ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

Designing a Crime-Prevention System by Converging Big Data and IoT

Jin-ho Jeon[1] Seung-Ryul Jeong[1*]

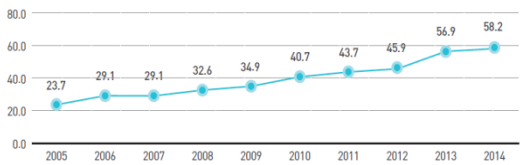
ABSTRACT

Recently, converging Big Data and IoT(Internet of Things)has become mainstream, and public sector is no exception. In particular, this combinationis applicable to crime prevention in Korea. Crime prevention has evolved from CPTED (Crime Prevention through Environmental Design) to ubiquitous crime prevention;however, such a physical engineering method has the limitation, for instance, unexpected exposureby CCTV installed on the street, and doesn't have the function that automatically alarms passengers who pass through a criminal zone.To overcome that, this paper offers a crime prevention method using Big Data from public organizations along with IoT. We expect this work will help construct an intelligent crime-prevention system to protect the weak in our society.

☞ keyword : Big Data, IoT, Crime Prevention, Public Sector

1. Introduction

Crime is a major concern for many urban residents [1]. Various theories have been applied to solve it, e.g., sociological crime theories and economic crime theories, such as rational choice theory and the conception of peer effects as positive and negative externalities[2], Furthermore, the advancement of ICT (information communication technology) has rendered a great service to the physical crime prevention area. Nevertheless, vicious crimes, sexual assault crimes in particular, have risen steadily in Korea during the past 10 years [3].



(Figure 1) Sexual-assault crime progress in Korea

The question in this work is how to reduce the sexual crime increase, as shown in (Figure1) We propose a solution to not only decrease sexual crime but also to defend citizens' rights through ICT, especially Big Data and IoT (Internet of Things). General criminology emphasizes that preventing a crime before it takes place is more efficient than powerfully executing the law after the crime occurs[4]. In such situations, crime prevention is gradually advancing with the increase in information technology.

In this paper, we present a crime-prevention system converging Big Dataand IoT. The methodology is as follows:

1) Collect and analyze Big Data from structured data in crime-related government organizations, like the Ministry of Justice, and unstructured social network data like tweets;
2) Send the results to IoT devices installed throughout the dangerous places; and
3) Alert a woman with a wearable device as she passes a hotspot so she can recognize that she is in a dangerous place and avoid becoming the victim of a sexual crime.

The remainder of the paper is organized as follows. Section 2 discusses related work on converging Big Data and IoT. Section 3 presents an intelligent crime-prevention system. Section 4 lists potential obstacles to the system and, finally, conclusions are presented in Section 5.

2. Related Work

2.1 Advances in Technical Crime Prevention

[1] Graduate School of Business IT, Kookmin Univ., Seoul, 136-702, Korea.
* Corresponding author (srjeong@kookmin.ac.kr)
[Received 24 April 2016, Reviewed 26 May 2016, Accepted 14 June 2016]

Technical crime-prevention methods traditionally depend on CPTED(Crime Prevention through Environmental Design) like CCTV (closed-circuit television) surveillance. However ,physically surveilling, controlling, and monitoring citizens is a privacy invasion[5].

The growth of ICT, e.g., electronic engineering, computerization, artificial intelligence, biochemistry, architecture, material science, etc., not only provides a technical basis to not violate a citizen's rights by analyzing and visualizing data related to society; it also serves as momentum for the crime-prevention industry and allows new technical crime-prevention systems to emerge. (Table1) shows how crime prevention has advanced technically from CPTED to ubiquitous. Further, to realize a safe society, such technical advancements in crime prevention must be constantly pursued.

On the other hand, according to the "Top five violent-crime locations in 2014," announced by the Supreme Public Prosecutors Office in Korea [49], the most frequent place where sexual crimes like rape or molestation occur proves to be 'in the street'. Even though a local government has installed and monitored surveillance devices like CCTV in the street to prevent sexual violence from happening, sexual crime occurs 'in the street' more often than in any other place (see Table 2).

(Table 1) Growth phases of technical crime prevention

Category	Phase 1	Phase 2
Crime prevention theory	CPTED (Environmental prevention)	Ubiquitous crime prevention
Crime prevention method	Crime prevention through environmental designs like CCTV	Backing all lifestyle sfaced by people in society
Year	Jeffery(1971)	Michel Walter(2012)

(Table 2) Violent crimes by place in 2014

Category		Apt	House	In the street	Shop	Accommodation	Entertainment	Office	Station	Transportation	Park	School	Bank	Others
2014 Year	Total	5,671	7,325	39,067	5,068	2,320	5,993	2,940	874	2,079	1,057	679	1,530	56,071
	Murder	22	36	33	-	8	6	9	-	1	1	-	-	42
	Robbery	37	36	92	45	10	17	11	2	1	5	-	-	87
	Rape/ molestation	296	404	1,007	102	471	364	127	144	694	48	26	1	1778
	Burglary	2,765	3,655	11,931	3,988	1,506	2,452	1,216	401	936	472	425	1,482	28,164
	Violence	2,551	3,194	26,004	933	325	3,154	1,577	327	447	531	228	47	26,000

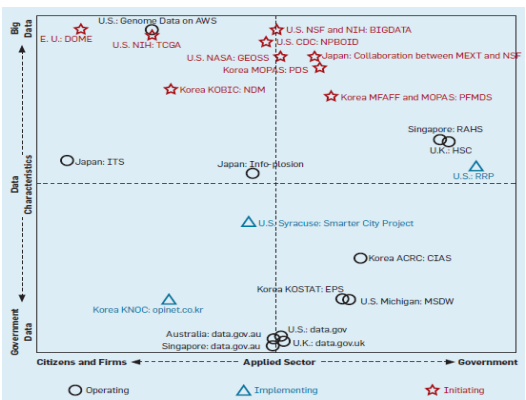
(Table 3) Literature history of ICT crime-prevention technology

