Thesis Project Portfolio

NLP Comparison between Movie Reviews from Critics and Consumers (Technical Report)

Societal Influences of Manipulating Customers by Faking Reviews in the Movie Industry (STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science University of Virginia • Charlottesville, Virginia

> In Fulfillment of the Requirements for the Degree Bachelor of Science, School of Engineering

> > Yuelan Sheng

Spring, 2020 Department of Computer Science

Table of Contents

Sociotechnical Synthesis

NLP Comparison between Movie Reviews from Critics and Consumers

Societal Influences of Manipulating Customers by Faking Reviews in the Movie Industry

Prospectus

Sociotechnical Synthesis

Today's world is a time when Internet use is ubiquitous, allowing people to access data from all over the world about almost everything and accomplish tasks that require in-person contacts before. One of the most convenient benefits brought by the Internet is to shop online, allowing people to glance through various products without moving. This action leads to the discussion about the common use of accessing large amounts of data through the Internet: to read reviews of all kinds of products and compare them. By going online, not only information about regular commodities is available on shopping websites such as Amazon or eBay, reviews about entertainment like movies or music are also present. Thus, customers are able to learn about the products from a variety of perspectives. In the STS research paper focused on only online reviews about movies, discussing the societal influences online movie reviews have brought to the market and its potential ethical issues. Meanwhile, the technical project aims to uncover the patterns and characteristics of movie reviews from different groups of commenters and how these reviews have affected the box office by using Natural Language Processing (NLP) techniques.

In a time in which data resources are abundant and applicable for analytics usages, it is becoming more and more important to identify what role data should play in people's lives. Studies of how data can change people's lives have increased significantly during the last decade. The STS research points out the transition from accessing data through papers and word of mouth to online platforms for movie reviews. The technical project has used multiple engineering and statistical techniques to discover the underlying characteristics of languages used in film reviews.

It is intuitive to point out that the group of professional critics and the group of regular consumers each has their own unique styles and characteristics in writing reviews. However, uncovering the common characteristics of groups is difficult to accomplish by hands due to a large amount of data and inefficient human productivity. Using machine learning and NLP techniques, it is possible to analyze large amounts of data in a reasonable time. Thus, both the STS and technical parts aim to discover the roles of digital data for movie reviews in the current society.

Additionally, the technical report also concludes some possible ideas relating to marketing strategies that will be helpful for the movie industry in future use. With the results of how people comment about movies and what information is useful in promoting ticket sales. At the same time, the STS research has talked about how companies have manipulated customers' purchasing decisions by inappropriate use of technology to adjust the comments of their products. It is possible that some companies have tried similar actions in analyzing the reviews in order to aid their promotions. However, this may lead to some possible socio-technical and ethical issues, which are also discussed in the STS paper.

The STS and technical papers both discussed how data in movie reviews have impacted people's daily lives and research focus, and the STS paper further brings up a discussion of possible issues these technologies can bring to the society. These two analyses may be the first step in understanding what people should deal with the increasing amount of data and advancing technologies in the next decade.