# Comparison Analysis of Danmaku and Comment on bilibili.com

For INST 741

By Yue Liu

**May 14 2017.**

## Table of Contents Page

## Introduction and Motivation

Compared to the traditional media, consumer-generated media, especially video-sharing website, are gaining a huge popularity. Young people are more inclined to gain information and express their opinions on such websites as YouTube and Facebook. However, although YouTube has various content, the lower user engagement and terrible content of comments make the comment function useless, so that YouTube become a pure video-sharing website without any interaction among audiences (Schultes, 2013). We observe the opposite situation in China, where bilibili.tv has three millions daily views and 350 thousand active daily users while having high user engagement, user retention and friendly interactions among viewers themselves and between uploader and viewer. The function named "danmaku" plays an important role for improving the user engagement and strengthening the participatory culture of bilibili.tv.

Therefore, this research focuses on the comparison analysis of traditional comment and danmaku on bilibili.tv. The motivation is to aims to evaluate the advantage and disadvantage of danmaku that can supplement the traditional comment function, thus improving the user engagement of such video-sharing websites.

Currently, there is no comparison analysis research between danmaku and traditional comment. Most Chinese informal articles and research papers only focus on the analysis of the effects of danmaku. One relevant knowledge that is relevant to this project is from Cohen & Lancaster (2014). The researcher found that audiences when using social media during television viewing made them feel like they were watching the shows in a group setting, in which the experience is highly similar to the usage of danmaku when watching the video on bilibili.tv

## Danmaku

In general, "danmaku", literally as "bullet curtain", is a kind of comment function only exist in certain video stream sites. Unlike the traditional video sharing sites, like YouTube, which put all comments below the video and take the comments in time order, some Japanese and Chinese video sharing sites with "Danmaku" function not only have the traditional comment box, but also enable audiences to "shoot" their comments immediately onto the video they're watching.

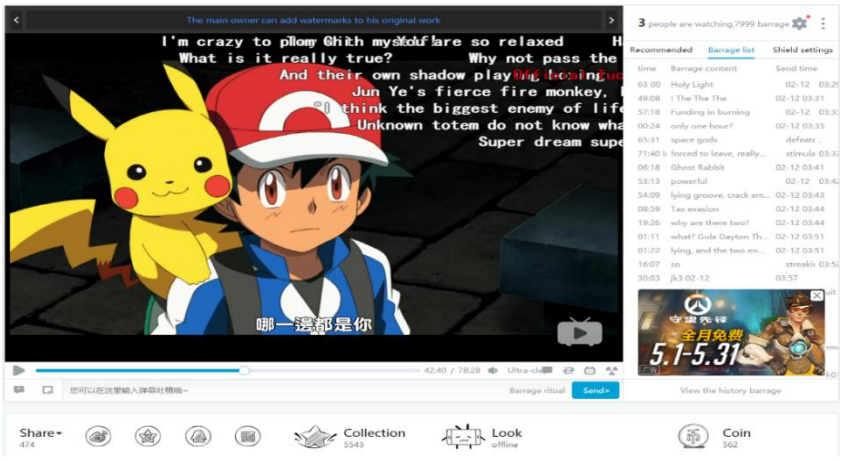



Figure 1. Danmaku

Compare to the traditional comment box, danmaku has the following advantages:

- *High synchronization*
  - When a video is played, previous danmaku will be displayed in accordance with the video time axis. Therefore, the content of danmaku go beyond the real-time, and the previous comments and later comments coexist together. Therefore, danmaku creates a chat room experience in which the audiences feel like watching and playing together with others.
- *Appropriate timing*
  - When using traditional comments box below the video, if audiences want to discuss a specific video plot, audiences have to describe what scene and context they are talking about before their real comments. On the contrary, when using danmaku, the contents are on the perfect timing and users don't need to describe the whole context.
- *Supplement of video content*
  - Danmaku can enrich the content of video through friendly ridiculing the video and plot, translating content into Chinese in form of subtitle, introducing the related knowledge and background, etc. Since not all videos can attract audiences' attention from the beginning to the end, danmaku can help to capture and attract users' distracted attention.

Compared to the traditional comment box, danmaku is more modern and interesting participation pattern that enables audiences absorb other people's thoughts toward the specific file-screenshot or scene in the first time.