Technology Category	Concept	Studies
HCI	Most of this research focuses on preventing face-to-face crimes and increasing the feeling of safety.	[6],[7],[8],[9],[10],[11],[12]
Data	Various data are used in crime prevention technologies. • Analyzing crime data produced by the government information system • Visualizinga crime map using Big Data techniques with social network data, e.g., online newspaper articles, Twitter, etc.	[13],[14],[15],[16],[17],[18],[19],[20], [21],[22],[23],[24]
Mobile	Prototyping mobile crime-prevention system, and emerging mobile platforms and crime maps.	[1],[25]
IoT	Appling wearable computing technologies in crime-prevention sectors	[26],[27],[28],[29],[30],[31]
Algorithm	Predicting and preventing criminal hot spots through algorithms.	[4],[32],[33],[34]

Additionally, Information Communication Technology has become an important element in crime prevention. Approximately 50 – 60 papers on ICT crime prevention from 2004 to 2015 were reviewed and categorized, as shown in(Table3). Through a literature review of ICT crime-prevention technology, five groups were categorized: HCI(Human Computer Interaction), Data, Mobile, IoT, and Algorithm. A summary of the literature review is also given in (Table 3).

2.2 Big Data in the Public Sector

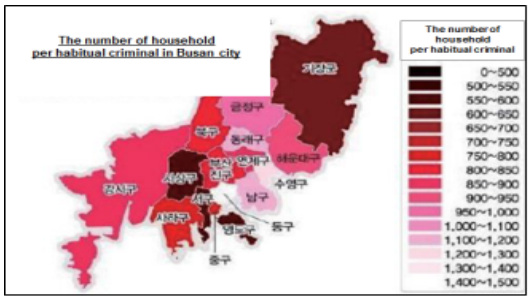
We have already entered an era of Big Data: data sets that are characterized by high volume, velocity, and variety[35, 36]. Furthermore, Big Data has had a huge impact on the public sector, and has begun to provide insight to help support real-time decision-making from fast-growing in-motion data from multiple sources, e.g., the Web, biological and industrial sensors, video, email, and social communications [37]. Korea is no different, and Big Data is being used in various areas like government agencies, e.g., the Ministry of Government Administration and Home Affairs, the Ministry of Health and Welfare, etc.[38].



(Figure 2)Usage of Big Data in the public sector

2.3 Big Data in Crime Part of the Public Sector

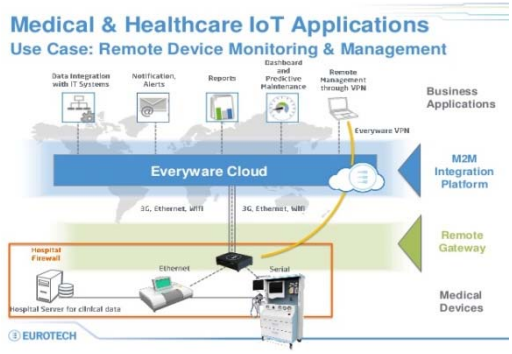
The use of Big Data in the public sector is actively spreading in Korea. As shown in (Figure 3), the Busan police agency analyzed habitual criminals' data (1,575) in each commune in Busan city along with geographical information, and found that the crime distribution in a low-income group in the west range of Busan was wider than the other areas. Therefore, it was recognized that the crime prevention system for the area should be reinforced [39].



(Figure 3)Analysis of habitual criminals' data in Busan city

2.4 IoT in the Public Sector

Gartner says that the Internet of Things Architecture and Platforms, which is one of the top 10 strategic technology trends for 2016[50], has remarkable worth. According to the EuroTech Group, one of the most important aspects of the IoT vision is that smart objects can communicate effectively with each other and with applications residing in data centers or the cloud. As shown in (Figure 4), EuroTech Group develops and offers medical and healthcare IoT[51]. The Korean government also tries to apply it in the public sector.

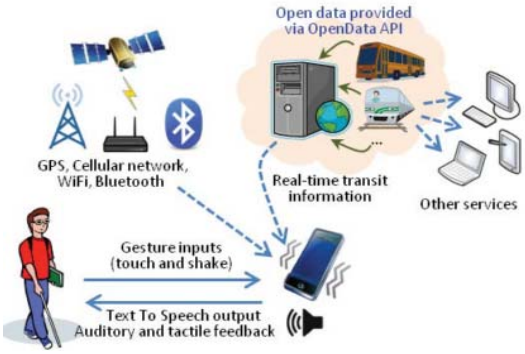


(Figure 4) Medical and healthcare IoT applications

2.5 Converging Big Data and IoT

Big Data and IoT have been variously re-interpreted and studied extensively in academia, at least in the last decade, with

very diverse proposals. For instance, there is a smart city, and due to theirevolution, the city environment is expected to make more effective[26]. (Figure 5) shows the convergence of users' Smartphones, open data infrastructure, and positioning infrastructure, including GPS (global positioning service), Wi-Fi, and cellular networks. The Talking Transit system allows the user to find his current location and search for nearby stops and stations[40].



