Weekly Report -
— Reviews Selection Sentiment Analysis

Mike Ma

September 22

1 Brief Summary

This is a weekly report of movietrailer research from Mike. I focused on comments selection and sentiment analysis this week. In this report, I would introduce a method of comments selection I used.

2 Major problem of other papers

The only paper focusing on trailer prediction in google scholar is "Predicting movie trailer viewer's "like/dislike" via learned shot editing patterns", However, they directly use "like/dislike" in YouTube to do the prediction. However, there is a big problem is their paper. Think about this:

Movie 1 is released 3 days ago, whose trailer has 400 likes.
Movie 2 is released 3 years ago, whose trailer has 4000 likes.

However, we couldn't say movie 2 is better than movie 1, while in their paper they ignore this problem.

3 My method

To challenge their result, I don't use like/dislike only. Rather, I use the sentiments of comments in a fixed period before the movie released to do the prediction. -
Here I need the "movie release time", "trailer comments time(review time of each comment)" "sentimentssubjectivity value of comments", "likes of comments" as essences. -
Because movie information and trailer information are different tables, while I need to match "movie release time" and its respondent "trailer comments time". That's why text matching is indispensable for our research. -
Then I will introduce my method and why it solves the problem of their paper.

You would know why our time on comments selection and sentiment analysis are indispensable and valuable.

4 Method introduction

1. Get "movie release time" and "trailer comments time" from the title-matched table and comments table.

2. Transfer "trailer comments time"(like 3 days ago, 6 days ago).
   0-24 hours ago — 2019-07-30 (1 days)
   1 days ago — 2019-07-29 (2 days)
   1 weeks ago - 2019-07-18(13 days)
   1 months ago - 2019-06-02(59 days)
   1 years ago - 2017-08-02(729 days)

3. Compare "movie release time" and "trailer comments time" to current data-collecting time(2019-08-01) individually. Calculate the differences:
   Difference 1 = 2019-08-01 - "movie release time"
   Difference 2 = 2019-08-01 - "trailer comment time"
   -
4. Prerequisites for comments selection
   (1)Guarantee the review of trailer before the movie released.
   Difference 2 - Difference 1>0
   (2)Set the comments selection period as 6 months(180 days).
   0 <Difference 2 - Difference 1 <180
   (3)Delate trailer info which is released less than 6 months.
   Difference 2 >180
   -
5. Comments selection -
Using the value of "difference 2 - difference 1", we could normalize each Trailer review in a fixed period of time(6 months)
   -
6. Sentiments analysis -
Using NLP to do sentiment analysis for each selected review. I have already used blobtext from python and it success, but I don't know what the function of value"objectivity".
   -
Besides, there are lots of popular sentiments analysis method, I'm still learning them. In this part, we could choose popular API for sentiment analysis, like IBM or Google, in this case we don't need to train data. Or rather, we could use NLTK to label parts of comments as training set and build a model ourselves.
   -
If there is any suggestion of sentiment analysis, please let me know. I sincerely hope we could have a little discussion of sentiment analysis. NLP is cool.