WHAT DO PEOPLE TALK ABOUT THE HACKERS GROUP "ANONYMOUS" IN 2015? A TOPIC MODELING APPROACH

[1]KYU HYUN LEE, [2]SUN GYU LEE, [3]JEE-EUN CHOI, [4]HEE-WOONG KIM

[1,2,3,4]Yonsei Graduate School of Informationa,
E-mail: [1]statkyu@yonsei.ac.kra, [2]chunkui89@yonsei.ac.krb, [3]piecastle@yonsei.ac.krc, [4]kimhw@yonsei.ac.krd

Abstract— Not only the formation of various hacker groups led the leak of personal information, but also caused the outbreak of war by paralyzing ICT technology of country or society through hacking. Although hacking personal or public information without permission is still an illegal act, some people refer to Anonymous as the "online avengers"
Therefore, this research aims at examining the different viewpoints of a famous hacker group "Anonyous" in 2015. This research extracted based on Text Mining and analyzed 251,253 different opinions by the utilization of Topic Modeling on Social Network Service (SNS) called Twitter in order to examine different viewpoints of Anonymous.

Index Terms—Anonymous, Hacker, Text Mining, Topic Modeling.

I. INTRODUCTION

Not only the development of ICT consolidated the foundation of turning all forms of information such as data, voice, video, multimedia, and etc into a digitalization, but also allowed people to get access to all technologies connected to ICT at anytime and anywhere. Thus, it is evident that cyber world called Internet is now dominating parts of the daily life of the physical world [1].
However, damage caused by hacking is also expanding along with the ICT development [4]. Not only the formation of various hacker groups led the leak of personal information, but also caused the outbreak of war by paralyzing ICT technology of country or society through hacking [10]. Among many hacker groups, the specific hacker group that this paper will discuss is a famous hacker group called Anonymous, which conducts Hacktivism in order to fulfill their political and social endeavors through hacking society, country, or even specific target that oppose them. People are starting to call this group the "online avengers" through their act of hacking antisocial phenomenon or subjects [7]. For example, Anonymous recently took an offensive behavior by having grand scale cyber attack on Islamic armed group called Daesh that simultaneously conducted terrors in Paris, France. However, although some refer to Anonymous as the "online avengers", hacking personal information without permission is still an illegal act. Therefore, this research extracted and analyzed 251,253 different opinions on Social Network Service (SNS) called Twitter in order to examine different viewpoints of Anonymous. It took an inspection of types of core subject of SNS users through the utilization of Topic Modeling on extracted Twitter data. Therefore, this research examined different types of core subjects discussed by SNS users and how these topics appear in different regions. With this as a foundation, the research will deeply apprehend and analyze opinions of people, and then will provide core subjects and concepts of the forward Anonymous research.
The composition of the research is as followed. The chapter 2 will examine the hacking, hacker, Anonymous and theoretical background of Topic Modeling that were utilized to deduct the research result. In chapter 3, it will discuss the design methodology of the study. In chapter 4, it will talk about the outcome of research analysis. Lastly in chapter 5, it will provide the discussion and implication of the research's conclusion.

II. THEORETICAL BACKGROUND

A. Hacking and Hacker

Hacking is an act that endeavors to access computer resources beyond the authorized boundaries, illegally trespassing computers connected to networks. Cracking is a specific name of a hacking that purposely tries to gain economic profit, paralyze hardware, steal software, or destroying stored data of others.
However, the current meaning of Hacking is comprehensively referred to as two cases written above. The act of purposely operating infringement accident through hacking is called the Hacking Attack. People who conduct hacking are called Hackers, and they are categorized into White-hat Hacker, and Black-hat Hacker [2]. White-hat hackers are referred to security professionals who utilize hacking technique to check vulnerability. On the other hand, Black-hat Hackers conduct hacking for their self-satisfaction or if they think it is a public interest although that hacking is illegal.

B. Anonymous Group

Anonymous is a point organization composed of internet hackers who are active across the world, and is estimated to have about 3,000 members. They are categorized as hactivist(hacking+activist), new type of behaviorists that employs the method of computer hacking to attack society, country, or specific target if they hold oppositions [12].