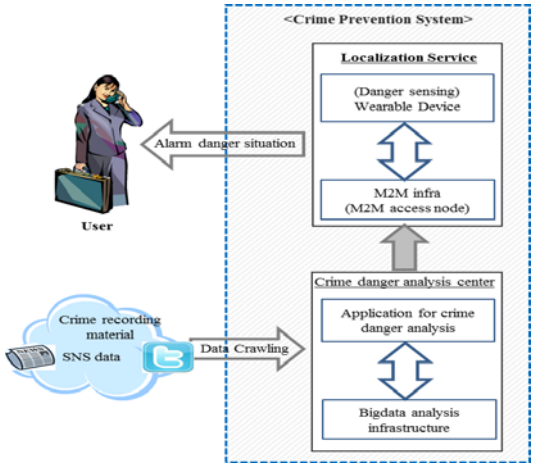
(Figure 5) Talking Transit system architecture

3. PROPOSED SYSTEM ARCHITECTURE DESIGN

According to (Table 2), most sexual crime takes place 'in the street'. As such, women or students passing a crime-ridden area cannot help being scared. To solve this problem, ICPS(Intelligent Crime-Prevention System) has been designed. The main concept is to prevent crimes against the weaker members of society, e.g., women, by alerting their smartphone if they approach a dangerous location, as noted by the crime-risk analysis center.

As illustrated in Figure 6, ICPS consists of a crime-danger analysis component and a localization component.

- Crime-danger analysis component; it discerns dangerous areas by analyzing Big Data obtained by crawling through related organizations and SNSs (social network services).
- Localization component; it notifies a user of a dangerous location through their wearable device as they pass the area, based on an M2M(Machine to Machine) Access Node.

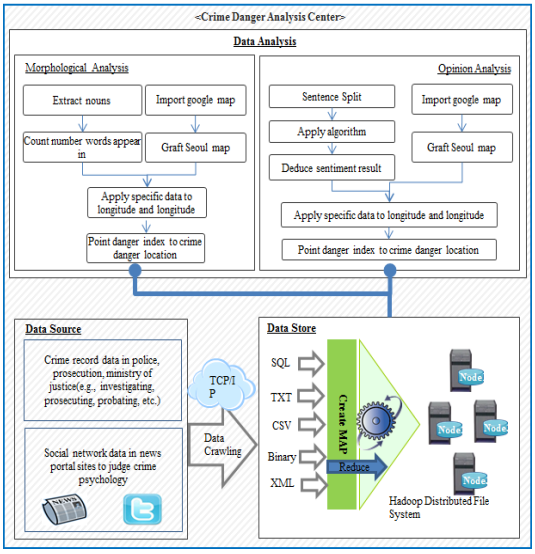


(Figure 6) Concept of crime-prevention system converging Big Data and IoT

3.1 Conceptual CDAC(Crime-Danger Analysis Center)

3.1.1 CDAC Concept

The crime-danger analysis center has the infrastructure to collect and analyze unstructured data produced by the Ministry of Justice, the national police, and social network services like crime-related newspapers.



(Figure 7) Image of the Crime Danger Analysis Center

3.1.2 CDAC Procedure

(Table 4) Data analysis procedure

Procedure	Contents
1) Data acquisition	Data source - Structured data:Crime data produced by crime-related government organizations like prosecution, police, and Ministry of Justice - Unstructured data:Crime-related SNS like tweets, news portals, etc. Method - System connection and data crawling
2) Data storage	Data type - SQL data acquired from a system - Unstructured data like text Method - Distributed storage using Map Reduce
3) Data analysis	Text analysis - Morphemic analysis - Predict dangerous area by applying DF-IDF and similarity algorithms - Combine the latitude and longitude of the relevant location - Deduct crime danger area Opinion analysis - Split sentence - Analyze sentiment by applying algorithm - Combine the latitude and longitude of the relevant location - Deduct crime danger area

3.1.3 Data Acquisition

To design and construct the proposed system, the most important question is how to collect crime-related data. The target and method are as follows.

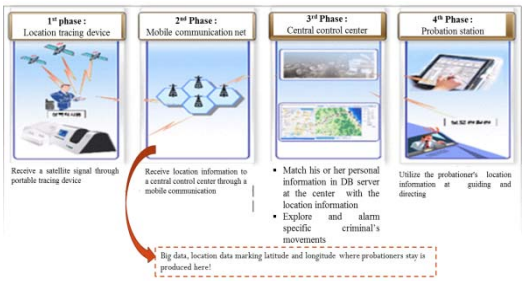
1) Acquisition target

Crime data are connected to government organizations, e.g., prosecutor, police, and Ministry of Justice, and SNS data generated by portal sites and tweets.

- Crime and criminal data in police and prosecution crime-information systems. (e.g., KICS crime data)

KICS(Korea Information System of Criminal-Justice Services) is an electronic work system by which the four criminal-justice agencies (police, prosecution service, courts, and the Ministry of Justice) perform investigation, indictment, trial, and execution work through the standard information system, and jointly use the resulting information and other documents[52].

In addition, the crime prevention bureau at the Korean Ministry of Justice has collected a huge amount of unstructured text data from monitoring and recording probationers' situations in detail,as well as data generated by the location-tracking system that monitors probationers using an electronic monitoring anklet[53].



(Figure 8) Producing unstructured data using an electronic anklet

- Crime-related SNS data: Various Korean news portal sites, e.g., naver.co.kr, daum.co.kr, nate.co.kr, etc. provide accident and incident news reports.

2) Collection method

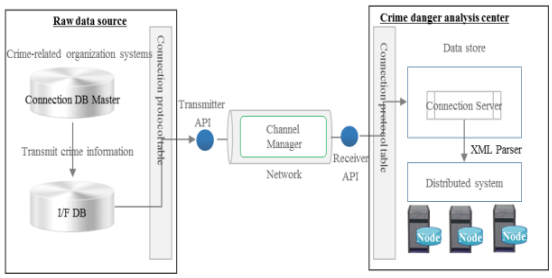
The data acquisition method differs depending on whether the data is structured or unstructured (Table5)[41].

