A Causality-Guided Prediction of the TED Talk Ratings from the Speech-Transcripts using Neural Networks

Abstract

Automated prediction of public speaking performance enables novel systems for tutoring public speaking skills. We use the largest open repository---TED Talks---to predict the ratings provided by the online viewers. The dataset contains over 2200 talk transcripts and the associated meta information including over 5.5 million ratings from spontaneous visitors to the website. We carefully removed the bias present in the dataset (e.g., the speakers' reputations, popularity gained by publicity, etc.) by modeling the data generating process using a causal diagram. We use a word sequence based recurrent architecture and a dependency tree based recursive architecture as the neural networks for predicting the TED talk ratings. Our neural network models can predict the ratings with an average F-score of 0.77 which largely outperforms the competitive baseline method.

Introduction

While the demand for physical and manual labor is gradually declining, there is a growing need for a workforce with soft skills. Which soft skill do you think would be the most valuable in your daily life? According to an article in Forbes BIBREF0 , 70% of employed Americans agree that public speaking skills are critical to their success at work. Yet, it is one of the most dreaded acts. Many people rate the fear of public speaking even higher than the fear of death BIBREF1 . To alleviate the situation, several automated systems are now available that can quantify behavioral data for participants to reflect on BIBREF2 . Predicting the viewers' ratings from the speech transcripts would enable these systems to generate feedback on the potential audience behavior.

Predicting human behavior, however, is challenging due to its huge variability and the way the variables interact with each other. Running Randomized Control Trials (RCT) to decouple each variable is not always feasible and also expensive. It is possible to collect a large amount of observational data due to the advent of content sharing platforms such as YouTube, Massive Open Online Courses (MOOC), or ted.com. However, the uncontrolled variables in the observational dataset always keep a possibility of incorporating the effects of the “data bias” into the prediction model. Recently, the problems of using biased datasets are becoming apparent. BIBREF3 showed that the error rates in the commercial face-detectors for the dark-skinned females are 43 times higher than the light-skinned males due to the bias in the training dataset. The unfortunate incident of Google's photo app tagging African-American people as “Gorilla” BIBREF4 also highlights the severity of this issue.

We address the data bias issue as much as possible by carefully analyzing the relationships of different variables in the data generating process. We use a Causal Diagram BIBREF5 , BIBREF6 to analyze and remove the effects of the data bias (e.g., the speakers' reputations, popularity gained by publicity, etc.) in our prediction model. In order to make the prediction model less biased to the speakers' race and gender, we confine our analysis to the transcripts only. Besides, we normalize the ratings to remove the effects of the unwanted variables such as the speakers' reputations, publicity, contemporary hot topics, etc.

For our analysis, we curate an observational dataset of public speech transcripts and other meta-data collected from the ted.com website. This website contains a large collection of high-quality public speeches that are freely available to watch, share, rate, and comment on. Every day, numerous people watch and annotate their perceptions about the talks. Our dataset contains 2231 public speech transcripts and over 5 million ratings from the spontaneous viewers of the talks. The viewers annotate each talk by 14 different labels—Beautiful, Confusing, Courageous, Fascinating, Funny, Informative, Ingenious, Inspiring, Jaw-Dropping, Long-winded, Obnoxious, OK, Persuasive, and Unconvincing.

We use two neural network architectures in the prediction task. In the first architecture, we use LSTM BIBREF7 for a sequential input of the words within the sentences of the transcripts. In the second architecture, we use TreeLSTM BIBREF8 to represent the input sentences in the form of a dependency tree. Our experiments show that the dependency tree-based model can predict the TED talk ratings with slightly higher performance (average F-score 0.77) than the word sequence model (average F-score 0.76). To the best of our knowledge, this is the best performance in the literature on predicting the TED talk ratings. We compare the performances of these two models with a baseline of classical machine learning techniques using hand-engineered features. We find that the neural networks largely outperform the classical methods. We believe this gain in performance is achieved by the networks' ability to capture better the natural relationship of the words (as compared to the hand engineered feature selection approach in the baseline methods) and the correlations among different rating labels.

Background Research

In this section, we describe a few relevant prior arts on behavioral prediction.

Predicting Human Behavior

An example of human behavioral prediction research is to automatically grade essays, which has a long history BIBREF9 . Recently, the use of deep neural network based solutions BIBREF10 , BIBREF11 are becoming popular in this field. BIBREF12 proposed an adversarial approach for their task. BIBREF13 proposed a two-stage deep neural network based solution. Predicting helpfulness BIBREF14 , BIBREF15 , BIBREF16 , BIBREF17 in the online reviews is another example of predicting human behavior. BIBREF18 proposed a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) based framework to predict humor in the dialogues. Their method achieved an 8% improvement over a Conditional Random Field baseline. BIBREF19 analyzed the performance of phonological pun detection using various natural language processing techniques. In general, behavioral prediction encompasses numerous areas such as predicting outcomes in job interviews BIBREF20 , hirability BIBREF21 , presentation performance BIBREF22 , BIBREF23 , BIBREF24 etc. However, the practice of explicitly modeling the data generating process is relatively uncommon. In this paper, we expand the prior work by explicitly modeling the data generating process in order to remove the data bias.

Predicting the TED Talk Performance

There is a limited amount of work on predicting the TED talk ratings. In most cases, TED talk performances are analyzed through introspection BIBREF25 , BIBREF26 , BIBREF27 , BIBREF28 , BIBREF29 .

BIBREF30 analyzed the TED Talks for humor detection. BIBREF31 analyzed the transcripts of the TED talks to predict audience engagement in the form of applause. BIBREF32 predicted user interest (engaging vs. non-engaging) from high-level visual features (e.g., camera angles) and audience applause. BIBREF33 proposed a sentiment-aware nearest neighbor model for a multimedia recommendation over the TED talks. BIBREF34 predicted the TED talk ratings from the linguistic features of the transcripts. This work is most similar to ours. However, we are proposing a new prediction framework using the Neural Networks.

Dataset

The data for this study was gathered from the ted.com website on November 15, 2017. We removed the talks published six months before the crawling date to make sure each talk has enough ratings for a robust analysis. More specifically, we filtered any talk that—