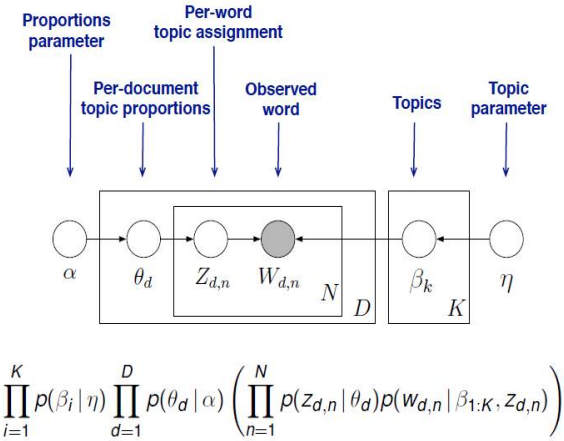
Anonymous is known to first appear in early 2000 on United States community called 4chan, and presented the support of Wikileaks that exposed leaks of diplomatically classified information of U.S. in late 2010. It also conducted a DDOS attack on the websites of Mastercard, Visa, and PayPal, which blocked donations to Wikileaks. In 2010, they also carried a DDOS attack on Iranian government due to the government's severe censorship of information on their citizens. Moreover, when there was the outbreak of Arab democratic movement, Anonymous declared the support of Arab demonstrators by attacking and paralyzing countries under autocracy such as Tunisia, Egypt and more. U.S. newsmagazine Times included Anonymous in the list of "the world's 100 most influential people: 2012" that year [3]. After the simultaneous terror attack on France in November 15, 2015, it conducted a grand scale of hacking attack on Daesh who were at the rear of the terror. In November 17, it recruited members for OpParis to hack Daesh, and presented hacking guidelines to attack Daesh. In November 22, it conveyed large scale of terror information scheduled on eight different places including France, U.S., Italy, Lebanon, and more to countries' information agencies that come under this list. It also hacked and blocked about 20,000 numbers of Twitter accounts related to Daesh. In December 7, 2015, Donald Trump, a front-runner of Republican presidential candidates, argued the need for the regulation of Muslims entering U.S. until there is a specific plan to prevent terrorism by stating that "U.S. should not be the victim of Islam believers' horrendous attack." Accordingly, Anonymous attacked the Trump Tower website, and on the video of their first attack on the Trump Tower Website, they declared that "If U.S. attacks all of Muslims, not the partial extremism of Muslims, the IS would actively utilize these to their social media campaign to interest and recruit new IS members". They also warned that" Trump, think once more before mentioning any matters in the future."

C. Topic Modeling

Topic Modeling is a type of statistical model to express topics of different documents into statistical mixture, and discovering the latent topics that occur in a collection of documents [5]. The most representative technique of Topic Modeling is called LDA(Latent Dirichlet Allocation) algorism that each topics of documents are composed of small numbers of mixtures, and presumes that the advent of vocabularies depend on the topics of documents [5].

In Topic Modeling, under the assumption that vocabularies are not independent, the deduction of posterior probability according to different conditions that generate vocabularies, and expression of Reference [5]'s statistical model can be possible (See Fig. 1). Through the words(W) and documents(D) what we know, we could predict latent topic's proprotions(θ), per-word topic assignment(Z), and topic(β) of a corpus. The Latent Topic(Z) of a word in the document is to be determined by θ value which follows Dirichlet Distribution [6], [9].

This research will utilize Topic Modeling to analyze different viewpoints of Anonymous by categorize them into related topics.



$$\prod_{i=1}^{K} p(\beta_i \mid \eta) \prod_{d=1}^{D} p(\theta_d \mid \alpha) \left(\prod_{n=1}^{N} p(z_{d,n} \mid \theta_d) p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \right)$$

Fig. 1 LDA as a graphical model [6]

III. METHOD

This research will utilize Topic Modeling to analyze different perspectives of Anonymous' declaration of hacking attacks. Moreover, it will categorize different opinions by topics and will indicate different references of viewpoints according to the regions. The procedure of this research was conducted according to the Fig. 2 below.

