# Stories Behind Song Comments: Case of NetEase Cloud Music

Natural Language Processing Final Project

Shuting Chen

March 22, 2019

## 1 Introduction

As an emerging and ubiquitous Chinese music streaming service focusing on building great music community, NetEase Cloud Music distinguishes itself by its abundant and emotional comments. As of April 2017, four years after it was launched in April 2013, the music service has 300 million users. In terms of NetEase Cloud Music's distinct features (e.g. users are highly active in posting comments and interacting with other users), conducting textual analysis for its comments is likely to uncover intrinsic properties of song comments and even to provide reasonable explanations for user behavior. Therefore, this project mainly analyzes comments of 56 songs using topic modeling, clustering, and sentiment analysis.

Specifically, this project aims to answer following two questions: 1. What are the differences between hot comments (i.e. comments with most liked counts) and normal comments? Can we find any tricks to write popular comments? 2. What can we tell about the features of songs given their comments? Do these comments correctly convey the emotional feelings of songs?

## 2 Dataset

Since user comments are transmitted by Ajax (asynchronous JavaScript and XML) with encryption, it is very difficult for me to figure out the encryption key by myself. Hence, I used sample codes found on GitHub with elaborate steps of how to get parsed parameters to request comments. Specifically, I modified the codes so that the scraped comments maintain the features and format that I need.

The dataset used in this project contains 15 hot comments which are identified by the platform and 1,000 most recent normal comments[1] for each of 56 songs included in the list of Top 100 New Songs 2018. Each of 56 songs originally have more than 10 thousand comments, however, I only scraped the most recent 1,000 comments for the sake of not being detected by the anti-scraping system. In total, there are 56,840 comments, including 840 hot comments and 56,000 normal comments. After removing emojis and some special punctuation, the raw length of total comments (in characters) is approximately 1.9 million.

## 3 Methods

To explore differences between hot comments and normal comments, I compared the distribution of comment length and 30 most frequently used tokens in two types of comments. These two preliminary analyses were conducted based on the hypotheses that the length of comments tends to

[1]These are most recent comments by February 5, 2019, when I scraped data on the web version of NetEase Cloud Music.

be longer in hot comments than in normal comments and the words used might be quite different between two types of comments. Further, after removing stopwords and high-frequency tokens such as “喜欢”(like), “真的”(really), “首歌”(a song), “爱心”(loving heart), and “希望”(hope), I built latent Dirichlet allocation (LDA) models for hot comments and normal comments, respectively. This might be helpful to compare the potential diversity in topics involved in both types of comments.

To examine whether comments reflect emotion which is consistent with their songs', I implemented sentiment analysis[2] and drew radar plots in terms of seven groups of emotional tokens (i.e. “乐”(le - happiness), “好”(hao - good), “怒”(nu - anger), “愁”(chou - worry), “惊”(jing - shock), “恶”(wu - hate), and “惧”(ju - fear)). Instead of conducting sentiment analysis for each song, which might produce less accurate results due to relatively small text size, I first clustered 56 songs into 5 groups and did sentiment analysis for each group. Particularly, I employed K-means algorithm with following preparation steps for clustering:

1. Select 8 most frequently used words to represent each song (further excluding words such as “大哭”(cry), “爱”(love), “加油”(come on/go for it) and some uncommon names such as “王琳凯” and “润玉”)

2. Create word vectors for each song based on a well-trained Chinese Word2Vec model, containing more than 6 millions vectors trained by gensim Word2Vec package using 268GB texts including more than 8 million entries of Baidu-baike (Chinese version of Wikipedia) and more than 4 million online news. Hence, the dimension for each song is $8 \times 128 = 1024$.

3. To visualize the results of K-means clustering, I used t-SNE to reduce the dimension into two for each song.

[2]For this purpose, I used hot comments and normal comments together.

