Analyzing the Reflection of Public Opinion in Online Petition Using Media Big Data

1st Ham Tae Wan, 2nd Hwang Jeong Min, Kwon Young Ohk
Kangwon Science High School

Abstract. Online petition such as 'We the people', 'Change.org' and 'Blue House petition' have become a new form of online democracy. Online petitions can be spread easily through various paths such as social media(Facebook, Twitter), online communities and news media. It can deliver new value of direct democracy and lower the barrier to entry of public political participation. Online petition systems are a new emerging form of online democracy, they have been studied little by researchers. We decided to research them and soon we reached a root question, 'Can we accept the number of assenters or signatures shown in the petition as public opinion?'. In this research, we answer this question by analyzing the relation between the media and the petition system using Big Data analysis. Among various online petition system, we decided to research about the South Korean online petition system, 'Blue House petition' because it is one of the most active petition systems. And it also arouses various social conflicts in the South Korean society. We compared the frequency of news articles keywords, agree trend and tweets, and also showed the difference of agree trend according to the emotions based on the opinion mining of news article comments. We were able to classify the petitions by Type N1, Type N2 based on the change of frequency of news articles keywords. Also, we classified the petitions by Type T1, Type T2 according to the change of tweets. Based on our classification, we discovered that there is high correlation between the media and the petition system. And we also found out that online petition seems to have journalistic characteristics, the online petitions create agendas and delivers them. We concluded that it is hard to say that the number shown in the petition system is not exactly the same as public opinion. Also, we discovered that the higher the negative emotion, the more erratic the tweets trends tend to be.

Keywords: Big data analysis, Online petition, Public Opinion, Tweet, Scraping, , Opinion Mining, Emotion analysis, Blue House Petition

I. Introduction

1.1 Motivation of research

Blue House Petition also known as National Petition of South Korea, was opened bu president Moon Jae-In in August 17, 2017. It is a communication platform that can communicate directly with the people. It is made under the philosophy "The government answers if the public asks". The overall system is similar to the White House's national petition site "We The People". Anyone can post a petition by simply logging into SNS, and there are no limits on posting a petition. If the posted petition gets more than 200,000 assenters within 30 days, government officials have to answer by video. Up to now, answers to 53 petitions have been posted on the website of Blue House Petition, and three are waiting to be answered. (October, 23rd, 2018)

It is being actively used since it is being processed faster than "Drum for the people", another petition system in Korea, which started its service in since 2008. Blue House Petition is interesting that the public is creating a new value of direct democracy, which proposes and expresses opinions.

This research began with a question, how much public opinion the petition reflects. In the current petition system, only the number of people who have agreed to the petition is shown. A petition is accepted if there are more than 200,000 assenter. We were curious whether, '200,000 assenter' is a reasonable criterion for judging public opinion. After looking into prior studies in Korea and abroad, no research analyzed the reflection of public opinion of online petition such as 'Blue House Petition', and 'We The People'. In order to achieve the purpose of Blue House Petition for reflecting public opinion on social issues, we thought research on the reflection of public opinion in online petition system was necessary.

The purpose of this study is to check whether the petition system can reflect public opinion. According to many previous studies on media and social big data analysis, frequency of specific keyword on the internet, comments and the number of SNS posts tend to be similar to the actual public opinion. [7] Therefore, we thought understanding the relationship between the petition system and the media was necessary for the research. We used Python to scrap specific keyword references in internet articles and related tweets. In addition, we figured out the public's emotional feelings toward the petition by analyzing emotion of the comments in the news article. In this study, we compare it with the change of assenters in Blue House petition to understand the relationship between the current petition system and the media. Then we will check if the petition system can reflect public opinion.

II. Literature review / Background research

2.1 Opinion Mining

Opinion mining is one of the technologies of text mining related to consumer‘s emotion. It extracts subjective tendencies and opinions of people from text. [6] Opinion mining uses a technique called Sentiment Analysis. It is now used to analyze the emotion of various opinion text such as goods, books and movie reviews. This Opinion Mining consists of data collection, subjectivity detection, and polarity detection. It collects from social media such as Twitter and Facebook and exclude parts (gender, age, text author) that are seemed to be non-subjective in the text collected. [4] In the polarity detection step, the emotion of the text is analyzed by giving scores according to the frequency of specific words or emotional properties such as positive and negative in which each word appears. There are two typical techniques of emotion analysis. First is the polarity detection based on machine learning. Second is the property unit analysis based on emotion lexicons. In case of machine learning, it is necessary to produce emotion judged data made by researchers. For this reason, we use opinion mining based on emotion lexicons, considering the subjective judgment of the researchers. Algorithms such as Naïve Bayes and SVM(Support Vector Machine) enable emotion analysis through emotion lexicons. It is necessary to extract emotion lexicon before emotion analysis. However, KOSAC(Korean Sentiment Analysis Corpus) and Korean Twitter Emotion Analysis are made by other researchers. Since only basic words exist in these emotional dictionaries, pre-processing operations such as morpheme analysis are required. Korean Twitter Emotion Analysis has an emotion lexicons including swear words and neologisms that appear in internet society these days, but KOSAC does not. Therefore, in this study, the Korean Twitter Emotion Analysis is selected as an appropriate model for our emotion analysis. [12]

