Correlation analysis between user's emotional comments and popularity measures

Wu, Zechen
Graduate School of Information Science and Electrical Engineering, Kyushu University

Ito, Eisuke
Research Institute for Information Technology, Kyushu University : Associate Professor

https://hdl.handle.net/2324/1463259

Correlation analysis between user's emotional comments and popularity measures

Zechen Wu

Department of ISEE,
Kyushu University
Fukuoka, Japan
wu.zechen.579@s.kyushu-u.ac.jp

Eisuke Ito

Research Institute for Information Technology
Kyushu University
Fukuoka, Japan
ito.eisuke.523@m.kyushu-u.ac.jp

Abstract—**On user generated contents sites, user's comments include the user's impression to the content. The authors believe that user's comments can be used as the data mining resource to evaluate the contents. In this paper, the authors focus on bilibili.tv, a Chinese video sharing site, and analyze user's emotional comments on the site. Toward development of content quality measurement based on emotional comments, they investigate co-relation between emotional comments and popularity measures such as number of replay and bookmark. They pick up the phrase "233", which is a Chinese network jargon which represents lot of laugh, and studied the number of "233" related comments and other popularity measures.**

Keywords—consumer generated media; collective intelligence; social media; user comment; sentiment analysis

I. INTRODUCTION

Consumer Generated Media (CGM) is gaining huge popularity. Particularly, the video-sharing websites such as YouTube (youtube.com) and Nicovideo (nicovideo.jp) are popular among young people than television. We observe a similar situation in China, where youku.com, tudou.com, sohu.com and bilibili.tv are most popular sites. A huge amount of movie contents is stored in these sites. It means that viewers have to use search engine for finding the videos they prefer.

We have proposed a ranking method of viewer comments-based with Japanese video site Nicovideo (nicovideo.jp) that is famous of a kind of unique viewer comment system [1,2]. This paper focuses on bilibili.tv and discuses assessment of video quality.

Bilibili is a video-sharing website based in China, where users can upload, view and add comments to videos. Mr. Xu Yi started a prototype website Mikufans.cn on June 26, 2009 [3,4], which was renamed as Bilibili and continues till the present. Bilibili does not keep the movie contents on the own site, but provides the meta-data to those contents for the users. Most of the subtitles are in Chinese. They provide not only Nicovideo but also movies of U.S. and Korea. Sometime, the site is called as a parasite site.

In recent years, Japanese culture was introduced to gain broad exposure of Japanese pop culture, such as game, manga (comic), animation, J-POP and idols. Actually, it is reported that there are deep-rooted fans abroad with respect to game, manga, anime, J-POP and idols [3]. Actually, Bilibili is strongly affected by Nicovideo and Vocaloid culture from Nicovideo. Not only Miku in prototype website Mikufans.cn is from the Hatsune Miku, a humanoid character voiced by a singing synthesizer which was popular in Nicovideo, but also Bilibili had the similar website structure of Nicovideo. The most characteristic feature is a real-time commentary subtitle. Those subtitles are called "danmaku" (literally bullets), which originally came from nicovideo.com. Many contents from Nicovideo are used in Bilibili.

We focus on numerical integration of emotional comment based on viewer comment, and use these measures to discuss the ranking method of video. This paper statistic analyzes the user's comments, and pay attention to the emotional comment "233", particularly, the correlation between population such as replay number, bookmark number, etc. and "233".

II. FREQUENCY STATISTICS OF TAGS OF BILIBILI

We analyzed the tags for the foreign animations [5]. The analysis is based on the meta-data tagged on Bilibili movies. We collected those meta-data submitted from November 2013 to January 2014. Each movie in Bilibili is assigned with an identifier AV-ID that consists of characters (av) and numbers. We generated the numbers from 1 to 990,000 and succeeded to collect about 480,000 meta-data in HTML format.

