Working Paper 2015-02

Social Big Data Analysis and Utilization Methodologies

- With Special Reference to Forecasting the dangers of sexting in Korea using social big data

Tamin Song · Juyoung Song

Korea Institute for Health and Social Affairs

Social Big Data Analysis and Utilization Methodologies

Tamin Song, Research Fellow

Korea Institute for Health and Social Affairs
Building D, 370 Sicheong-daero, Sejong city 30147 KOREA
http://www.kihasa.re.kr
ISBN: 978-89-6827-292-9 93510

Contents

List of Tables

List of Figures

1 Social Big Data Analysis and Utilization Methodologies

1 Social Big Data Analysis and Utilization Methodologies[1]

I. Background

With the rapid distribution of smart phones, smart TV, radio frequency identification, and sensors and the diffusion of mobile internet and social media, data volumes have increased exponentially, causing big changes in production, distribution, and consumption of data. This has given rise to the age of big data, in which data have become an economic asset[2]. Governments and businesses around the world expect big data to be the source of new economic value that will determine the success of nations and businesses. Organizations like The Economist, Gartner, and McKinsey are putting forward case studies that explain the effects of economic value creation, such as predictions about market movements and development of new business lines. Specifically, big data is expected to have significant influence on national competitiveness in the future. In addition, on a national scale, big data is being implemented on a priority basis in anticipation of global factors that threaten

1) It is hereby disclosed that portions of this paper have been published in Song TM (2015). Social Big Data Analysis and Utilization Methodologies. Health and Welfare Policy Forum. Vol 227.
2) Song TM (2012)The Effective Utilization Methodology of Big Data from Health and Welfare. Health and Welfare Policy Forum. Vol 193, pp. 68-76.

safety, such as terrorism, natural disasters, diseases, and crises.

For example, Google Flu Trends (www.google.org/flutrends) researches the frequency of keyword queries related to the flu, such as “flu” and “influenza,” to provide an early warning system for outbreaks of the flu. To prepare against an uncertain future from terrorism and infectious diseases, Singapore set up the Risk Assessment & Horizon Scanning[3] system to manage big data in 2004. The United Kingdom established the Foresight Horizon Scanning Centre[4] to establish strategies for using big data technology to combat various social problems, including obesity, potential dangers, such as coastal erosion and climate change, and infectious diseases. The European Union is responding to the uncertainties of global change through the Interconnect Knowledge project for future studies, including natural disasters, like high-magnitude earthquakes and tsunamis, terrorism, participation and network, and global crises. The Organization for Economic Co-operation and Development (OECD) has recognized big data as a new asset type that provides for business efficiency, and has selected the measurement of the economics of big data as an agenda item for the 15th session of the Working Party on Indicators for the Information Society (WPIIS)[5]. In Korea, the usefulness of big data in various

3) http://www.rahs.gov.sg/public/www/home.aspx.
4) https://www.gov.uk/government/groups/horizon-scanning-centre.
5) OECD, 15th Meeting of the Working Party on Indicators for the Information Society, June 7-8, 2011.

sectors is being emphasized for the execution of the Government 3.0 and the Creative Economy projects, and for supporting key policy considerations of the current government.

Big data consists of various forms of data in large volumes, which are rapidly created. Therefore, it requires a new management and analysis methodology. Moreover, as social media platforms emerge as the source of information about the feelings and sentiments of the current times with messages on politics, economics, society, and culture, policy agendas set up in public spheres can be identified from social media. Countless comments between individuals and society evolve as a log of information, which continues to evolve as an asset for public policy[6]. As such, many nations and businesses actively strive for new economic effects, job creation, as well as solving social problems through the utilization and analysis of social big data created through social networking services (SNS).

Existing studies that focus on cross-sectional and longitudinal research related to predetermined variables can be used to examine relationships between individuals and groups, but are limited in terms of shedding light on the correlation between information discussed in individualized transactions (buzz) found in cyberspace[7], and the cause thereof. By con-

---

6) Song YJ (2012)Age of Big Data! The Evolution of SNS and Public Policy. National Information Society Agency.

7) Song, JY, Song TM (2014) Predicting Threat Recognition Factors relating to North Korea through Using Social Big Data. The Studies of International Affairs. Fall. pp. 209-243.

trast, the analysis of social big data utilizes much larger data volumes, confirms the thoughts and opinions of a diverse range of participants, and is able to reveal the complex correlation in the prediction and phenomena of social issues in a more accurate manner. This study proposes study methodologies and utilization strategies of social big data that can create value and predict the future by gathering and analyzing social big data from various fields.

II. Overview of Big Data

A. Definition of Big Data

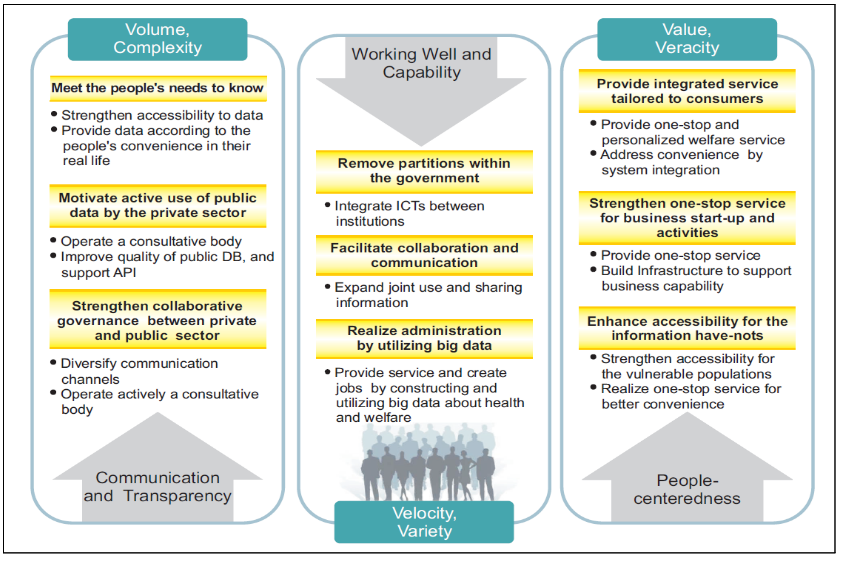
According to Wikipedia (2015.8.5.), big data is the "large structured or unstructured data sets that previous database management tools do not have the capacity to gather, store, manage, or analyze, and the technology that extracts value from such data and analyzes the results." Gartner[8] defines big data as high-volume, high-velocity, and/or high-variety information assets that demand innovative forms of information processing and that enable enhanced insight, decision making, and process automation. McKinsey[9], on the other hand, defines big data as datasets whose size is beyond the ability of typical database software tools to cap-

8) Gartner (2012) www.gartner.com/newsroom/id/2124315.
9) McKinsey Global Institute (2011) Big Data: The Next Frontier for Innovation, Competition, and Productivity.

ture, store, manage, and analyze. These definitions indicate that big data refers to an extremely large amount of data that escapes normal quantitative definitions, and is used as a concept that includes data analysis and utilization (Song, 2012).

Korea has established a big data action strategy at government level for the effective execution of Government 3.0, a life-cycle-based tailored services project for the well-being of citizens. The main characteristics of big data are summarized as "3V" (volume, variety, and velocity), and often include "2V" (value and veracity), or "1C" (complexity) in the definitions (Figure 1). Specifically, in the field of health and welfare, big data deals with information that directly links with the lives of citizens, and thus, value and veracity are of enormous importance(Figure 1).

[Figure 1] Characteristics of Big Data and Government 3.0 Execution Strategy



B. Current State of Public Big Data

As the paradigm for information disclosure shifts from being supplier-focused to citizen-centric, many countries are disclosing public information in order to create an ecosystem in which public information can be merged with private innovative ideas to create new businesses. Korea has disclosed data of government and public institutions since August 2011, and is carrying out public data disclosure that enables citizens and businesses to utilize the information freely for commercial purposes and to share it between institutions. Public data refers to all databases or electronic files produced or acquired by institutions, and its development and utilization are being encouraged in a government-wide manner regardless of purpose, whether for-profit or otherwise. As the Act for the Promotion of the Provision and Use of Public Data came into force in October 2013, each government department is working towards the disclosure of public data by sector and the development of effective strategies for the utilization of big data (Song TM, 2014).

According to an analysis of public data in 86 countries in the Open Data Barometer 2015 by the World Wide Web Foundation, the United Kingdom was ranked first (100 pts) in open data, the United States second (92.66 pts), Sweden third (83.7), France fourth (80.21), and New Zealand fifth (80.01), while Korea tied in 17th place with the Czech Republic and Japan was 19th. Compared to the 2013 evaluation results, Korea and Japan have dropped five places in the ranking (Figure 2).

[Figure 2] World Wide Web Foundation Ranking of Open Data: 2014

Country	Barometer Rank	ODB Scaled	Readiness (Scaled)	Implementation (Scaled)	Impact (Scaled)	2013 ODB	ODB Change	2013 Rank	Rank Change
UK	1	100	98	100	100	100	0	1	0
US	2	92.66	96	88	100	93.38	-0.72	2	0
Sweden	3	83.7	100	76	88	85.75	-2.05	3	0
France	4	80.21	91	75	84	63.92	16.29	10	6
New Zealand	4	80.01	81	88	55	74.34	5.67	4	0
Netherlands	6	75.79	95	76	57	63.66	12.13	10	4
Canada	7	74.52	90	75	58	65.87	8.65	8	1
Norway	7	74.59	88	73	64	71.86	2.73	5	-2
Denmark	9	70.13	94	54	95	71.78	-1.65	5	-4
Australia	10	68.33	92	69	43	67.68	0.65	7	-3
Germany	10	67.63	85	67	53	65.01	2.62	9	-1
Finland	12	66.49	93	54	78	49.44	17.05	14	2
Spain	13	59.89	78	60	42	48.19	11.7	17	4
Estonia	13	60.18	84	51	64	49.45	10.73	14	1
Austria	15	58.52	83	42	84	46.03	12.49	18	3
Chile	15	58.7	69	73	8	40.11	18.59	25	10
Czech Republic	17	58.07	64	61	46	43.18	14.89	22	5
Korea	17	57.65	79	54	48	54.21	3.44	12	-5
Japan	19	53.58	81	53	30	49.17	4.41	14	-5
Israel	20	52.97	70	51	43	45.58	7.39	18	-2

Source: World Wide Web Foundation (2015). Open Data Barometer Global Report

Public data in Korea are disclosed in the Open Data Portal (data.go.kr), and as of July 24, 2015, 18,014 data sets have been provided. Among these, data related to health and welfare accounts for 10.57% with 909 data sets in the health sector and 995 data sets in the welfare sector, totaling 1,904 sets (Table 1).

〈Table 1〉 State of public data provision in the health and welfare sectors (based on the date of initial registration)

Category		2013	2014	2015	Total (percentage of subtotals)
Health	Public institutions	23	88	4	115 (12.7)
	National administration agencies	13	82	8	103 (11.3)
	Regional administrative organizations	142	519	30	691 (76.0)
	Subtotal	178 (19.6)	689 (75.8)	42 (4.6)	909 (100.0)
Welfare	Public institutions	23	83	1	113 (11.4)
	National administration agencies	12	48	10	70 (7.0)
	Government-invested institutions	1	7	1	9 (0.9)
	Regional administrative organizations	219	550	34	803 (80.7)
	Subtotal	219 (26.2)	668 (67.1)	46 (4.6)	995 (100.0)
Total		433 (22.7)	1,377 (72.3)	88 (4.6)	1,904 (100.0)

Source: 1) Public institutions (15institutions): Health Insurance Review & Assessment Service, National Cancer Center, National Medical Center, National Health Insurance Service, Korea Labor Welfare Corporation, Korea Red Cross, Korea Health Promotion Foundation, Korea Institute of Science & Technology Information, Korea Foundation for International Healthcare, Korea Health and Welfare Information Service, Korea Health Industry Development Institute, National Health Personnel Licensing Examination Board, Korea Occupational Safety and Health Agency, Korea District Heating Corporation, and Korea Environment Corporation.