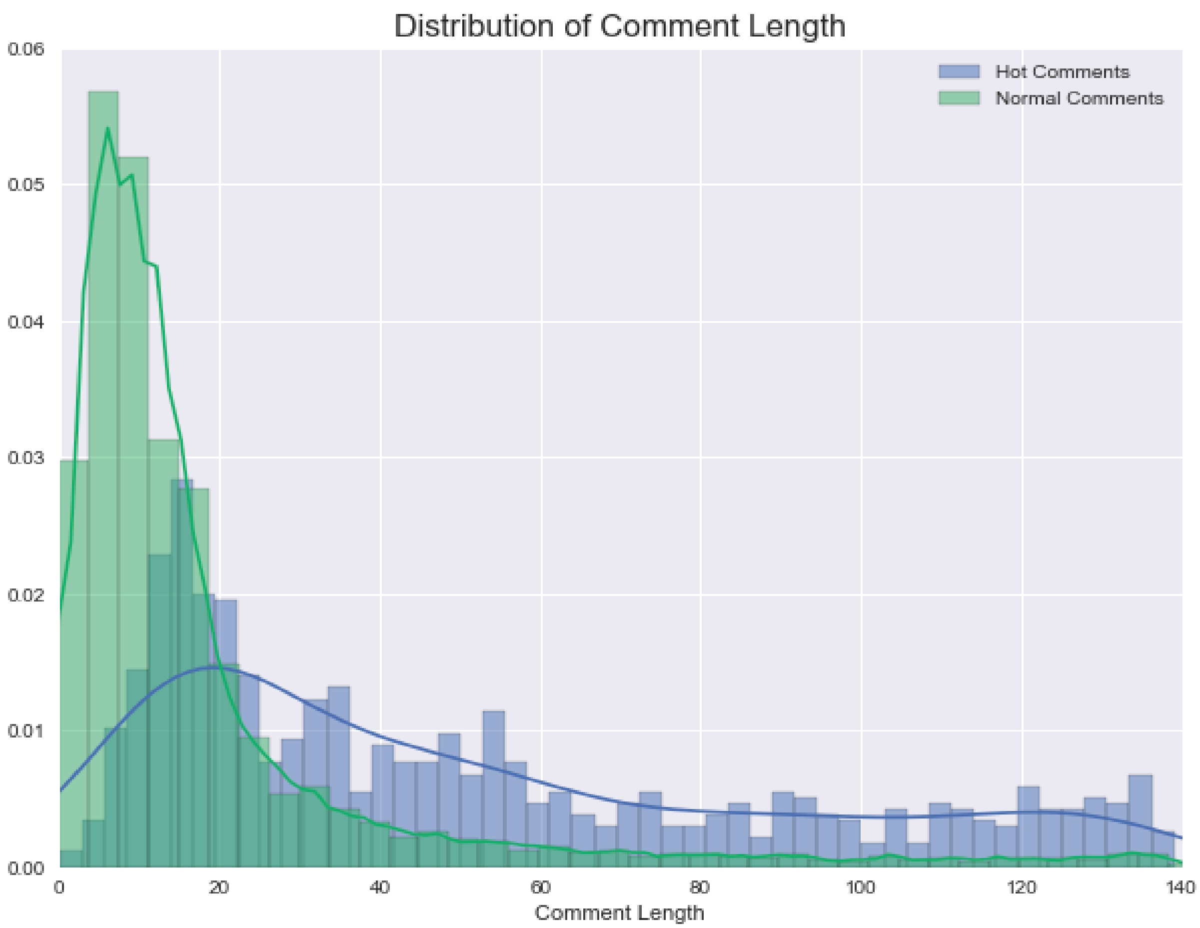
4. Implement K-means with K = 5.

Besides, I extracted 20 tags based on TF-IDF statistics for each cluster and computed average polarity and average subjectivity using TextBlob.

# 4 Results

## 4.1 Comparison between hot comments and normal comments

As shown in the plot of "Distribution of Comment Length", the distribution of comment length for normal comments (green one) is much more skewed to the right than that for hot comments (blue one). More than half of normal comments have length fewer than 20 characters while approximately



40 percent of hot comments are longer than 40 characters. This verifies the hypothesis that hot comments generally tend to be longer than normal comments.

