Joe Standerfer, Spriha Gupta, Jasmine Kaur, Daniel Lesser

Marketing Analytics 95-832

Assignment 2: Movie Topic Modeling

3/29/2019

Gotham Group is preparing to launch Maze Runner, a new dystopian film based on James Dashner’s 2009 novel. Our group has been tasked with identifying the most opportune week of 2014 to launch the movie. Heuristic approaches have traditionally been used to choose release dates, taking advantage of industry knowledge regarding what time of the year is ideal for certain movie genres. In this case study, we have employed a topic modeling approach. Using movie reviews, our analysis identifies which movies are categorically similar to Maze Runner, what the optimal release weeks have been for the genre over time, and what specific week Gotham Group should release Maze Runner in 2014. Our analysis shows that releasing in week xx will both maximize expected revenue and minimize competition from similar movies.

Our group has used Latent Dirichlet Allocation (LDA) to generate topic probabilities for a subset of movies released between 2006 and 2014. Ten topics were generated using tags from movie reviews, with each movie having a probabilistic distribution across the topics. The probability distribution of movies across the topics was relatively uniform (Figure 1). However, if one were to take a deterministic approach and assign only the highest probability topic to each movie, the distribution is more skewed (Figure 2). In this case, we can see that Animation, Drama, Comedy, and Action are the most frequent topics.

LDA does not provide topic labels by default, so one required step in our approach was to assign qualitative labels to each topic based on the output of the model. To determine the appropriate labels, we reviewed the conditional probability of each tag showing up given each topic as well as the probability of each topic showing up given each movie (Figure 3 and 4). Based on the term and topic frequencies, we assigned the following labels to topics one through ten: Thriller Drama, Action, Animation, Fantasy, Sci-Fi, Dark Comedy, Dystopia, Drama, Comedy. Maze Runner falls within the Dystopia topic, which is often described with words such as dystopia, post-apocalyptic, zombies and horror. This topic includes movies such as I am Legend, The Road, The Book of Eli and Limitless.

We curated a list of weekly launch dates for 2014 and for each week, computed the average similarity score using a decaying function. Euclidean distance was used as the similarity measure. For each week, the distance of each movie was penalized depending on whether it was released two weeks before or one week after that week.

The penalty function used was: **Penalty = exp(-(weeksApart)^2)**

where weeksApart = Difference between a week and the movie release week

This version of a decaying function was chosen because it looks closest to how movie demand would behave.

The distance was then updated: **Penalized distance = Distance \* exp(-(weeksApart)^2)**

[The decay function is referenced in Figure 6]

\*\*\*\*\*\*\*\*\*\*\*\* Question 4 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Along with computing a similarity for each week, we also took several other industry factors into consideration. From the assignments prompt, it is clear that movie preference and theater attendance varies seasonally. One such example, is the abundance of horror films released in the Fall, particularly around Halloween. When deciding on a release date, we felt that it was important to take these industry cycles into consideration. In Figure 5 below, you can see the all-time, average opening weekend revenue for dystopian movies released in each week of the year. There are some holes in the data, but overall the trend can provide some indication of how the genre performs throughout the year rather than relying on the prompt’s broad generalizations of seasons (also shown in Figure 5).

As a result of our analysis, we feel the best week to release “The Maze Runner” would be week 37, Labor Day weekend. There were a few factors that lead to this decision. For one, week 37 had a very low similarity score compared to the rest of the year, meaning it wouldn’t face much competition from other movies released around that date. Second, historically that time of year has shown elevated opening weekend revenues for the genre. That, along with it being a holiday weekend, should lead to a high initial viewership. Third and last, there are no other “block buster’s” being released around that date. The only other movies on the docket are “No Good Deed” and “Dolphin Tale 2” which, frankly, shouldn’t pose much of a threat. The other top selections were week 50, 51, and 22. Weeks 50 and 51 were attractive because they showed very low similarity scores as well as high historical revenues. However, our hesitation with these weeks was that “The Maze Runner” does not appear to be a family-oriented film, which is what typical performs well in the holiday season. Also, there are no other large releases booked for those weeks but, given past years, we can expect a large amount of competition for those two weeks. Lastly, week 22 is a good candidate because it shows a low similarity score, falls right at the beginning of the summer, and does not have any large releases planned. As indicated in the prompt, summer release dates show consistently high revenues for the genre. Several other studios appear to already be capitalizing on this fact for the summer of 2014. Consequently, week 22 is one of the only weeks that does not have a block buster coming out.