2) National Administration Agency (three agencies): Ministry of Public Safety and Security, Ministry of Health and Welfare, and Ministry of Government Administration and Home Affairs.

3) Regional Administrative Organizations (17 provinces and cities): Gangwon Province, Gyeonggi Province, South Gyeongsang Province, North Gyeongsang Province, Gwangju Metropolitan City, Daegu Metropolitan City, Daejeon Metropolitan City, Busan Metropolitan City, Seoul Special City, Sejong Metropolitan Autonomous City, Ulsan Metropolitan City, Incheon Metropolitan City, South Jeolla Province, North Jeolla Province, Jeju Special Self-governing Province, South Chungcheong Province, and North Chungcheong Province.

※ Research of the current state of public data as of July 24, 2015.

The highest-frequency keywords by year for public data in the health and welfare sector are shown in Table 2. In the health sector, public data related to the keyword "hospitals" were provided frequently in 2013 and 2014, while in 2015, public data related to the keyword "cosmetics" were provided frequently (Table 2).

〈Table 2〉 High frequency keywords in the health and welfare sector by year

Rank	Health			Welfare		
	2013	2014	2015	2013	2014	2015
1	Hospital	Hospital	Cosmetic	Senior	Facility	Senior
2	Medical institution	Branches	Hospital	Current state	Disabled	Town center
3	Current state	Medical institution	Medical institution	Disabled	Current state	Children
4	Pharmacy	Pharmacy	Public sanitation	Welfare facilities	Senior	Current state
5	Doctor	Medical	Skin	Welfare	Welfare	Senior welfare
6	Branches	Current state	Doctor	Facility	Children	Free school meal
7	Vaccination	National examination	Hair styling	Pharmacy	Children	Free school meal
8	Yongin Province	Health and medical	Hair styling salons	Old age home	Youth	Welfare
9	Public sanitation	Public sanitation	Laundry	Children	Center	Welfare facilities
10	Information	Gwangju Metropolitan City	Sanitary center	Center	Welfare facilities	Funeral

In the welfare sector, public data related to senior citizens, the disabled, and welfare facilities were provided most frequently, and in 2015, public data related to free school meals were uploaded, which became a high-frequency keyword(Table 2).

C. Methods for privacy protection in big data

The most important issue in the utilization of big data, regardless of the country in question, is the protection of individual privacy and personal information[10]. The focus on protecting private information may hinder the utilization of big data. While the purpose of the Personal Information Protection Act is "to increase the rights and benefits of the citizens by protecting the personal information from gathering, leaks, misuse, and abuse of personal information," it is difficult to define personal and non-personal information strictly. Even if the data automatically gathered by businesses may be non-personal, such information may invade privacy (Song TM, 2014). Specifically, personal information revealed on social media can be easily forged, modified, misused, or abused, and may be exposed to information collection activities for commercial use. Thus, the potential for such problems as privacy invasion is very high (Song TM, 2014). The Korean Communications Commission engaged in opinion gathering via a discussion on big data personal information protection on December 18, 2013, and an Online Personal Information Protection Seminar

10) Song TM et al. (2014) Studies of Methodologies of Effective Management of Big Data for Health and Welfare. Korea Institute for Health and Social Affairs.

on March 19, 2014. In addition, the Commission issued Big Data Personal Information Protection Guidelines on December 23, 2014. Key features of the guidelines, as mentioned in Table 3, include measures to achieve non-identificability of personal information starting from the big data-gathering process, and measures to forbid processing of sensitive information, such as personal ideologies, beliefs, and political views, including the merging and analysis of such data. The most important measure to protect individuals from big data disclosure is anonymization to prevent identification of a specific individual, and restrictions on accessing and processing data. However, if such restrictions are strong, the utilization of information cannot be cultivated, and therefore, effective policies for the "balance of utilization and protection" of big data need to be provided[11].

〈Table 3〉 Key contents of the Big Data Personal Information Protection Guidelines

Category	Contents
Non-identifiability measures	- Take strict measures to make personal information non-identifiable in the gathering process (Articles 3, 4, 5, and 10). - Disclosed information and information on its usage patterns, including personal information that can be gathered, stored, merged, analyzed, and provided to third parties after measures to make the information non-identifiable.
Acquire transparency for disclosure	- Acquisition of transparency through disclosure of the actual big data-processing activity and purposes thereof (Articles 4, 5, and 9).

11) Song TM (2013) The Current State of Health and Welfare Big Data in Korea and Utilization Methodologies. Science and Technology Policy. Vol 192, Science and Technology Policy Institute.

Category	Contents
	- After measures taken to make the information non-identifiable through personal information policies, reveal the processing activity, purposes, gathering source, and right to reject information utilization by users in a transparent manner. - **(Personal Information Policy)** After measures taken to make the information non-identifiable, reveal the act of and purposes of big data processing to users and provide a link to "reject information utilization" and allow users to exercise rejection rights. - When processing personal information gathered from people other than users, notify the users of the right to request the "gathering, source, purpose, and right to cease the processing of personal information."
Take measures to make the information non-identifiable in event of re-identification	- When personal information is made personally identifiable again, destroy the information and take measures to make it non-identifiable (Articles 3 and 6). - When personal information is made identifiable again during the big data-processing process and information is created therefrom, destroy the information gathered in this way and take additional measures to make the information non-identifiable.
Forbidding sensitive information processing	- Forbid the processing, including gathering, utilization, and analysis, of sensitive information and communication secrets (Articles 7 and 8). - Forbid processing of information, including gathering, utilization, storage, and combining processing, for the purposes of creating sensitive information, such as ideologies, beliefs, and political views of specific individuals. - Forbid processing of communications information, including gathering, utilization, storage, and combining processing, such as emails and text messages.
Protective technological and management measures	- Enforcement of "technological and management protective measures" during the storage and management processing of gathered information (Article 3, Section 2). - Application of technological and management protective measures for information processing systems that store and manage information, about which measures have been taken to make it non-identifiable. - (Protective measures: installation of access restriction devices, such as anti-invasion systems, preventive measures for forging and falsification of access history, and preventive measures for break-ins by malicious programs, including installation and operation of anti-virus software.

Ⅲ. Analysis methods of social big data

In the big data sector, social analytics quickly analyze unstructured data gathered from Facebook, Twitter, and SNS. The methods of extracting and analyzing data from social media largely can be divided into three methods[12]. First, text mining utilizes natural language-processing technologies on the unstructured text produced by humans in languages to extract useful data, and identify, classify, group, or summarize connectivity to identify hidden useful information within big data. Second, opinion mining targets the text sentences in social media and applies natural language-processing technologies and sentiment analysis to analyze users' opinions. In marketing, this is called buzz (word-of-mouth) analysis. Third, network analytics analyzes network connection structure and connection strength to identify which messages are propagated through which medium and whom they influence.

The social big data analysis process and methods are as follows. First, analytical modeling is applied to a transaction related to the topic in question (e.g., Middle East Respiratory Syndrome, MERS), and the targets and range of data collection are set, after which crawlers and other collection engines (robots) are utilized in target channels (news, blogs, cafes, bulletin boards, SNS, etc.) to collect data. At this stage, stop words

12) Song, TM, Song JY (2013) Methodology of Big Data Analysis. Hannarae Academy.

(e.g., “Mers Benz” or “Mercedez-Benz”) are selected to prevent errors in the collection process, and the MERS-related keyword group (“MERS virus,” “Middle Eastern Respiratory Syndrome,” “MERS Coronavirus,” and “Maers”) is selected. Second, the raw data collected on MERS are gathered in the form of a texter, and are unstructured. In addition, there are difficulties for the researcher to analyze the data as they are gathered in original form. Therefore, the collected unstructured data undergo text mining and opinion mining in a process of classification and refining. Buzz analysis, keyword analysis, sentiment analysis, and account analysis are used to analyze unstructured data. Third, the unstructured big data needs to be converted to structured big data. According to the case studies of subject analyses related to MERS, each transaction for MERS buzz needs to be codified with an identification and the keyword that occurs within the buzz also needs to be codified. Fourth, to allow for analysis in connection with social phenomena, the structured big data needs to be linked to the offline statistical research data. Offline statistical research data are often provided for free or at a fee from the government and public institutions, along with the data to be linked. Upon confirmation of the linkable identifier (daily, monthly, yearly, and regional), the offline data can be gathered and linked. Fifth, the analysis of the structured big data linked to the offline statistical research data is carried out through a structural equation model that can analyze the correlations between factors or time-series changes. This is performed via a multi-level model that can analyze the relationship with factors related to daily, monthly, or yearly regional social phenomena, and data mining analysis or visualization that can provide insights into a new phenomenon through the classification of the collected keywords.

There are two methods of big data linkage: exact matching and statistical matching. Exact matching is used when unique identifying information is available, and statistical matching is used when similar objects are located to merge respective data, as there is no unique identifying information available. The linkage between social big data and public big data can be analyzed through exact matching by utilizing time and regional variables as unique identifiers (Figure 6).

Ⅳ. Methodologies of social big data collection and sorting

On the one hand, the collection and sorting of social big data can be done through the top-down method where the theoretical background of the chosen topics are analyzed and ontology is developed, after which the keywords of the ontology are collected and sorted. On the other hand, the bottom-up method can be used, in which the chosen topics are collected using web crawls, and are sorted using a universal dictionary or a user dictionary.

A. Top-down method[13)]

As the language used in social media from unstructured data is made up of colloquial sentence structures that people use in everyday life (Noh JS, 2012)[14)], an analytical framework is needed to collect and analyze these data more effectively (Song TM et al., 2014). As the contents of the analytical framework require conceptual domains in related subjects and definitions of the relationship between each concept, the ontology that reflects this information needs to be developed (Song TM et al., 2014). Ontology is a computer-interpretable knowledge model that formalizes and represents shared concepts (Kim et al., 2013)[15)]. As the collected social big data are expressed diversely in an unstructured manner, the terminology that explains the concepts that make up the ontology and the synonyms need to be defined and described to form a system of terminologies (Song TM et al., 2014). This study is the analytical framework for sorting and uti-

13) The contents in this section are a part of Song TM et al., (2014) Studies of Methodologies of Effective Management of Big Data for Health and Welfare, which was carried out jointly with the Park HA Research team at the College of Nursing, Seoul National University. It is hereby disclosed that the results have been published in Ae Ran Kim, Tae Min Song, Hyeoun-Ae Park (2014) Development of an Obesity Ontology for Collection and Analysis of Big Data Related to Obesity. TBC.

14) Noh JS (2012). Big Data and Social Analysis: Looking for "Meaning" in the Sea of Big Data. http://www.imaso.co.kr/?doc=bbs/gnuboard.php&bo_table=article&wr_id=40725

15) Kim HY,Park HA,Min YH, Jeon E (2013) Development of an Obesity Management Ontology based on the Nursing Process for the Mobile-Device Domain. Journal of Medical Internet Research. 15(6), e130. doi: 10.2196/jmir.2512

lizing the big data gathered about obesity from online mediums. It sorts subjects related to obesity management, and develops the obesity management ontology and a system of terminologies thereof. As an example of ontology related to obesity, the diagnosis, prevention, and cure for obesity are influenced by the demographic characteristics of the target, risk factors, major observable symptoms, and existence of complications.