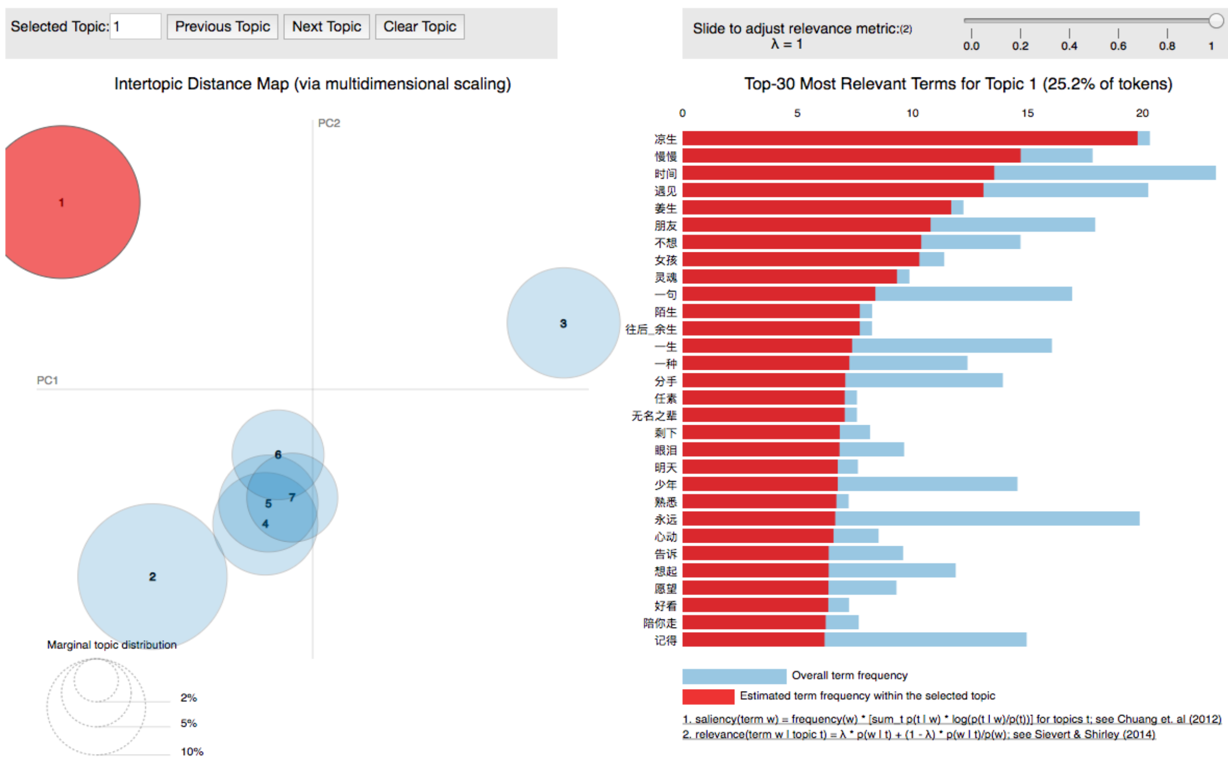
For 30 most common tokens displayed in following word clouds, tokens such as “喜欢”(like), “真的”(really), “首歌”(a song), and “希望”(hope) are most frequently used in both types of comments. This makes sense to me as “love” and “hope” are always popular topics among human beings. One notable finding is that words appearing in normal comments are more likely to convey

