Analysis of Celebrity Reputation Based on Metadata from Internet Tabloids

Daewon Kim
School of Computing
20180061

Youngil Yoon
School of Computing
20180416

Hyunho Lee
School of Computing
20190531

Abstract

This document contains the overall information, including research objectives, research process, and anlysis on celebrity reputation from metadata from internet tabloids. This research team used metadata from internet tabloids to analyze celebrity reputation. Specifically, we analyzed attention from the public and impression celebrity gives using *R*. Statistical techniques such as regression, moving average, STL, etc. are used to analyze and predict. Results suggests that analysis using internet tabloids metadata is meaningful.

1 Introduction

In the 20th and the early 21st century, newspapers were the main media to give people information. In modern days however, internet tabloids have taken their place. Internet tabloids does more than publish news, but they also offer some functions for people to react to the news, and those reactions that people gave are shown to other people as the form of metadata.

This research team has started this research based on a simple question while observing the metadata of internet tabloids; Can we Analize the reputation of a celebrity from the metadata of internet tabloids?

Figure 1: Article from Internet Tabloid *NAVER*

We drew up some research plans, including crawlling the internet tabloids for the metadata, pre-processing those metadata with natural language processing techniques, and statistical research methods involving *R*, which we learned from class.

2 Objectives

Our first goal is to plot the impression certain celebrity gives, and the interest from public verses time and observe the transition.

Our second goal is to present the magnitude of issues which gives dramatic changes to the impression that certan celebrity give.

Our Third goal is to present the keywords related to certain celebrity in the form of wordcloud.

Our fourth goal is to predict the transition of impression certain celebrity gives, and the interest from public in the form of time.

3 Background Knowledge

This research team has mainly used the internet tabloids website *NAVER Newspapers* to extract metadata(crawlling). We used natural language processing techniques to extract meaningful data(preprocessing), and made statistical approaches via *R*.

The following illustrates how this research team approached each components.

3.1 NAVER Newspapers

NAVER Newspapers is the biggest internet tabloids website in Korea. Thousands of posts and articles are uploaded everyday in NAVER Newspapers, including the three biggest newspapers: Chosun, Donga, and Joongang. Articles are categorized in 'politics', 'cultural', 'society', and 'entertainment' etc. Most significantly, *NAVER Newspapers* offers readers some functionalities to express and react to the article.

Comment Section Firstly, there is comment section, where readers write down their expressions freely, and people can either 'like' or 'dislike' the comment. Due to regretful incidents happend in entertainment industry, involving suicide of celebri-

Figure 2: Internet Tabloids Platform *NAVER*

ties, comment sections has been removed from those articles of categories entertainment.

Emotional Expressions Secondly, and most importantly, there is a emotional expressions where people can pick from 'like', 'warm', 'sad', and 'angry' etc. emotional expressions are most easily accessible as they are exposed at the top of the article, right below the title.

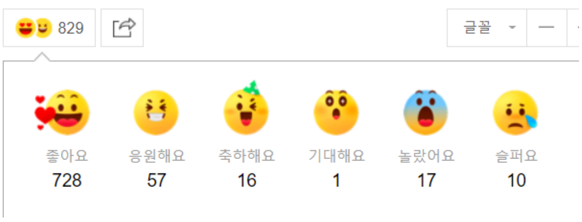


Figure 3: Emotional Expressions in *Naver TV/Entertainment News*

This research team has used *NAVER Newspapers* to scrap data as it contains variaty of metadata including comment section, emotional expressions, and the date of time the article is released etc.

3.2 Bigkinds

Bigkinds is a website which offers easy access to internet tabloid platforms such as *NAVER Newspapers*. The website makes it easy to search internet tabloid articles with keywords such as person, the journalist, event etc.

This research team used *Bigkinds* to access *NAVER Newspapers* with celebrity name as a keyword.

3.3 koNLPy

koNLPy is a python package to process korean literal informations. This package offers high-precision text pre-processing, such as word tokenizing, part-of-speech tagging etc.

This research team used this package to preprocess the title of the article, hence signifying meaning of the title.

Figure 4: News analysis site *Bigkinds*

4 Assertions

Before starting our research, this research team had to make some assertions in order to make our analysis valid. As concepts such as 'reputation', 'attention', 'opnion', 'preference' are abstract and possibly vague, we had to define or assume those concepts in the form of statistical variables.

Impressions We assume that people have impression for a certain celebrity as a number between 1 and 0, where the prefrence gets more intense as the number gets larger.

Emotinal Expressions We assume that people with negative impression of a certain celebrity will give negative emotional expressions, such as *Angry*, *Sad*, *surprized*. We also assume that people with positive impression of a certain celebrity will give positive emotional expressions, such as *Like*, *Congratulations*, and *Looking forward to*. These are emotional expressions *NAVER* provides: Like, Warm, Sad, Angry, Want further inquaries, Cheer, Congratulations, Looking forward to, Surprized

Interest We assert that any individual either has interest in certain celebrity, or not. Those with interest in certain celebrity will give a emotional expression to the celebrity related article, those with no interest will not.

Press We assume that press does not give any influence to individual reader's any choice.

Independancy We assume that every individual's opinion is independant.

etc We assume that most of the inflow to any article is through the most-searched keyword section from the website.

5 Dataset

As we mentioned earlier on the material part, we used *NAVER News* and *Bigkinds* to gather the data we need. Also, we had to some preprocessing on the dataset to handle exceptions. In this section, we will introduce with some of our methods and preprocessing.

5.1 Dataset Gathering

As we need specific articles related to specific celebrities, keyword searching for articles were necessary. For keyword searching, using *Bigkinds* was one option, as it provides keyword searching for various articles. We coded a crawler to gather all the articles related to celebrities, containing the date the article was posted, article title, and emotional expressions.

5.2 Dataset

The dataset contains approximately 200 csv files containing the date the article was posted, article title, and emotional expressions for past year searched on celebrity name. We additionaly acquired 3years amount of dataset for those celebrities with significantly resulting data.

5.3 Pre-processing

There were some anomalities where negative article about a celebrity was given positive emotional expressions, mainly for mocking purposes.

A pre-processing based on NLP(natural language processing) technique was used to handle this exception. We used a python NLP package *koNLPy*, iMDB package, kkma and okt class, and tensorflow to determine the positive/negative sentiment in the article title, together with a manualy made list of negative words.

iMDB iMDB is a movie review website that collects thousands of movie review data from public. In 2011, a research done by Stanford university NLP research team yields approximately 90 percent in sentiment prediction. We also used iMDB package dataset to train our model.

Tokenize Using koNLPy default okt class, we tokenized and stored raw data from iMDB package dataset. Using pos(part of speech) tagger from kkma, pos tag was added to tokenized data. As tokenized data is over 4Gb, we made .json files to store them. We seperated the data in to two equally sized parts: train data and test data.

Figure 5: Korean NLP package *koNLPy*

Tensorflow A Tensorflow learning technique was used with train data and test data. xtrain data includes tokenized words, and ytrain data includes the sentiment represented by ratings. xtest data includes tokenized words, and ytest data includes the sentiment represented by ratings. We used these data to train the model with tensorflow.

predictor In our model, predict-pos-neg predicts the sentiment based on the train data. predict-pos-neg takes a input sentence, and returns 1 or -1 depending on which the sentence has a positive sentiment or a negative sentiment.

It takes a sentence, tokenize it with okt class, and based on tensorflow-trained model, introduced a number between 0 and 100, depending on its positivity.

6 Methods

These are some of the methods we used to achieve our objectives. Firstly, we categorized the emotional expressions such as Like, Warm, Want further inquaries, Cheer, Congratulations, Looking forward to as positive or neutral expressions, and Sad, Angry, Surprised as negative expressions.

6.1 Analyzing Attention from Public

- Observe the transition from attention from public by using moving average and STL.

- Find correlation in idol groups and its members by using scatterplot
- Find the member in a group that influences the group's public profile the most.
- Level the public profile of idol groups by using clustering.

6.2 Impression Celebrity Gives

- Observe the transition of attention from public by using plot()
- Present the magnitude of the issues that makes significant change in the transition from observation.
- Categorize the negative articles and use the data to analyze idol fandom.

6.3 Wordcloud

- Present the most deeply connected words recognized by public in a visible scale.
- Change the threshold for attention for which we collect the articles and observe the change.

7 Analysis

7.1 Attention from Public

Techniques We computed the attention from public as the net number of emotional expressions.