2.2 Public Opinion

The word "Public Opinion" was first used by Michel Eyquem de Montaigne in 1588. [10] Since then, there have been many definitions of public opinion. However, there are difficulties on defining public opinion because it is an abstract concept. This study used the concept of public opinion defined by Blumer. H and Hennessy B. C. According to Blumer, public opinion is "the opinion of the public formed by the public with issues". [9] In addition, Hennessy defined public opinion as "various opinions expressed by a number of people on issues that has importance." [11] A text-mining analysis of tweets generated during the second TV debate of the 18th presidential election on Dec 10th, 2012, produced tweets reflecting the public's interest, such as "high income bracket," "boosting the underground economy", "free discussion" and "circular investment". Also, 'PR&Korea', a specialized company for public opinion analysis, analyzed the amount of big data buzz(referred to related keywords on SNS) excluding retweets from November 27th to December 18th, 2012(the election campaign period). As a result, the amount of related buzz for each candidate was 1,648,264 with the candidate, Park Geun-hye, and 1,646,717 with the candidate, Moon Jae-in. In the actual 18th president election, Park Geun-hye was elected by 51.6 percent, which is very similar to the results. In other words, you can analyze public opinion defined by Blumer and Hennessy from social media, especially on Twitter. Therefore, in this study, Twitter was used as a sample group to understand public opinion. [8]

III. Research Method

3.1 Scraping Petitions

First we gathered various information about the top 1000 petitions on a specific period (starting from August 15th, 2018 to September 14th, 2018). Among those, we selected petitions for our analysis based on the two criteria. The first criteria is, petitions have to get more than 100,000 assenters if the petition ends before September 14th. The second criteria is, petitions have to get more than 10,000 assenters on September 14th. Eventually, we were able to select 24 petitions. After this, in order to scrape internet news article and tweets, we extracted keywords from the petition. The keywords were extracted carefully so that no objection was raised from all the researchers, considering the context and related news article. The result is shown in **Table 1** Title, Starting Date, Keyword and classification codes(which we made it for convenience) are shown in **Table 1**.

Table 1. Selected Blue House Petition

Case	Petition Title	Starting Date	Keyword
A	Seongnam Mayor Eun Soo-mi and Gyeonggi Governor Lee Jae-myung, who are involved in the illegal gangster Comatrade, resign immediately.	2018.07.22.	Comatrade
B	We call for a special investigation into File Hosting Service cartel and the digital sex abuse industry.	2018.07.29	File Hosting Service cartel
C	I urge the Ministry of Gender Equality and Family to disclose the results of its budget use.	2018.08.15.	Ministry of Gender Equality and Family / Budget
D	Gangneung Cat torso murder	2018.08.17	Cat / torso murder
E	Please abolish the performance in the army, which is filled with "sex commercialization"	2018.08.17.	Sex commercialization / Performance in army
F	We require the government to act on 540 bear problem.	2018.08.17	Breed bear
G	I strongly argue the Korean government to ban the Ministry of Gender Equality and Family!!	2018.08.18.	Banning of the Ministry of Gender Equality and Family
H	Perpetrator of suicide of Incheon middle school girl must be punished	2018.08.19.	Suicide of Incheon middle school girl
I	I'd like to ask for investigation into the suspicions surrounding the Ministry of Justice’s response to the lawsuit against Elliot.	2018.08.22	Elliot / lawsuit
J	Please close schools nationwide.	2018.08.22.	Storm / Closure of school
K	Ban lending stocks of the national pension	2018.08.24	National pension
L	Investigate the dog abuse incident of Cheongju animal shelter which locked the dog in the refrigerator alive	2018.08.25.	Cheongju animal shelter
M	I want the right of the Ministry of Health and Welfare to immediately punish an abortion surgeon.	2018.08.28	Ministry of Health and Welfare / abortion surgery
N	We ask for greater victim protection and punishment of perpetrators in the school 'hidden camera' incident.	2018.8.29.	School hidden camera
O	Please withdraw your nomination for Education Minister Yoo Eun-hye.	2018.08.30.	Yoo Eun-hye / Minister
P	Deprive International soccer referee, Kim Dae-yong	2018.09.01	Kim Dae-yong / International soccer referee
Q	President Moon Jae-in, Keep your promise	2018.09.03	Moon Jae-in / Promise

R	Is director of kindergarten a thief?	2018.09.04	Director of Kindergarten
S	Guarantee the rights of labor of 80,000 dental hygienist	2018.09.05.	Dental Hygienist / Labor
T	Let go of my husband's resentment!	2018.09.06	Sexual Harassment / Resentment / False Charge
U	I'm calling to discipline the judge of "Bobae-Dream sexual harassment" case.	2018.09.08.	Bobae-Dream / Sexual Harassment
V	The sexual minority hate rallies that started at the sexist minority rallies, should not be any longer.	2018.09.08	Sexual Minority Rally
W	Even though I was not guilty on sex crime, I got a court ruling. It is really unfair.	2018.09.09	Sex Criminal / Court Ruling
X	Please strongly punish prostitution as an anti-humanistic act to buy and sell people's human rights.	2018.09.10	Prostitution

3.2 Scraping Related Internet News Article and Tweets

With keywords extracted from each petition, we scraped related internet news article and tweets. We were able to gain keyword reference data which shows how often the keyword was mentioned in news article and tweets. We also gained data for emotion analysis by adding codes that collects comments of the news article.