Fig. 1. Popularity measures (in av689970)

Fig.1 shows an example of meta-data, where we can see the title, the contributor, the view count, the amount of coins, the number of mylist (or bookmarked users) and the number of "danmaku". The word "danmaku" came from Japanese network jargon, and it means bullets of literature.

Fig. 2. Example of tags for a movie in bilibili.tv (av689970)

Fig.2 shows the tag part of a movie in bilibili. Tags are written to the 2nd line from the bottom with each movie by the video-player as shown in Fig.2. Viewers can assign at most 10 tags to a movie. Fig.2 displays the word "sm21443197" at the bottom line. "sm21443197" is a movie ID in Nicovideo.jp, and it linked to the URL of nicovideo.jp. Then, this word and web link represent that this movie came from other site.

We extracted 345,140 unique tags from the collected meta-data. Table I shows the number of tags with respect to their frequencies. About 250,000 tags were assigned to only one movie.

TABLE I.   NUMBER OF UNIQUE TAGS

Freq.	Num. of tags	Ratio
1	249,825	72.4%
2	34,013	9.9%
3	14,595	4.2%
4	8,579	2.5%
More than 5	38,128	11.0%
Total	345,140	100.0%

III.   USER'S COMMENTS OF BILIBILI

Bilibili has two kinds of user's comment. One is danmaku, and the other is long-comment. As we mentioned before, the word "danmaku" came from Japanese network jargon, and it means bullets of literature.

In case of video replay, danmaku comments appear at the posted period of the video replay, and they are overlaid onto the video, according to the post time. The danmaku comment floats from right to left. The posted user name of a danmaku comment is randomized, so it is impossible to bind a danmaku comment to the posted user.

On the other hand, long comments of bilibili are similar to many BBS (bulletin board system). All long comment are sorted with posted time, and recent long comments are listed at the below part of the video player. Fig.3 shows the long comments for AV689970. As shown in Fig.3, posted user name are shown, then it is easy to bind a long comment and the posted user.

We extracted all texts of comment from the colleted meta-data files. Table II shows the number of comments and their averages.

Fig. 3. Long comments (in av689970)

TABLE II.   NUMBER OF COMMNETS

Comment type	Num. of comments	Average
Danmaku	72,622,153	167.4
Long comment	5,173,862	11.92
Total	77,796,015	179.33

IV.   EMOTIONAL COMMENTS "233"

User's emotional comment contains the feeling about the video content, especially the Danmaku, and it has many Text-Pics and phrases.

When we analyze the user's comments, we found that there are many "233" and 3's loop like "23333" in interesting video content's comments. The expression "233" is commonly used by Internet users in China to express a meaning similar to that of the Internet discourse LOL. It originated from the 233rd emoticon from the popular Chinese website mop.com, which portrays a figure laughing while banging on the floor. [4] "233" is usually used for expressing the emotion of smiling. The more "3" in "233", the stronger emotion is.

The way of using "233" is similar to the way of using Japanese Internet slang "wwww", "wwww" come from warai (smile). It has shown that the more "w", the stronger emotion is in Nicovideo.

We analyze the expression "233" as the first step of analyzing emotional comments. We called the comment "233comment" which comments contain the expression "233", and called the video "233 video" which comments contain the "233 comment". In all the collected 440,849 videos, there are 134,411 "233 videos", about 30.5% of videos are "233 video".

TABLE III.   NUMBER OF "233 VIDEO"

	Num. of Videos
All Video	440,849
"233 Video"	134,411

The number of comments which contain "233" is 1,403,180. Average frequency of "233" comment in "233 video" is 10.44.