Transition We used moving average for 15 weeks(MA15) and STL(Seasonal and Trend decomposition using Loess) to effectively observe the transition of attention from public. The observation was meaningful based on the issues that appeared on significant weeks and days. (Folder : Trend Graph)

Correlation in Team We used scatter plots(Figure 6) to observe the correlation between an idol group and its members. The observation turned out to be meaningful, as the transitions were highly alike. From this observation, we could make a justification about our methodology; detaild analysis on group members can be deducted from analyzing in units of groups as we did in this research. (Folder : Scatter)

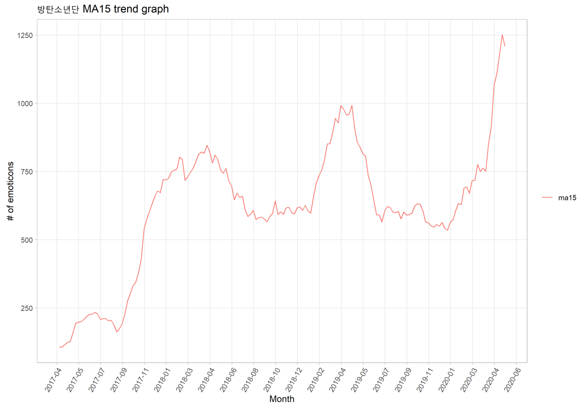


Figure 6: BTS transition graph of attention from public (*MA15*)

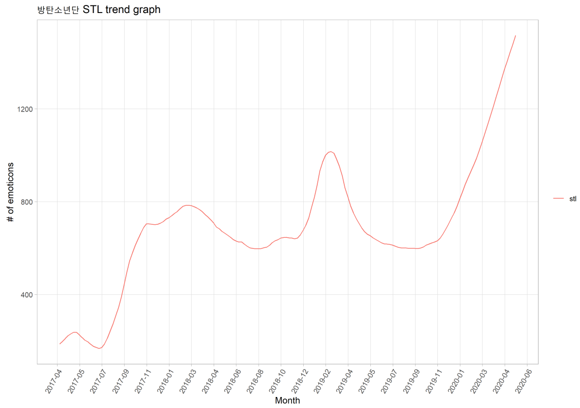


Figure 7: BTS transition graph of attention from public (*STL*)

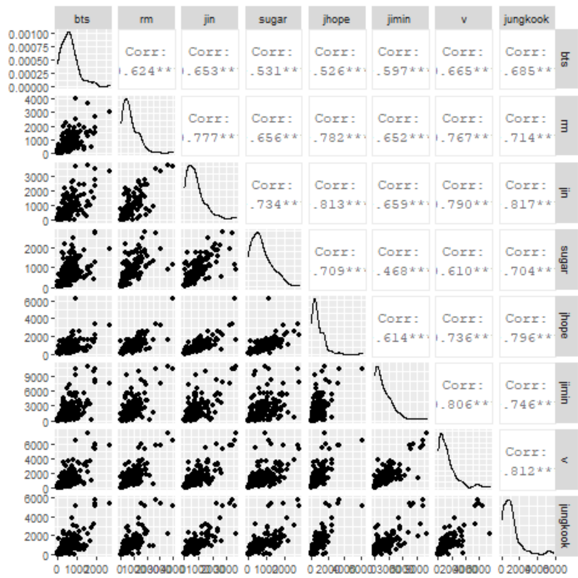


Figure 8: Scatter plot matrix of BTS and its members

Member with high influence We tried to find the member from group whom influences the group's public profile using multiple regressions. In trying regression on a idol group by the data of each member from the group, we figured out that as the data has a trend we have to remove the trend and then do a regression. In which case, we can try two different options: first-difference method and doing a regression after subtracting trend acquired by STL.

We could deduce the member who influences the BTS's public profile. The significance probability was observed in both first-difference method and STL. The common significant members are observed as Jungkook, J-Hope, Jin, RM. The STL using method suggests that V is significant too. We analyzed based on articles that Sugar is a member who does individual work a lot, and Jimin has a large individual fandom.

The result presents that ACF is very high, which suggests that the regression was not appropriate. There will be a need to add informations other than metadata from internet tabloids. (Folder : multi)

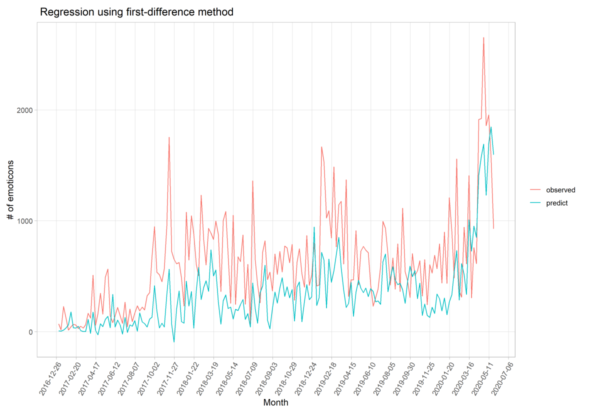


Figure 9: Multiple regression graph of public attention in BTS using *first-difference* method (red: observed, blue: predict)

```
Call:
tslm(formula = bts ~ rm + jin + sugar + jhope + jimin + v + jungkook,
    data = mean_total_table2)

Residuals:
     Min       1Q   Median       3Q      Max
-1083.70  -266.57    -5.86   190.42  1017.03

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.11426   27.87260  -0.148  0.88283
rm           0.17903    0.05972   2.998  0.00313 **
jin          0.17179    0.06034   2.847  0.00496 **
sugar       -0.10057    0.06480  -1.552  0.12252
jhope       -0.15695    0.05849  -2.683  0.00801 **
jimin        0.02310    0.02094   1.103  0.27149
v            0.03755    0.03773   0.995  0.32108
jungkook     0.16778    0.05202   3.225  0.00151 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 369.8 on 169 degrees of freedom
Multiple R-squared:  0.2535,	Adjusted R-squared:  0.2226
F-statistic: 8.198 on 7 and 169 DF,  p-value: 1.369e-08
```

Figure 10: Multiple regression summary of public attention in BTS using *first-difference* method (red: observed, blue: predict)

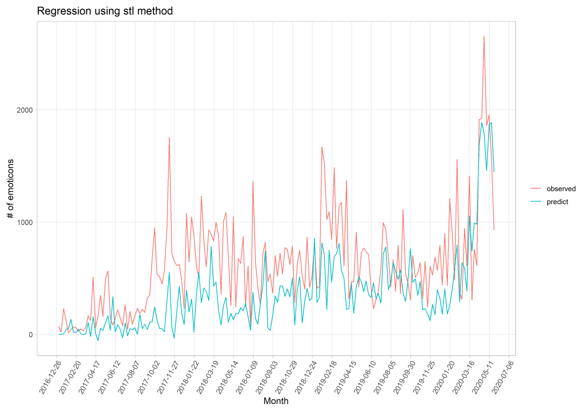


Figure 11: Multiple regression graph of public attention in BTS using *STL* method (red: observed, blue: predict)

```
Call:
tslm(formula = bts ~ rm + jin + jhope + v + jungkook + sugar +
    jimin, data = mean_total_table3)

Residuals:
    Min      1Q  Median      3Q     Max
-705.50 -154.39   -7.15  152.11 1041.27

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.72458   19.81763  -0.440 0.660320
rm           0.11286    0.06303   1.791 0.075152 .
jin          0.10559    0.05999   1.760 0.080170 .
jhope       -0.20894    0.05897  -3.543 0.000511 ***
v            0.09794    0.04010   2.443 0.015598 *
jungkook     0.20760    0.05232   3.968 0.000107 ***
sugar       -0.04347    0.06414  -0.678 0.498803
jimin        0.02476    0.01935   1.280 0.202423
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 263.4 on 170 degrees of freedom
Multiple R-squared:  0.2744,	Adjusted R-squared:  0.2445
F-statistic: 9.183 on 7 and 170 DF,  p-value: 1.302e-09
```

Figure 12: Multiple regression summary of public attention in BTS using *first-difference* method (red: observed, blue: predict)

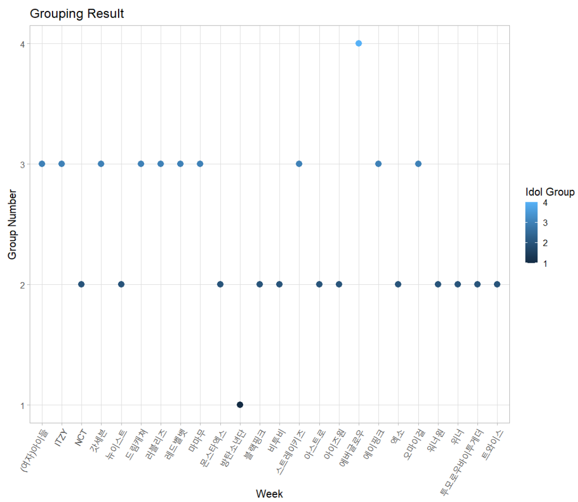


Figure 13: Idol group clustering result

Clustering We used clustering for the attention versus time in the unit of groups. For idol groups, it turned out a meaningful result of groups clustered by their public profile(Figure. 11). Groups 1 to 4 are in order of public profile. BTS, a global phenomenon is in group1, and those idols with giant fandoms such as EXO, BlackPink, and TWICE are in group2. Those idols that are not well-known are in group3 and 4. (Folder : Cluster)

7.2 Impression Celebrity Give

Techniques We computed the Impression certain celbrity gives from the ratio of Positive/Negative article titles for time and the ratio of emotional expressions per net number of emotional expressions for time. We also computed severity with negative ratio with net number of emotional expressions. As we mentioned in dataset pre-processing, for some controversial subjects there were a lot of positive expressions even for the negative titles, mostly for mocking purposes. We handled the case by processing natural language on the titles.