3.3 Emotion Analysis

Based on 'Emotion Lexicon' from 'Korean Twitter Emotion Analysis' [12], which sorted the emotion of a human in to 6 types (Anger, Disgust, Fear, Happiness, Sadness and Surprise), we analyzed the emotion of the scraped comments. By going through the process of opinion mining, we were able to gain emotion score for each petition.

To use emotion score for our analysis we defined an index named γ. γ can be also called 'negative emotion index', since it shows how negative the comments are. The higher γ, the more negative the emotion is toward the petition. We calculated γ by adding the emotion score of Anger, Disgust, Fear and Sadness and subtracting the emotion score of Happiness. We excluded Surprise since it can be sorted as neutral emotion. The equation is shown below.

$$\gamma = Anger + Disgust + Fear + Sadness - Happiness$$

3.4 Analyzing the Reflection of Public Opinion

We graphed the data of the trend of assenters and keyword reference in news and twitter using Pyplot. We identified the relation between data by overlapping the graph of assenters of petition (Agree) with graph about the frequency of related internet news articles (NKF, News Keyword Frequency) and related tweets (Tweets).

IV. Result

4.1 Comparative analysis of Agree and NKF

According to the analysis, there are two types based on the trend of the graphs. We classified the petitions into two types and the result is shown in **Table 2** We named Type N1 if the peak of NKF comes after the peak of Agree. Type N2, on the other hand, is a type when related keywords in the news article are found before the peak of Agree and NKF have the maximum value after the peak of Agree.

Table 2. Two Types of Results (Relation between Agree and News Keyword Frequency)

Type	Case
N1	A, B, D, F, H, J, S, U
N2	C, G, I, K, L, M, N, O, P, Q, R, T, V, W, X

We were able to classify 8 petitions in to Type N1 and 15 petitions in to Type N2.

Fig 1 is the graph of Case B ('Digital Sexual Abuse in File Hosting Service') which is classified in to Type N1. **Fig 2** is the graph of Case X ('Punishment of Prostitution') which is classified in to Type N2. Both figures allow you to compare Agree and NKF. In the figures, Agree is marked as blue circle marker and NKF is marked as red triangle marker. The horizontal axis of the graph shows time(day) and the vertical axis shows keyword references.

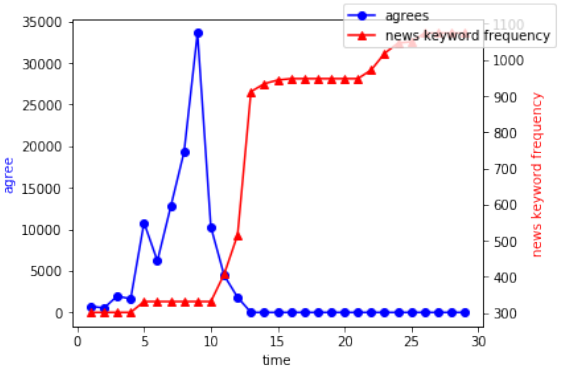


Figure 1. Graph of Agree and NKF in Type N1

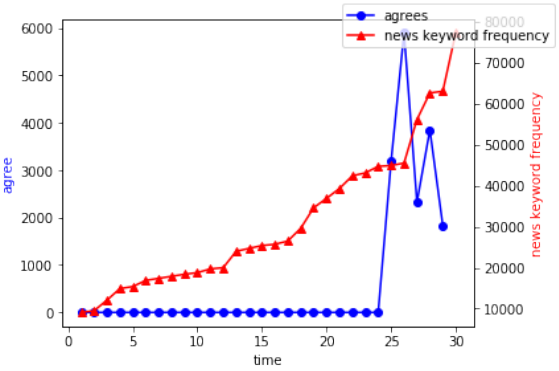


Figure 2. Graph of Agree and NKF in Type N2

4.2 Comparative Analysis of Agree and Tweets

According to the analysis, there are two types based on the trend of the graphs. We classified the petitions into two types and the result is shown in **Table 3** We named Type T1 if the peak of Tweets comes after the peak of Agree. Type T2, on the other hand, is a type when related keywords on twitter are found before the peak of Agree and Tweets have the maximum value after the peak of Agree.

Table 3. Two Types of Results (Relation between Agree and Tweets)

Type	Case
T1	B, D, E, F, H, J, L, O, P, S, T, U, V, W
T2	A, C, G, I, K, M, N, Q, R, X

We were able to classify 14 petitions in to Type T1 and 10 petitions in to Type T2.

