khan, S. A., and Akhter, F. (2019). Regression-based transmission power control for green healthcare. *2019 2nd International Conference on*, 1-9.

[44] Zawadzki, P., and Krzysztof, Z. (2016). Smart product design and production control for effective mass customization in the Industry 4.0 concept. *Management and Production Engineering Review, 7*(3), 105-112.

◆ About the Authors ◆

Hyuna Lim

Hyuna Lim is a Master of Business Administration at Soonchunhyang University. Her research interests include Consumer Behavior, Consumer Marketing, Digital Marketing, Social Media, Data Mining, and Big data analysis.

Pham Duong Thuy Vy

Pham Duong Thuy Vy is a master's degree student at Soonchunhyang University. Her research interests include Consumer Behavior, Tourism Business Management, Behavioral Economics and Consumer Marketing.

Jaewon Choi

Jaewon Choi is an assistant professor of Business Administration, Global Business School, Soonchunhyang University. His research areas are investigating big data analysis, social network analysis, block chain, personalized intelligent agents in e-commerce and m-commerce. He published papers on Journal of Electronic Commerce Research, International Journal of Electronic Commerce, Cyberpsychology Behavior and Social Networking, and other journals.

Submitted: May 31, 2019; 1st Revision: August 16, 2019; Accepted: September 16, 2019