Issues We lined up those issues that showed great downward change in impressions celebrities give. In the article, the higher the proportion of certain negative emotions, the more the events in the article have a pure character. In other words, the higher the ratio, the more the public thinks the same. Through the ratio of negative emotions, we can accurately extract the issues that have occurred to celebrities. (Folder : list)

	nth_week	title	neg_ratio	sad	surprise	angry	maxval	newschar
1	7	"대놓고 똑같다" 엑소 로고 표절 논란에 휩싸인 일본 브랜드	0.00000000	0	0	61	61	angry
2	51	[단독] '마이클잭슨 헌정앨범' 제작사, 정산금 미지급...가창자...	0.00000000	3	46	0	46	surprise
3	48	엑소 수호, 5월 14일 입대 "엑소엘, 정말 보고싶을 것"	0.02777778	70	0	0	70	sad
4	31	베트남 공항 직원이 엑소 찬열·세훈 여권 사진 유출	0.03236246	5	0	593	593	angry
5	31	엑소 여권 정보 유출한 '베트남 공항 직원'...비난 거세져	0.04000000	0	0	72	72	angry
6	35	엑소 첸·쿨 이재훈...깜짝 '결혼 고백'이 황당한 팬들	0.04464286	3	0	104	104	angry
7	31	베트남 공항 직원이 엑소(EXO) 찬열·세훈 여권 사진 유출 '파...	0.04891304	2	0	173	173	angry
8	33	엑소 첸 결혼 놓고 팬들 "탈퇴" vs "잔류" 갈등 심화	0.05405405	1	0	34	34	angry
9	10	[단독]BTS·엑소 공연티켓 구해준다면서 먹튀한 '메이다니' 구속	0.05421687	3	0	311	311	angry
10	49	[뉴스체크 | 문화] 엑소 리더 수호, 훈련소 입소	0.05681818	81	0	2	81	sad

Figure 14: Top 10 negative tabloids on EXO

Transition We observed the transition of impression versus time plot. Interestingly, the change in the transition of related celebrities was more than significant in the week when the critical issues such as 'Me Too'movement and 'Debt Too'movement. Also, in the week the 'Burning Sun Gate' was exposed, the related celebrities' impression literally plummeted. First is the Me Too movement. About the 60th week of the Me Too movement, which resulted in a significant decrease in the favorability of the celebrities involved. In particular, Cho Min-ki showed a very low level of favoritism for a while after the suicide incident. Since then, the percentage of people expressing their feelings of sadness due to innocence, suicide, etc. is not small.

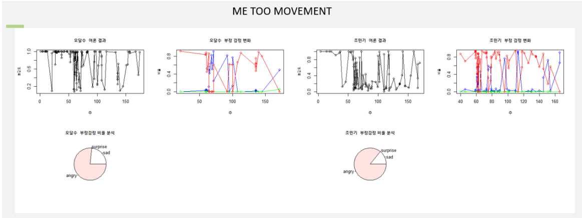


Figure 15: Transition of celebrities involved in Metoo movement

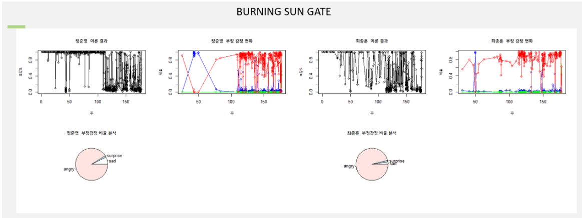


Figure 16: Transition of celebrities involved in Burning Sun gate

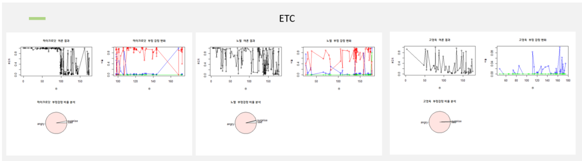


Figure 17: Transition of other controversial celebrities

The following are figures related to Burning Sun Gate, especially Jung Joon-young. It can be seen that Jung Joon-young's case broke out in the 110th week and his favorability decreased significantly. Unlike the people involved in the Me Too movement, these figures have little room for public sympathy. In fact, we found that the percentage of angry emotions was very high compared to the previous person. The last one is guitar celebrities. Each is associated with debt-ridden, drunk driving, various controversies, and sexual assault of minors. In particular, Ko Young-wook's favorability is very low even after a long period of time, and almost 99 percent of the negative expressions have been confirmed to be anger. (Folder : Plots)

Categorizing As we assumed three of the expressions *Angry*, *Sad*, and *Surprized* as negative expressions, we could categorize the negative articles in to three categories.

Fandom For idol groups and their individual members, the existence of large-scale fandom gives a lot of influence to the impression. For non-idol celebrities, there are a very few number of negative expressions unless there is a highly severe incident. However for idols, the change in the transition for emotional expressions are very significant

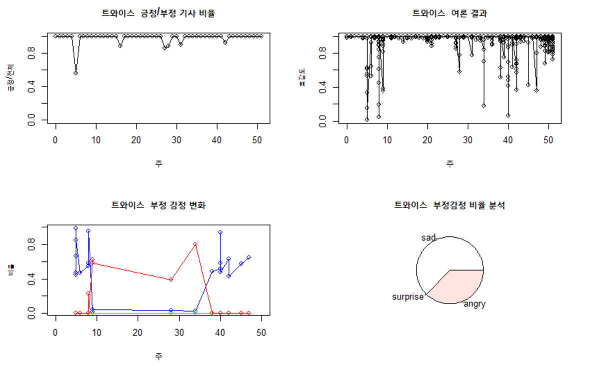


Figure 18: Fandom analysis plots on Twice

even for a minor incident; which makes it possible for us to verify the characteristics of idol fandoms. Thus, fandom anlysis will enable idol agencies to understand and predict how the fandom will react. (Folder : Fandoms)

7.3 Wordcloud

See folder 'Wordcloud' for images.

Techniques We wordclouded those articles related to a certain celebrity and over a certain threshold of attention. We seperated the article title to 9 lexical categories, and stored them as tibble. For nouns, we excluded a stopword and stored them as frequency distribution. We used same technique for those articles with bad issues about celebrity, resulting in the wordcloud of negative keywords.

keywords Through wordclouding, we could get gather the keywords which the public recognize them of in a visible scale, and those keywords which the public got most interested when a bad issue happend to the celebrity.

Idol Groups For idol groups, we could distinguish the member who got the most interest from the public in a visible scale.

Threshold We initially set the threshold for the attention at 30, but as we changed the threshold to 50 and 100, the generated wordcloud barely changed, which must mean we have sampled the dataset pretty well.

Figure 19: Jungjoonyeong Threshold 30

Figure 20: Jungjoonyeong Threshold 50

Figure 21: Jungjoonyeong Threshold 100

Figure 22: IU Threshold 30

Figure 23: IU Threshold 50

Figure 24: IU Threshold 100

8 Conclusion

Through this project, we have succeeded in following: Firstly, analyzing the severity of issues. Secondly, the finding common trend in the publice profile using moving average, and predicting idol public profile for groups using clustering. Finally, we could analyze the characteristics of fandoms, which is in larger sence, the public.

Through the series of analysis, we have discovered it is not only possible to analyze the metadata from internet tabloids, but we can also find the pattern in the processed data and predict the transition. If we use improved preprocessing and use dispersion to calculate the difference with original data, we will be able to calculate the general reputation factor for celebrities quite accurately.

There are metadata in other media, such as SNS, and the plain text in community replies. If we use these data, I strongly believe analysis will be more precise, and we could produce better output.

9 Github

The github repository this research team has worked apon is as follows.

git@github.com:askme143/CS492E_R.git https://github.com/askme143/CS492E_R

Datasets as .csv formats are included.

10 Appendix

The followings are folders containing plots we created and analyzed.