Fig 3 is the graph of Case D ('Gangneung cat torso murder') which is classified in to Type T1. **Fig 4** is the graph of Case I ('Elliot's response of Ministry of Justice') which is classified in to Type T2. Both figures allow you to compare Agree and Tweets. In the figures, Agree is marked as blue circle marker and Tweets is marked as red triangle marker. The horizontal axis of the graph shows time(day) and the vertical axis shows the number of related tweets.

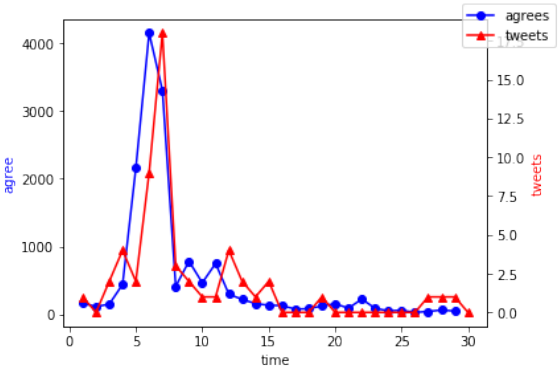


Figure 3. Graph of Agree and Tweets in Type T1

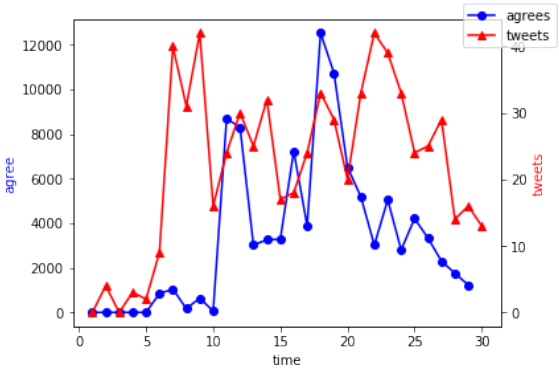


Figure 4. Graph of Agree and Tweets in Type T2

4.3 Comment Emotion Analysis in Related News Articles

We analyzed the comments for each petition based on six emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise). The result is shown in **Table 4**.

Table 4. Relation between Petition Case and Emotion Score (Top Three and Bottom Three Values are Highlighted in Bold.)

Case	Anger	Disgust	Fear	Happiness	Sadness	Surprise
A	2.661241	3.08568	1.180943	2.112604	2.234615	1.512811
B	3.595928	3.96724	1.65058	2.918553	3.302004	2.20732
C	3.490000	3.898500	1.281600	2.606700	2.796600	1.879100
D	2.681900	2.812600	1.083800	2.433300	2.568900	1.673200
E	3.287900	3.412500	1.533900	2.770300	2.971000	**2.395700**
F	4.548266	3.506015	1.498487	3.383937	**4.260066**	2.009714
G	3.310630	3.780302	1.224048	2.388109	2.517987	1.710839
H	3.167200	3.694000	1.231000	2.410000	2.897600	1.571500
I	3.503800	4.470600	1.425700	2.641400	2.763200	1.966000
J	2.122955	2.085251	0.883998	1.663268	1.986814	1.117968
K	3.986444	4.764295	1.388749	2.809165	2.849549	2.050786
L	2.189280	2.281794	1.000110	2.008651	2.048374	1.320458
M	3.391519	3.672687	1.349910	2.832540	3.166632	1.989006
N	2.960129	3.227061	1.476860	2.417974	2.801733	1.932903
O	3.135242	3.659654	1.211545	2.158868	2.252164	1.581725
P	2.574348	2.873426	1.407509	1.723200	2.082024	1.696652
Q	3.307675	3.979046	1.203120	2.444616	2.512916	1.812147
R	3.233408	3.264042	1.349408	2.637906	2.978165	1.887235
S	**4.616024**	**4.834544**	**1.833649**	**3.798045**	4.223539	2.090840

T	2.950920	3.261900	1.160143	2.336133	2.528942	1.668981
U	3.307391	3.614120	1.529735	2.552690	2.853595	2.023185
V	3.681620	3.866936	1.602447	3.090309	3.095191	2.038000
W	3.369267	3.685397	1.678812	2.614833	2.897774	2.008866
X	3.072668	3.416291	1.433892	2.484098	2.783315	1.894843

Table 5. Relation between Petition Case and γ (Top Three and Bottom Three Values are Highlighted in Bold.)

Case	γ	Case	γ
A	7.049875	M	8.748208
B	9.597197	N	8.047809
C	8.860243	O	8.099737
D	**6.713995**	P	7.214107
E	8.435239	Q	8.558141
F	**10.428897**	R	8.187117
G	8.444858	S	**11.709711**
H	8.579953	T	7.565772
I	9.521993	U	8.752151
J	**5.415750**	V	9.155885
K	**10.179872**	W	9.016417
L	**5.510907**	X	8.222068

In **Table 4**, you can see that Case S shows the highest Anger, Disgust, Fear and Happiness. Case F showed the highest Sadness and Case E showed the highest Surprise.