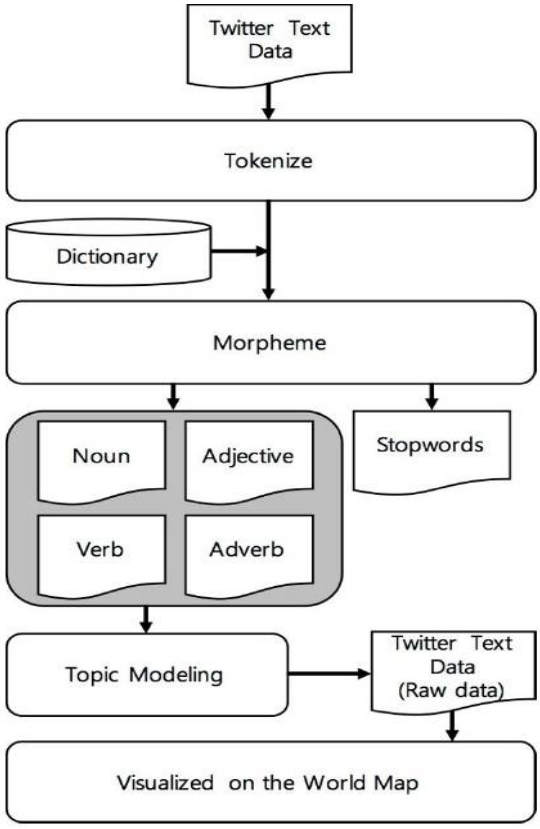


Fig. 2 Research Procedure

This research sectionalizes the procedure into four big steps; Data Crawling, Cleansing, Analysis of Twitter based on Topic Modeling, and marking the map according to the topics found through Topic Modeling, and related viewpoints occurred in different regions.

First, it collected different opinions on hacking organization Anonymous on SNS Channel called Twitter(https://twitter.com) through Twitter API. Twitter users can express their ideas all over the world. Dates of collected data of Anonymous were from December 6, 2015 to December 31, 2015(total of 26 days), and the number of posted Tweets were 251,253, and posts that had location information were 3,156. Second, there is an operation of preprocessing procedures in order to analyze collected data through Topic Modeling. Preprocessing procedures was progressed with Natural Language Process using Standford NLP, and unified data forms by deleting stopwords. Third, through the application of Topic Modeling on collected data, find latent topics of different tweets on Anonymous. This would allow the understanding of topics discussed among Twitter users about Anonymous and the core key words according to different topics. Lastly, express how different topics extracted from Topic Modeling are discussed in different countries on the map.

IV. RESULTS

A. Utilization of Topic Modeling to categorize different opinions on Anonymous

The research analyzed different perspectives of Anonymous through the use of Twitter API by placing collected data into a structural frame for Topic Modeling analysis. It conducted an analysis and classifications of opinions of different Twitter users on Anonymous using Mallet Topic Modeling Package embeded with JAVA to process Topic Modeling analysis [6]. Since there was no preliminary information on Anonymous topic of twitter users, this research assumed the number(η) of topics as 5.

It analyzed total number of 251,253 different Tweets using LDA of Topic Modeling. LDA classifies documents into themes through the use of vocabularies and structures of text format data. Each title of the topics is decided according to the most frequently used vocabularies of researchers [9].

As a result, it summarized top 15 vocabularies that were frequently used on Tweets corressponding to each topic.

'Topic 1' is consist of the words such as trump, Donald, and Muslims that reveals warning of Anonymous to the US presidential candidate Donald Trump's utterance. Also, the attack on his website actually occurred in 2015.

The words like IS, daeshbags, paris which are related to the Paris Terror on November 13th, 2015 are listed in 'Topic 2'. Instead of representing the personal emotion of the individual, the words in this topic indicate distinguishing themes about terrors.

In 'Topic 3's case, this topic consists of the words such as say, youtube, and twitter. We can figure out that this topic has the theme about the SNS channels that Anonymous mainly uses.