(Table 5) Data collection methods

Data type	Data kinds	Technology
Structured	RDB, spreadsheet	EAI, ETL, FTP, Open API
Semi-structured	HTML, XML, JSON, web document, web log, sensor data	Crawling, RSS, OpenAPI, FTP
Unstructured	Social data(text), image, audio, video, IoT	Crawling, RSS,Open API, Streaming, FTP

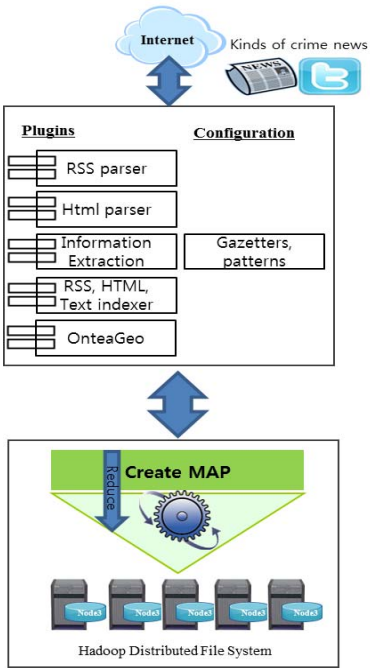
The data-collection method for the proposed design must account for structured data from crime-related government organizations and unstructured data from web sites. The former is collected by system connections and the latter by web crawling.

- Acquisition from government crime-information systems synchronizes mutual protocols, transmits the data to the connection server at the CDAC, and stores it to a distributed system through an XML parser (Figure 9).



(Figure 9) Data Acquisition - EAI(Enterprise Application Integration)

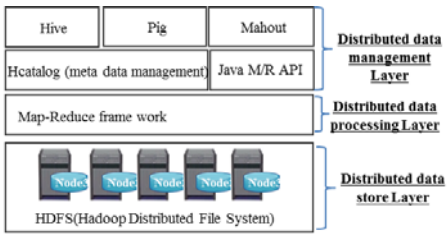
- Acquisition of SNS data is accomplished by web crawling because accident and incident data comes from portal sites with varying data types; the configuration is as follows (Figure 10) [54].



(Figure 10) Data Acquisition - Crawling

3.1.4 Data Storage

Various technologies have been introduced to store big data. The proposed system utilizes Apache Hadoop technology, and the mechanism is as follows.



(Figure 11) Data Storage

- Distributed process layer and storage layer: Materials collected through web crawling are treated through Map-Reduce, and stored in the HDFS(Hadoop Distributed File System). HDFS is not an OS(Operating System) that stores, reads, and writes the files on the local disk of the distributed server, but a file system that handles files with an API. The data storage infrastructure consists of a distributed treatment method like HDFS because of the immense amount of data from crime-related systems and portal sites[55].

3.1.5 Data Analysis

Text mining, opinion analysis, social network analysis, and cluster analysis are Big Data analysis techniques[42, 43]. We decided that text mining and opinion analysis were the proper techniques for crime-danger analysis, for the reasons listed in Table 6.

1) Text mining (Morphological analysis)

The collected data is analyzed using the R tool, and the procedure and processing are as (Table 7).

Algorithm 1 applied in morphological analysis

In this design, the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is used to omit words that cannot discriminate against danger-related words[45]. This technique determines how important a word is to a document in a collection, and scores the importance of words or terms in a document based on how frequently they appear across multiple documents[46].

DF is the counterpart of TF in TF-IDF. The difference is that DF only counts the number of documents containing the word; this is necessary in the context of the writer who writes the document, since multiple appearances of the same word are usually associated with the same event in a single document.

(Table 6) Suitability of analysis techniques for crime-danger analysis

Analysis technique	Technique purpose	Technique utilization for crime danger analysis	Utilization Suitability (1 - 5)
Text Mining	Purpose Extracting and treating useful information based on natural language processing technology Method Using statistical algorithms to analyze natural languages Applied fields Document classification, document clustering, information extraction, document summarization, etc.	Filtering words about crime dangerthat are missing from the assessment function using TF-IDF	5
Opinion Mining	Purpose Assigning positive, negative, or neutral preferences to social media unstructured data Method Classifying the polarity of a given text at the document or sentence level Applied fields Reviews and social media for a variety of applications, ranging from marketing to customer service	Using analysis algorithms like the Naïve Bayesian algorithm to predict potential criminals and the dangerous areas where they act	5
Social Network Analysis	Purpose Discovering viral objects by measuring a user's reputation or influence based on their connection structure or the connection strength of their social network Method Connectivity-based clustering, centroid-based clustering, density-based clustering, etc. Applied fields Biology, computational biology, bioinformatics, etc.	Improve analysis accuracy through information reliability	2
Cluster Analysis	Purpose Grouping a set of objects such that objects in the same group are more similar to each other than to those in other groups Method Connectivity-based clustering, centroid-based clustering, density-based clustering, etc. Applied fields Biology, medicine, business, and marketing	Distinguish latent criminal clusters	2