In **Table 5**, Case F, K, S shows the highest γ and Case D, J, L shows the lowest γ in which γ represents how negative the comments are.

V. Conclusion

5.1 Relation of petition and media and its effect on the formation of public opinion

In this research, we have classified the petitions based on the peaks of the graph in the Type N1, Type N2 and Type T1, Type T2. In all types, after the peak of Agree came the increasing trend of NKF and Tweets. Other words, after the petition starts, public opinion is formed based on the petition.

However, there are difference between Type N2, Type T2 and Type N1, Type T1. In Type N2 and Type T2, public opinion existed before the petition formed a petition. Unlike the previous types, in Type N1 and Type T2, public opinion formed after the petition. Through this, we were able to found out there is high correlation between the media and the petition system that the petition and the media induces each other. There are petitions that starts from social issues, there are petitions that starts from personal issues. In Type N1, personal issue became public by the news media after the start of petition.

However, it needs to be addressed specifically if the comments in the news continue, and the petition begins after reaching a certain level as in the case of Type N2. At this time, it is more reasonable to explain that large public opinion which had already been formed by the news media, results in a petition. However, the petition is also an important factor in this time. We were able to found out that public opinion which had already been formed through news media, has increased even more after the petition. In this aspect, the petition amplifies public opinion.

5.2 Consideration of emotional analysis on comments of related news articles

Based on our emotional analysis, we compared the trend of the graphs of the top 3 and the bottom 3 in **Table 5** based on gamma index. Case F, K, S showed the highest gamma index and Case D, J, L showed the lowest gamma index. Gamma index is a relative value of how negative the comments are. We predicted that the higher the γ, the more likely there will be differences in Agree, NKF, and Tweets due to the intense emotion. According to our result, petition with high gamma index (Case F, K) appears to have more turbulence in Tweets relatively to petition with low gamma index (Case D, J, L). Unlike Tweets, we could not find any differences in NKF and Agree.

Since emotional analysis is not a perfect technology and using Bag of Words method which has poor accuracy, the results of emotional analysis need to be carefully reviewed. Also, the Korean Twitter Emotion Analysis's Emotion Lexicons used in this study show approximately 60% performance. [12] Therefore, we cannot rule out the possibility that the emotional analysis of the comments is not accurate.

5.3 Implication of Research

According to another recent study in Korea, public opinion that forms on the internet can be classified in to 3 groups. First is the Pure internet-based public opinion in which no cases have been mentioned in the media, and only been mentioned on the internet. It has been set as agenda on the internet bulletin board and numerous people exchanged their opinion eventually forming public opinion. Second is the Amplified internet-based opinion in which people exchanged their opinions on the internet after it has been mentioned in the media, eventually forming a public opinion. Third is Time-spaced internet-based opinion, in which it has been forgotten by other social issues and time after it have been mentioned in the media. However, in Time-spaced internet-based opinion, it again forms public opinion by internet users. [2]

The journalistic characteristics of the Internet, which have so far been raised by many Korean researchers, are as follows: Emotional and inaccurate information, Setting an agenda without a presenter, Setting an agenda that doesn't depend on time and space, Limitless of agenda setting, Two way communication, Indistinctness between private and public areas, Anonymous communication space, Powerful cohesion and so on. [1][3]

Comparing with our research, Type N1 and T1 corresponds to Pure internet-based opinion and Type N2 and T2 corresponds to Amplified internet-based opinion. In other words, online petition forms Pure internet-based opinion and Amplified internet-based opinion. And furthermore, petition has most of the characteristics of internet journalism. We reached to the conclusion the Blue house petition has a journalistic characteristic that produces and delivers news. Therefore, the number of agree or signatures shown in the petition cannot be accepted as public opinion. It is more reasonable to understand online petition system as a new form of media rather than a tool that shows public opinion.

The recent 'Isu station incident' has triggered a gender conflict as 360,000 agreed (2018.11.18) to the related article in the Blue House Petition. However, it has been confirmed by the police that the article contains distorted facts. It implies that online petition can trigger unnecessary social conflicts due to the journalistic characteristic. There is no effective system in the online petition system to block this negative impact of journalistic characteristic. Therefore, online petition system must provide an institutional strategy that blocks fake information.

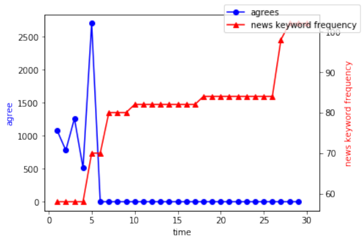
VI. Appendix

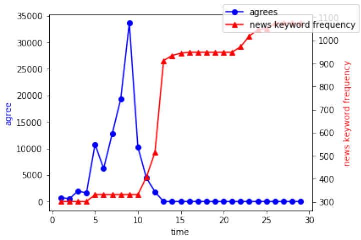
6.1 Abbreviation List

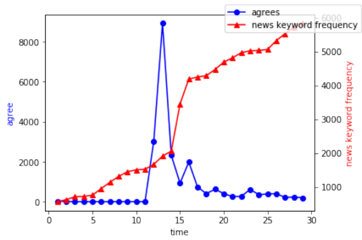
Table 6. Abbreviation List

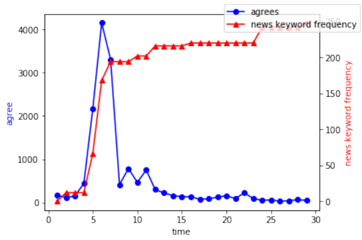
Abbreviation	Meaning
Agree	Number of assenters on petition
NKF	Frequency of specific keyword in news article (News Keyword Frequency)
Tweets	Frequency of related tweets