30 most frequently used tokens in hot comments


30 most frequently used tokens in normal comments

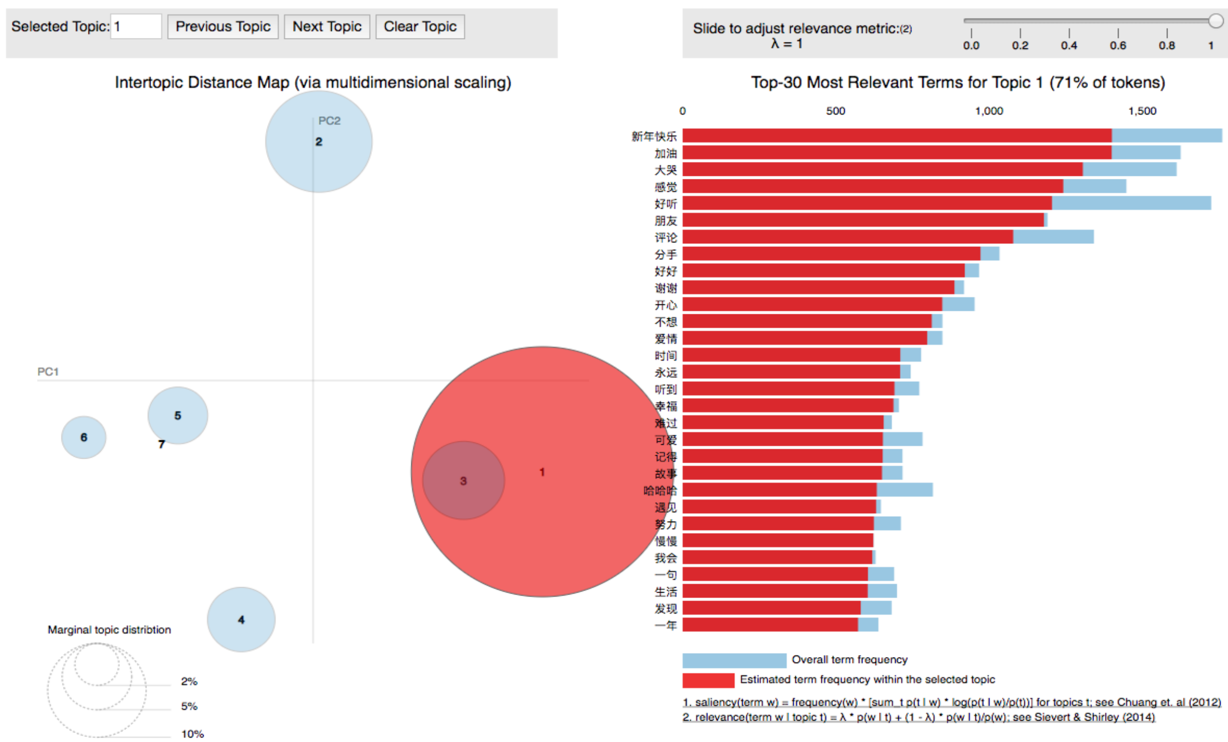

users’ instantaneous feelings while words in hot comments are more related to some reflections on life or personal stories. The most representative words in normal comments are “新年快乐”(happy new year), “哈哈哈”(laughing), and “加油”(come on). Since the date I scraped the comments coincided with Chinese New Year, “新年快乐”(happy new year) surprisingly becomes a popular token in normal comments. This provides evidence that in most of the time people write song comments based on their instantaneous mood about things or events around them. Nevertheless, hot comments contain more tokens such as “世界”(world), “朋友”(friend), and “青春”(youth) which are more likely to arouse emotional resonance as long as you have some similar personal experience.

For topic modeling, I built LDA models with seven topics for both types of comments respectively. Here, I only display the topic including most tokens among seven topics for each type of comment and the results of other six topics can be found in the jupyter notebook. As shown in the first visualization of Topic 1 (for hot comments), two of top-5 relevant tokens, “凉生” and “姜生” are the name of two main characters in a popular Chinese television series. Since the emotional




theme song of this Chinese television series is included in the dataset and this television series is adapted from a very popular romantic novel, it is perceivable that the hot comments of that song would frequently mention the romantic stories between these two main characters. Hence, it might be proper to conclude that the main topic of Topic 1 is about stories in that television series. For the rest of Top-30 most relevant tokens, it is hard to draw a concordant conclusion to a specific topic, though tokens such as “时间”(time) and “女孩”(girl) are frequently used when we talk about personal stories or life insights.