A Trend Graph

A.1 Idol trend graph with MA15

Figure A.1: Trend graph of Redvelvet

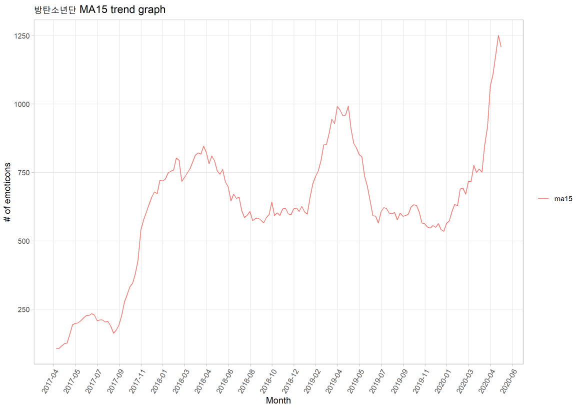


Figure A.2: Trend graph of idol BTS

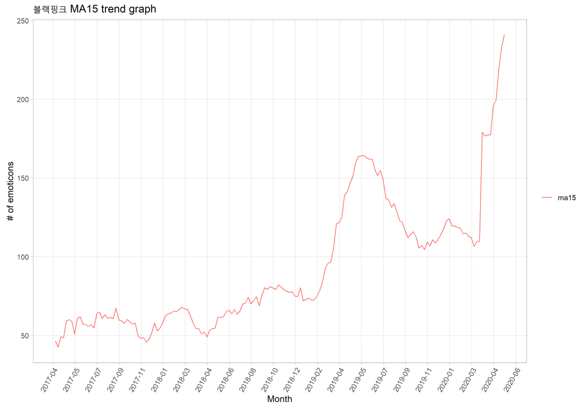


Figure A.3: Trend graph of idol BLACKPINK

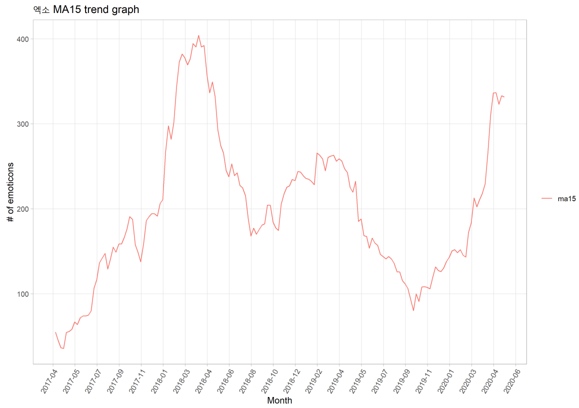


Figure A.4: Trend graph of EXO

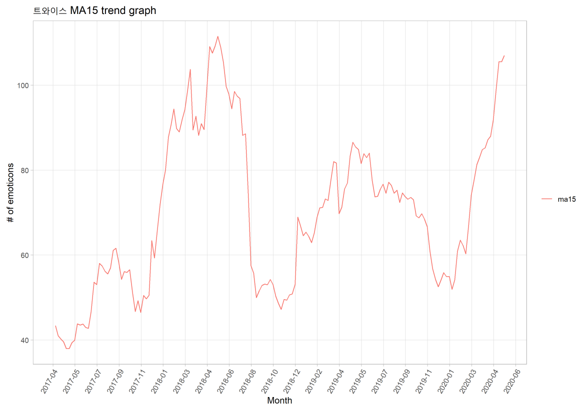


Figure A.5: Trend graph of Twice

A.2 Idol trend graph with STL

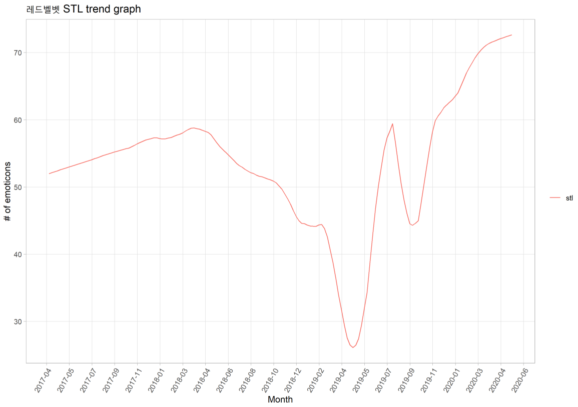


Figure A.6: Trend graph of Redvelvet

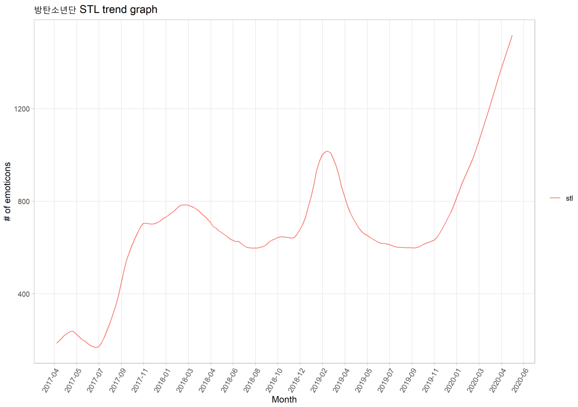


Figure A.7: Trend graph of idol BTS

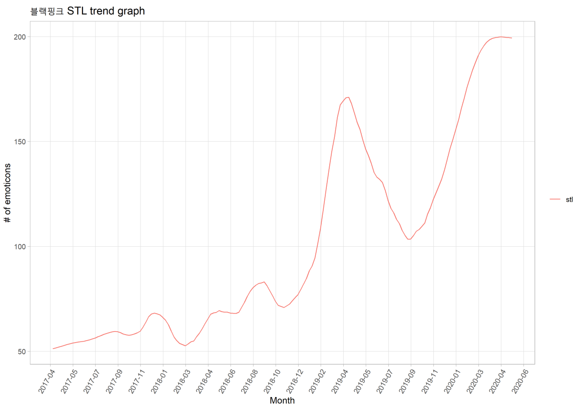


Figure A.8: Trend graph of idol BLACKPINK

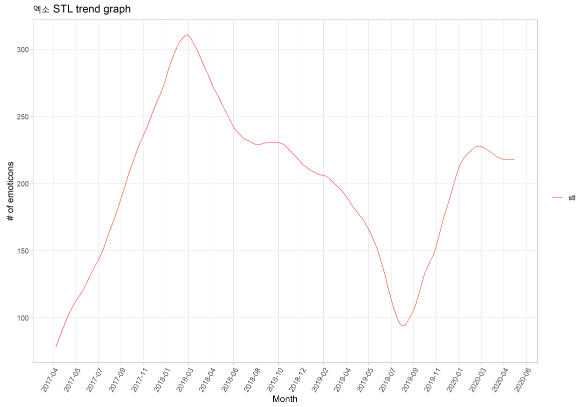


Figure A.9: Trend graph of idol EXO

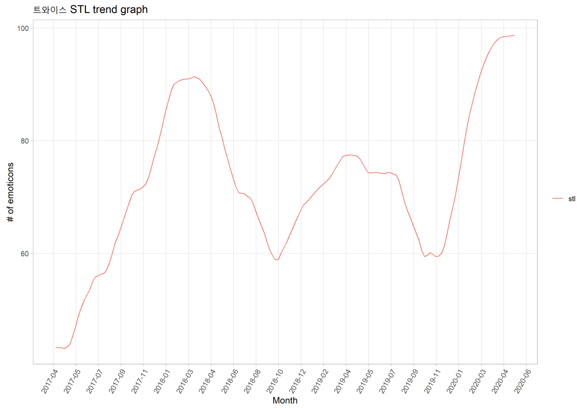


Figure A.10: Trend graph of idol TWICE

A.3 Compare MA15 and STL within integrated plots

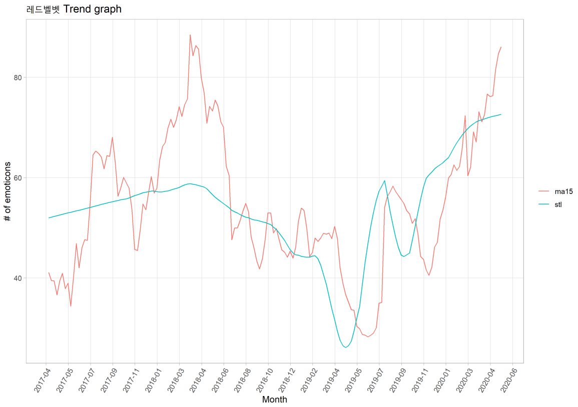


Figure A.11: Trend graph of Redvelvet

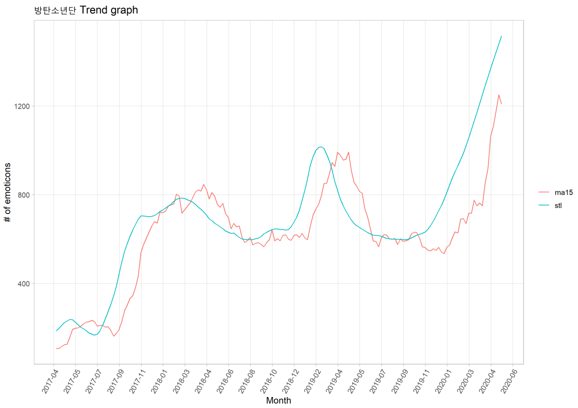


Figure A.12: Trend graph of BTS

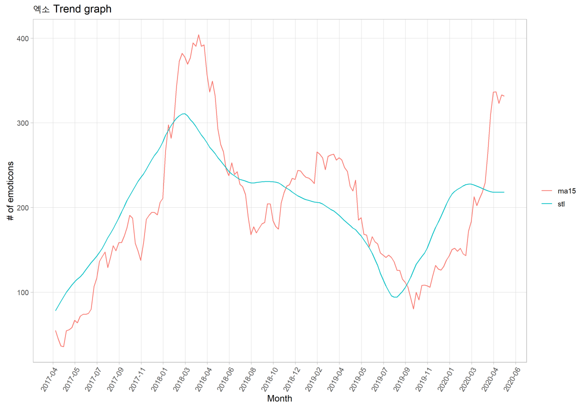


