Are Rom-Coms Dead?
Project Proposal

Naren Prakash[a], Fathima Shaikh[a], Caleb Williams[a]

[a]*University of California, Los Angeles, Department of Statistics and Data Science, Los Angeles, 90024*

Abstract

Romantic comedies, or "rom-coms", have been claimed to be worse now than before. Is this truly the case? This project will use 30 popular rom-com movie scripts from 1980 to 2020 in an attempt to answer the question: Have romantic comedies declined in quality over time (past 40 years)? The methods employed are lexical diversity, topic modeling, and sentiment analysis. The important features from these methods will help to develop a random forest model that predicts the decade in which each rom-com was published based on the original data and additional engineered features.

1. Introduction

Rom-coms have captivated audiences for decades, becoming a staple in popular cinema. However, in recent years, critics and casual viewers alike have argued that the quality of rom-coms has declined, referencing factors such as weaker storytelling and formulaic plots. Additionally, many claim that recent rom-coms are more predictable and cliche than those from years before. This project explores whether or not rom-coms have indeed declined in quality, using text mining techniques such as lexical diversity, sentiment extraction, correlation analysis, and topic modeling to examine transcripts of popular movies since 1980. Through this approach, this paper aims to determine whether the perceived decline in rom-com quality is supported by measurable changes in film content, or if it is simply a result of changing public perceptions such as the effect of nostalgia.

2. Objective

This paper seeks to answer whether romantic comedies have declined in quality over time, specifically the last 40 years. Quality will be defined using quantitative metrics.

3. Research Questions

- Are recent romantic comedies more correlated with each other than those of the past?
- Have romantic comedy scripts become less complex (in terms of vocabulary)?
- Can the era of a romantic comedy be easily predicted based on aspects of the script?

4. Methods

The methods used in this project aspire to quantify how rom-coms have changed over time through different metrics. One test of this involves lexical richness and linguistic complexity. This involves using Type-Token Ratio (TTR), which measures the diversity of a script's vocabulary, and readability scores, which evaluate the ease of comprehension for each script.

Another type of method used will be sentiment analysis, specifically using lexicons like AFINN, NRC, and Bing to analyze how each script's tone changes over time. This will help to see if romantic comedies have a more positive or more negative sentiment as well as note similarities in sentiments across time periods.

Also, this project will employ Topic Modeling through Latent Dirichlet Allocation (LDA) to find recurring topics in the scripts. LDA will contribute to exploring whether or not rom-com themes have become more formulaic in the past 40 years.

Additionally, correlation analysis will be employed to investigate similarities in script content among romantic comedies of different time periods.

Lastly, after important features have been engineered, a Random Forest model will help illustrate the similarity or dissimilarity of romantic comedies across time periods. In forming a classification model, the aim is to evaluate if scripts from different time periods are truly different in some aspect or if there is another factor not considered in this paper.