(Table 7) Morphological analysis process using the R tool

Variable identifier	Analysis processing	Explanation
danger_analysis	<- file(big data storage location)	Put the big data in physical storage device to variable
danger_analysis.lines	<- readLines(danger_analysis)	Read texts in variable 'danger_analysis'
danger_analysis.nouns	<- sapply(danger_analysis.lines, extractNoun, USE.NAMES = F)	Extract the crime words from each line of text
danger_analysis.data	<- data(danger_analysis.nouns)	Deduct the nouns
danger_area_analysis	<- colnames(danger_analysis.data.datacounts) [order(sdev(tfidf(danger_analysis.data.datacounts))) [1:10]][44]	Deduce the ten highly dangerous areas from the TF-IDF terms
danger_analysis.wordcount	<- table(unlist(Cosine_sim(danger_area_analysis)))	Apply vector space modelto compare the similarity among given word clusters
seoul_map	<- danger_analysis.wordcount	Find the most dangerous areas on the Seoul map to determine their longitude and latitude
mapIt	<- geom_point(aes(x = lon, y = lat, size = freq), data = seoul_map)	Apply the longitude and latitude of the relevant dangerous areas to the danger index

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

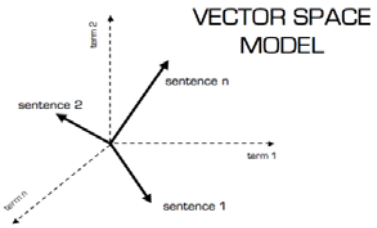
$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

Algorithm 2 applied in morphological analysis

First, we compare the similarity among given word clusters; e.g., a place similar to the clusters where dangerous words appeared. We call this the Cosine Similarity (Vector Space Model). The similarity formula is as follows.

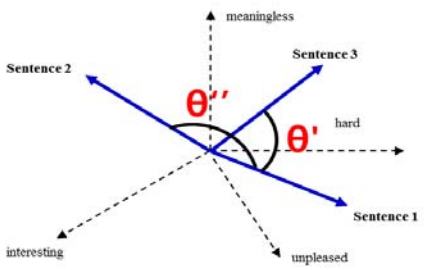
$$\text{similarity} = \cos(\vartheta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Next, we apply the vector space model(Figure 12). The closer the distance between sentence1 and sentence2 is, the more the term similarity is maximized.



(Figure 12) Vector space model

Sentence 1: "I don't have any pleasure in my life anymore."
Sentence 2: "I don't have anything pleasurable."
Sentence 3: "My life is meaningless and hard."



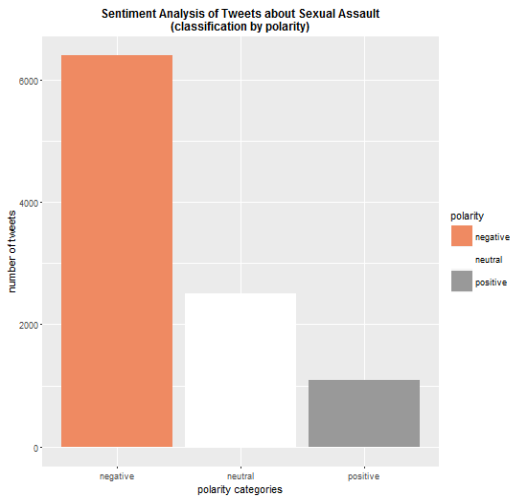
(0 < cosine similarity < 1)

Figure 13) Application of vector space model

2) Opinion mining (sentimental analysis)

- ○ Data source:Twitter
- ○ Data acquisition method: App. from https://apps.twitter.com
- ○ Data count: 10,000 tweets
- ○ Period of data acquisition: January 01, 2016 until January31
- ○ Acquisition keyword:Sexual assault

The steps are shown as (Table 8).



(Figure 14) Classification by polarity

- class_assess<- classify.naivebayes(crime_txt)

POS	NEG	POS/NEG	SENT	
[1]	"9.47547003995745"	"0.445453222112551"	"21.2715265477714"	"positive"
[2]	"1.03127774142571"	"9.47547003995745"	"0.108836578774127"	"negative"
[3]	"67.1985217685598"	"35.1792261323723"	"1.9101762362738"	"positive"

The algorithm considers the negative tweets, traces the place where the most tweets are associated, and reanalyzes the locations of the relevant areas connected to the negative tweets.

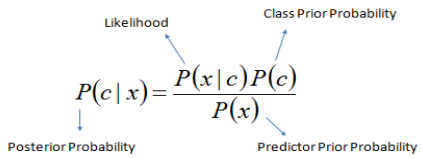
Algorithm applied in sentimental analysis

A Naïve Bayesian algorithm is used to analyze the tweets in the sentimental method, which is the computational study of people's opinions, attitudes, and emotions toward an entity [47].

(Table 8) Sentimental process analyzed by using the R tool

Variable identifier	Analysis processing	Explanation
require (devtools, sentR, twitteR, etc)		Include package 'devtools', 'sentR', etc. for sentimentally analyzing dangerous situation in R
crime_txt	<- gsub()	Acquire and clean the data using 'gsub', 'sapply', etc.
class_emo	<- classify_emotion(crime_txt, algorithm="bayes",prior=1.0)	Apply the Bayesian sentimental algorithm
ggplot(sent_df, aes(x=polarity))	<- data(danger_analysis.nouns)	Analyze the crime-related sentiment to divide tweet sentences into positive, negative, or neutral

- P(c|x) is the posterior probability of a class (target) given a predictor (attribute).
- P(c) is the prior probability of a class.
- P(x|c) is the likelihood,i.e., the probability of the predictor, of a given class.
- P(x) is the prior probability of a predictor.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

(Figure 15) Naïve Bayesian algorithm[56]

3) Analyze result values

We filter the words with no danger discrimination capacity against danger-related words using the TD-IDF and Naïve Bayesian algorithms, and connect the words to dangerous areas, obtaining the results in (Table 9).