Also, in 'Topic 4', there are words like daech, terrorists as well as war. The words comprising this topic show subjective themes that can be attracted by Anonymous.

'Topic 5' is consist of the words such as gift, santa, and angel, and this indicates people's different perspectives on Anonymous.

So, the research deducted five related topics of Anonymous corressponding with Table 1 through the use of LDA such as "Anonymous' warning to Donald Trump's remark on Islam", "Declaration of war against Daesh, culprit of Paris terror", "Channel used by Anonymous", "Target of Anonymous", "Different perspectives of Anonymous".

Table 1. Utilization of Topic Modeling to categorize different opinions on the Anonymous Community in 2015

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
trump	ISIS	say	rape	life
donald	trolling	world	allegation	gift
war	daeshbags	system	daech	santa
declared	opis	youtube	terrorists	secret
launches	declare	twitter	trolls	help
declare	december	account	sociaux	angel
target	friday	people	war	justice
youtube	war	opkkk	gayboy	army
muslims	hacker	like	spotlight	salvation
guerra	terror	kkk	incognition	businessman
message	hacktivist	look	muslims	donor
attack	paris	time	reveal	hampshire
website	media	know	target	world
hacker	news	thing	group	christmas
aim	online	take	no	new

Topic 1	Anonymous' warning to Donald Trump's remark on Islam
Topic 2	Declaration of war against Daesh, culprit of Paris terror
Topic 3	Channels used by Anonymous
Topic 4	Target of Anonymous
Topic 5	Different perspectives of Anonymous



Fig. 3 Overall Opinion Map

B. Distribution of perspectives on Anonymous by regions according to the result of Topic Modeling

The research comprehended different viewpoints of Anonymous in different regions through Topic Modeling. It classifed each of the Tweets by categorizing different topics produced by Topic Modeling into keywords, and demonstrated the 3,156 data that had locations on the map using the Data Visulization program called Tableau.

Fig. 3 represents all of the Tweet data that had location information. As a result, there were mentionings of Anonymous all over the world. Especially, for the five related topics of Anonymous that were deducted by the reuslt of Topic Modeling, many tweets were seen about "Donald Trump" in U.S., "Consecutive terror of Daesh" in Euroep region including France, and "Daesh supporting countries" in Some of Southeast Asian countries.

Fig. 4 demonstrates the distribution of "Anonymous' warning to Donald Trump's remark on Islam" through Topic Modeling, especially in Eastern parts of the United States. The eastern United States is the region that the Democratic Party is politically predominant [13]. Thus, the warning against the representative politician of the Republican Party, Donald Trump's aurgument about regulation of Muslims entering the U.S. ingenerated a keen interest in this for the political supporters of the Democratic Party in the Eastern U.S. People expressed ideas on such topic by commenting "The Anonymous just warned Donald Trump for his stupidity. Can't wait to see what they have on" and "I wonder what Anonymous gonna do to Trump. Hacked into his bank account?" and more.



Fig. 4 Opinion Map about Anonymous' warning to Donald Trump's remark on Islam

Fig. 5 shows the distribution "Declaration of war against Daesh, culprit of Paris terror" through Topic Modeling, espcially in France, the victim of the terror, Europe, U.S., and Southeast Asian countries. In the case of Southeast Asian countries, the 42% of the total population(about 240 million people) is a Muslim [8]. Hence, the support forces in IS could proliferate their forces in Southeast Asia [11]. As a result, we can find out that the support forces' responses to the announcement of hacking attack to the Daesh have revealed through the SNS in Southeast Asian countries. People represented ideas on such topic by making a remark like "Anonymous hacker declare Friday "Troll ISIS Day" Pray for Paris" and "How does Anonymous knows ISIS twitter account? How would anyone know?" and more.



Fig. 5 Opinion Map about Declaration of war against Daesh, culprit of Paris terror

V. DISCUSSION AND IMPLICATIONS

The research utilized the Topic Modeling to analyze SNS channel called Twitter to examine different perspectives of worldwide famous hacker organization Anonymous, and demonstrated distributions of viewpoints according to the regions on map.

