

Look At The Barrage From Another Angle

—— Visualization of Barrage Data
With Python & Text Sentiment Analysis



| | 发送时间 | 会员等级 | 弹幕点赞 | 情感分值 |
|----|---------------------|----------|---------|----------|
| 0 | 2020-12-24 00:00:00 | 0.760598 | 5.36026 | 0.597022 |
| 1 | 2020-12-24 00:15:00 | 0.755379 | 7.26792 | 0.539171 |
| 2 | 2020-12-24 00:30:00 | 0.751851 | 6.81649 | 0.512125 |
| 3 | 2020-12-24 00:45:00 | 0.806384 | 6.3929 | 0.520323 |
| 4 | 2020-12-24 01:00:00 | 0.786389 | 7.26516 | 0.510811 |
| 5 | 2020-12-24 01:15:00 | 0.798296 | 6.7297 | 0.504133 |
| 6 | 2020-12-24 01:30:00 | 0.770896 | 7.53894 | 0.530489 |
| 7 | 2020-12-24 01:45:00 | 0.82215 | 8.24078 | 0.508483 |
| 8 | 2020-12-24 02:00:00 | 0.725625 | 7.96195 | 0.51156 |
| 9 | 2020-12-24 02:15:00 | 0.754767 | 7.89367 | 0.585757 |
| 10 | 2020-12-24 02:30:00 | 1.04552 | 10.4143 | 0.562774 |

Look At The Barrage From Another Angle

———— Visualization of Barrage Data
With Python & Text Sentiment Analysis

I. Background and Significance:

"Exciting offer Season 2" is launched by Tencent video a career observation reality show, by he Jiong, Sabine as a regular refueling group members. Eight law school students with a lawyer's dream, under the guidance of four lawyers, experienced a one-month internship and competed for the offer of two law firms.

The second season brings a new career experience on the basis of the last season's legal career theme. After a rigorous and intense interview, eight law interns of different backgrounds and personalities will enter the "Junhe Law Office" of China's Red Circle, where they will embark on an internship adventure. The first season of the program scored 8.3 in Douban, with a very high degree of attention and a good reputation. This variety show heat and controversy gathered, including a variety of labels. Combined with the above considerations, the final "exciting offer season 2" has become our topic.

The significance of this topic is to start with the most popular variety show, based on a large number of program screen data, the audience's preferences, concerns for a reasonable analysis. Through

analysis, we can understand the current trend of popular program appreciation, but also avoid market misjudgment caused by lack of information. By understanding the audience's preferences, concerns and other aspects of the situation, can more accurately judge the market demand, commercial analysis, so as to maximize the benefits of all aspects, has a wide application prospects.

II. Functions and Techniques to be Implemented:

We use a Jupyter Notebook platform, language for the Python, to achieve the functions of data crawler, data cleaning, visual analysis (chart analysis, text emotional analysis).

By crawling through the seven bullet screens of offer second season (interview six), At present, a total of 261724 bullet screen data. We crawled through the data, data cleaning, The specific steps of cleaning data include : 1. delete duplicate data; Delete special characters (because special characters can not be expressed for emotional analysis); Filter missing values (because data containing missing values can not be analyzed); 4. Selection of data to be analysed, And the selected data for the corresponding visual analysis and a series of operations. Among them, In the above steps to delete duplicate data, we applied the mechanical compression deweighting function. Because we need to analyze the text emotion of the corresponding screen, We need word segmentation, word frequency statistics, But when we do word frequency statistics, Just one count of words, Therefore, we need to use the "mechanical compression weight removal function ".

After cleaning the data obtained by the crawler, we can visualize the analysis. Exciting offer Season 2 has been broadcast seven times (including interview articles). The first phase due to the upgrade of the rules, interns face high pressure assessment, video time is longer, the plot has twists and turns, so the number of screen, up to 46367. Because the intern interview was tortured by the soul, the audience was depressed, and the video length was shorter than the positive film, so the number of screens was the least, only 18743. We have charted the specific data for each issue. We counted the number of screens sent by users. According to the chart, the user name ID "想太多de猫" sends the most data, reaching 375. We also counted the distribution of membership levels. From the pie chart, we can see that most of the viewers watching this variety are members of 0, so we know that the audience is because the program meets their preferences, not because of "membership stickiness".