(Table 9) Crime-danger area result values

DA	NC	Latitude	Longitude
Hwagok	430	37.538118	126.847102
Mok	327	37.537019	126.873530
Deungchon	319	37.558653	126.853251
Yeouido	317	37.528573	126.928974
Bulgwang	152	37.618309	126.939045
Banghwa	134	37.576456	126.813873
Junggok	131	37.562620	127.085367
Singil	120	37.506249	126.913952
Sillim	116	37.462618	126.938325
Bon	98	37.512442	126.953403

※ DA: Dangerous Area, NC: Number of counts

Continuing, we connect a specific criminal with a dangerous area by assessing the crime-danger similarity analysis; the results are shown in (Table10).

(Table 10) Crime-danger area and criminal result values

DA	Criminal	Latitude	Longitude
Hwagok	Criminal 1	37.538118	126.847102
Mok	Criminal 2	37.537019	126.873530
Deungchon	Criminal 3	37.558653	126.853251
Yeouido	Criminal 4	37.528573	126.928974
Bulgwang	Criminal 5	37.618309	126.939045
Banghwa	Criminal 6	37.576456	126.813873
Junggok	Criminal 7	37.562620	127.085367
Singil	Criminal 8	37.506249	126.913952
Sillim	Criminal 9	37.462618	126.938325
Bon	Criminal 10	37.512442	126.953403

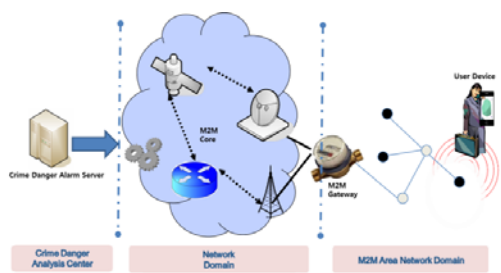
※ DA: Dangerous Area

3.2 Localization Service

The values, words (sorted), crime-danger area information, specific criminal information, and other CDAC results are transmitted to the M2M gateway in each place through an AP (access point), which is a TCP/IP transmitter and receiver.

3.2.1IoT Crime-alarm service architecture

The architecture of the crime-alarm service using IoT depends on the M2M standard technology established by the ESTI (European Telecommunications Standards Institute)[57]. The M2M configuration is as follows.



(Figure 16) M2M Architecture

(Table 11) IoT Application algorithm for constructing ICPS

IoT Applications	Algorithm	Explanation
AP transmitting the crime-danger analysis results	{main begin 1) wait param_src:cm:00 lop:== param_src:m2m_connect_num:0 rop:then 2) load param_dest:m2m_connect: param_src:scrrot_num:0 3) put param_dest:aux:0:00 param_src:dir_num:0000 end} end}	1) Wait until being connected to M2M gateway of relevant crime danger area 2) Connect to the M2M gateway of the relevant crime-danger area 3) Send the danger value to the M2M gateway
M2M Gateway AP	{main begin 1) standby param_dest:aux:0:00 param_src:dir_num:0000 2) {if param_src:user_device:000:000 lop:== param_src:smsnum_num:1 rop:then end} end} end}	1) Stand by to transmit crime danger information to a user passing a relevant danger area 2) Transmit the information when the user passes the relevant area
User AP	{main begin 1) load param_dest:m2m_connect: param_src:scrrot_num:0 2) {if param_src:user_device:000:000 lop:== param_src:smsnum_num:1 rop:then load param_text : "This is a dangerous area! The risk ratio is over 80%." alarm end} end} end}	1) Connect wearable user device to the M2M gateway 2) If there is crime-danger information at the M2M gateway of the relevant area, display an alarm danger signal. "This is a dangerous area! The risk ratio is over 80%."

3.2.2 IoT Application

The elements of IoT Application consist of AP transmitting the crime-danger analysis results, M2M Gateway AP, and User AP. (Table 11) is application algorithm for constructing them.

When a user passes a dangerous area, a crime-danger signal is displayed and an alarm is sent to the user's wearable device as a picture.

(Figure 17) Concept of User Application

4. Potential Obstacles

4.1 Big Data Considerations

1) Personal information leakage and privacy invasion in the public sector

As the crime-recorded data being accumulated at the government-organization information system is personal data, it may leak out and violate someone'sprivacy.

2) Data copyright in the private sector

The proposed design uses social network service data, such astweets or news at portal sites. Twitter data is freely available with the Twitter API, so anyone may evaluate it academically; however, portal sites or newspaper companies that produce news related to accidents may take a different position on ownership.

Reprocessing Big Data from their sites may lead to data copyright problems.

3) Big Data quality

In actual fact, some public organizations and private organizations consider data as one of their most important assets, and the data is constantly issued as Big Data. However this might be irrespective of the quality of data available to such organizations[48].

4) Big Data infrastructure performance

In the process of collecting huge structured or unstructured data, the data may overflow its bandwidth and cause a blackout. In addition, gathering data from government organizations or private companies may activate the site security.

4.2 IoT Considerations

1) Security issues of IoT technology

IoT technology is constantly improving. However ,as it is still in its early stages, security issues may occur; i.e., fake or modified data may be possible and cause unexpected results; e.g., mismatches between an area that claims to be a 'crime danger' and a real one.

2) Instability of IoT infrastructure

When constructing a system utilizing IoT, network noise or data losses are possible, and should be compensated.

5. Conclusion and Future Work

Sexual crime has been increasing in Korea, as well as in many other parts of the world. In Korea, diversified methods are being used to reduce the soaring sexual crime rate. In particular, information systems are very useful in the field of crime prevention.