6.2 Other Graphs of Agree and NKF (alphabetical order based on Case)

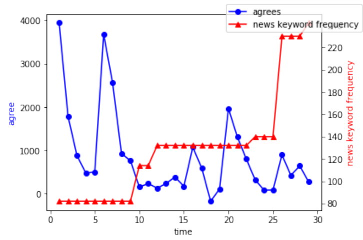


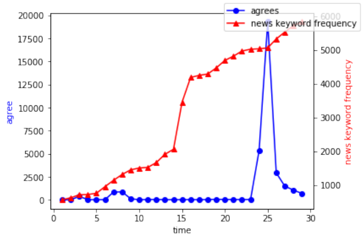


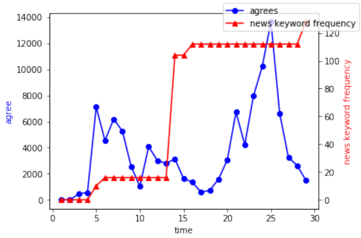


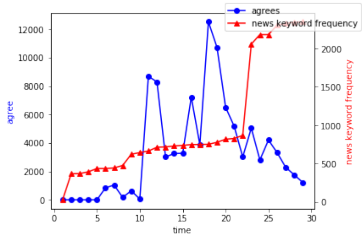


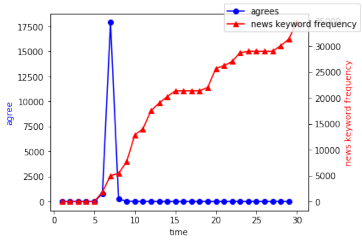
Case E has been excluded from our analysis since the data was too small.