Of all the screens, Ding Hui was mentioned by the audience far more than the other seven interns ,17698 times, followed by Zhan Qiuyi ,4675 times by the audience, and Liu Yucheng was least mentioned by the audience, only 1016 times. We then drew a picture of eight interns and found that many viewers recognized Ding Hui ," refueling "," like "," optimistic "," support "and other words appear more frequently; for the more introverted Zhan Qiuyi, the audience also very much like, from" beautiful "," Liu Yifei "," good-looking "and other high-frequency words can be seen, many people like her is based on face value; as Wang Xiao from Stanford, the audience presents the situation on both sides, some people say" Wang Xiaohao ", others think he is "Versailles "; Zhu Yixuan also, some people think

she is" lovely ", others "hate" her; Qu Zelin was praised for his high EQ and loveliness; Li Jinye's "handsome" was praised by the audience, and many even thought he was like the popular intern he Yunchen in the first season; Wang Yingfei, who graduated from the National people's Congress, was praised by the audience for his "good-looking" and "beautiful "; Liu Yucheng, who had passed the examination, was praised by the audience for his" good professional knowledge ".

Through the screen content we also carried on the text emotion analysis. We use Baidu open source mature NLP model Senta to calculate the emotional score of screen content. Senta is Baidu open source Chinese emotion analysis model, which can be used for emotional analysis of Chinese sentences. The output is one of the positive/neutral/negative. Code reference official document <https://github.com/PaddlePaddle/PaddleHub/tree/release/v0.5.0/demo/senta>. and we found that the overall emotional score of "exciting offer season 2" was higher than 0.5, and the audience showed a higher positive tendency. The more members of the higher the audience can stick to the end, the number of screen likes from the video play began to show an increase trend, in the last 15 minutes suddenly fell. The emotional score is high and low in the middle. Finally, we resample the processed data. Added the attribute of an emotional score.

III. CONCLUSIONS:

Based on our visual analysis, we draw a lot of conclusions. And we learned some basic information about the program, such as membership level distribution, users TOP10 the number of screen

launches, the distribution of each screen, the hot topics that the audience likes to discuss, and the number of references of each intern in the program. Finally, based on this information, we got eight interns' word cloud pictures, intuitively showing the audience's evaluation and attention to the interns. After resampling, the average value of member grade, bullet screen likes and emotional score is arranged, and the project is completed.

In this project, we still need to improve, we will follow up the program update development, real-time expansion of data, so that our conclusions more perfect. At the same time, in the future work life, when we are engaged in some work about program evaluation or content hot spot extraction, we can apply the technology involved in this project to practice, based on data speaking, more accurate judgment of market demand, commercial analysis, so as to maximize the benefits of all aspects.

Group 5 Division of Labor

Mainly Divided into Three Parts :

1.Code Part

| | | |
|------------|-------------------------|-----|
| 2030223131 | Zhao LUO (罗钊) | 40% |
| 2030223118 | Sunqingnian WANG (王孙卿念) | 30% |
| 2030223158 | Shuang PENG (彭爽) | 30% |

2. Text Composition Part :

| | | |
|------------|-------------------------|-----|
| 2030223131 | Zhao LUO (罗钊) | 35% |
| 2030223118 | Sunqingnian WANG (王孙卿念) | 35% |
| 2030223158 | Shuang PENG (彭爽) | 30% |

3.Making PPT (PowerPoint) Part :

| | | |
|------------|-------------------------|-----|
| 2030223131 | Zhao LUO (罗钊) | 30% |
| 2030223118 | Sunqingnian WANG (王孙卿念) | 40% |
| 2030223158 | Shuang PENG (彭爽) | 30% |