In this paper, we presented a solution to decrease sexual crime by using Big Data and IoT technology. In the future to build up this system completely, many political issues (i.e., the policy of personal information protection for big data utilization, etc.) or technical factors (i.e., instable IoT infra, etc.)must be settled, and particularly on the basis of ICPS design crime-related government organizations must construct better intelligent crime prevention systems, e.g., installing IoT devices and smartphone applications, collecting and analyzing the crime data in crime-related government organization information systems and SNS data (Twitter, portals, and homepages), and applying the results. Through ICPS, the Korean government expects to reduce sexual violence crimes and take another step toward a secure and safe society.

References

[1] Cvijikj, I. P., Kadar, C., Ivan, B. and Te, "Towards a crowd sourcing approach for crime prevention", UBICOMP/ISWC '15, 2015, pp.1367-1372. http://dx.doi.org/10.1145/2800835.2800971

[2] Wilcox, S. P., "AGENTIZING THE SOCIAL SCIENCE OF CRIME", Winter Simulation Conference, 2011.

[3] Prosecutors's office, "Statistics of Sexual Crime during 10years in 2015 in Korea", Prosecutors' office, 2015. http://www.sppo.go.kr

[4] Tibor Bosse, C. G., "An Agent-Based Framework to Support Crime Prevention", International Foundation for Autonomous Agents and Multiagent Systems, 2010. http://www.few.vu.nl/~{tbosse, cg}

[5] Kyung, J-H, "Meaning and Limitation of technical crime prevention", Korean Criminological Review, 2013.

[6] Erete, S. L., "Engaging Around Neighborhood Issues", 2015, pp1590-1601. http://dx.doi.org/10.1145/2675133.2675182

[7] Erete, S. L., Miller, R. and Lewis, D. A., "Differences in technology use to support community crime prevention", The Powers of Co-location, 2014, pp.153-156. http://dx.doi.org/10.1145/2556420.2556499

[8] Erete, S. L., "Empowerment Through Community Crime-Prevention Technologies", FORUM COMMUNITY + CULTURE, 2014. http://dx.doi.org/DOI:10.1145/2517444

[9] Dillahunt, T. R., "Fostering social capital in economically distressed communities", 2014, pp.531-540. http://dx.doi.org/10.1145/2556288.2557123

[10] Erete, S. L., "Protecting the Home, exploring the Roles of Technology and Citizen Activism from a Burglar's Perspective", Changing Perspectives, 2013.

[11] Lewis, S., "Examining and Designing Community Crime Prevention Technology", ACM, 2012.

[12] Erete, S. L., "Crime Prevention Technologies in Low-Income Communities", XRDS, 2012, Vol .19, no.2. http://dx.doi.org/DOI: 10.1145/2382856.2382867

[13] Guo, B., Wang, Z., Yu, Z., Wang, Y., Yen, N. Y., Huang, R. and Zhou, X., "Mobile Crowd Sensing and Computing", Computing Surveys, 2015, pp.1-31. http://dx.doi.org/10.1145/2794400

[14] Andrew Garbett, J. K. W., Ben Kirman, Conor Linehan,Shaun Lawson, "Anti-Social Media: Communicating Risk through Open Data, Crime Maps and Locative Media", Proceedings of HCI KOREA, 2015.

[15] Rexy Arulanandam, B. T. R. S., Maryam A., "Extracting Crime Information from Online Newspaper Articles", The Second Australasian Web Conference, 2014.

[16] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F. and Pentland, "A Once Upon a Crime", International DOI Foundation, 2014, pp.427-434. http://dx.doi.org/10.1145/2663204.2663254

[17] John (Jong Uk) Choi, S. A. C., Dong Hwa Kim, Angelos Keromytis, "SecureGov: Secure Data Sharing for Government Services", the 14th Annual International Conference on Digital Government Research, 2013.

[18] Youzhong Ma, X. H., JiaRao, Yu Zhang, Weisong Hu, Yunpeng Chai, Xiaofeng Meng, Chunqiu Liu, "An Efficient Index for Massive IOT Data in Cloud Environment", CIKM'12, 2012.

[19] Toole, J. L., Eagle, N. and Plotkin, J. B., "Spatiotemporal correlations in criminal offense records", ACM Transactions on Intelligent Systems and Technology, 2011, pp.1-18. http://doi.acm.org/10.1145/1989734.1989742

[20] Mohammad A. Tayebi, M. J., Martin Ester, Uwe Glässer, Richard Frank, "Crime Walker: A Recommendation Model for Suspect Investigation", RecSys'11, 2011.

[21] Victor Raskin, J. M. T., Christian F., Hempelmann, "Ontological Semantic Technology for Detecting Insider Threat and Social Engineering", ACM, 2010.

[22] Thomas Heverin, L. Z., "Twitter for City Police Department Information Sharing", The College of Information Science and Technology Drexel University, 2010.

[23] Fatih Ozgul, J. B., Hakan Aksoy, "Mining for offender group detection and story of a police operation", Australian Computer Society, Inc., 2007.

[24] Barros, C. P. and Alves, F. P., "Efficiency in Crime Prevention: A Case Study of the Lisbon Precincts", International Advances in Economic Research, 2005, pp.315-328. http://dx.doi.org/DOI: 10.1007/s11294-005-6660-z

[25] Kasper L. Jensen, H. N. K. I., Sebastian Mukumbira, "Toward an mPolicing Solution for Namibia: Leveraging Emerging Mobile Platforms and Crime Mapping", SAICSIT, 2012.

[26] Yokoyama, T., Akiyama, T., Kashihara, S., Kawamoto, Y. and Gurgen, L., "Considerations towards the construction of smart city test bed based on use case and testbed analysis", 2015, pp.1623-1630. http://dx.doi.org/10.1145/2800835.2801633

[27] Cranshaw, J.,"Whose City of Tomorrow" Is It? On Urban Computing, Utopianism, and Ethics", UrbComp, 2013

[28] Amit Sheth, P. A., "Physical Cyber Social Computing for Human Experience", WIMS, 2013.