Additionally, We used K-means on our data to make this decision. Our analysis converged to 10 clusters based on similarity between two movie vector representations. Euclidean distance was used to this purpose. The distance of “The Maze Runner” was calculated from the centroid of each other cluster. The farthest cluster from this group was chosen. The release dates of the movies in this segment were our potential release dates for the target movie. However, according to this analysis, our recommendation would be to release the movie during week 27 or the first week of July. This week has the least similarity with other movies and also contains a long weekend , owing to the 4th of July (Independence Day) holiday. This may act as a sales booster.

We also tried using 15 topics and found that the results for the topic 10 weeks are somewhat similar.

[Figure 8]

Week 52: This week has least similarity among the movies being released, however we don’t have the next weeks’ data available in this case.

Week 50: This week has a higher similarity score, yet the average revenue earned is greater than that earned in week 52. This week being closer to Christmas, indicates the possibility of other large releases.

Week 36: This week has a higher similarity score and fairly good past average revenue. Also contains a long weekend owing to Labour day.

Week 37: This week has an even higher similarity score and good enough revenue.

In this case, we would recommend releasing the movie in week 36 as it will have less possible number of clashes with similar movies and a fair revenue as well. Additionally, this week contains a long weekend and will possibly lead to increase in customers.

**Appendix**

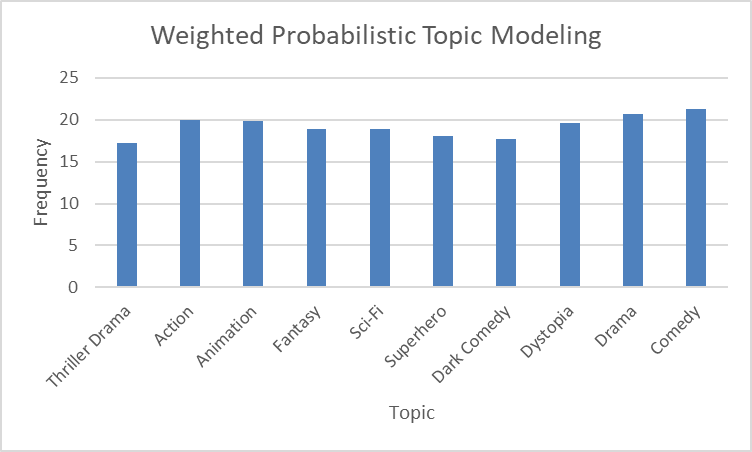


Figure 1

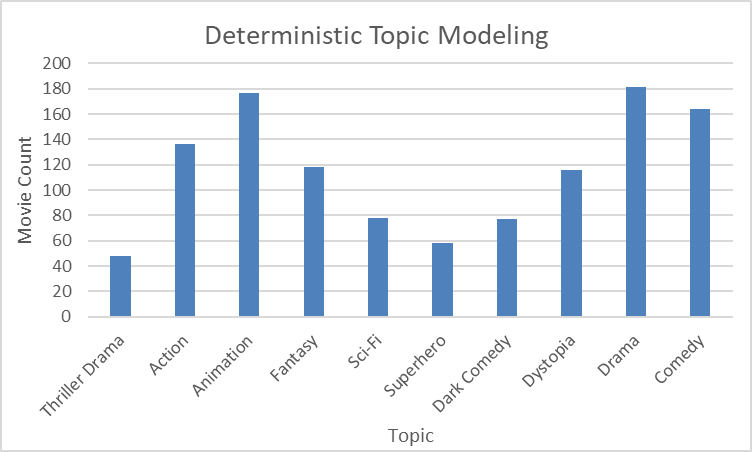


Figure 2

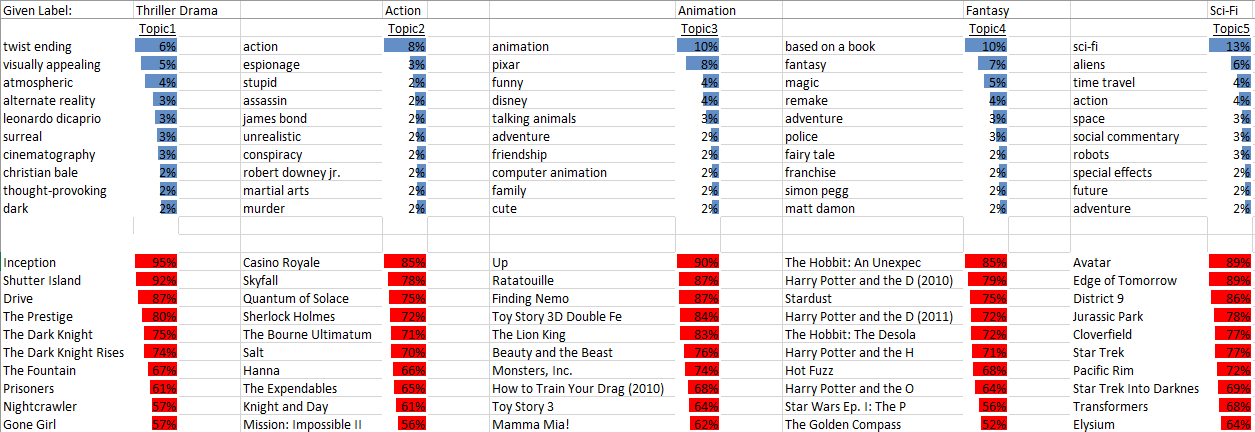


Figure 3

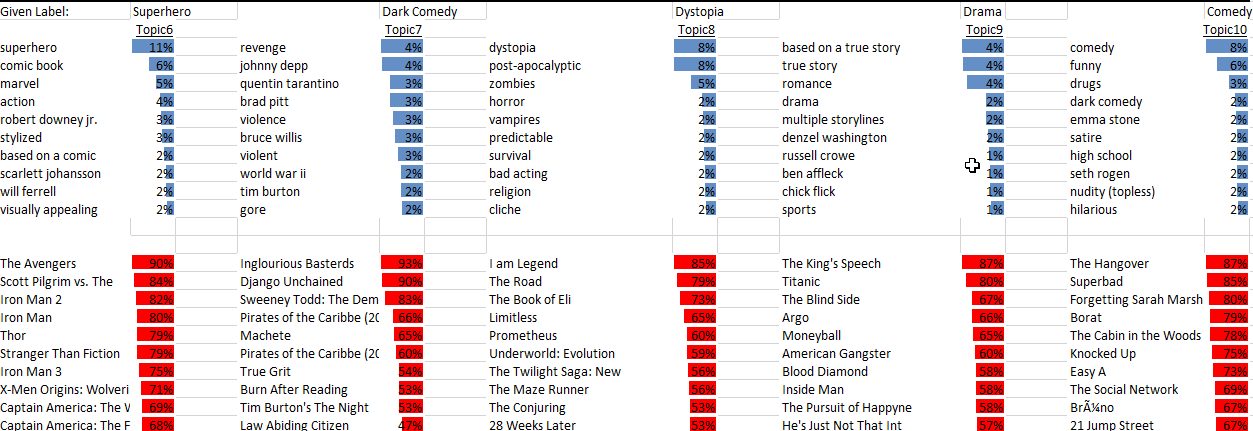


Figure 4

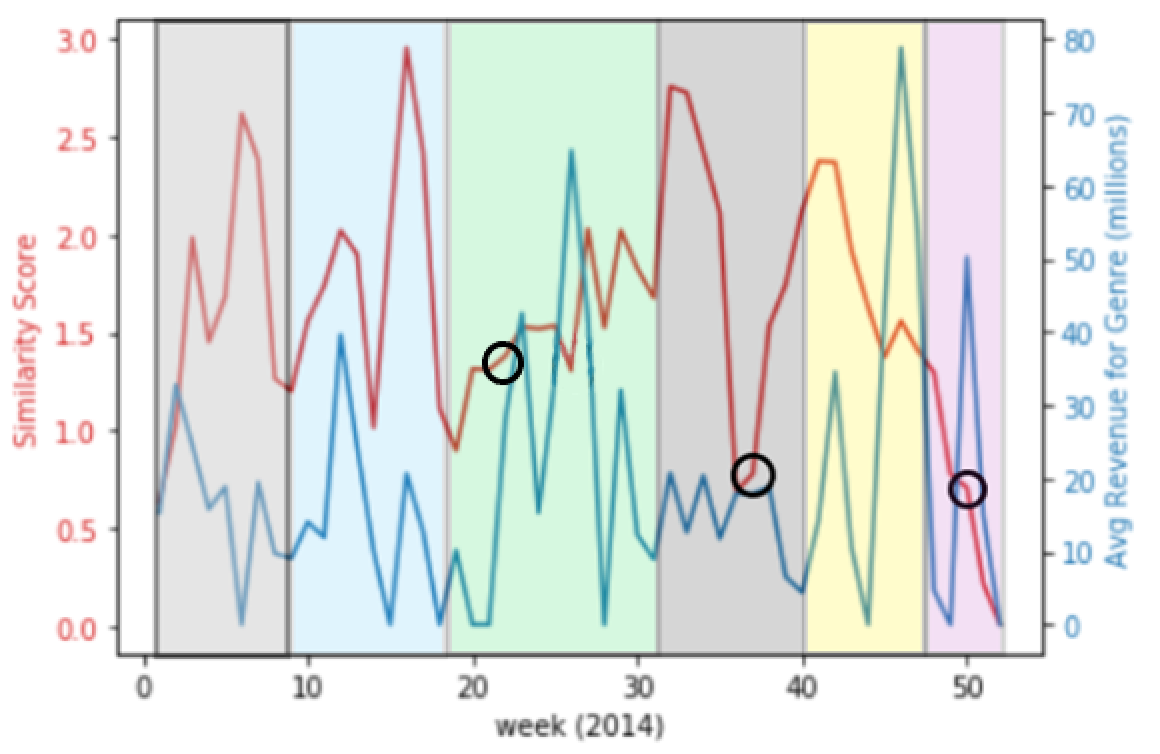


Figure 5

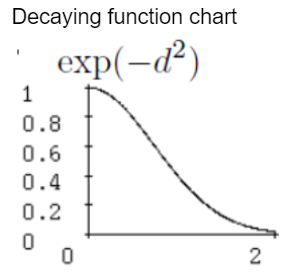


Figure 6

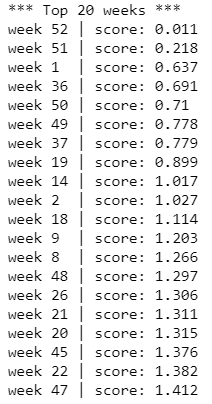


Figure 7

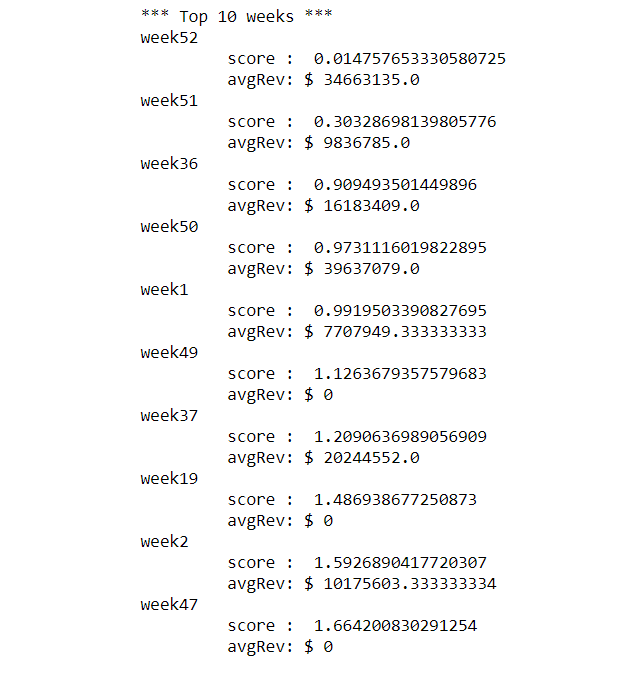


Figure 8