**Hypothesis and Expectation**

Because of the above advantages of danmaku, we expect that audiences should have different behavior patterns when using danmaku and traditional comment box from the perspectives of attitude, word frequency, the length of comment, etc. Specifically, we made the following hypothesizes:

- *Hypothesis 1*: Compared to traditional comment, the length of danmaku should be short.
- *Hypothesis 2*: Danmaku should have a higher sentiment analysis score than traditional comment
- *Hypothesis 3*: The number of danmaku should be larger than the number of comment for each video
- *Hypothesis 4*: Danmaku and comment should have different patterns from the perspective of word frequency analysis.

**Data Collection**

Since bilibili.tv does not provide any formal API connector to acquire their data, I developed Python scraping code that can automatically download the comment and danmaku of certain video by given video URL. The basic methodology is as follows:

- Each video in Bilibili is assigned with an identifier AV-ID that consists of characters (av) and numbers. The URL for each video will be the combination of "www.bilibili.com/video/" and the corresponding AV number. For example, if a video's AV number is 10591987, the URL will be www.bilibili.com/video/av10591987. The number of AV-ID is also called by aid.

- However, the video content, comment and danmaku cannot be directly identified by only given AV number. Bilibili uses another unique ID number (cid) within their own databases and servers in order to prevent web scraping. Each cid is a unique identifier that has one-to-one relationship with AV number. Although cid cannot be directly observed, it can be captured through analyzing the source code of each video's site.
- Through using Chrome Developer Tools to analyze the web structure of bilibili.tv and requests from the browser, we found that the content and comments can be obtained through the following URL: http://api.bilibili.cn/feedback?aid= and "comment.bilibili.com". For example, comments and danmaku of the video (av10575757) can be contained through "http://api.bilibili.cn/feedback?aid=10575757" and "comment.bilibili.com/17459077.xml", while "10575757" is the aid and "17459077" is the cid for this video. In addition, through using the above URLs, the danmaku will be displayed as XML format, while the traditional comment will be represented as JSON format of Unicode, not Chinese character.


Figure 3. Scraping of Comment

Figure 2. Scraping of Danmaku

Therefore, through deploying the above methodology, I scraped the comment and danmaku for 18 popular videos under gaming category from the past 5 months (Jan 2017 – May 2017). Because of the display limitation of danmaku on bilibili.tv that each video can only display up to 3000 danmaku, the program I developed can only scrap the latest 3000 danmaku. As a result, in total there are 56,655 comments and 108,459 danmakus for 18 videos.

## Building Sentiment Analysis Package

Since there is no open source sentiment analysis package for Chinese language, we developed a dictionary-based sentiment analysis method for comparison analysis in Chinese. The basic unit for this package is a sentence clause that consists of privative word (optional), adverbs of degree and sentiment word that has been manually labelled by previous research. For example, "not very good" can been a sentence clause, in which "not" is "not" word, "very" is adverbs of degree and "good" is the sentiment word that represent positive emotion. Therefore, the score for "not very good" is (-1) ^1 * 1.25 * 0.747127733968.

For package building, we used the following dictionary:

- ***Sentiment word dictionary***

Boson LLC (2014, December 29). BosonNLP sentiment word dictionary [TXT]. Retrieved by May 5, 2017, from http://bosonnlp.com/dev/resource

The dictionary contains the Chinese words and the corresponding score that are manually labelled by previous researches and other intuitions. Since sentiment word dictionary plays an important role for the dictionary-based sentiment analysis package, the sentiment word dictionary should have the same context as the danmaku and comments on bilibili.tv The dictionary is automatically generated from millions sentiment records from news, Twitter-like and Reddit-like social media. Therefore, the dictionary can capture a number of cyber-speak, emoji and abbreviation.