[29] Bly the, M. A., Wright, P. C. and Monk, A. F., "Little brother: could and should wearable computing technologies be applied to reducing older people's fear of crime?", Personal and Ubiquitous Computing, 2004, pp.402-415. http://dx.doi.org/DOI 10.1007/s00779-004-0309-4

[30] Daniele Quercia, L. M. A., Rossano Schifanella, "The Digital Life of Walk able Streets", The International World Wide Web Conference Committee, 2015. http://dx.doi.org/10.1145/2736277.2741631.

[31] Ghose, A., Sinha, P., Bhaumik, C., Sinha, A., Agrawal, A. and Dutta Choudhury, "A UbiHeld - Ubiquitous Healthcare Monitoring System for Elderly and Chronic Patients", UbiComp'13, 2013, pp1255-1264. http://dx.doi.org/10.1145/2494091.2497331

[32] Theresa L. Hillenbrand-Gunn, M. J. H., Pamela A. Mauch, and Hyun-joo Park, "Men as Allies: The Efficacy of a High School Rape Prevention Intervention", Journal of Counseling & Development, 2010.

[33] Schneider, R. M., "A Comparison of Information Security Risk Analysis in the Context of e-Government to Criminological Threat Assessment Techniques", InfoSecCD, 2010.

[34] Kiyoshi Murata, Y. O. Japanese,"Risk Society: Trying to Create Complete Security and Safety Using Information and Communication Technology", SIGCAS Computers and Society, 2010, Volume 40, No. 2.

[35] Kitchin, R., "Big data and human geography: Opportunities, challenges and risks", National University of Ireland, 2013. http://dx.doi.org/DOI: 10.1177/2043820613513388

[36] Lazar, N., "The Big Picture: Data, Data, Everywhere ….", CHANCE2013, 2013. http://www.tandfonline.com/loi/ucha20

[37] Kayode Ayankoya, A. C., Jean Greyling, "Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication", SAICSIT2014, 2014. http://dx.doi.org/10.1145/2664591.2664619

[38] GANG-HOON KIM, S. T., AND JI-HYONG CHUNG, "Big-Data Applications in the Government Sector", COMMUNICATIONS OF THE ACM2014, 2014. http://dx.doi.org/10.1145/2500873

[39] Kim, J.-P., "Busan crime prevention system using the Big Data", Provincial administrative informative study, 2013. http://www.sasang.go.kr

[40] Jee-Eun Kim, M. B., Noboru Koshizuka,Ken Sakamura, "Enhancing Public Transit Accessibility for the Visually Impaired Using IoT and Open Data Infrastructures", Urb-IoT 2014, 2014. http://dx.doi.org/DOI:10.4108/icst.urpb-iot.2014.257263

[41] Ministry of Science, "Utilization manual for big data usage by stages (Version 1.0)", Ministry of Science, 2014.

[42] Dimitropoulos, E. G. S. B. A. X.,"Visualizing big network traffic data using frequent pattern mining and hypergraphs" First IMC Workshop on Internet Visualization, 2014. http://dx.doi.org/DOI 10.1007/s00607-013-0282-8

[43] Mehdi Mirakhorli, H.-M. C., Rick Kazman, "Mining Big Data for Detecting, Extracting and Recommending Architectural Design Concepts", 1st International Workshop on Big Data Software Engineering, 2015. http://dx.doi.org/DOI 10.1109/BIGDSE.2015.11

[44] Wild, F.,"Latent Semantic Analysis",SnowballC, 2015.

[45] Kazunari Sugiyama, K. H., Masatoshi Yoshikawa, Shunsuke Uemura, "Refinement of TF-IDF Schemes for Web Pages using their Hyperlinked Neighboring", HT' 03, 2003.

[46] Claire Fautsch, J. S., "Adapting the tf idf Vector-Space Model to Domain Specific Information Retrieval", SAC'10, 2010.

[47] Alexey Miroshnikov, E. M. C.,"Parallel MCMC combine: An R Package for Bayesian Methods for Big Data and Analytics", PLOS ONE2014, 2014. http://www.plosone.org

[48] Ayankoya, K., Calitz, A. and Greyling, J. "Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication", 2014, 192-198. http://dx.doi.org/10.1145/2664591.2664619

[49] http://www.sppo.go.kr

[50] http://www.gartner.com

[51] http://www.eurotech.com

[52] http://www.kics.go.kr

[53] http://www.cppb.go.kr/

[54] https://hadoop.apache.org/

[55] http://wiki.apache.org/nutch/

[56] http://www.saedsayad.com/ naive_bayesian.htm

[57] http://www.etsi.org/

◐ Authors ◑

Jin-Ho Jeon
1997 B.S in theology, University of Manila Theological College, Philippines
2005 M.S. in Computer Science, Univ. of Kwangwoon, Korea
2016 Ph.D. in Business IT, Graduate School of Business IT, Kookmin Univ., Korea
1997~Present: Project Manager in Public Sector, LIG System, Korea
Research interests : Big Data, Data Mining, e-Government, Project Management, etc.
E-mail : jhjeon@ligcorp.com

Seung Ryul Jeong
1985 B.A. in Economics, Sogang Univ., Seoul, Korea
1989 M.S. in MIS, Univ. of Wisconsin, WI, U.S.A.
1995 Ph.D. in MIS, Univ. of South Carolina, SC, U.S.A.
1997~Present: Professor, Graduate School of Business IT, Kookmin Univ., Korea
Research Interests: System Implementation, Process Innovation, Project Management, Information Resource Management etc.
E-mail; srjeong@kookmin.ac.kr