Utilizing the Topic Modeling, there are five different viewpoints about Anonymous in 2015. Such as "Anonymous' warning to Donald Trump's remark on Islam", and "Declaration of war against Daesh, culprit of Paris terror" etc. This respresents that because of the the antisocial speech of Donald Trunmp and the terror in Paris, Anonymous announced that hack attack, and there are specific regions like U.S., Europe, and Southeast Asian Countries that mainly discuss about it on SNS (See Table 2).

Table 2. Generated Opinions by Regions about the Anonymous Community in 2015

Topics	Regions
Anonymous' warning to Donald Trump's remark on Islam	The eastern U.S.
Declaration of war against Daesh, culprit of Paris terror	Europe, U.S., Southeast Asian Countries
Channels used by Anonymous	-
Target of Anonymous	-
Different perspectives of Anonymous	-

The Implications of the research was conducted in an endeavor to deeply study and analyze the thoughts of Anonymous and provide core topics and concepts for future Anonymous studies. And by examining the present condition of SNS users' opinions about Anonymous, we could get the point of the concern degree difference about hacker groups between countries. Also, we could find out the new information considered in IT penetration as well as the political orientation of regions and countries through this. In future studies, we will be able to do the in-depth analysis of the hacking issues based on specific regions, and the users' opinion differences between the regions.

The limitation of the research is as followed. First, although there were 250,000 numbers of different data collected through Crawling, only about 3,000 had their location information. There are hardships of collecting data with location since the number of SNS exposing users' location information is decreasing. Second, existence of the word "Anonymous" brought unrelated information to this research.

In the future, not only improving the research's conclusion of finding different perspectives of Anonymous by searching related topics through the utilization of Topic Modeling, but also it will further the study to seek both the positive and negative significance of different perspectives through Opinion Mining.

REFERENCES

[1] A. Gekker, "Legionnaires of chaos: "Anonymous" and Governmental Oversight of the Internet," Digital Formations, Vol. 85, pp. 178-192, 2012.

[2] A. Simmonds, P. Sandilans, "An Ontology for Network Security Attacks," Applied Computing, Vol. 3285, pp. 317-323, 2004.

[3] B. Gellman, "The Worlds 100 Most Influential people : 2012," http://content.time.com/time/specials/packages/article/0,28804,2111975_2111976_2112122,00.html, 2012.

[4] D. B. Parker, "The stratagic values of information security in business," Computer & Security, pp. 572-582, 1997.

[5] D. M. Blei, T. L. Griffiths, and M. I. Jordan, "The nested Chinese restaurant process and bayesian nonparametric inference of topic hierarchies," Journal of the ACM(JACM), Vol. 57, No. 2, pp. 7, 2010.

[6] D. M. Blei, "Probabilistic topic models," Communications of the ACM, Vol. 55, No. 4, pp. 77-84, 2012.

[7] E. Bazelon, "The online Avengers," The New York Times Magazine, 2014.

[8] I. Yusuf, "The Middle East and Muslim Southeast Asia: Implications of the Arab Spring," Oxfordislamicstudies, http://www.oxfordislamicstudies.com/Public/focus/essay1009_southeast_asia.html, 2012.

[9] J. H. Park, and M. Song, "A study on research trens in library & information science in Korea using topic Modeling." Korea Society for information Management, Vol. 30, No. 1, pp. 7-32, 2013.

[10] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and R. Richardson, "CSI/FBI computer crime and security survey," pp. 1-32, 2005.

[11] R. C. Paddock, And I. M. Sentana, "Fears Grow Over Islamic State's Influence in Southeast Asia", THE WALL STREET JOURNAL, http://www.wsj.com/articles/fears-grow-over-islamic-states-influence-in-southeast-asia-1437106924, 2015.

[12] Samuel, A. Whitney, "Hacktivism and the future of political participation," Harvard University, pp. 1-268, 2004.

[13] Wikipedia, "Red states and blue states," https://en.wikipedia.org/wiki/Red_states_and_blue_states, 2015.

★★★