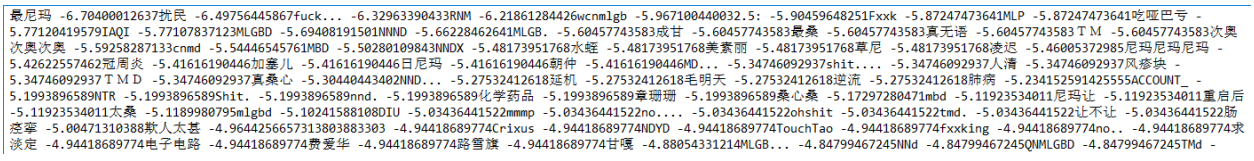

最尼玛 -6.70400012637扰民 -6.49756445867fuck... -6.32963390433RNM -6.21861284426wcnmlgb -5.967100440032.5: -5.90459648251Fxxk -5.87247473641MLP -5.87247473641吃哑巴亏 -5.77120419579IAQI -5.77107837123MLGBD -5.69408191501NNND -5.66228462641MLGB. -5.60457743583成甘 -5.60457743583最桑 -5.60457743583真无语 -5.60457743583ＴＭ -5.60457743583次奥次奥次奥 -5.59258287133cnmd -5.54446545761MBD -5.50280109843NNDX -5.48173951768水蛭 -5.48173951768美素丽 -5.48173951768草尼 -5.48173951768凌迟 -5.46005372985尼玛尼玛尼玛 -5.42622557462冠周炎 -5.41616190446加塞儿 -5.41616190446日尼玛 -5.41616190446朝仲 -5.41616190446MD... -5.34746092937shit.... -5.34746092937人清 -5.34746092937风瘘块 -5.34746092937ＴＭＤ -5.34746092937真桑心 -5.30440443402NND... -5.27532412618延机 -5.27532412618毛明天 -5.27532412618逆流 -5.27532412618肺病 -5.2341525914255555ACCOUNT_ -5.1993896589NTR -5.1993896589Shit. -5.1993896589nnd. -5.1993896589化学药品 -5.1993896589章珊珊 -5.1993896589桑心桑 -5.17297280471mbd -5.11923534011尼玛让 -5.11923534011重启后 -5.11923534011太桑 -5.1189980795mlgbd -5.10241588108DIU -5.03436441522mmmp -5.03436441522no.... -5.03436441522ohshit -5.03436441522tmd. -5.03436441522让不让 -5.03436441522肠痉挛 -5.00471310388欺人太甚 -4.964425665731380388330 3 -4.94418689774Crixus -4.94418689774NDYD -4.94418689774TouchTao -4.94418689774fxxking -4.94418689774no.. -4.94418689774求淡定 -4.94418689774电子电路 -4.94418689774费爱华 -4.94418689774路曾旗 -4.94418689774甘嘎 -4.88054331214MLGB... -4.84799467245NNd -4.84799467245QNMLGBD -4.84799467245TMd -

Figure 4. BosonNLP sentiment word dictionary

- ***"Not" word***

  The dictionary contains 45 words that represent the similar meaning as "not". The appearance of "not" word will directly turn the positive clause to the negative. The example, "not very good", can reflect the effect of this type of word.

| 不 | 不曾 | 不用 | 决不 | 未尝 | 从不 | 一无 | 从来不 | 切莫 |
|---|---|---|---|---|---|---|---|---|
| 非 | 没 | 无须 | 休想 | 未曾 | 从未 | 并 | 从没 | 永不 |
| 无 | 没有 | 并非 | 永不 | 毋 | 从未有过 | 未 | 绝非 | 休想 |
| 未 | 请勿 | 毫无 | 不要 | 莫 | 尚未 | 尚无 | 远非 | 绝不 |
| 拒绝 | 杜绝 | 否 | 弗 | 木有 | 毫不 | 不必 | 禁止 | 忌 |

- ***Adverbs of degree***

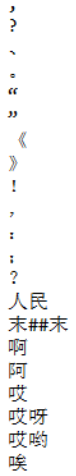
  The dictionary contains six levels of adverbs of degree words with the corresponding score. The score represents the intensity of the emotion or action.

| Level | Score | Example | Number |
|---|---|---|---|
| Extreme | 2 | 非常 绝对 十分 要命 异常 满 极 等 | 69 |
| Very | 1.75 | 颇为 实在 特别 尤其 着实 太 甚 等 | 42 |
| More | 1.25 | 那么 愈加 这样 足足 更加 还 多 等 | 37 |
| -ish | 0.5 | 略微 略加 稍许 有点 相当 些 蛮 等 | 29 |
| Insufficiently | -0.5 | 半点 不大 轻度 丝毫 相对 微 弱 等 | 12 |
| Over | -0.75 | 过度 超额 出头 何止 开外 溢 忒 等= | 27 |

- ***Stop word***

  Chinese stop word lists. (2012). Retrieved May 8, 2017, from http://www.datatang.com/data/43894

  The dictionary contains 1208 stop words that do not contain important significance, including punctuation mark, modal particle, pronoun, etc.