Compared to tokens in Topic 1 for hot comments, tokens in Topic 1 for normal comments are more diverse and even harder to end up with a concrete topic. As mentioned previously, most of

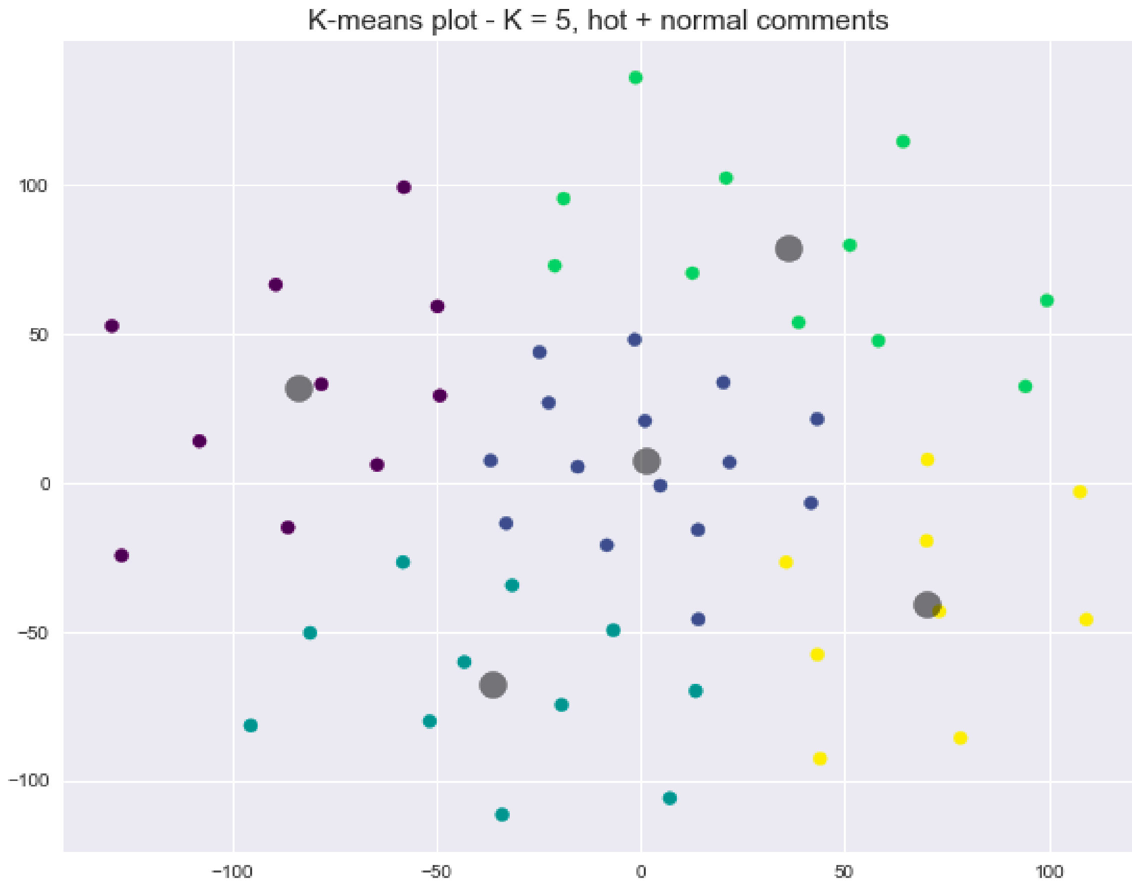



tokens are just to express people's instantaneous feelings without having deeper meanings. This probably accounts for why these comments remain as normal comments and cannot widely arouse emotional resonance.

### 4.2 Sentiment analysis using both types of comments

The plot of "K-means plot - K = 5, hot + normal comments" (displayed below) visualizes 5 groups of 56 songs following the methods described previously. Since dots are almost evenly distributed in the plot, there is no apparent clustering among these 56 songs. This implies that there is probably no strong similarity which could push some of them quite close to each other. Moreover, the relatively small number of songs likely makes it harder to achieve clustering phenomenon. Nevertheless, it is still worth trying to implement sentiment analysis for comments within the same group.

Table 1 displays average polarity and subjectivity of 20 extracted tokens, which are selected




based on TF-IDF statistics, for each cluster. In terms of average polarity, all clusters are slightly positive, though Clusters 1, 2, and 5 are bit more positive than Clusters 3 and 4. For average subjectivity, Cluster 1 is most subjective with the highest average subjectivity of 0.36 while Cluster 4 is least subjective with lowest average subjectivity of 0.108. It seems reasonable that all five