Figure A.13: Trend graph of EXO

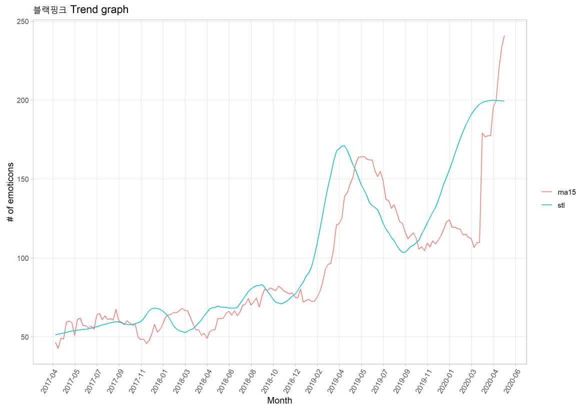


Figure A.14: Trend graph of BLACKPINK

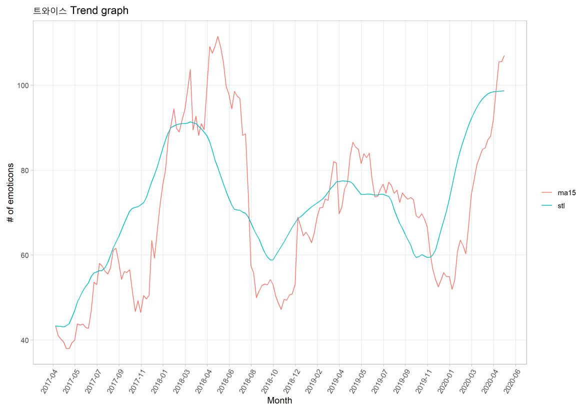
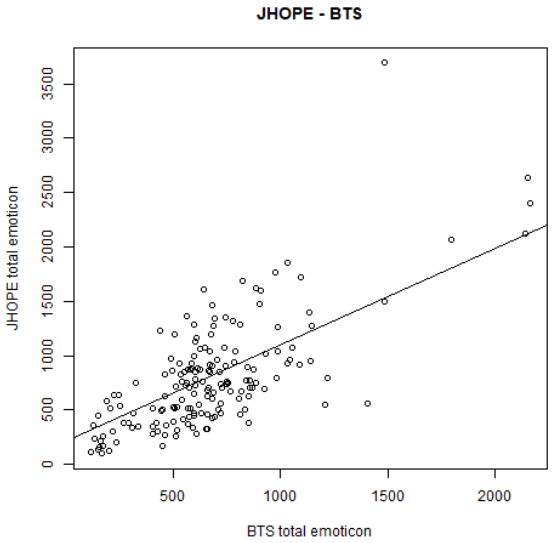
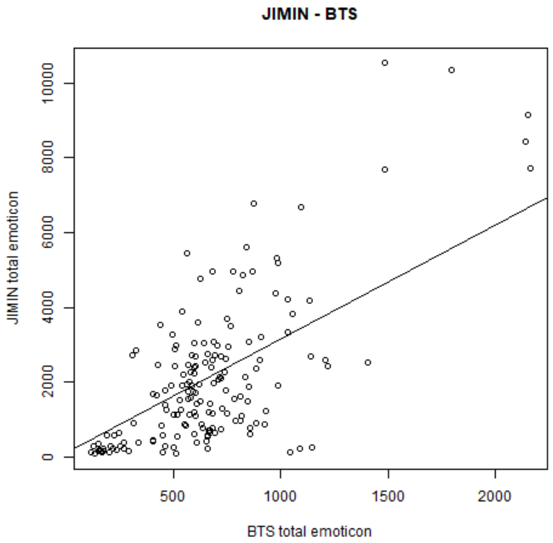


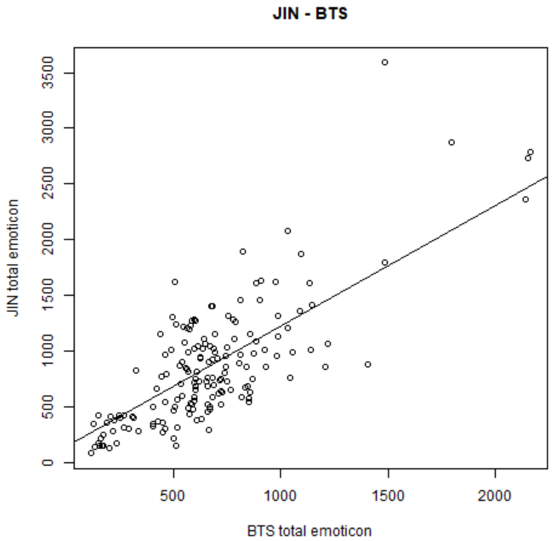
Figure A.15: Trend graph of Twice

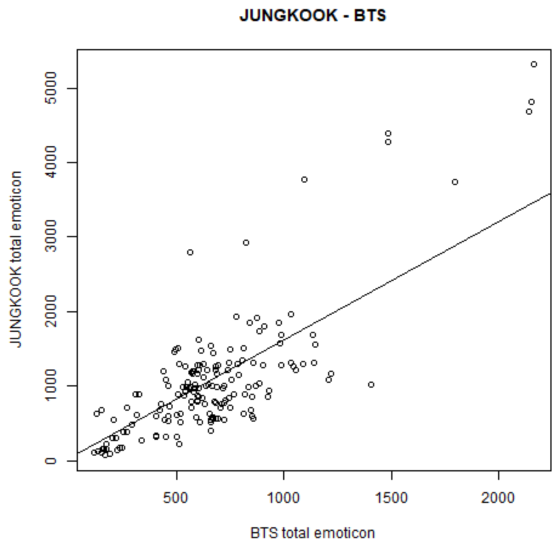
B Scatter

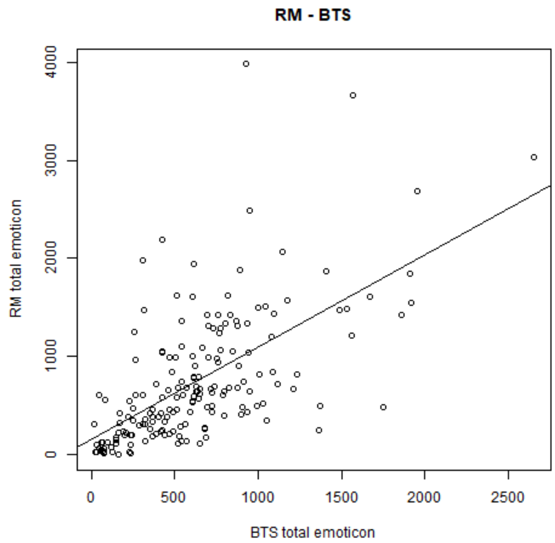
B.1 Member-group scatter plots

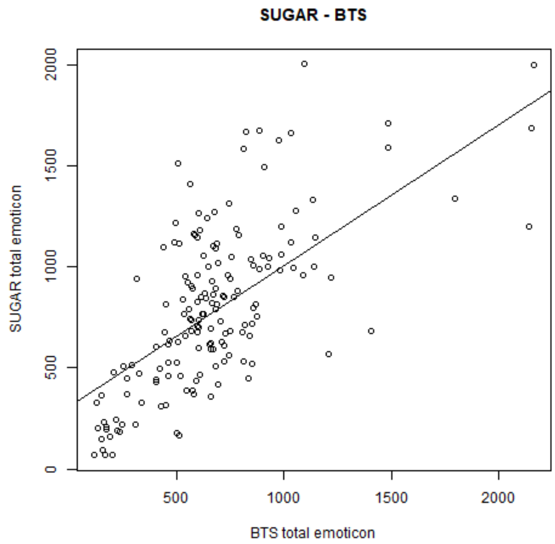












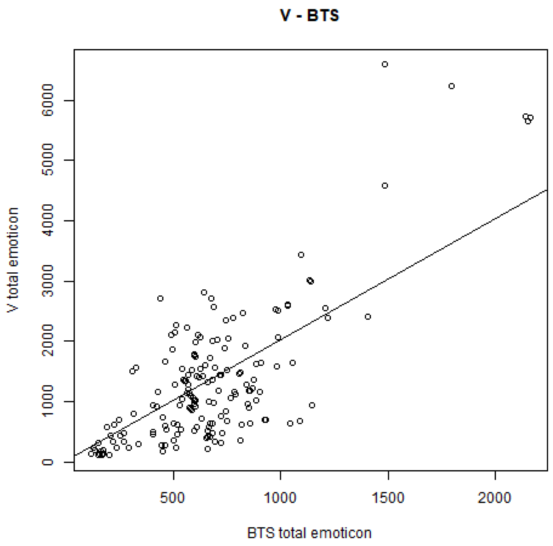


Figure B.1: Trend graph of idol groups

B.2 Member-group scatter plots

```
Call:
tslm(formula = jhope ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-1236.2  -316.8  -134.8   247.9  5139.8 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 218.1336    82.4906   2.644  0.00892 ** 
bts           0.8818     0.1075   8.205   4.7e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 631.5 on 176 degrees of freedom
Multiple R-squared:  0.2767,	Adjusted R-squared:  0.2726 
F-statistic: 67.33 on 1 and 176 DF,  p-value: 4.696e-14
```