Therefore, by employing the above four dictionary, the basic methodology is as follow:

- Using “Jieba” python package to split the danmaku or comment into a word list
- Removing all stop words, e.g. “this”, “that”, etc.
- Categorizing the remaining words into sentiment words, “not” words and adverbs of degree and the corresponding score if exist.
- Calculating the sentiment score for the clause

However, the main disadvantage of the above methodology is that the sentiment score for comment that has multiple sentences will be the sum of all sentences’ score. In other words, there is a positive relationship between the length of the comment (or the number of sentences in that comment) and its score.

**Data Analysis & Results**

- *Hypothesis 1*: Compared to traditional comment, the length of danmaku should be short.

  Descriptive Analysis

| | Total number | Mean | SD |
|---|---|---|---|
| Comment | 56655 | 15.83 | 22.63688 |
| Danmaku | 108459 | 7.836 | 5.221389 |

```
        Welch Two Sample t-test

data:  length by category
t = -82.925, df = 59823, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.184329 -7.806377
sample estimates:
  mean in group danmu mean in group comment
             7.83603              15.83138
```

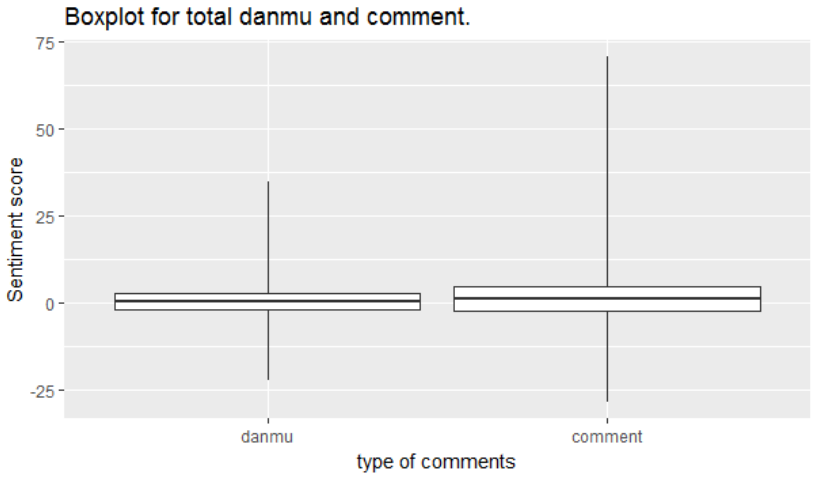
From the above table and the result of two sample t-test, since p-value is smaller than 2.2e-16, we reject the null hypothesis and conclude that the average difference between the length of danmaku and comment is significant and danmaku is usually shorter.

- *Hypothesis 2*: Danmaku should have a higher sentiment analysis score than traditional comment

| | Total Number | Mean of Sentiment Score | SD |
|---|---|---|---|
| Comment | 52683 | 1.2227 | 2.746119 |
| Danmaku | 108459 | 0.4024 | 2.211201 |

```
        Welch Two Sample t-test

data:  score by category
t = -49.076, df = 73583, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to
95 percent confidence interval:
 -0.8530243 -0.7875044
sample estimates:
  mean in group danmu mean in group comment
            0.4024075            1.2226718
```




Before doing the analysis, we removed all comment records that are greater than the third quartile, because of the disadvantage of the dictionary-based sentiment analysis package we mentioned above. In other words, the comparison analysis should exclude those comments that contain multiple sentences and thus compare the sentence-level content of comments and danmaku.

From the above table and the result of two sample t-test, since p-value is smaller than 2.2e-16, we reject the null hypothesis and conclude that the average difference between the sentiment score of danmaku and comment is significant. However, the result is opposite to our expectation that danmaku may have a higher sentiment score.

- *Hypothesis 3*: The number of danmaku should be larger than the number of comment for each video

| | Number of video | Mean of number of records | SD |
|---|---|---|---|
| Comment | 18 | 4413.278 | 5932 |
| Danmaku | 18 | 15023.56 | 13882 |

```
        Welch Two Sample t-test

data:  number by category
t = -2.9818, df = 23.008, p-value = 0.006668
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17971.243  -3249.312
sample estimates:
mean in group comment   mean in group danmu
             4413.278             15023.556
```

From the above table and the result of two sample t-test, since p-value is smaller than 2.2e-16, we reject the null hypothesis and conclude that the average difference between the number of danmaku and comment for each video is significant. And, the number of danmaku is usually larger than the number of comment for each video.

- *Hypothesis 4*: Danmaku and comment should have different patterns from the perspective of word frequency analysis.