The development of ontology, related to the terminology that belongs to the sorting framework that explains the subjects related to obesity management, requires that terminologies are extracted and classified into major, medium, or minor classification systems. Therefore, as shown in Table 4, each terminology needs to be defined with synonyms and analogues via such methods as internet searches and literature surveys[16].

〈Table 4〉 Domain levels based on the classification of obesity ontology

Major classification	Medium classification	Minor classification 1	Minor classification 2	Synonyms / Analogues	Domain level
Risk factor				Risk aspect, Risk element	Major classification (Risk factor)
	Dietary and eating habits			Meal, Diet, Dietary treatment, Eating habits	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits)
		High fat diet		High fat meals, Fat overdose	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits) 〉 Minor classification 1(High-fat diet)

16) For details on the development of ontology relating to obesity, refer to Song TM et al. (2014) Studies of Methodologies of Effective Management of Big Data for Health and Welfare. pp. 230-252.

Major classification	Medium classification	Minor classification 1	Minor classification 2	Synonyms / Analogues	Domain level
			Fat	Oil, Oily portions, Fat and oils	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits) 〉 Minor classification 1(High-fat diet) 〉 Minor classification 2(fat)
			Sweets	Sweets, Biscuits, Snack	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits) 〉 Minor classification 1(High-fat diet) 〉 Minor classification 2(sweets)
			Roasted pork belly	Roast pork, Pork belly	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits) 〉 Minor classification 1(High-fat diet) 〉 Minor classification 2(roasted pork belly)
			Fried food	Fried dishes, Fried food	Major classification (Risk factor) 〉 Medium classification (dietary and eating habits) 〉 Minor classification 1(High-fat diet) 〉 Minor classification 2(fried food)
			…		This may include other classifications similar to the above.

B. Bottom-up approach

To collect and sort social big data, universal dictionaries are available, such as the 21st Century Sejong Plan, but typically, a user-constructed dictionary is used for the analysis. For example, if social big data were gathered for the purposes of predicting health and welfare policy demand, the collection criteria for the web crawl would be “health, welfare, health and welfare."There are 116 online news sites, 4 blogs, 2 internet cafes, 2 SNSs, and 4 bulletin boards, totaling 128 online channels available for collection for keywords of health and welfare. The online transactions on health and welfare gathered from the channels available for collection are classified using a universal dictionary or user dictionary, as per Figure 8 below. These data need to be turned into structured big data by confirming the existence of the relevant keywords.

Upon completion of the sorting and converting to structured big data of the collected social big data, sentiment analysis is performed on the sorted keywords in order to extract factors (variable reduction). Sentiment analysis is divided into two methods: one requires the user to develop a sentiment word dictionary and analyze the sentiment of the transaction[17]; the other method requires sentiment analysis through factor analysis and subject analysis. For demand predictions of health and welfare, sentiment analysis needs to confirm agreement or disagreement with the pertinent transaction. Therefore, factor analysis, on the 59 keywords classified as emotions/attitude and variable reduction must be performed. The keywords are

17) The sentimental analysis for obesity was performed using a sentiment word dictionary to analyze positive emotions ("prevent obesity,""escape from lower-body obesity,""go on a diet,""solve,""manage,""lose,""escape,""effective diet,""right,""fast," and "recommend"), that is, the positive emotion of escaping obesity was considered as success. Negative and neutral emotions ("metabolic obesity,""cause,""excess,""serious,""increase,""dangerous,""excessive diet,""fail,""go wrong,""give up,""not good," etc.) that is, the negative and neutral meanings of escaping obesity were considered failures.

"support," "need," "problem," "disagree," "execute," "manage," "possible," "progress," "happiness," "plan," "assertion," "magnify," "attention," "help," "visit," "implement," "utilize," "provide," "diverse," "attempt," "confirm," "improve," "participate," "present," "benefit," "point," "important," "controversy," "donate," "use," "best," "abolish," "regulations," "enforce," "ready," "apply," "expected," "reinforce," "introduce," "burden," "justice," "criticism," "stop," "realize," "recommend," "lie," "minimize," "worry," "increase," "insufficient," "difficulty," "welfare party," "accuse," "ignore," "dear," "turn away," "speedy," "foremost," and "tears"[18].

From the first factor analysis, the number of factors was reduced to 18, and from the second factor analysis on the 18 factors, the number of factors was reduced to 5. In the second factor analysis, the meanings of the subject words determined by the 5 factors were identified and sentiment analysis was performed to check for agreement or disagreement. The results were classified as agreement factors ("manage," "support," "plan," "schedule," "reinforce," "implement," "magnify," "progress," "use," "utilize," "enforce," "execute," and "participate") and disagreement factors ("problem," "point out," "disagree," "lie," "controversy," "criticism," "worry," "accuse," and "turn away").

18) For details on the sentimental analysis of health and welfare, refer to Song TM, Song JY (2015) Completing Big Data Research with One Book. Hannarae Academy, pp.415-418.

2

Forecasting the dangers of sexting in Korea using social big data

2 Forecasting the dangers of sexting in Korea using social big data[19)]

Ⅰ. Background

As of 2014, 99.7% of Korean teenagers possessed smartphones (Korea Internet and Security Agency, 2014, p.19), 95.2% of teenagers used the Internet, and 78.1% of high school students utilized social networking services (SNS) (Ministry of Science, ICT and Future Planning and Korea Internet and Security Agency, 2014). This indicates that internet usage through smartphones has become a critical means of communication for the young generation. As internet and smartphone use in the everyday lives of the young generation increases, side effects, such as internet addiction, are being highlighted, along with positive effects. In 2014, the rate of smartphone addiction among the young generation was 29.2%, almost 2.6 times more than the rate among adults of 11.3% (Ministry of Science, ICT and Future Planning and Korea Internet and Security Agency, 2014). In addition, as of 2014, 52.6% of middle school and high school students had utilized obscene material (sexting) (Ministry of Gender Equality and Family, 2014).

19) This study was jointly conducted by and Song JY (Pennsylvania State University Assistant Professor) and Song TM (KIHASA, Research Fellow) for the purposes of publishing in a foreign academic journal.

Furthermore, some teenagers had exhibited delinquent behavior by filming their own sexual behavior and sharing it in real-time internet broadcasting services, such as UCC broadcasting, or on file-sharing websites, such as webhards. This phenomenon has highlighted the need to block obscene content from the young generation at a government level.

Sexting is a compound word of "sex" and "text" that refers to obscene content in which teenagers below 18 years of age expose certain body parts with non-specific individuals of the opposite gender, whom they have come across through their cellphones and the internet (Lounsbury et al.,2009, p.1; Walker et al., 2011, p.8). The main challenges of sexting include legal action that needs to be taken for child pornography if images of teens exposing their bodies are kept on mobile phones. Moreover, sexting may cause emotional and physical damage to the sexting partner (Chalfen, 2009, p.263). For teenagers, accessing obscene content online is often viewed as a type of status delinquency in the sense that it is a type of delinquency without victims, much like alcohol or smoking, in which they may engage tosatisfy their curiosity in their teenage years. However, obscene content has the potential to influence perceptions about sex, values, and sexual attitude of teenagers, and is dangerous because it may lead to teenagers accepting distortions of sex as reality (Kim JG, 2012, p.2). Moreover, teenage experience with pornography has the potential to create obsession over pornography through ongoing curiosity (Joo and Kim HI, 2010, p.18). Despite the seriousness of the problems related to sexting, the scientific study of sexting in Korea remains insufficient.

The dispersion of mobile internet and social media has led to exponential increases in data volumes, leading to substantial changes in the production, distribution, and consumption systems of data, which has brought about the era of big data in which data can be an economic asset. Governments and businesses around the world expect that solving social problems and executing government policies would be more effective through the utilization and analysis of social big data created through SNS as they forecast ripple effects of big data in the public and private spheres. To execute and achieve Korea's Government 3.0 and Creative Economy projects, Korea continues to seek methods to utilize big data effectively in various sectors. The analysis of social big data involves analyzing the meaning of a transaction left behind by a user. It begins with text mining, natural language-processing technology, opinion mining, and sentiment analysis technology, and is followed by network analysis and statistical analysis. Existing studies that focus on cross-sectional and longitudinal research can be useful for examining relationships between individuals and groups related to predetermined variables, but are limited in shedding light on the correlation and cause thereof for information discussed in individualized transactions (buzz) in cyberspace (Song TM et al., 2014). By contrast, the analysis of social big data utilizes a much larger volume of data and confirms the thoughts and opinions of a diverse range of participants. Thus, it is more accurate in forecasting social issues. This study forecasts the risk factors of sexting in Korea based on the social big data gathered from domestic online news sites, blogs, internet cafes, SNS, and bulletin boards.

Ⅱ. Theoretical Background

The teenage years are typically a period of high sexual curiosity, inability to distinguish between the facts and fantasies of sex, and these days, are marked by easy exposure to online sexual (性) content (Shin, 2013,p.276). “Obscene online content” is an overarching term that covers images, articles, and multimedia that explicitly describe human sexual behavior to give a lewd and wanton impression. It typically indicates material in the form of writing, photography, cartoons, and magazines that, for commercial purposes, focuses on the viewing of sexual organs and sexual intercourse, and is intended to lead the reader or viewer to be sexually aroused (Kim MN, 2012, p.52). Specifically, child or teenage pornography in Korea is defined as the depiction of sexual acts, including acts that involve the viewing of sexual organs, the anus, or sexual objects,” where objects are defined as parts of the body, result in sexual humiliation and disgust by normal citizens, and that feature children, teenagers, or indicators that relate to children or teenagers, in the format of films, videos, games, or other forms of multimedia through computers or other communication mediums (Act for the Protection of Children and Juveniles against Sexual Abuse, Article 2, Section 5)[20]. In the case of the United States, a federal decree, the PROTECT Act, specifically defines child pornography as that which depicts minors engaging in sexually explicit conduct. The definition of sexually explicit conduct in this context is "sexual intercourse, including genital-genital, oral-genital, anal-genital, or oral-anal, whether between persons of the same or opposite sex, bestiality, masturbation, sadistic or masochistic abuse, or lascivious exhibitions of genitals or pubic area of any person" (US Department of State, 2003). In Japan, the concept of child pornography is informed by the Act on Punishments for Acts Related to Child Prostitution and Child Pornography(Section 2, Subsection 3) and the Tokyo Teenager Protection Amendments(Article 3, Section 7, Subsection 2). It is defined as arousing sexual feelings for teenagers that are thought to be and/or depicted as under the age of 18 years, sexual intercourse with a child, or engaging in acts similar to sexual intercourse with a teenager (東京都議会, 2010). The social background that has given rise to pornography as an issue in

20) http://www.law.go.kr/lsInfoP.do?lsiSeq=150720&efYd=20140929#0000. Accessed on June 16, 2015.

modern society is that while the principal consumers of media were the older generation, teenagers are now given the choice to consume various media. This has led to their greater ability to access obscene content on the internet and engage in other problematic actions (Gruber and Thau, 2003, pp.441-443).