Figure B.2: Summary of Jhope-BTS regression

```
Call:
tslm(formula = jimin ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-5264.7 -1069.7  -361.8   833.6  7341.5 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 120.5899   235.3900   0.512    0.609    
bts           3.0305     0.3067   9.882   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1802 on 176 degrees of freedom
Multiple R-squared:  0.3569,	Adjusted R-squared:  0.3532 
F-statistic: 97.66 on 1 and 176 DF,  p-value: < 2.2e-16
```

Figure B.3: Summary of Jimin-BTS regression

```
Call:
tslm(formula = jin ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-1471.9  -295.3  -120.5   174.1  2498.4 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 141.95889   72.45377   1.959   0.0517 .  
bts           1.08091    0.09439  11.452   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 554.7 on 176 degrees of freedom
Multiple R-squared:  0.427,	Adjusted R-squared:  0.4237 
F-statistic: 131.1 on 1 and 176 DF,  p-value: < 2.2e-16
```

Figure B.4: Summary of Jin-BTS regression

```
Call:
tslm(formula = jungkook ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-1681.7  -349.9  -106.0   272.0  3641.1 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  40.5097    97.5745   0.415    0.679    
bts           1.5846     0.1271  12.466   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 747 on 176 degrees of freedom
Multiple R-squared:  0.4689,	Adjusted R-squared:  0.4659 
F-statistic: 155.4 on 1 and 176 DF,  p-value: < 2.2e-16
```

Figure B.5: Summary of Jungkook-BTS regression

```
Call:
tslm(formula = rm ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-1323.4  -280.2  -119.9   192.7  2948.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 159.8641    67.8601   2.356   0.0196 *  
bts           0.9366     0.0884  10.594   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 519.5 on 176 degrees of freedom
Multiple R-squared:  0.3894,	Adjusted R-squared:  0.3859 
F-statistic: 112.2 on 1 and 176 DF,  p-value: < 2.2e-16
```

Figure B.6: Summary of RM-BTS regression

```
Call:
tslm(formula = sugar ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-1218.7  -320.8  -111.3   280.5  1636.5 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 312.66022   64.19690   4.870 2.47e-06 ***
bts           0.69444    0.08363   8.303 2.60e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 491.5 on 176 degrees of freedom
Multiple R-squared:  0.2815,	Adjusted R-squared:  0.2774 
F-statistic: 68.95 on 1 and 176 DF,  p-value: 2.597e-14
```

Figure B.7: Summary of Sugar-BTS regression

```
Call:
tslm(formula = v ~ bts, data = mean_total_table)

Residuals:
    Min      1Q  Median      3Q     Max 
-2539.0  -519.0  -130.2   359.4  4622.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   23.862    129.715   0.184    0.854    
bts            1.999      0.169  11.827   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 993.1 on 176 degrees of freedom
Multiple R-squared:  0.4428,	Adjusted R-squared:  0.4397 
F-statistic: 139.9 on 1 and 176 DF,  p-value: < 2.2e-16
```

Figure B.8: Summary of V-BTS regression

C multi

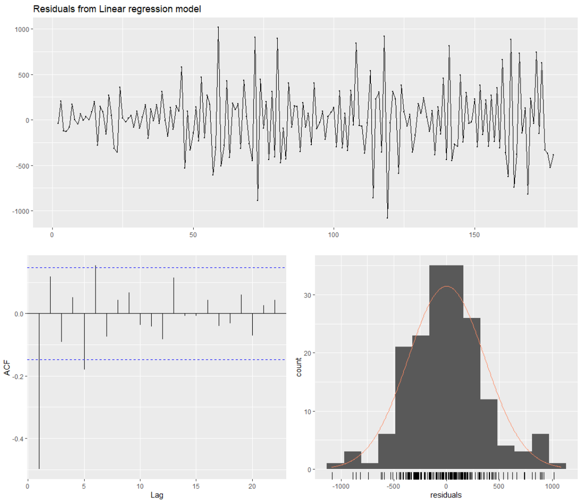


Figure C.1: Multiple regression residual check of public attention in BTS using *first-difference*

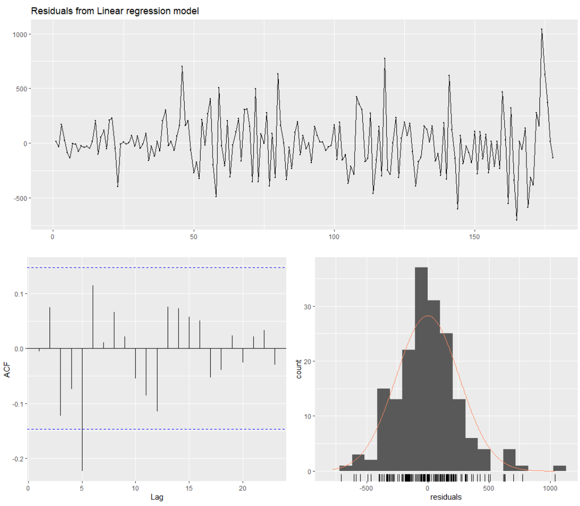


Figure C.2: Multiple regression residual check of public attention in BTS using *STL*

D Cluster

E list

E.1 High negative ratio tabloids on idol

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	38	방탄소년단 콘서트도 취소... 가요계, 코로나19 직격탄	0.02777778	36	34	0	1	34	sad
2	42	방탄소년단, 코로나19 여파로 북미투어 결국 연기	0.05714286	35	32	0	1	32	sad
3	38	방탄소년단, 우한 코로나에 4월 서울콘서트 취소	0.06250000	32	29	0	1	29	sad
4	38	방탄소년단, 코로나19 확산으로 4월 서울콘서트 취소	0.15515610	1057	892	0	1	892	sad
5	49	방탄소년단 지민, 5月 보이그룹 개인 브랜드평판 1위...2위 백...	0.37172012	686	55	375	1	375	surprise
6	40	방탄소년단, 월드투어 美 공연장 폐쇄에 "변동사항 있으면 공...	0.42431193	436	250	0	1	250	sad
7	50	"방탄소년단 지민, 듣고 있나요?" 팬덤 이벤트 화제	0.42708333	480	17	257	1	257	surprise
8	40	방탄소년단 월드투어 난항...빅히트 "일정 변동? 전달받은 바 ...	0.43828715	397	222	0	1	222	sad
9	42	방탄소년단, 코로나 확산에 북미투어 결국 연기(종합)	0.50000000	326	162	0	1	162	sad
10	9	방탄소년단 日 싱글, 한국 가수 최초 100만 장 돌파	0.52380952	105	1	0	49	49	angry

Figure E.1: High negative ratio tabloids on BTS

	nth_week	title	neg_ratio	sad	surprise	angry	maxval	newschar
1	7	"대놓고 똑같다" 엑소 로고 표절 논란에 휩싸인 일본 브랜드	0.00000000	0	0	61	61	angry
2	51	[단독] '마이클잭슨 헌정앨범' 제작사, 정산금 미지급...가창자...	0.00000000	3	46	0	46	surprise
3	48	엑소 수호, 5월 14일 입대 "엑소엘, 정말 보고싶을 것"	0.02777778	70	0	0	70	sad
4	31	베트남 공항 직원이 엑소 찬열·세훈 여권 사진 유출	0.03236246	5	0	593	593	angry
5	31	엑소 여권 정보 유출한 '베트남 공항 직원'...비난 거세져	0.04000000	0	0	72	72	angry
6	35	엑소 첸·클 이재훈...깜짝 '결혼 고백'이 황당한 팬들	0.04464286	3	0	104	104	angry
7	31	베트남 공항 직원이 엑소(EXO) 찬열·세훈 여권 사진 유출 '파...	0.04891304	2	0	173	173	angry
8	33	엑소 첸 결혼 놓고 팬들 "탈퇴" vs "잔류" 갈등 심화	0.05405405	1	0	34	34	angry
9	10	[단독]BTS·엑소 공연티켓 구해준다면서 먹튀한 '메이다니' 구속	0.05421687	3	0	311	311	angry
10	49	[뉴스체크 | 문화] 엑소 리더 수호, 훈련소 입소	0.05681818	81	0	2	81	sad