	Num. of Comments	Average Frequency
"233 Video"	1,403,180	10.44

Fig.4 shows the distribution of "233 comment". The vertical axis is the number of "233 video", and horizontal axis is the rank of the video order by "233 comment" frequency.
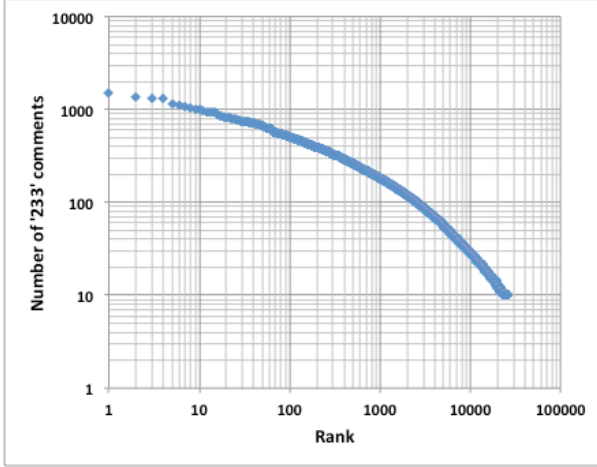

Fig. 4.   Distribution of "233comment"

V.   ANLYSIS OF "233 COMMENT"

In this section, we analyze the correlation between population such as replay number, bookmark number, etc. and "233 comment".

A.   The popularity measures

The more "233" in video comment, the stronger happy emotion video has, and the video is popular. For proving that, we calculate the correlation between popularity measures and "233 comment".

The popularity measures of video can shown by video meta-data, such as replay number, bookmark number, score, "coin". Score and coin is the unique user evaluation system provided by Bilibili. This system is only for registered user. Score point is increasing with account online time, and coin is increasing with account video-watching time. 40 score point when login each day, and 0.1 coin when video-watching time could be gained. User could give most 2 coins to original video, and most 1 coin to other video.

Replay number is increased not only registered user, but also visitor or any access. In the other hand, score point and coin only for registered user, because of the limited point, the number of score and coin is the characteristic of video's popularity measures.

B.   The Correlation Between Population And "233Comment"

For proving the more "233comment", the more popular video is, we calculate the correlation between popularity measures and "233Comment". The results are shown as the Table V- VIII. In the table, P: number of replay, N: number of long comment, D: number of danmaku, M: number of mylist, H: score, C: coin, 233: number of "233comment".

Replay has strong correlation with the score, and number of mylist is correlated to score and coin. In the other hand, the correlation between score and coin is 0.56, neither strong nor weak. Number of "233 comment" is only strongly correlates with number of danmaku; its correlation with others is weak. It is hard to say the more "233 comment" has, the more popular video is.

TABLE V.   VIDEO CORRELATION

	P	N	D	M	H	C
N	0.38					
D	0.35	0.22				
M	0.61	0.37	0.26			
H	0.78	0.43	0.30	0.70		
C	0.47	0.51	0.21	0.70	0.56	
233	0.31	0.16	0.56	0.18	0.23	0.19

We think that coin is the most important indicator can show the popularity measures, so we also calculate the correlation of Top100, Top1000 and Top10000 videos order by coin. The result is shown as Table VI – VIII.

TABLE VI.   COIN ORDER TOP10000 VIDEO CORREALTION

	P	N	D	M	H	C
N	0.28					
D	0.12	0.11				
M	0.50	0.27	0.05			
H	0.70	0.34	0.04	0.61		
C	0.37	0.45	0.10	0.63	0.47	
233	0.15	0.06	0.50	-0.02	0.03	0.06

TABLE VII.   COIN ORDER TOP1000 VIDEO CORREALTION

	P	N	D	M	H	C
N	0.17					
D	-0.05	0.03				
M	0.31	0.16	-0.13			
H	0.61	0.30	-0.17	0.49		
C	0.21	0.45	0.01	0.47	0.34	
233	0.12	-0.003	0.47	-0.17	-0.05	-0.03

TABLE VIII.   COIN ORDER TOP100 VIDEO CORREALTION

	P	N	D	M	H	C
N	0.23					
D	-0.17	-0.04				
M	0.60	0.14	-0.31			
H	0.70	0.51	-0.35	0.71		
C	0.33	0.54	0.01	0.41	0.53	
233	-0.06	-0.06	0.53	-0.23	-0.22	-0.12

Replay has strong correlation with the score as expected, and number of mylist is correlated to score and coin too. The correlation between score and coin is neither strong nor weak. It is almost the same result, and also hard to say the more "233 comment" has, the more popular video is.