Meanwhile, sexting is an issue in modern society as teenagers are now given sexting choices before the age of 18 years through their meeting via their mobile phones, similar devices, or the internet (Lounsbury et al., 2009, p.1; Walker et al., 2011, p.8). The term sexting has been used by researchers since it was first registered on the Macquarie online dictionary[21] in 2010 (Walker et al., 2011, p.8). The term is used to refer to exchanges of pictures only and excludes the activity of exchanging SMS text messages with sexual meaning (Lounsbury et al., 2010, p.2). Korea defines sexting as the act of teenagers using mobile phones to produce or distribute (sending and receiving) sexually stimulating text messages, voice messages, or pictures and videos of themselves or their friends, without commercial intent (Lee and Kim EG, 2009). Study into the state and reasons thereof of pornographic material distribution by teenagers using mobile phones(Korean Institute of Criminology, p. 43). The word "sexting" is often not used or recognized by teenagers and a majority of them does not consider the exchange of pictures of exposed bodies as sexting (Ringrose et al., 2012, p.8).

21) https://www.macquariedictionary.com.au

Sexting has been studied mainly in foreign countries. In Korea, only studies related to the use of pornographic content by teenagers using the internet and influence thereof have been conducted. In the United States, it has been found that 67% of male teenagers and 71% of female teenagers have experienced sending or receiving messages with nude photos or sexual content (Chalfen, 2009, pp.258-259). In the case of sexting pictures, taking photos of themselves for sexting begins at the age of 15 years, with participation rates increasing for the ages of 16 and 17 years. In addition, female teenagers engage in the production of pictures with bodily exposure 1.56 times more than male teenagers (Mitchell et al., 2011, pp.17-18). Sexting usually occurs with teenagers aged 13-18 years, and results in such problems as discontinued friendships, bullying, and verbal violence when the pictures with bodily exposure are sent to close friends and are shared with groups of friends of similar age. This may ultimately result in serious deterioration of the mental health of the teenager, who becomes a victim of the distribution of the picture of bodily exposure (Jaishankar, 2009, pp.21-22).

Personal and environmental factors are both at play for teenagers who engage in sexting: personal attitude is impulsive, there may be low self-monitoring, or there are situational constraints, in which there is a decreased level of attitude-behavior consistency through anonymity (Sherman and Fazio,1983, pp.

311-320).

For female teenagers, problems, such as sexual harassment, may occur as their boyfriends request pictures of exposed body parts as evidence of their romantic relationship or continuation thereof. In addition, they may boast of close relationships with friends of the opposite gender by uploading pictures of their opposite-gender friends through anonymous pictures to the public, which may lead to the sexual humiliation of many teenagers (Ringrose et al., 2012, p.29).

The percentages of teenagers mutually exchanging pictures of exposed bodies are similar for both genders. However, while sexting is in progress, male teenagers are 2.5 times more likely to be on the requesting side of pictures of exposed bodies, and female teenagers are 1.75 times more likely to be requested for pictures of exposed bodies (Temple et al.,2012, p.830).

The US Supreme Court defines the act of distributing and exchanging sexting pictures as exchanging pictures of exposed bodies that induce sexual feelings. With regard to child pornography, the court clearly states that the distribution and exchange of pictures of teenagers whose bodies are exposed is punishable, even though the image may not fit the definitions of obscenity. Moreover, the taking of pictures with bodily exposure may be a factor that hinders healthy mental growth for teenagers and could be a factor in potential child abuse (McLaughlin, 2010, pp.14-15). On the other hand, when the teenagers recognize that their parents are aware of their exchange of pictures of exposed bodies using their mobile phones, the likelihood of them continuing sexting decreases. Therefore, parental attention has been identified as a good remedy to decrease teenage participation in sexting and prevent participation altogether (Lenhart, 2009, p.15).

In studies of sexting in Korea, 20% of all teenagers (13t their parents are aware of their exchange of pictures of exposed bodies using their mobile phones, the likelihood of them continuing sexting and sending pictures of legs, underwear, and the act of undressing themselves or their friends (Lee and Kim EG, 2009,P. 100). As internet pornography can be accessed easily, anonymously, and free of charge, the internet has become a point of access for pornography for teenagers who are curious about sex and have a high tendency to seek sexual content (Joo and Kim HI, 2013, p.12). Internet pornography sites include sites that utilize hidden cameras to satisfy voyeuristic needs and hentai, consisting of animation videos. These websites allow viewing, copying, and downloading of pornographic content for a small upfront fee or a subscription fee, as well as free tours/ previews, allowing teenagers below the ages of 18 years to view content without limits (Kim M and Kwak JB, 2011, pp.294-295). A popular method of accessing pornographic content among teenagers is the peer-to-peer file-sharing network, allowing teenagers to forego accessing specific websites and to use file-sharing network programs to receive and send pornography from anonymous people (Greenfield, 2004, p.746). In addition, smartphone pornography is being distributed to teenagers, and among 41,206 items of obscene content found on smartphones, 90.2% were discovered to be applications that enable obscene chatting and the viewing of adult videos (Oh, 2013). Teenagers who purposefully search for pornography using smartphones were doing so more often if they had a low degree of self-control or faced discriminatory relationships with peer groups (Choi and Jung, 2014, p.451).

Among the television programs widely viewed by teenagers, it was found that if teenagers were to watch programs with sexual content repeatedly, it would be difficult for them to form a comprehensive value framework on sex, and they would be likely to develop a twisted concept of sexual issues (Shim, 2010, p.79). This includes the characteristics of lower humility or guilt in seeking sexual pursuit online, which extended to offline activities (Kim M and Kwak, 2011, p.303). In the case of pornography that features children or teenagers, the main content involves sexual intercourse with sexually mature children or teenagers. As a result, the sexual perceptions of children or teenagers may become distorted and potentially be the main factor that justifies sexual wrongdoing involving children by pedophiliac individuals (Itzin, 1997, p.96).

If exposure to online pornography is frequent, the rape myth phenomenon may be found in teenagers, in which they mistake as reality the enjoyment of females featured in rape pornography, and a wrong sexual concept of females wanting to engage in unilateral sexual intercourse with males may develop in teenagers and adults (Malamuth and Check, 1985, pp.314-315). Teenagers whose pornography-viewing frequency is high are more likely to engage in gage in featured in pornographic content who enjoy the pornography viewing frequency. In addition, if alcohol-consuming teenagers view pornography, the chances of delinquency through sexual issues are much higher (Ybarra and Mitchell, 2005, p.483).

Pornography is divided into soft porn, containing relatively lighter content, and hard-core pornography, containing sadistic content. The volume of hardcore pornography, which includes videos and images that are shocking and surprising, is rising rapidly on the internet. This may strengthen wrong perceptions about sex (Eberstadt and Layden, 2010, p.21).

Ⅲ. methods

1. Study target

This study targeted social big data gathered through the internet, including domestic SNS and online news sites. This study defined text-based web transactions (buzz) collected through 156 online channels, consisting of 146 online news sites, 9 bulletin boards, and 1 SNS (Twitter). The collection of topics relating to sexting[22] was undertaken from January 1, 2011 to March 31, 2015 (4 years and 3 months) without considering weekdays, weekends, and holidays. Among 65,611 hits, 13,774 text transactions thought to be related to teenagers (1,086 hits in 2011, 5,352 hits in 2012, 3,983 hits in 2013, 2,319 hits in 2014, and 1,034 hits in 2015) were included in the study. In order to gather all transactions related to sexting, topic analogues include "pornography distribution," "adult material distribution," "sexting" (English), "pornographic distribution," "spreading pornography," "pornography upload," "pornography download," "porn sharing," "porn chatting," "porno distribution," "porno spreading," "adult video distribution," "adult video spreading," "adult video upload," and "adult video download." The collection of social big data for this study[23] was completed using crawlers, and subject analysis was performed to sort the noun-form words, which were then categorized and set as analysis variables.

22) A topic is defined as the subject keyword that forms the target of social analysis and monitoring, and transactions that contain the topic are collected.

23) The collection and topic sorting of the social big data for this study was carried out by SK Telecom Smart Insight.

2. Study tools

The transactions gathered that relate to sexting underwent a process of subject analysis[24] for codification into structured data as follows.

A. Sentiments relating to sexting

This study identified 100 sentiment keywords related to sexting through subject analysis after gathering the transactions. They were“trouble,” “sympathize,” “fear,” “positive,” “happiness,” “controversy,” “tears,” “severance,” “warm,” “question,” “oppose,” “opposition,” “signature,” “unrest,” “uncomfortable”, “blame,” “social evil,” “wound,” “sign,” “stress,” “fail,” “serious,” “bad influence,” “difficulty,” “concern,” “depression,” “admission,” “freedom,” “wrong,” “fun,” “addictiveness,” “exhilaration,” “obsession,” “best,” “worst,” “give up,” “tired,” “sigh,” “pathetic,” “bliss,” “curiosity,” “regret,” “interest,” “hope,” “attention,” “urgent,” “strict,” “slander,” “fortify,” “arouse,” “love,” “shock,” “expect,” “ask,” “firm,” “insult,” “importance,” “focus,” “threat,” “assistance,” “review,”

24) To collect and sort social big data, universal dictionaries are available, such as the 21st Century Sejong Plan, but typically, a user-constructed dictionary is used to fit the purposes of the analysis. The subject analysis relating to grouping in this study was performed through constructing a user dictionary with the top 2,000 keywords from the raw data being categorized after SKT collected the related transactions.

"solve," "burden," "danger," "criticize," "joke," "blatant," "obscene," "abase," "sensational," "embarrassment," "blind end," "seduce," "invade," "swear," "stimulus," "garbage," "secret," "expectation," "lie," "confusion," "weary," "inappropriate," "dazzlement," "appeal," "favorable treatment," "jeer," "poor," "threatening," "humility," "cruel," "brutal," "distort," "debauchery," "betrayal," "devil," "pissed off," "eradication," "abhorrence," "decadent," "mental stress," "shocking." To determine the extent of the sentiments in the sentiment keywords among the 110 pornography distribution keywords, a second factor analysis was performed to shorten the list to 11 factors (67 variables), which were subjected to a sentiment analysis. Normally, sentiment analyses require the use of a sentimental word dictionary made up of positive and negative words. However, this study identified the meaning of the subject words from the results of factor analysis for utilization in the sentiment analysis. The subject meanings were identified for about 11 factors determined from the factor analysis, which were then divided to objected to a sentimenton," or category" for sentiment analysis. The 27 words in the normal category wereafirm, or about 11 factors determined from factor analysis, "unrest," "uncomfortable", "bl," "sign," "abase," "insult," "lie," "betrayal," "social evil," "confusion," and "poor,". On the other hand, the 28 words in the danger category were" in the other hand, the 28 words in the danger category were about 11 factors determined from factor analysis, in the other hand, the 28 words in category were about 11 factors determined from factor analysis, which were then divided to objected to a sentimenton", er category" for sentimentnature," "unrest," "uncomfortable", "bl," "sign," "abase," "insult." The danger category refers to the sentiment that looks favorably on sexting, and the normal category to sentiment that looks negatively on sexting.

B. Regulations related to sexting

There are four kinds of regulations related to sexting based on the factor analysis and subject analysis process, including additional punishment, the Act for the Promotion of Information and Communications Network Utilization and Data Protection, the Act for the Protection of Children and Juveniles against Sexual Abuse. When the transaction contained a regulation, the transaction was codified as '1', and the transaction did not contained a regulation was codified as '0'.

C. Institutions related to sexting

The eight institutions related to sexting based on the subject analysis process are the Korea Communications Commission, National Police Agency, Parliament, Blue House, Government, the judicial authority, civic group, and international agencies When the transaction contained a institution, the transaction was codified as '1', and the transaction did not contained a institution was codified as ‘0'.

D. Harmful effects of sexting

Based on factor analysis and subject analysis, the five harmful effects of sexting are defamation, sexual crimes, fraud, drinking, and social issues. When the transaction contained a harmful, the transaction was codified as '1', and the transaction did not contained a harmful was codified as ‘0'.