Figure E.2: High negative ratio tabloids on EXO

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	30	[공식입장] JYP "갓세븐 유겸, 건강상의 이유로 'MBC 가요대...	0.01063830	94	93	0	0	93	sad
2	30	갓세븐 유겸, 건강 상 이유로 'MBC 가요대제전' 불참	0.01694915	59	58	0	0	58	sad
3	30	JYP "갓세븐 유겸, 건강문제로 'MBC 가요대제전' 불참" [공식...	0.02325581	43	42	0	0	42	sad
4	27	갓세븐 영재, 사생팬 피해 호소 "밤낮 없이 전화하니까 돌아...	0.03030303	33	32	0	0	32	sad
5	50	방탄소년단 지민, 5월 아이돌 개인 브랜드 1위	0.18571429	490	39	360	0	360	surprise
6	50	방탄소년단 지민, 5월 보이그룹 개인 브랜드평판 '1위'	0.19599109	449	50	311	0	311	surprise
7	49	방탄소년단 지민, 보이그룹 개인 브랜드평판 1위..2위 백현-3...	0.24064171	374	21	263	0	263	surprise
8	49	방탄소년단 지민, 5月 보이그룹 개인 브랜드평판 1위...2위 백...	0.24825175	572	55	375	0	375	surprise
9	49	"압도적 인기"..방탄소년단 지민, 5월 보이그룹 브랜드평판도...	0.27659574	329	22	216	0	216	surprise
10	41	JYP, 트와이스 日활동 막혀...목표가↓-현대차	0.31372549	51	11	0	24	24	angry

Figure E.3: High negative ratio tabloids on GOT7

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	38	방탄소년단 콘서트도 취소... 가요계, 코로나19 직격탄	0.02857143	35	34	0	0	34	sad
2	45	'솔로데뷔' 마마무 솔라, 삭발 변신...'뱉어' 티저 공개	0.03529412	85	82	0	0	82	sad
3	5	마마무 화사 '노브라' 공항 패션에 네티즌 '시끌'	0.08695652	46	42	0	0	42	sad
4	7	"천만 원이면 이영자 맛집·마마무 병원"...방송가 사칭주의보	0.45754717	212	2	0	113	113	angry
5	5	마마무 팬들, 화사 지지 성명문 발표 "'프로 불편러'들의 과잉 ...	0.57377049	61	1	0	25	25	angry

Figure E.4: High negative ratio tabloids on MaMaMoo

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	5	트와이스 미나, 건강 문제로 월드투어 불참 "큰 불안감 겪는 ...	0.01408451	71	70	0	0	70	sad
2	8	'심리적 긴장' 트와이스 미나, 日휴식→韓입국 중 눈물	0.04545455	44	42	0	0	42	sad
3	40	트와이스 쯔위, 고향 대만에서 자가격리 중...코로나19 여파	0.06493506	77	72	0	0	72	sad
4	5	트와이스 미나, 월드투어 불참 "극도의 심리적 긴장" [전문]	0.15217391	920	780	0	0	780	sad
5	34	트와이스 나연 스토커, 이번엔 영상 공개 "사랑해, 정말이야"	0.17777778	990	25	0	789	789	angry
6	8	[공식입장]'활동 중단' 트와이스 미나, 입국길에 '눈물'..."힘들...	0.19210245	937	542	0	215	542	sad
7	5	트와이스 팬들, 미나 월드투어 불참 소식에 응원 "건강하게 ...	0.33712121	264	175	0	0	175	sad
8	47	트와이스 정연 "할머니, 내 꿈 이루는 것 못 보고 돌아가셔" ...	0.35546039	467	301	0	0	301	sad
9	9	슈퍼주니어 김희철·트와이스 모모, 열애설 보도에 "사실 무근"	0.36111111	648	8	0	406	406	angry
10	42	트와이스, 日 도쿄돔 콘서트 또 연기..."건강·안전 최우선"(공식)	0.36619718	71	45	0	0	45	sad

Figure E.5: High negative ratio tabloids on Twice

E.2 High negative ratio tabloids on singer

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	21	'놀면 뭐하니?' 土 저녁 완전 휩쓸었다...'유플래쉬' 드럼 독주...	0.2500000	36	27	0	0	27	sad
2	36	'슈가맨3' 더크로스 소환, 휠체어 타고 등장→'돈크라이' 열창....	0.2751678	149	108	0	0	108	sad
3	2	[취재후 Talk] 검찰총장 후보자는 '윤서결'일까? '윤성렬'일까?	0.4413408	179	1	0	99	99	angry

Figure E.6: High negative ratio tabloids on Hahyunwoo

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	27	볼빨간사춘기 안지영, 반려견 떠나보낸 후 쓴 편지 "자꾸 ...	0.02564103	39	38	0	0	38	sad
2	27	"너 가는 길, 외로울까봐 자꾸 눈물이 나"...볼빨간사춘기 안...	0.05000000	40	38	0	0	38	sad
3	44	볼빨간사춘기 우지윤, 탈퇴 발표→소속사 계약 만료→새로...	0.05882353	34	32	0	0	32	sad
4	43	볼빨간사춘기, 우지윤 탈퇴로 안지영 1인 활동 "의견 존중"...	0.12218650	311	273	0	0	273	sad
5	43	우지윤, 볼빨간사춘기 떠난다...안지영 홀로 활동	0.18468468	666	543	0	0	543	sad
6	44	국민의당 "영탁·볼빨간사춘기 등 음원 차트 조작 확인"	0.24070022	914	15	0	679	679	angry
7	43	[공식입장] 볼빨간사춘기, 우지윤 탈퇴→안지영 1인 체제 ...	0.24796748	246	185	0	0	185	sad
8	43	볼빨간사춘기, 안지영 1인 체제 활동..."우지윤 의견 존중하...	0.48258706	201	104	0	0	104	sad
9	44	[공식입장] 우지윤, 볼빨간사춘기 탈퇴→쇼파르뮤직과 결...	0.54726368	201	91	0	0	91	sad

Figure E.7: High negative ratio tabloids on Bol4

E.3 High negative ratio tabloids on entertainer

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	41	'골목식당' 원주 칼국숫집 사장 암투병, 백종원-시청자도 울...	0.02000000	50	48	0	1	48	sad
2	3	원주 미로예술시장 칼국수집에 대한 백종원 솔루션은?	0.02040816	49	47	0	1	47	sad
3	28	"말도 안되는 일 벌어져" 백종원 밝힌 돈가스집 이전 이유	0.02040816	49	47	0	1	47	sad
4	40	'골목식당' 백종원 "세상 거지 같다" 눈물...원주 칼국수집에 ...	0.02222222	45	43	0	1	43	sad
5	8	이대 백반집 불시점검, 결국 눈물 보인 백종원 [SSEN리뷰]	0.02857143	35	33	0	1	33	sad
6	44	'맛남' 홍진영-양세형 실수에 '골목식당' 되나, 백종원 분노 [...	0.03225806	31	29	0	1	29	sad
7	48	"하루 매출 0원"...'골목식당' 백종원, 떡튀순집 맛+위생에 '경...	0.03225806	31	29	0	1	29	sad
8	51	'맛남' 백종원, 용인 청경채밭 갈아엎자 "비극의 현장, 아까워"	0.03225806	31	29	0	1	29	sad
9	41	'골목식당' 원주 칼국숫집, 암 수술→항암치료ing 근황..백종...	0.04761905	42	39	0	1	39	sad
10	3	'골목식당' 원주 미로 예술시장 개선한 백종원의 원 포인트 ...	0.06250000	32	29	0	1	29	sad
11	33	"지금이라도 다른 일 해"...'골목식당' 백종원, 감자탕집 아들...	0.06250000	32	29	0	1	29	sad
12	50	MBC 측 "백종원 신규 프로그램 출연 확정, 편성 미정" [공식...	0.06250000	32	29	0	1	29	sad
13	17	[종합]"사장으로서 책임감 필요"...'골목식당' 백종원, 주객전...	0.09433962	53	47	0	1	47	sad
14	30	'골목식당' 백종원 "도시락집 맛 달라져.. 날 속인 거냐"	0.13043478	92	38	0	42	42	angry
15	39	"요리는 장비빨이쥬" 주방·가전도 '백종원 효과'	0.14634146	41	21	0	14	21	sad
16	22	백종원 채용공고 '조기마감'...운전기사 1명 경쟁률 '300대 1'	0.20000000	70	2	0	54	54	angry