Table 1: Average Polarity and Subjectivity

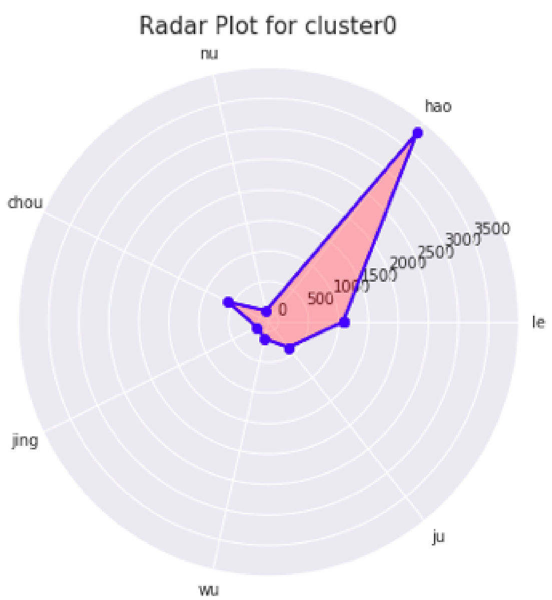
| Cluster | Average Polarity | Average Subjectivity |
|---|---|---|
| Cluster 1 | 0.125 | 0.360 |
| Cluster 2 | 0.155 | 0.248 |
| Cluster 3 | 0.091 | 0.245 |
| Cluster 4 | 0.069 | 0.108 |
| Cluster 5 | 0.147 | 0.272 |

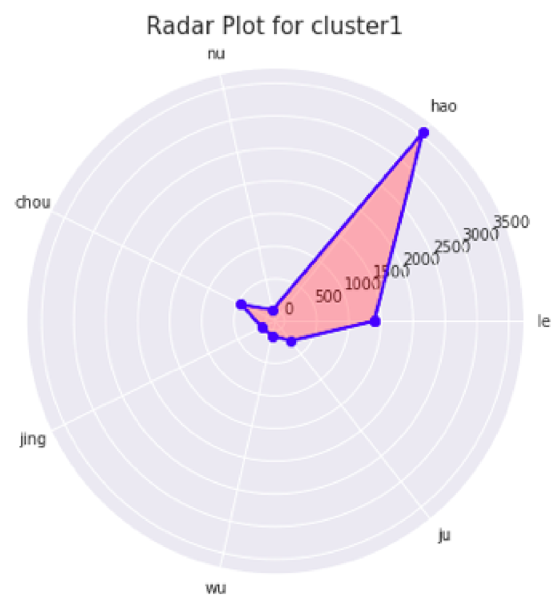
clusters are subjective because user comments heavily rely on personal emotions and experiences. Overall, Cluster 4, which is most neutral and least subjective, is quite distinct from other four clusters. The following picture shows all 20 selected tokens for Cluster 4. Since most of tokens

['好听', '单色', '润玉', '成全', '感觉', '原版', '评论', '谢谢', '歌手', '打歪', '分手', '作词', '记得', '哈哈哈', '不想', '流泪', '旭凤', '网易', '摩天大楼', '歌词']

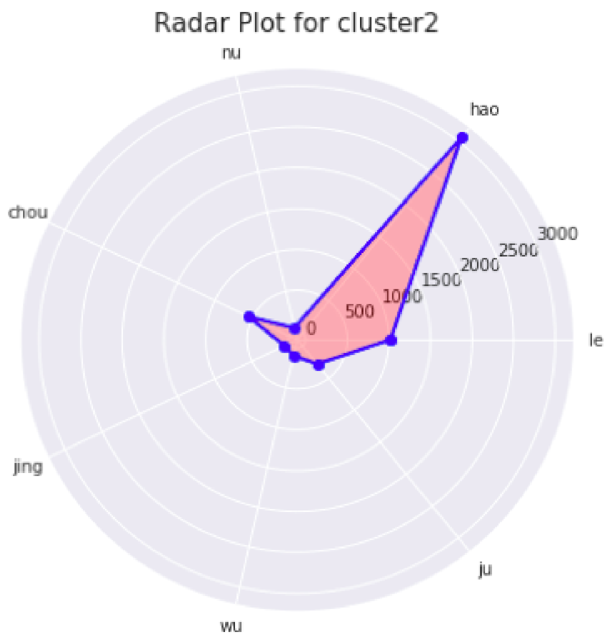
are nouns without any emotional feelings, it makes sense to achieve almost zero in average polarity and 0.1 in average subjectivity, implying that Cluster 4 is almost neutral and only a bit subjective.