E. Influence from sexting

Based on the subject analysis process, the six influential factors that affect sexting are study, health, social relationships, cost, ethics, and libido. When the transaction contained a influential, the transaction was codified as '1', and the transaction did not contained a influential was codified as ‘0'.

F. Aid for sexting

Based on factor analysis and subject analysis, the five aid measures for sexting are preventive education, expert counsel, encouraging integrity in everyday life, control, and love. When the transaction contained a aid, the transaction was codified as '1', and the transaction did not contained a aid was codified as ‘0'.

G. Types of sexting

Based on factor analysis and subject analysis, sexting was classified into four types, namely, adult pornographic content, harmful advertisements, smishing[25], and child pornography. When the transaction contained a type, the transaction was codified as '1', and the transaction did not contained a type was codified as ‘0'.

H. Contents of sexting

Based on factor analysis and subject analysis, there are five types of sexting content, namely, nude, sexual intercourse, statutory rape, promiscuous acts, and violence. When the transaction contained a content, the transaction was codified as '1', and the transaction did not contained a content was codified as ‘0'.

Smishing is a compound word of SMS and Phishing, and defines phishing attacks that are sent through text messages in mobile phones.

I. Distribution methodology of sexting

Based on the subject analysis process, there are three types of distribution methodologies, namely, demand, supply, and sharing. When the transaction contained a distribution, the transaction was codified as '1', and the transaction did not contained a distribution was codified as ‘0'.

J. Sexting channels

Based on factor and subject analysis, there are three sexting channels, namely, SNS, online communities, and file-sharing channels. When the transaction contained a channel, the transaction was codified as '1', and the transaction did not contained a channel was codified as ‘0'.

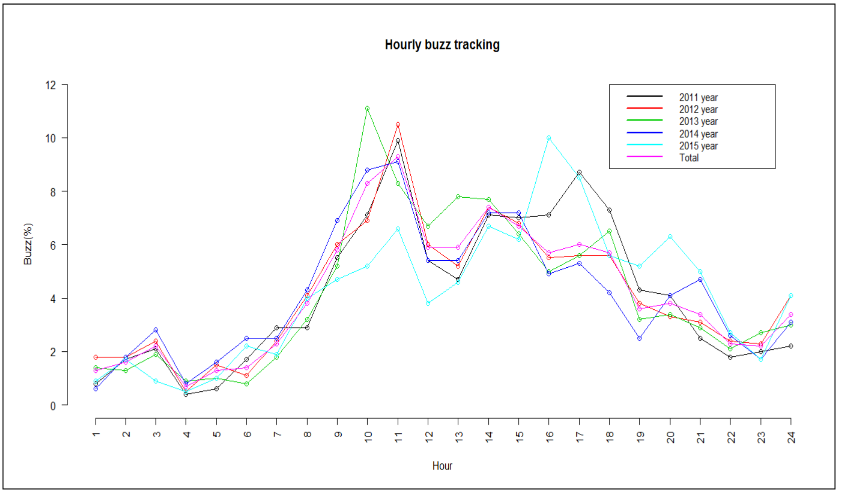
3. Analysis methodology

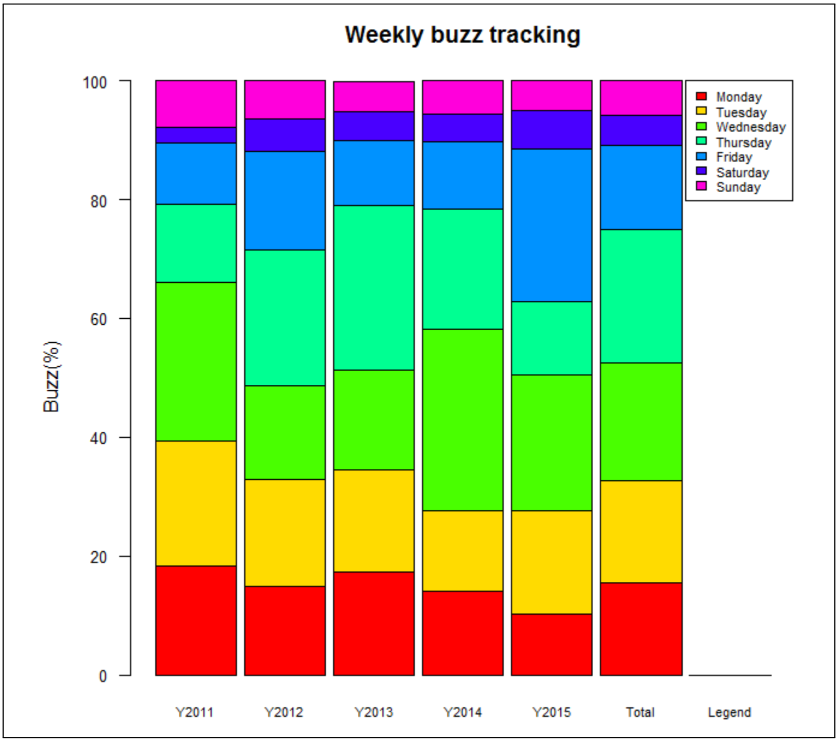
This study utilized the association analysis of data mining and the decision tree method, which do not require special statistical assumptions in order to construct the most effective prediction model explaining the dangers of sexting in Korea. In social big data analysis, association analysis is used to discover correlations in two or more words in an online transaction, and is utilized to discover the conditions and association rules for a group of words that occur together. The association analysis of this study utilized the apriori principle algorithm. The standard of measurement used in the association analysis for predicting the dangers of sexting was based on a support value of 0.02 and confidence of 0.2. The decision tree analysis of data mining automatically calculated the prediction model that best explains the independent variable amidst a sea of data, and made it easier to identify sexting-related factors that have different characteristics. The analysis algorithm for decision tree creation is Chi-squared automatic interaction detection (CHAID). CHAID (Kass, 1980) is a sorting standard for two-sided independent variables, which utilizes a chi-squared (test and searches through all possible combinations to identify the optimal division. Through the stopping rule, the amount of data observed was sufficient and the minimum number of cases for the upper node (parent node) was set at 100, while the minimum number of cases for the lower node (child node) was set at 50, with three levels of tree depth. The technological analysis, multi-level response analysis, and decision tree analysis were performed using SPSS version 22.0, the association analysis and visualization through R version 3.2.1.

Ⅳ. Study results

1. State of sexting-related transactions (buzz)

Even though it differed by year, buzz related to sexting shows a pattern of increasing from 10 am and decreasing after 11am, increasing again from 1pm and decreasing after 3pm, and increasing after 11pm and decreasing after 3am. The sexting-related buzz was the highest on Thursdays and Wednesdays, and decreased over the weekend.





As shown in Table 2, the buzz indicating positive sentiment (danger) on sexting was 38.3% of the total buzz (2011: 51.7%, 2012: 32.4%, 2013: 36.1%, 2014: 46.3%, and 2015: 45.5%). The harmful effects of sexting were, in descending order, sexual crimes (71.2%), defamation (9.5%), and fraud (7.6%). The types of sexting, in descending order, were adult pornography (71.3%), child pornography (16.1%), harmful advertisements (3.8%), and smishing (3.8%). The contents of sexting, in descending order, were sexual intercourse (52.7%), nudity (25.3%), and violence (12%). The help measures for sexting were expert counsel (33.8%), control (29.2%), and preventive education (17.7%). The most popular channels of distribution of sexting were supply (58.2%), demand (22.8%), and sharing (19%), in descending order. Sexting was found to influence studying (61.7%), health (15.8%), and libido (8.1%). The regulations on sexting were the Act for the Promotion of Information and Communications Network Utilization and Data Protection, etc. (61.2%), additional punishment (13.6%), fines (13.3%), and the Act for the Protection of Children and Juveniles against Sexual Abuse (11.9%). Institutions related to sexting were the National Policy Agency (49.6%), Korea Communications Commission (15%), the government (13%), and parliament (9.2%). Channels of sexting were SNS (56.1%), file-sharing channels (33.3%), and online communities (10.6%), in descending order.

〈Table 5〉 State of sexting-related buzz

Type	Item	N(%)	Type	Item	N(%)
Sentiment	Danger	5,277(38.3)	Harmful effects	Defamation	657(9.5)
	Normal	8,497(61.7)		Sexual crimes	4,931(71.2)
	Total	13,774		Fraud	529(7.6)
Channel	SNS	5,820(56.1)			
	Online community	1,094(10.6)		Drinking	417(6.0)
	File sharing channel	3,454(33.3)		Social issues	393(5.7)
	Total	10,368			
Type	Adult pornography	7,440(71.3)		Total	6,927
	Harmful advertisements	919(3.8)	Influence	Study	1,432(61.7)
	Smishing	398(3.8)		Health	367(15.8)
	Child pornography	1,676(16.1)		Social relationships	77(3.3)
	Total	10,433		Cost	133(5.7)

Type	Item	N(%)
Content	Nudity	1,893(25.3)
	Sexual intercourse	3,943(52.7)
	Statutory rape	94(1.3)
	Obscene acts	659(8.8)
	Violence	897(12.0)
	Total	7,486
Aid measures	Preventive education	2,044(17.7)
	Expert counsel	3,916(33.8)
	Encouraging integrity in everyday life	708(6.1)
	Control	3,382(29.2)
	Love	1,530(13.2)
	Total	11,580
Distribution	Demand	2,681(22.8)
	Supply	6,829(58.2)
	Sharing	2,232(19.0)
	Total	11,742

Type	Item	N(%)
	Ethics	123(5.3)
	Libido	188(8.1)
	Total	2,320
Regulations	Additional punishment	2,105(13.6)
	Act on Promotion of Information and Communications Network Utilization and Data Protection	9,455(61.2)
	Fine	2,060(13.3)
	Act On The Protection Of Children And Juveniles Against Sexual Abuse	1,838(11.9)
	Total	15,458
Institution	Korea Communications Commission	1,642(15.0)
	National Police Agency	5,442(49.6)
	Parliament	1,006(9.2)
	Blue House	528(4.8)
	Government	1,427(13.0)
	Judicial authority	590(5.4)
	Civic groups	244(2.2)
	International agencies	82(0.7)
	Total	10,964

2. Analysis of social networks relating to sexting

As shown in Table 3, the association with highest confidence related to the danger sentiment of sexting is {obscene acts, adult pornography} => {danger}.The association of these three variables shows a support value of 0.031, confidence of 0.765, and a lift value of 1.996. This indicates that if an online transaction contains the keywords “obscene acts” and “adult pornography,” the chance of a positive sentiment of sexting (danger) is 76.5%.This means that compared to transactions without these keywords, the likelihood of sexting becoming dangerous is 1.996 times higher. Specifically, the association of the two variables of {child pornography} => {normal} show a support value of 0.074, confidence of 0.609, and support value of 0.987. This indicates that if an online transaction contains the danger sentiment of sexting is 60.9%, and compared to transactions without "child pornography" as a keyword, the chances of sexting appearing negative are lowered by 0.98 times.