Figure E.8: High negative ratio tabloids on Baekjongwon

	nth_week	title	neg_ratio	Total	sad	surprise	angry	maxval	newschar
1	19	'놀면뭐하니' 유재석, 드럼 독주회 대성공→故신해철 미발표...	0.05042017	119	112	0	1	112	sad
2	12	[단독] 김철민 "유재석·조세호 후원금, 박명수 라디오...눈물나...	0.10000000	80	71	0	1	71	sad
3	40	새롭게 출발한 '유 퀴즈' 유재석이 오열한 까닭	0.30927835	97	66	0	1	66	sad
4	12	[단독] 유재석·조세호, '암투병' 김철민 깜짝 병문안 '거액 응...	0.38549618	262	160	0	1	160	sad
5	33	유산슬(유재석) ""놀면 뭐하니?' 하면서 외로움이 가장 힘들...	0.39828080	349	209	0	1	209	sad
6	28	가세연 "유재석은 민주당 지지자"...이유는 '파란 옷' 입어서	0.53571429	56	4	0	22	22	angry

Figure E.9: High negative ratio tabloids on YuJaeseok

F Plots

G Fandoms

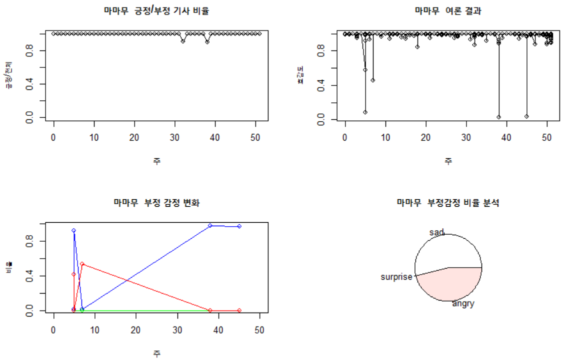


Figure G.1: Fandom analysis plots on MaMaMoo

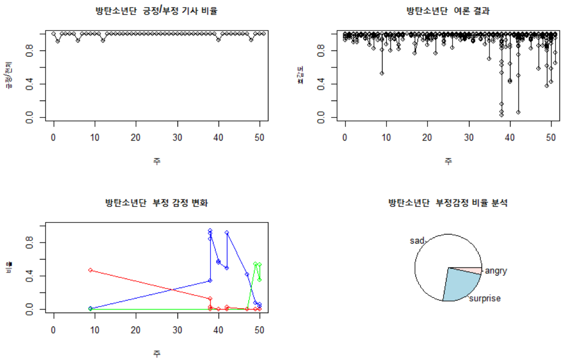


Figure G.2: Fandom analysis plots on BTS

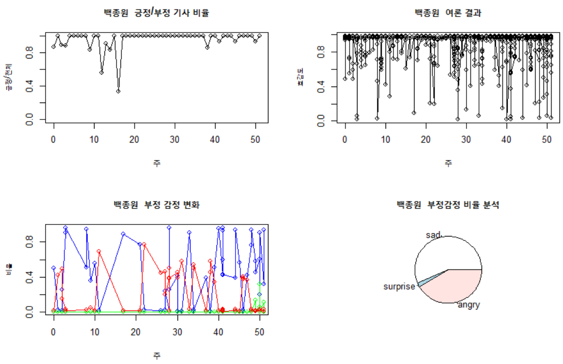


Figure G.3: Fandom analysis plots on Baekjongwon

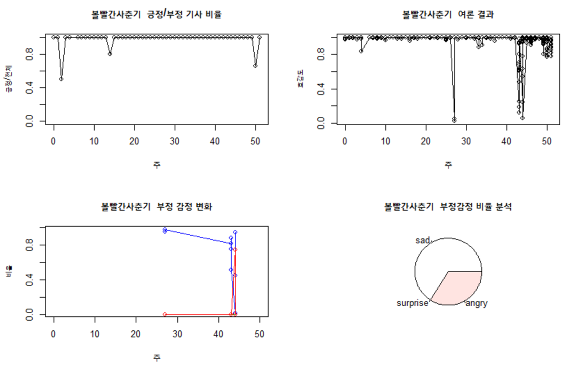


Figure G.4: Fandom analysis plots on Bol4

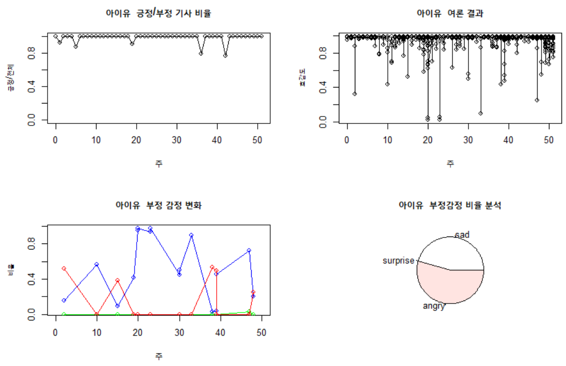


Figure G.5: Fandom analysis plots on IU

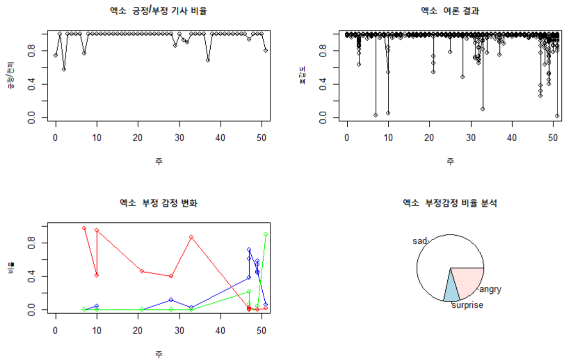


Figure G.6: Fandom analysis plots on EXO

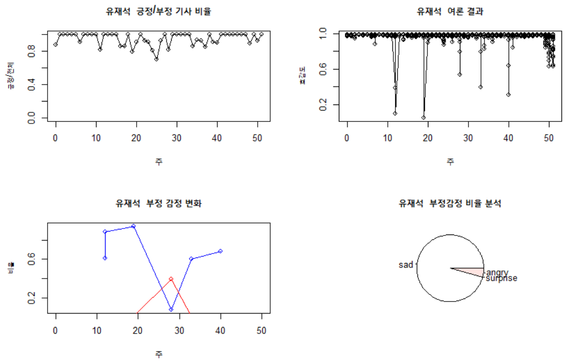


Figure G.7: Fandom analysis plots on YuJaeseok

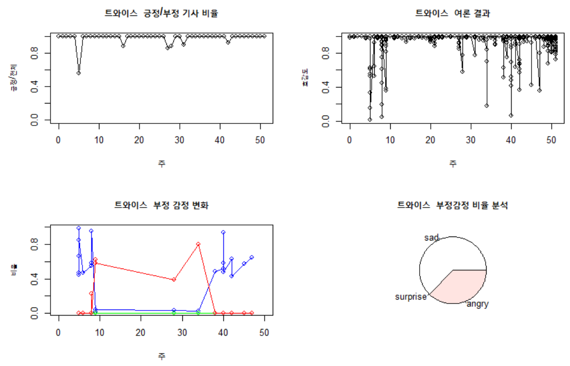


Figure G.8: Fandom analysis plots on Twice

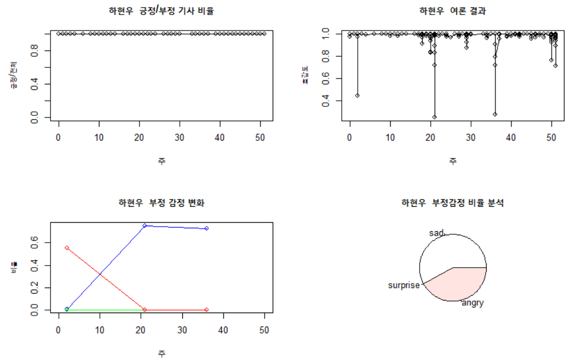


Figure G.9: Fandom analysis plots on Hahyunwoo

H Wordcloud

H.1 Idol word cloud

Figure H.1: Word clouds of MaMaMoo

Figure H.2: Word clouds of BTS

Figure H.3: Word clouds of EXO

Figure H.4: Word clouds of Twice

H.2 Singer word cloud

Figure H.5: Word clouds of Bol4

Figure H.6: Word clouds of IU

H.3 Entertainer word cloud

Figure H.7: Word clouds of Baekjongwon

Figure H.8: Word clouds of YuJaeseok

Figure H.9: Word clouds of Hahyunwoo

H.4 Controversial celebrities word cloud

Figure H.10: Word clouds of Goyoungwook

Figure H.11: Word clouds of Guhara

Figure H.12: Word clouds of GuHyeseon

Figure H.13: Word clouds of NOEL

Figure H.14: Word clouds of MicroDot

Figure H.18: Word clouds of Anjaehyun

Figure H.15: Word clouds of Parkyoocheon

Figure H.19: Word clouds of Jungjoonyeong

Figure H.16: Word clouds of BI

Figure H.20: Word clouds of Jominki

Figure H.17: Word clouds of sulli

Figure H.21: Word clouds of Choijonghoon

Figure H.22: Word clouds of PRODUCE

I Threshold

Figure I.1: EXO Threshold 30

Figure I.2: EXO Threshold 50

Figure I.3: EXO Threshold 100

Figure I.4: YuJaeseok Threshold 30

Figure I.5: YuJaeseok Threshold 50

Figure I.6: YuJaeseok Threshold 100