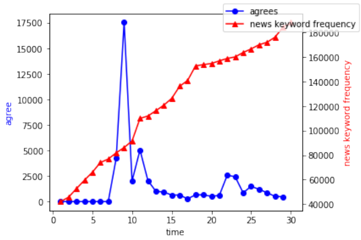


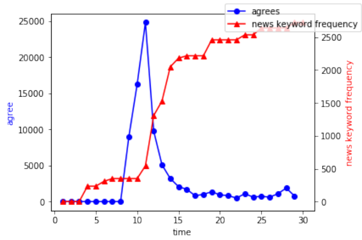


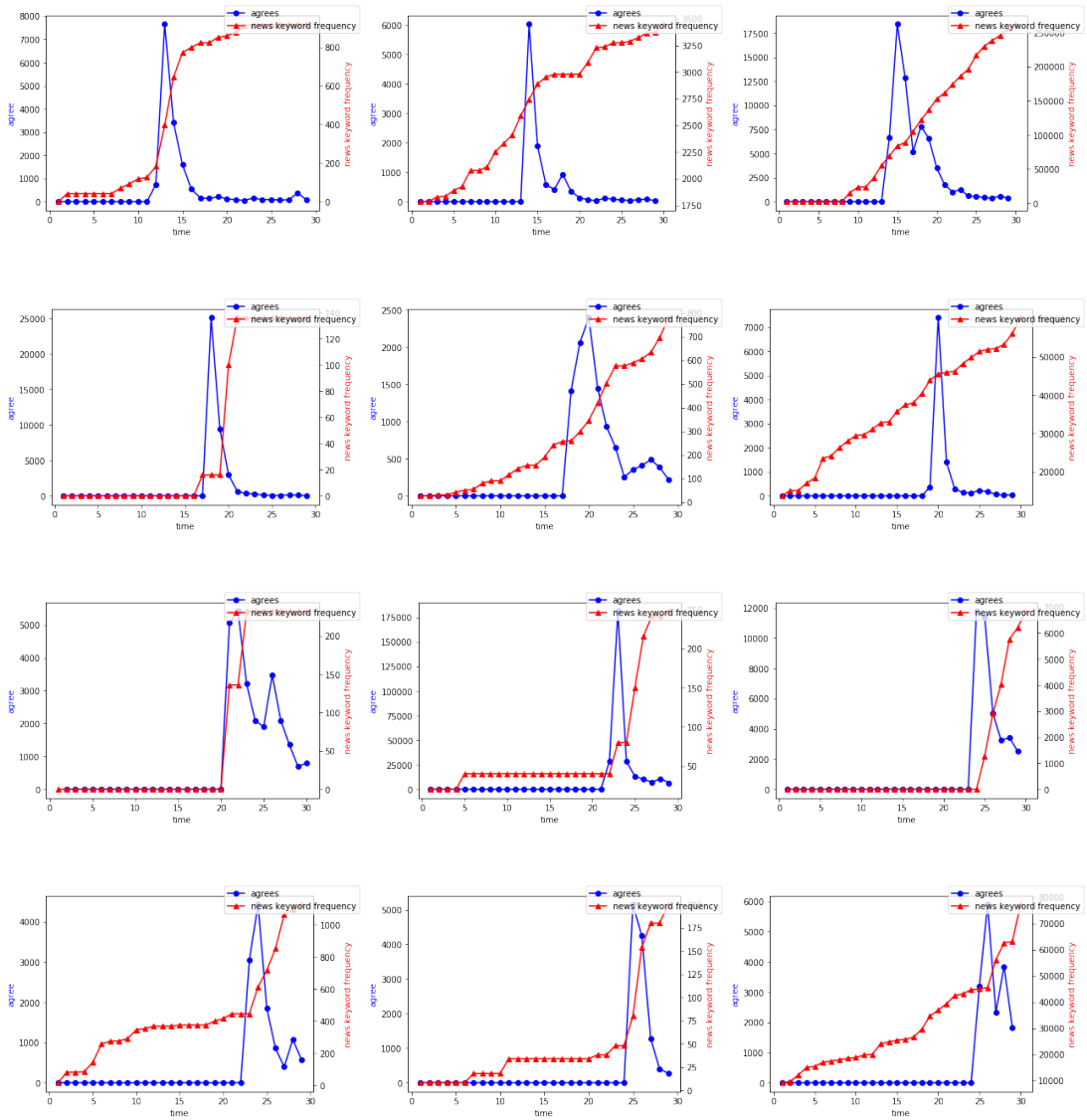




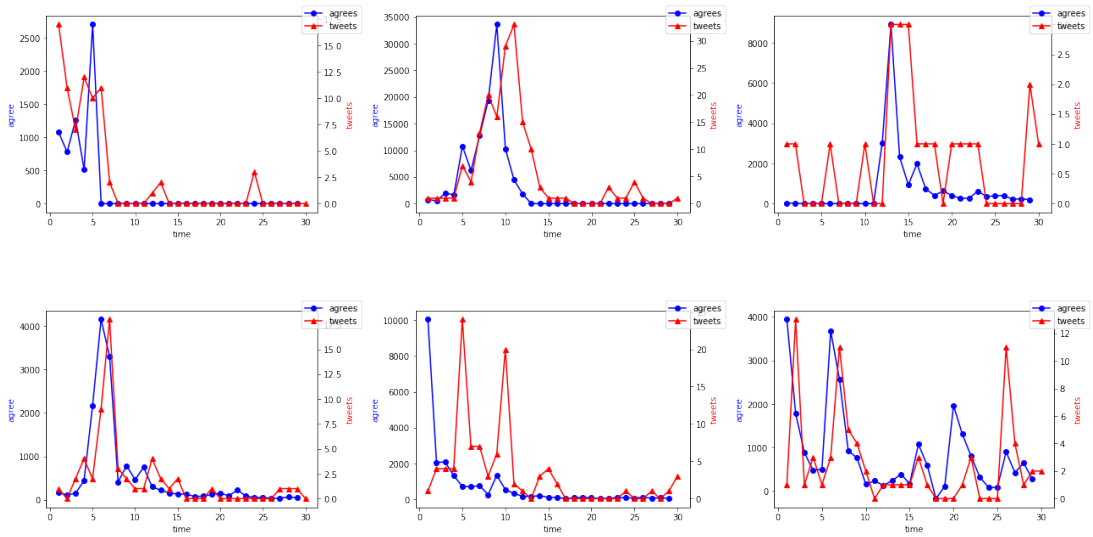


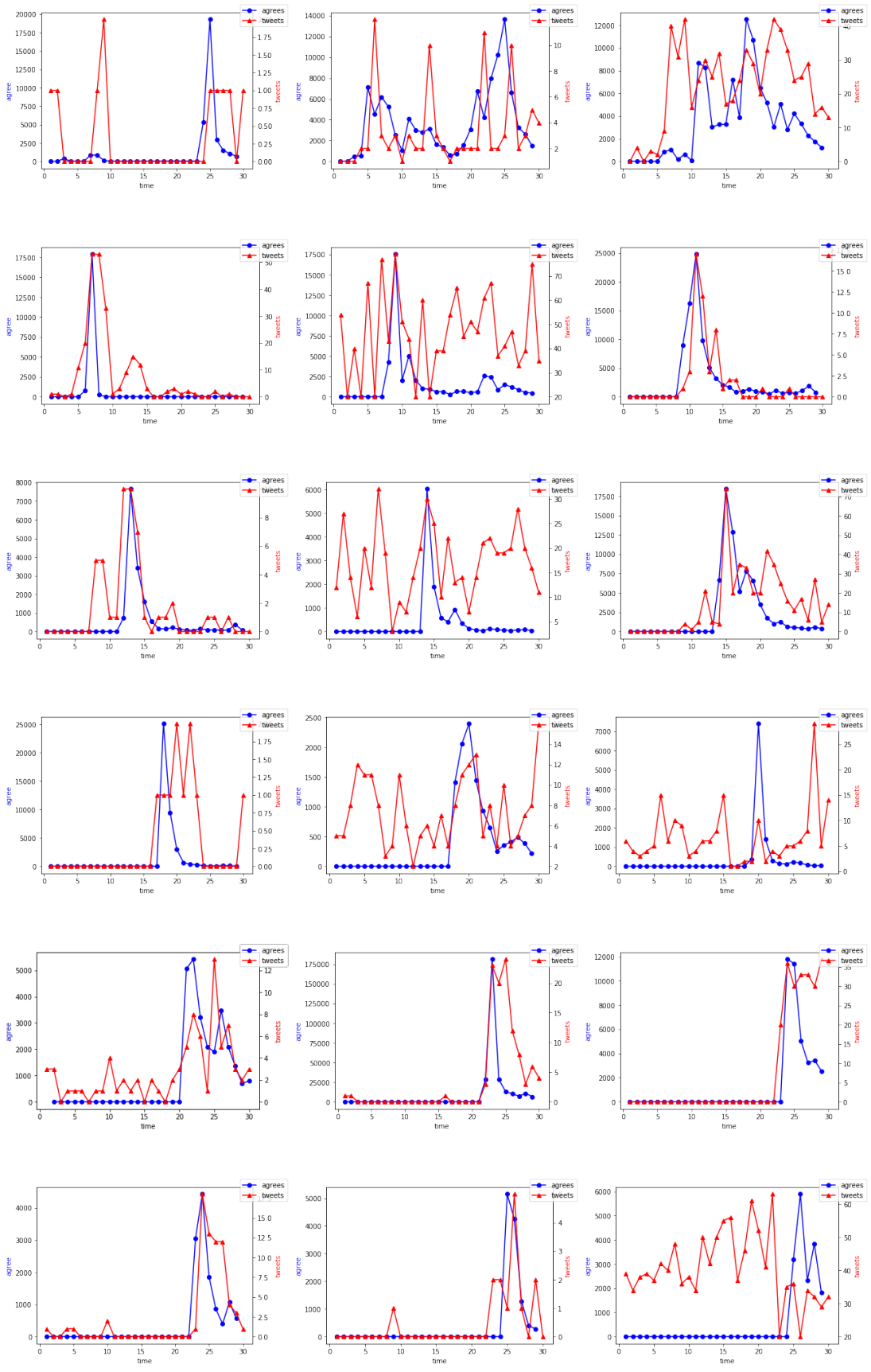






6.3 Other Graphs of Tweets and Agree (alphabetical order based on Case)





6.4 Code

You can see the full code used in the research in the following link. (Github)

https://github.com/wani-ham/KSHS_RnE

VII. References.

[1] Kang, M. E. (2001). *Internet Journalism and Public Opinion**. Seoul: CommunicationBooks.

[2] Park, S. H. (2005). *On the Journalistic Characteristics and Social Impacts of Internet Bulletin Board as a Public Opinion Space*. Journal of Communication Science. 5(3), 191-226

[3] Park, S. H. (2002). *The Understanding of Internet Media and its application**. Seoul: CommunicationBooks

[4] Shin, S. J. (2014). *'Reading emotion from text' Understanding of Emotion Analysis*. IDG Tech Report.

[5] Oh, M. Y., Jung, I.S. (2005). *Core Theory of Communication* (pp. 222-239). Seoul: CommunicationBooks.

[6] *"Opinion Mining"*. Glossary Telecommunications Technology Association*. 2012.

[7] Chang, S.B., Sung, Y.H. (2017). *Online Opinion Analysis of Performing Arts through Social Big Data*. Journal Korea Society of Visual Design Forum, 57, 79-90.