〈Table 6〉 Predictions of danger of sexting based on types and content factors

Rule	Support	Confidence	Lift
{Obscene acts, Adult pornography} => {Danger}	0.03092784	0.7648115	1.9963073
{Sexual intercourse, Violence} => {Danger}	0.02279657	0.7511962	1.9607686
{Obscene acts} => {Danger}	0.03455786	0.7223065	1.8853610
{Nudity, Sexual intercourse, Adult pornography } => {Danger}	0.04450414	0.7078522	1.8476324
{ Nudity, Sexual intercourse } => {Danger}	0.05227240	0.6792453	1.7729628
{ Violence, Adult pornography } => {Danger}	0.03071003	0.6438356	1.6805366
{ Nudity, Adult pornography } => {Danger}	0.06780892	0.6401645	1.6709543
{ Sexual intercourse, Adult pornography } => {Danger}	0.13329461	0.6232179	1.6267204
{} => {Normal}	0.61688689	0.6168869	1.0000000
{Child pornography} => { Normal }	0.07412516	0.6091885	0.9875207
{ Adult pornography, Harmful advertisements} => {Danger}	0.02686220	0.6055646	1.5806419
{ Nudity } => {Danger}	0.08269203	0.6016904	1.5705295
{ Violence } => {Danger}	0.03818789	0.5863991	1.5306161
{ Sexual intercourse } => {Danger}	0.16589226	0.5795080	1.5126290
{ Harmful advertisements } => {Danger}	0.03833309	0.5745375	1.4996551
{ Sexual intercourse, Child pornography } => {Danger}	0.02243357	0.5638686	1.4718071
{ Adult pornography, Child pornography } => { Normal }	0.03760709	0.5504782	0.8923487
{ Adult pornography } => { Normal }	0.27152606	0.5026882	0.8148790

Rule	Support	Confidence	Lift
{ Adult pornography } => {Danger}	0.26862204	0.4973118	1.2980809
{ Adult pornography, Child pornography } => {Danger}	0.03071003	0.4495218	1.1733396
{ Harmful advertisements } => { Normal }	0.02838682	0.4254625	0.6896928
{ Sexual intercourse } => { Normal }	0.12037171	0.4204920	0.6816355
{ Violence } => { Normal }	0.02693480	0.4136009	0.6704647
{ Nudity } => { Normal }	0.05474082	0.3983096	0.6456768
{ Child pornography } => {Danger}	0.04755336	0.3908115	1.0200942
{} => {Danger}	0.38311311	0.3831131	1.0000000
{ Sexual intercourse, Adult pornography } => {Normal}	0.08058661	0.3767821	0.6107798
{ Nudity, Adult pornography } => { Normal }	0.03811529	0.3598355	0.5833087
{ Nudity, Sexual intercourse } => { Normal }	0.02468419	0.3207547	0.5199571

3. Factors influencing the dangers of sexting

As shown in Table 4, all factors related to aid measures were found to have a positive influence on the dangers of sexting. In descending order, encouraging integrity in everyday life, love, expert counsel, and control and preventive education were found to be helpful for reducing the dangers of sexting. All influential factors were found to have a positive influence on the dangers of sexting. In descending order, ethics, social relationships, health, and studying were found to have an influence on the dangers of sexting. Among the content factors, all factors except statutory rape had a positive influence on the dangers of sexting. In descending order, obscene acts, and sexual intercourse were found to have an influence on the dangers of sexting. Among the type factors, all factors except child pornography were found to have a positive influence on the dangers of sexting, and in the descending order of adult pornography, smishing, and harmful advertisements. Among the factors that had a harmful influence, all factors except social issues were found to have positive influences on sexting, in the descending order of drinking and sexual crimes. All the factors related to regulations had a positive influence on the dangers of sexting, in the order of the Act for the Promotion of Information and Communications Network Utilization and Data Protection, and additional punishment. Among the factors of distribution, sharing had the most influence on the dangers of sexting.

〈Table 7〉 Factors influencing sexting

Variable		Danger			
		b†	S.E.‡	OR§	P
Aid	Preventive education	.568	.057	1.765	.000
	Expert counsel	.835	.043	2.306	.000
	Encouraging integrity in everyday life	1.326	.103	3.767	.000
	Control	.607	.045	1.834	.000
	Love	.881	.062	2.414	.000
Influence	Studying	1.061	.060	2.891	.000
	Health	1.136	.123	3.114	.000
	Social relationships	1.589	.316	4.901	.000
	Cost	.800	.179	2.225	.000
	Ethics	2.081	.268	8.016	.000
	Libido	.978	.173	2.659	.000

Variable		Danger			
		b†	S.E.†	OR§	P
Content	Nudity	.739	.054	2.095	.000
	Sexual intercourse	1.028	.041	2.797	.000
	Statutory rape	-.425	.242	.654	.079
	Obscene acts	1.595	.092	4.927	.000
	Violence	.664	.075	1.942	.000
Type	Adult pornography	1.067	.037	2.906	.000
	Harmful advertisements	.689	.072	1.992	.000
	Smishing	.992	.112	2.695	.000
	Child pornography	.053	.056	1.054	.341
Harmful effect	Defamation	.412	.086	1.509	.000
	Sexual crimes	1.162	.038	3.197	.000
	Fraud	.632	.096	1.881	.000
	Drinking	1.334	.120	3.795	.000
	Social issues	-.090	.110	.913	.411
Regulation	Additional punishment	.528	.050	1.695	.000
	Act for Promotion of Information and Communications Network Utilization and Data Protection	.643	.042	1.903	.000
	Fine	.438	.051	1.550	.000
	Act forthe Protection of Children and Juveniles against Sexual Abuse	.439	.053	1.551	.000
Distribution	Demand	.170	.044	1.186	.000
	Supply	.006	.035	1.006	.876
	Sharing	.662	.047	1.939	.000

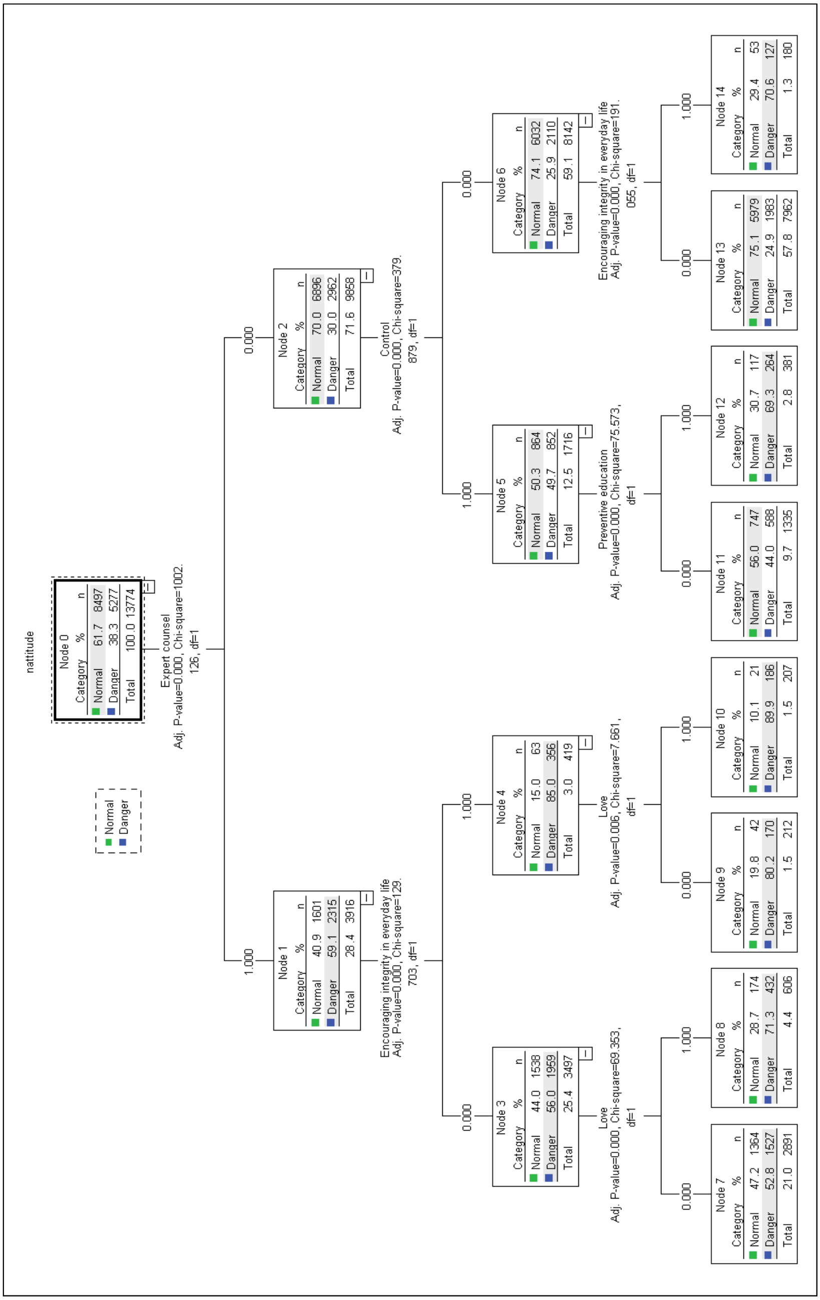
Note: * Normal ranges: Normal, †Standardized coefficients,†Standard error, §Adjusted odds ratio

4. Prediction model for dangers related to sexting

In order to predict the dangers related to sexting, this study carried out data mining analysis on aid factors, content factors, and type factors of sexting. The influence of aid factors of sexting on the danger prediction model of sexting are shown in Figure 5. The square located on the top of the tree structure is the root node, and indicates the frequency of the dependent variable (danger or normal) without considering predicted variables (independent variables). In the root node, the frequency of danger was 38.3% (5,277 hits) and the frequency of normal was 61.7% (8,497 hits). The topmost factor in the lower part of the root node is the most influential (highly associated) factor in predicting the dangers of sexting, and the expert counsel factor was found to be most influential. In the presence of the expert counsel factor, the danger of sexting increased from 38.3% to 58.1%, and the normal aspect decreased to 40.9% from 61.7%. When there are expert counsel factors and integrity encouragement factors, the danger of sexting increased to 85% from 58.1%, and the normal aspect decreased to 15% from 40.9%. As per the gains chart of the danger prediction model of aid factors related to sexting in Table 6, the most influential factor combination for the danger of sexting is the expert counsel-integrity encouragement-love factors. In other words, the index of the 10^{th} node is 234.5%, and the group with the conditions in the 10^{th} node had a 2.34 times higher chance of being exposed to the dangers of sexting compared to the root node. The groups without the expert counsel factor, control factor, and integrity encouragement factors were found to be the most influential forthe normal aspect. In other words, the index of the 13th node is 121.7%, and compared with the root node, the group with the conditions in the 13th node had a 1.22 times higher chance of being normal.

The influence of content factors of sexting on the danger prediction model of sexting is shown in Figure 6. The most influential content factor was found to be the sexual intercourse factor. When this factor was high, the danger of sexting increased to 58% from 38.3%, and the normal aspect of sexting decreased to 42% from 61.7%. When the sexual intercourse factor and the nudity factor were both high, the dangers of sexting increased from 58% to 67.9%, and the normal aspect decreased to 32.1% from 42%. As per the gains chart of the danger prediction model of content factors related to sexting, as shown in in Table 7, the most influential factor combination was found to be the sexual intercourse–nudity–violence factors. In other words, the index of the 12th node is 208.2%, and compared with the root node, the group with the conditions in the 12th node had a 2.08 times higher chance of being exposed to the dangers of sexting. The combinations without the sexual intercourse, obscene acts, and nudity factors were found to be the most influential to the normal aspect. In other words, the index of the 7th node was found to be 119.3%, and compared with the root node, the group with the conditions in the 7th node had a 1.19 times higher chance of being normal.