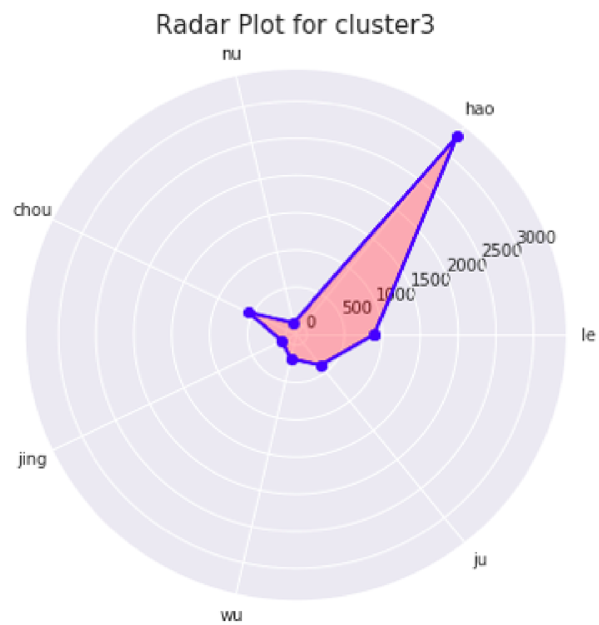
Further, I implemented another sentiment analysis by counting the total number of tokens in each of seven emotional groups (defined above) for a single cluster. As shown in following five radar plots, all clusters have most tokens in “好”(hao) emotional group including tokens of “赞扬”(praise), “喜爱”(like), “感动”(moving), “好听”(fair-sounding), “喜欢”(like), “爱”(love). “乐”(le) and “愁”(chou) are the second and third largest emotional groups for all clusters, while

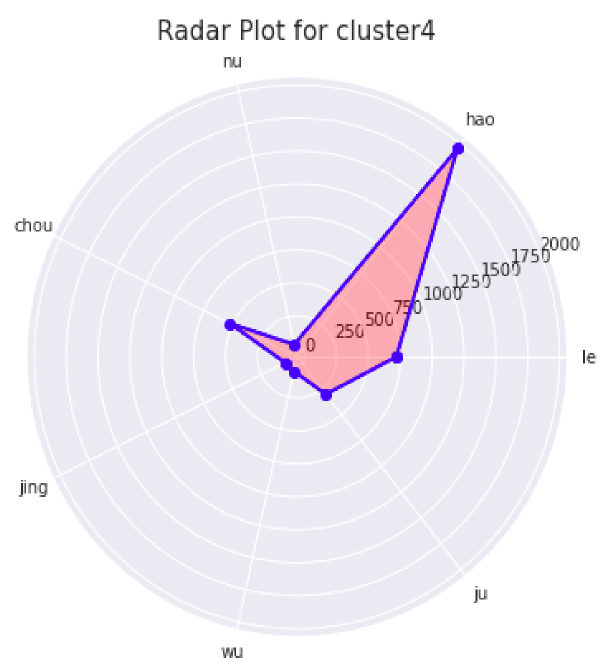



Cluster 2 (corresponding to cluster1 in Radar Plot) has most tokens (close to 1,500) in “乐”(le) across all five clusters. To verify whether these findings are consistent with emotional feelings embedded in the songs, I looked at all lyrics for songs in Cluster 2. Based on my own feelings, these songs are a mixture of happiness and sadness with a bit more positive feeling. Hence, to conduct more rigorous analysis, I would scrape lyrics as well and compare the results of sentiment analysis between lyrics and comments in the future.



# 5 Conclusions

In summary, hot comments are generally inclined to have more words than normal comments, and comments either describing concrete personal stories or expressing some insights on life are most likely to arouse public emotional resonance. Since the number of songs in this project is not large, the results of K-means clustering might not be accurate enough. This probably has a non-negligible impact on the quality of sentiment analysis for each cluster, implying that we might have better results as the number of songs increases.