[8] Han, C. J., Kim, K.S. (2013). *Twitter's impact on the election of TV debates - 18th presidential election TV debates.* Journal of Digital Contents Society, 14(2), 207-214

[9] Blumer, H. (1966). *The mass, the public, and public opinion*. Reader in public opinion and communication, 2, 43-50.

[10] Braatz, K. (1988). *Friedrich Nietzsche: eine Studie zur Theorie der Öffentlichen Meinung*. Berlin: Walter de Gruyter.

[11] Hennessy, B. C. *Public opinion*. Belmont, California: Wadsworth Publishing Company.

[12] Hyo Jin, Do., Ho Jin, Choi. (2015). *Korean Twitter Emotion Classification Using Automatically Built Emotion Lexicons and Fine-Grained Features*. 29th Pacific Asia Conference on Language, Information and Computation, 142-150.

[13] Manyika, J., Chui, M. (2011). *Big data: The next frontier for innovation, competiotion, and productivity*, McKinsey Global Institute.

[14] Douglas, L. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*. Gartner

[15] Snijders, C., Matzat, U., Reips, U, D. (2012). *'Big Data.': Big gaps of knowledge in the field of Internet*. International Journal of Internet Science, 1-5.

*Titles with * are translated by Ham, T.W., Hwang, J.M. since they have no English title for them.*