The influence of type factors of sexting on the danger prediction model of sexting are shown in Figure 7. The most influential type factor was found to be the adult pornography factor, in the presence of which the danger of sexting increased to 49.7% from 38.3%, and the normal aspect of sexting decreased to 50.3% from 61.7%. When the adult pornography factor and harmful advertisement factor were both present, the danger of sexting increased from 49.7% to 60.6%, and the normal aspect decreased to 39.4% from 50.3%. As per the gains chart of the danger prediction model of type factors related to sexting, as shown in Table 8, the most influential factor combination was found to be that without the adult pornography-smishing factors. In other words, the index of the 6th node is 178.9%, and compared with the root node, the group with the conditions in the 6th node had a 1.78 times higher chance of high dangers of sexting. The combination without the smishing factor and harmful advertisement factor were found to be the most influential on the normal aspect. In other words, the index of the 10th node was 125.3%, and compared with the root node, the group with the conditions in the 10th node had a 1.25 times higher chance of being normal.

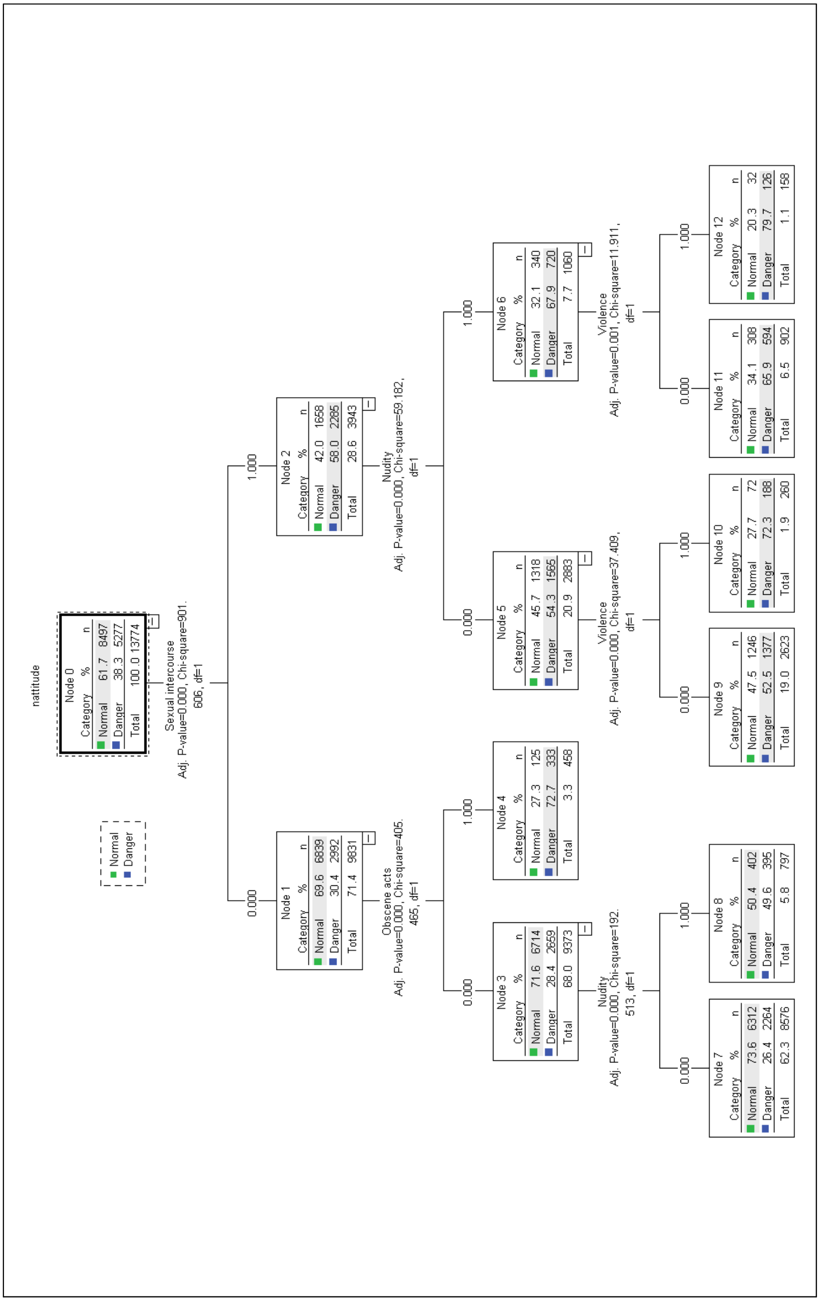


[Figure 3] Danger prediction model of sexting related to aid factors

〈Table 8〉 Gains chart on the danger prediction model of sexting related to aid factors

Type	Node	Gains index				Cumulative index			
		Node(n)	Node(%)	Gains(%)	Index(%)	Node(n)	Node(%)	Gains(%)	Index(%)
Danger	10	207	1.5	3.5	234.5	207	1.5	3.5	234.5
	9	212	1.5	3.2	209.3	419	3.0	6.7	221.8
	8	606	4.4	8.2	186.1	1025	7.4	14.9	200.7
	14	180	1.3	2.4	184.2	1205	8.7	17.3	198.2
	12	381	2.8	5.0	180.9	1586	11.5	22.3	194.0
	7	2891	21.0	28.9	137.9	4477	32.5	51.3	157.8
	11	1335	9.7	11.1	115.0	5812	42.2	62.4	147.9
	13	7962	57.8	37.6	65.0	13774	100.0	100.0	100.0
Normal	13	7962	57.8	70.4	121.7	7962	57.8	70.4	121.7
	11	1335	9.7	8.8	90.7	9297	67.5	79.2	117.3
	7	2891	21.0	16.1	76.5	12188	88.5	95.2	107.6
	12	381	2.8	1.4	49.8	12569	91.3	96.6	105.8
	14	180	1.3	.6	47.7	12749	92.6	97.2	105.0
	8	606	4.4	2.0	46.5	13355	97.0	99.3	102.4
	9	212	1.5	.5	32.1	13567	98.5	99.8	101.3
	10	207	1.5	.2	16.4	13774	100.0	100.0	100.0

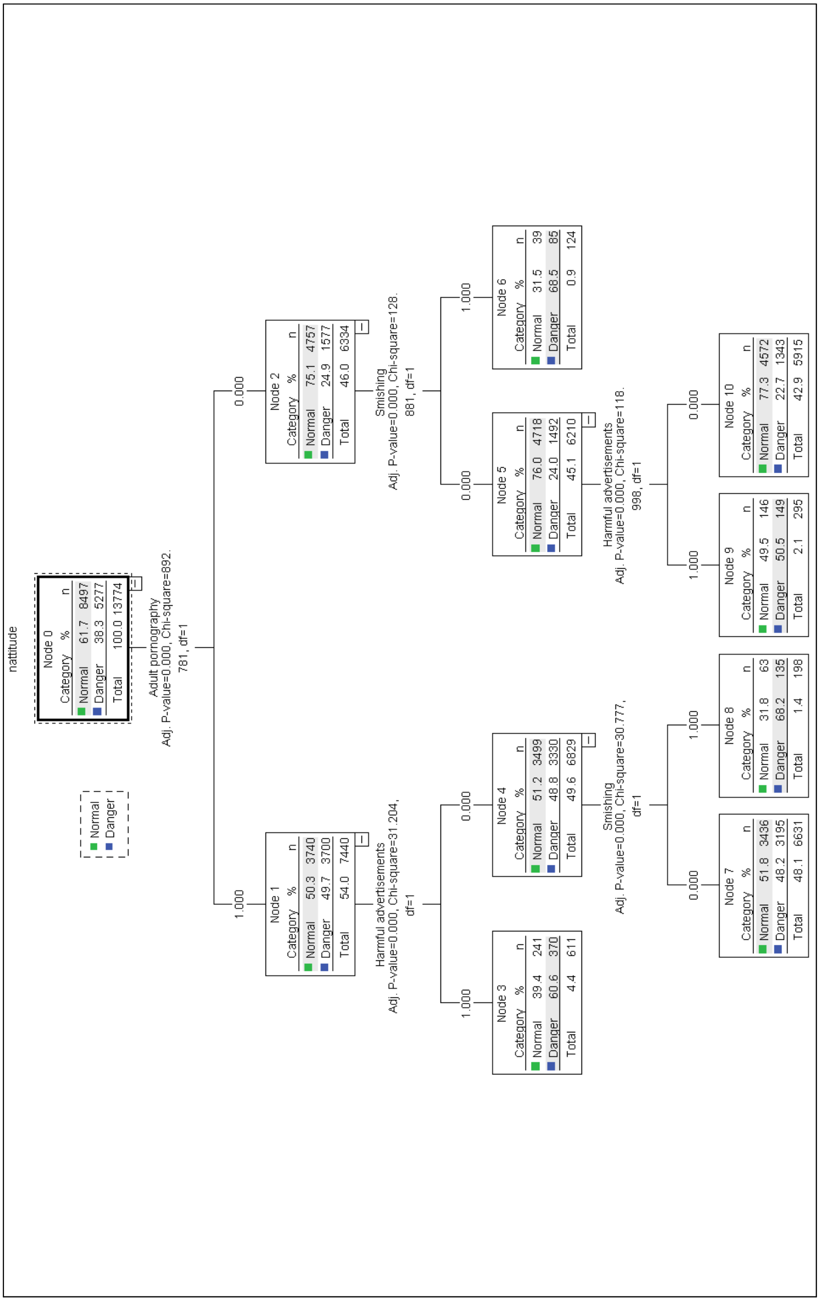
[Figure 4] Danger prediction model of sexting related to content factors



〈Table 9〉 Gains chart on the danger prediction model of sexting related to content factors

Type	Node	Gains index				Cumulative index			
		Node(n)	Node(%)	Gains(%)	Index(%)	Node(n)	Node(%)	Gains(%)	Index(%)
Danger	12	158	1.1	2.4	208.2	158	1.1	2.4	208.2
	4	458	3.3	6.3	189.8	616	4.5	8.7	194.5
	10	260	1.9	3.6	188.7	876	6.4	12.3	192.8
	11	902	6.5	11.3	171.9	1778	12.9	23.5	182.2
	9	2623	19.0	26.1	137.0	4401	32.0	49.6	155.3
	8	797	5.8	7.5	129.4	5198	37.7	57.1	151.3
	7	8576	62.3	42.9	68.9	13774	100.0	100.0	100.0
Normal	7	8576	62.3	74.3	119.3	8576	62.3	74.3	119.3
	8	797	5.8	4.7	81.8	9373	68.0	79.0	116.1
	9	2623	19.0	14.7	77.0	11996	87.1	93.7	107.6
	11	902	6.5	3.6	55.4	12898	93.6	97.3	103.9
	10	260	1.9	.8	44.9	13158	95.5	98.2	102.7
	4	458	3.3	1.5	44.2	13616	98.9	99.6	100.8
	12	158	1.1	.4	32.8	13774	100.0	100.0	100.0

[Figure 5] Danger prediction model of sexting related to type factors



〈Table 10〉 Gains chart on the danger prediction model of sexting related to type factors

Type	Node	Gains index				Cumulative index			
		Node(n)	Node(%)	Gains(%)	Index(%)	Node(n)	Node(%)	Gains(%)	Index(%)
Danger	6	124	.9	1.6	178.9	124	.9	1.6	178.9
	8	198	1.4	2.6	178.0	322	2.3	4.2	178.3
	3	611	4.4	7.0	158.1	933	6.8	11.2	165.1
	9	295	2.1	2.8	131.8	1228	8.9	14.0	157.1
	7	6631	48.1	60.5	125.8	7859	57.1	74.5	130.7
	10	5915	42.9	25.5	59.3	13774	100.0	100.0	100.0
Normal	10	5915	42.9	53.8	125.3	5915	42.9	53.8	125.3
	7	6631	48.1	40.4	84.0	12546	91.1	94.2	103.5
	9	295	2.1	1.7	80.2	12841	93.2	96.0	102.9
	3	611	4.4	2.8	63.9	13452	97.7	98.8	101.2
	8	198	1.4	.7	51.6	13650	99.1	99.5	100.4
	6	124	.9	.5	51.0	13774	100.0	100.0	100.0

Ⅴ. Conclusions and Considerations

This study predicts the dangers of sexting in Korea by analyzing the social big data gathered from domestic online news sites, blogs, internet cafes, SNS, and bulletin boards through network analysis, association analysis of data mining, and decision tree analysis.