We think the correlation of all videos probably is disturbed by the videos not "233 video". So we calculate the same correlation of "233 video". The result is shown as the Table IX- XII.

TABLE IX.   "233 VIDEO" CORRELATION

	P	N	D	M	H	C
N	0.36					
D	0.31	0.20				
M	0.60	0.35	0.22			
H	0.77	0.42	0.26	0.69		
C	0.47	0.50	0.19	0.69	0.56	
233	0.28	0.15	0.54	0.15	0.20	0.18

TABLE X.   COIN ORDER TOP10000 "233VIDEO" CORREALTION

	P	N	D	M	H	C
N	0.28					
D	0.12	0.11				
M	0.50	0.27	0.04			
H	0.70	0.34	0.04	0.61		
C	0.37	0.45	0.10	0.63	0.47	
233	0.15	0.06	0.50	-0.03	0.03	0.06

TABLE XI.   COIN ORDER TOP1000 "233VIDEO" CORREALTION

	P	N	D	M	H	C
N	0.17					
D	-0.05	0.03				
M	0.30	0.17	-0.14			
H	0.62	0.30	-0.16	0.49		
C	0.21	0.45	0.01	0.47	0.33	
233	0.12	-0.002	0.46	-0.18	-0.05	-0.03

TABLE XII.   COIN ORDER TOP100 VIDEO CORREALTION

	P	N	D	M	H	C
N	0.23					
D	-0.19	-0.04				
M	0.60	0.14	-0.32			
H	0.70	0.51	-0.36	0.71		
C	0.33	0.55	0.01	0.42	0.53	
233	-0.04	-0.06	0.51	-0.22	-0.21	-0.13

The result is almost the same as the all video's result. Replay has strong correlation with the score as expected, and number of mylist is correlated to score and coin too. The correlation between score and coin is neither strong nor weak. It is almost the same result, and also hard to say the more "233 comment" has, the more popular video is. We think the population affects the number of coin; we want to remove the affection from the number, so we also calculate the correlation between "233 comment" and coin/replay. The result is shown as Table XIII. In the table 233/CP: correlation between "233 comment" and coin/replay.

TABLE XIII.   "233 COMMENT" CORRELATION WITH C/P

Correla-tion	"233 video"	Top 10000	Top100	Top100
233/CP	0.04	-0.12	-0.13	-0.07

The result shows that "233 comment" is not correlated to coin/replay.

VI.   CONCLUSION

This paper is one step of analyzing the user's emotional comment. We collected meta-data of the Bilibili website. We analyzed the emotional comment, "233" to know the relationship between "233 comment " and popularity measures. The correlation was calculated to show the relationship, it is hard to say the more "233 comment" has, the more popular video is.

Future work is necessary. Making a new method of ranking videos base on user's emotional comment, and checking the effectiveness of it.

REFERENCES

[1] Naomichi Murakami, Eisuke Ito, Emotional video ranking based on user comments, Proc. of iiWAS2011, ACM, pp.499-502, 2011.

[2] Naomichi Murakami, Eisuke Ito, Video weighting method based on viewer's comments and its evaluation, Proc. of DEIM2012, F8-3, 2012. (in Japanese)

[3] Bilibili (Jan.20,2014). In Wikipedia: The Free Encyclopedia. Retrieved from http://zh.wikipedia.org/wiki/Bilibili

[4] Biliili (Jan.20,2014). In Wikipedia: The Free Encyclopedia. Retrieved from http://en.wikipedia.org/wiki/Bilibili

[5] T. Sakurai: "Chinese girls loving Japanese too much", PHP, ISBN-10: 4569812422, 2013. (in Japanese)