After using Jieba Python package to split content into words, we counted the top words for danmaku and comment. We list the top meaningful words in the following table. (For better review, I translated them into English)

| **Comment** | **Number** | **Danmaku** | **Number** |
|---|---|---|---|
| The first | 6174 | 23333 and 3's loop | 11620+ |
| "aaa" | 6126 | LOL | 9176 |
| game | 3230 | Game | 1746 |
| video | 1804 | Die | 1334 |
| update | 1349 | +1 | 618 |
| Finally | 927 | Danmaku | 554 |
| comment | 837 | BGM (background music) | 362 |

The expression "2333" and 3's loop like "23333333" is commonly used by Chinese users and represents the similar meaning as LOL. However, the iterative "3" can represent their intensity of the emotion of smiling. The more "3" in "233", the stronger emotion is. "+1" is used by Chinese internet users to represent that they agree with the precede danmaku. For example, if someone leave comment like "I really like this uploader", other people can directly use "+1" to represent that they have the same thought but don't need to waste time and type again.

From the difference of top words in comment and danmaku, we found that audiences always use danmaku to express their emotion for a particular scene and plot. For example, when game character who is controlled by uploader died because of some accidence, audiences may leave a danmaku and friendly laugh at uploader, e.g. "23333", "You finally die, 2333" or "No

one can defeat me in my BGM". On the other hand, audiences typically use comment to represent their opinion and emotion for the whole video. For example, when uploader always uploads their video in a very low frequency, audiences may leave a comment like "Finally, I waited for this video", "I'm the first to leave the comment" or "the video is finally uploaded in my remaining life".

**Conclusion**

From the above analysis, we found that the first, third and the fourth hypothesis are approved and the second hypothesis about sentiment analysis is opposite to our analysis. Therefore, we can conclude that because of the characteristics of danmaku, like high synchronization with other audiences and appearing on appropriate timing, audiences tend to leave a relatively short danmaku to express their emotion for a particular scene and interact with uploader and other audiences. Also, because audiences do not need to describe the context of their danmaku, they usually tend to leave more danmaku than traditional comments. However, if audiences hold various emotions and opposite opinions toward a particular plot, it is also much easier for audiences to have a pointless quarrel with each other, leading to a low sentiment score

**Limitation and Next Steps**

During the research, there are three limitations that more or less affect the reliability of the analysis results and conclusion.

- Biased data collection:

Because of convenience sampling, we only scraped and analyzed the comments and danmakus from game video and popular updaters. Different video topics may affect the word frequency and relative conclusion.

- Chinese word segmentation

All words in English have been naturally spitted by space, like XXX YYY ZZZ, while words in Chinese are aggregated together, like XXXYYYZZZ. Therefore, to guarantee the reliability of sentiment analysis result, the main work is to appropriately split the sentence into words. Although we have added some words that are frequently used by audiences on bilibili.tv

into the "Jieba" package, the package still cannot perfectly split the sentence, e.g. the name of uploader.

- Dictionary-based sentiment analysis

Compared to machine learning-based sentiment analysis method, one of the main disadvantage of dictionary-based method is that the evaluation is highly depending on polarity lexicon that was derived from human raters and previous researcher's intuitions. Although the polarity lexicon we found is generated from Twitter-like and Reddit-like social media, there is still a difference between the context of dictionary and bilibili.tv.

For next steps, we want to build a machine-learning based method and compare these two results from dictionary-based and machine learning based methods. In addition, it is also important to find more dictionaries and corpus for such context.

All codes and relative datasets can be found in this link: https://github.com/rex1517liu2368/INST741

# References

Schultes, Peter; Dorner, Verena; and Lehner, Franz, "Leave a Comment! An In-Depth Analysis of User Comments on YouTube" (2013). Wirtschaftsinformatik Proceedings 2013. 42.

Wu, Z., & Ito, E. (2014). Correlation Analysis between User's Emotional Comments and Popularity Measures. 2014 IIAI 3rd International Conference on Advanced Applied Informatics. doi:10.1109/iiai-aai.2014.65

Cohen, E. L., & Lancaster, A. L. (2014). Individual Differences in In-Person and Social Media Television Coviewing: The Role of Emotional Contagion, Need to Belong, and Coviewing Orientation. Cyberpsychology, Behavior, and Social Networking,17(8), 512-518. doi:10.1089/cyber.2013.0484

What's Danmaku? (n.d.). Retrieved May 17, 2017, from http://danmaku.weebly.com/