The results of this study can be summarized as follows. First, sexting-related buzz increases from 10am, decreases sharply after 11am, increases again after 1pm, decreases after 3pm, increases after 11pm, and decreases sharply after 3am. The buzz related to sexting was highest on Wednesdays and Thursdays, and decreased during weekends. Second, the descriptive words of positive sentiment related to sexting (danger) were focused on “ask,” “shock admission,” “freedom,” “importance,” and “invade.” In the prediction of association of danger sentiment with keywords about sexting, the danger sentiment was found to be strongly associated with “abhorrence,” “obsession,” “stimulation,” “obscene,” “freedom,” and “ask.” Third, the buzz indicating positive sentiment about sexting danger was found to be 38.3% of the total buzz related to sexting. Fourth, the influence of sexting was found to be dangerous for ethics, social relationships, libido, health, studying, and cost, in descending order.Among the distribution methods, the dangers of sharing were higher than that of demand. Fifth, in the network analysis related to sexting, child pornography was found to be closely linked to sexual intercourse, nudity, and violence, and smishing was found to be closely associated with sexual intercourse. Sixth, the association analysis of the content of sexting revealed that if the online transaction mentioned obscene acts and adult pornography, the chance of positive sentiment about sexting dangers was higher. Seventh, the aid factor that helped ease the dangers of sexting was the combination of expert counsel, encouraging integrity in everyday life, and love. For the content factors that influence the dangers of sexting, the factor combination included sexual intercourse and nudity. For the type factors, the combination excluded adult pornography and contained the smishing factor.

Based on this study, the following policy considerations can be derived to address the problem of sexting in Korea. First, as online sexting transactions are concentrated between 11pm and 3am, methods to prevent teenage sexting in the late hours are required. Currently, in order to prevent game addiction among teenagers, the Ministry of Gender Equality and Family operates a shut-down policy, and the Ministry of Culture, Sports, and Tourism operates the a time-choice policy. The shut-down policy restricts teenagers under the age of 16 years from accessing internet games from 12am to 6am, and the game time-choice policy enables the guardians of teenagers to select their game times and limit them from playing in order to prevent teenage game addiction. To prevent sexting, a unification method to ensure consistency between regulations and policies must be considered. Given that sexting takes place in the late hours, it should be unified under the shut-down policy operated by the Ministry of Gender Equality and Family. Second, from 2011 to March 2015, the danger of sexting in Korea was measured at 38.3%. This was found to partially support an average of 28.5%, as researched by the Ministry of Gender Equality and Family (2014), measured at 12.3% in 2011, 20.5% in 2012, and 52.6% in 2014, which averages 28.5%. The fact that the dangers of sexting have continued to increase, except in 2011, is believed to be related to smartphone addiction due to the spread of smart devices among teenagers. Along with developing a solution through preventive education and cure/counseling at government level, various applications should be developed to prevent teenagers from accessing harmful information. Third, the danger of sexting to ethics, social relationships, libido, health, studying, and costs thereof is linked to teenagers sexting to seek attention among friends, and teenagers who engage in sexting have the potential to become sexters in pursuit of mental and emotional satisfaction who do not consider falling grades, humility, or ethics as important (Lee and Kim EG, 2009, p. 101). Fourth, sharing was found to be more dangerous than demand among the methods of distribution sexting. This is because the access point of pornographic content shared by teenagers does not have to pass through a specific internet site and they are able to receive content directly through a one-on-one file sharing network. Therefore, continuous monitoring of the distribution of pornographic content on webhards and file-sharing websites is required on the part of the government. Fifth, the danger prediction of content factors of sexting reveals that transactions without “adult pornography” and with “smishing” sharing websites for distribution of pornographic content. Fifth, the danger prediction of content factors of sexting reveals that transactions without “adult pornography” and with a combination of expert counsel, encouragement of integrity, and love were the most influential for addressing the dangers of sexting. This indicates that a combined school and household approach for combatting sexting is required. Therefore, it is important for schools to teach students to read and understand media messages critically, to communicate to parents the importance of parental monitoring and understanding, and to provide alternatives to receive education about sex and relationships with the opposite gender in an appropriate manner (Flood, 2009 p.143). Moreover, attempts to encourage parental regulation at home are believed to be effective in preventing teenage sexting.

In addition, a government-level response to monitoring sexting is required. Currently, the Korean National Police Agency Cyber Bureau (http://cyberbureau.police.go.kr) has a child/teenage pornography team in place to eliminate pornographic content and focuses on cracking down on webhards, P2P sites, and internet pornography, but monitoring of real-time sexting remains difficult. Therefore, the risk factors of sexting should be analyzed to develop an application that can monitor and prevent dangerous online sexting transactions.

As this study did not analyze individual characteristics, but analyzed data on the overall group, ecological fallacy may occur when this study is applied to individuals (Song et al., 2014). Moreover, the dangers of sexting were defined through sentiment analysis of sentimental words occurring on online transactions. Thus, the definition of danger related to sexting may differ from that in previous offline research. Moreover, the accuracy of the target of analysis may be limited as the online transactions (buzz) that are thought to involve teenagers may include transactions from adults that have teenage keywords (under the age of 10 years, elementary school students, middle school students, and high school students). Despite this limitation, this study is meaningful in its novel research methodology as it quickly and effectively gathered real content on the dangers of texting from social big data, and overcame the limits of previous cross-sectional research that targets certain schools with the precondition of anonymity. Lastly, as sexting is distributed through text messages or smartphones, if the previous cross-sectional methodology through sample extraction is utilized in tandem with the utilization and analysis of big data gathered from social media, the predictions of sexting dangers could be even more reliable.

References

Chalfen.R(2009). it's only a picture. the previous cross-sectional methodology through *Visual Studies,* 24(3); 258-269.

Choi, JY & Chung, DH(2014). Teenagers with Smartphones Exposed to Sexual Content, The Journal of the Korea Contents Association, 14(4); 445-455.

Eberstadt, M. & Layden, M.A,(2010). *The social costs of pornograpy, A statement of Findings and Recommendations*, The Witherspoon Institute.
http://www.lifeissues.net/writers/may/may_17pornographycost.html

Flood, M. (2009). Youth, Sex, and the Internet, *Counselling, Psychotherapy, and Health,* 5(1); 131-147.

Greenfield, P.M.(2004). Inadvertent exposure to pornograpy on the Internet : Implications of peer-to-peer file-sharing networks for child development and families, *Applied Development Psychology*, 25; 741-750.

Gruber, E. & Thau, H.(2003). Sexually Related Content on Television and Adolescents of Color : Media Theory, Physiological Development, and Psychological Impact, *Journal of Negro Education*, 72(4); 438-456.

Itzin, C.(1997). Pornography and the Organization of Intra familial and Extra familial Child Sexual Abuse: Developing a Conceptual Model. *Child Abuse Review*, 6: 94-106.

Jaishankar.K(2009). Sexting: A new form of Victimless Crime?, *International Journal of Cyber Criminology,* 3(1); 21-25.

Joo, JH & Kim, HI(2013). Exploration of relationship among Korean adolescents' sexual orientations, exposure to internet pornography and sexual behaviors after exposure : focused on PLS path modeling analysis. *The Journal of Digital Policy & Management*, 11(6); 11-21.

Lounsbury, K., Mitchell, K.J. & Finkelhor, D.(2011). *The true prevalence of 'sexting',* Crimes against Children Research Center.

Kim, JG(2012). The Predictive Factors on Adolescents` Exposure to Sexually Explicit Online Materials and Adolescents`s Sexuality, *Social Science Studies*, 24(1); 1-33.

Kim, M & Kawk, JB(2011). You Cybersex Addiction in the Digital Media Era. Soon Chunhyang Journal of Humanities, 29; 283-326.

Kim, MN(2012). A Study on understanding of Adolescent Sex Offenses and Pornography Addiction, Korean Association of Addiction Crime, 2(1); 47-71.

Korea Internet and Security Agency(2014). 2014 Mobile Internet use Survey, 19.

Lee, CH & Kim, EG(2009). A Study of Sexting Activites among south Korean Youths. *Korean Institute of Criminology Research series*, 09-17; 11-131.

Lenhart, A.(2009). *Teens and Sexting: How and why minor teens are sending sexually suggestive nude or nearly nude images via text messaging,* Pew Internet & American Life Project.

Malamuth, N. M. & Check, J.V.P.(1985). The Effects of Aggeressive Pornograpy on Belief in Rape Myths: Individual Differences. *Journal of Research in Personality,* 19; 299-320.

McLaughlin, J.H(2010). *Crime and Punishment: Teen Sexting in*

Context, Florida Coastal School of Law.

Ministry of Science, ICT and Future Planning & Korea Internet and Security Agency(2014). 2014 Internet use Survey.

Ministry of Science, ICT and Future Planning & National Information Society Agency(2014). 2014 Internet Addiction Survey.

Ministry of the Gender Equality & Family(2014). 2014 Youth Harmful environment Contact survey.

Mitchell, K.J., Finkelhor, D., Jones, L.M & Wolak, J.(2011). Prevalence and Characteristics of Young Sexting: A National Study, *Pediatrics*, 129(13); 13-20.

Ringrose, L., Gill, R., Livingstone, R. & Harvey, L.(2012). A qualitative study of children, young people and d tative study of children, young people and *National Society for the Prevention of Cruelty to children*, London, UK.

Sherman, S.J. & Fazio, R.H.(1983). Parallels between attitudes and traits as predictors of behavior, *Journal of Personality*, 51(3); 308-345.

Shin, SM.(2013). Associations of Demographic and Psycho-Social Characteristics with Frequent Watching Pornography Material or the Adults-Only Internet Chatting, The Korean Journal of Stress Research, 21(4); 275-281.

Shim, JW(2010). The Role of Timing to Exposure to Pronography in What Adolescent Boys and Girls Think About Sexual Issues, Media, Gender& Culture, 16; 75-105.

Song TM, Song J, An JY, Hayman LL & Woo JM(2014).). Psychological and Social Factors Affecting Internet Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends, Yonsei Med Journal, 55(1); 254-263.

Temple, J.R., Paul, J.A., Berg, P.V.D., Le, V.D., McElhany, A., Temple, B.W.(2012). Teen Sexting and Its Association With Sexual Behaviors, *Archives of Pediatrics and Adolecent Med*, 166(9); 828-833. doi:10.1001/archpediatrics.2012.835

Tokyo Congress(2010). Tokyo Congress C Youth Protection Ordinance revised.

U.S. Department of state(2003). Prosecutorial Remedies and Other Tools To end the Exploitation of Children Today Act of 2003. Title V, Obscenity and Pornography.

Walker, S., Sanci, L. & Temple-Smith, M.(2011). Sexting and young people, *Youth Studies Australia*. 30(4); 8-16.

Ybarra, M.L. & Mitchell, K.J.(2005). Exposure to Internet Pornograpy among Children and Adolescents: A National Survey. *Cyber Psychology & Behavior*, 8(5); 473-486.