Predicting the Popularity of TED Talks Using Composite Measures of Popularity and Mixed Models

Marcos Cardenas Zelaya

March 29, 2020

Abstract

Objectives: Determine what predicts the popularity of TED talk videos using average views per day and a composite measure of popularity.

Design: Poisson and Linear Mixed-Model Regressions are used to determine the popularity of TED talk videos and whether the popularity varies by when the video was published and its theme.

Data: Web scraped data from the TED talks website. The data contains descriptions of videos created during 2006 to September 21st, 2017. The original data set contained 2550 observations.

Predictors and Response: Average views per day and a composite measure of popularity that encompass number of languages the video has been translated to, number of comments, a string dictionary of the ratings given to the talk (e.g., inspiring, fascinating, jaw dropping, etc.) and their frequency, and the number of related talks. The predictors in this analysis are a talks duration, number of speakers in the video, how old the video is, when the video was published, the "label" of the title and themes associated with the video, and the length of the title.

Results: We find that popularity varies by when it was published and its by its theme. Further, talks duration, talks film age in days since published, and the talks age group (when the talk was published) are the strongest predictors of popularity.

Conclusion: Although there is variation among themes, a talks duration, talks film age in days since published, and the talks age group (when the talk was published) are the strongest predictors of popularity. This suggests that the best way to create a popular video is by frequently publishing long videos, regardless of the content. Popularity seems to be a function of quantity not quality.

References

[1] Mi-Kyung Song, Feng-Chang Lin, Sandra E Ward, and Jason P Fine. Composite variables: when and how. *Nursing research*, 62(1):45, 2013.