Screenplay Quality Assessment

Ming-Chang Chiu
University of Southern California
mingchac@usc.edu

Abstract

Movie production is a complex process which depends on key stakeholders at various stages. In particular, given a new movie screenplay, green-lighting its production depends on studio executives who must evaluate the potential for success of the movie based on factors such as market landscape, movie storyline, etc. So there is a need to build a tool to aid the process. In this work, we present a method to evaluate the quality of a screenplay based solely on linguistic cues. First, we choose a metric for quality, and then extract lexical, movie-related features, and then some domain-specific features for quality prediction. Our definition for "quality" comes from movies nominated at major film festivals and awards. We will perform experiments with the features using Support Vector Machines (SVMs), Tree-based algorithms and Neural Networks (NN). We believe our work can potentially shed light on the screenplay development and the foci of a quality movie script.

1 Introduction

The motion picture industry is a multi-billion dollar business worldwide (Lash and Zhao, 2016). Decisions in selecting movies to be produced are critical to the profitability and reputation of a movie studio (Vogel, 2014). However, determining factors for the financial and artistic success of a film is a non-trivial task, with a large subjective element. Particularly, the selection of the screenplay, which happens in the crucial initial development phase of movie production, has a large influence on the financial budget and quality of the final movie production. Thus, an objective and reliable tool to evaluate the quality of a screenplay is of vital importance to aid the green-lighting process.

Here, we want to highlight that domain knowledge in screenplays may bring numerous benefits in selecting the most relevant textual properties for the prediction of script quality (Gunter, 2018). Following ideas inspired by writing guidelines (Brown), we explore features along three specific aspects – linguistic diversity, emotion, and writing style. Such interpretable features may also offer opportunities for an enhanced understanding of the essential elements in high-quality movie scripts. Besides, we hope to automatically extract these important characteristics to help obtain a comprehensive assessment of screenplays in a more efficient way, i.e., to support the green-lighting process in the industry.

```
STERLING => I hate to drop in unexpectedly.
STERLING => Time for a slow boat to China.
STERLING => Oh bloody hell!
SAM TANNICK => Come.
GIBBONS => Evening, Sam.
```

Figure 1: Script excerpt from "xXx," an American action movie released in 2002.

Unlike existing literature that primarily investigates using screenplays to predict the box office performance (Eliashberg et al., 2006), our work focuses on measuring quality as defined as whether or not a movie will be nominated for a film festival. To the best of our knowledge, this is the first approach that aims at measuring screenplay quality, and there is a need for approaches which could systematically evaluate the quality of the screenplay.

2 Dataset

In this work, we plan to use the movie screenplays collected by (Ramakrishna et al., 2017). It contains 945 Hollywood movies, from 12 different genres (1920 – 2016). We choose this corpus because it is larger than others (Banchs, 2012) in size and also provides actors utterances parsed from the scripts, as shown in Figure 1, which makes the preprocessing relatively light-weight.

Another dataset that we will be using is

```xml
<script title="10 Things I Hate About You">
    <scene count="0">
        <stageDirection count="0">PADUA HIGH SCHOOL – DAY</stageDirection>
        <description count="0">
            <text>
                <sentence id="0" type="normal">
                    <wordList>
                        <word id="t_0" lemma="welcome" ne="O" pos="VB" stem="welcom">Welcome</word>
                        <word id="t_1" lemma="to" ne="O" pos="TO" stem="to">to</word>
                        <word id="t_2" lemma="Padua" ne="ORGANIZATION" pos="NNP" stem="padua">Padua</word>
                        <word id="t_3" lemma="High" ne="ORGANIZATION" pos="NNP" stem="high">High</word>
                        <word id="t_4" lemma="School" ne="ORGANIZATION" pos="NNP" stem="school">School</word>
                        <word id="t_5" lemma="," ne="O" pos="," stem=",">,</word>
                        <word id="t_6" lemma="," ne="O" pos="," stem=",">,</word>
                        <word id="t_7" lemma="you" ne="O" pos="PRP$" stem="your">your</word>
                        <word id="t_8" lemma="typical" ne="O" pos="JJ" stem="typic">typical</word>
                        <word id="t_9" lemma="urban" ne="O" pos="JJ" stem="urban">urban</word>
                        <word id="t_10" lemma="–" ne="O" pos=":" stem="–">–</word>
                        <word id="t_11" lemma="suburban" ne="O" pos="JJ" stem="suburban">suburban</word>
                        <word id="t_12" lemma="high" ne="O" pos="JJ" stem="high">high</word>
                        <word id="t_13" lemma="school" ne="O" pos="NN" stem="school">school</word>
                        <word id="t_14" lemma="in" ne="O" pos="IN" stem="in">in</word>
                        <word id="t_15" lemma="Portland" ne="LOCATION" pos="NNP" stem="portland">Portland</word>
                        <word id="t_16" lemma="," ne="O" pos="," stem=",">,</word>
                        <word id="t_17" lemma="Oregon" ne="LOCATION" pos="NNP" stem="oregon">Oregon</word>
                        <word id="t_18" lemma="." ne="O" pos="." stem=".">.</word>
                    </wordList>
```


Figure 2: Script excerpt from “10 things I hate about you,” an American action movie released in 1999.

ScriptBase-J (Gorinski and Lapata, 2018) which contains 906 movies after preprocessing, ranging from 1920 – 2017. An example can be referenced in Figure 2.

2.1 Screenplay Quality Metric

On top of the parsed scripts, we label the screenplays nominated for film festivals as quality ”ground truth”. The venues where we collect from are the most well-known and professional prizes in honoring top quality screenplays over the years, which include “Writers Guild of America Award” (WGA), “Academy Awards” (Oscar), “Golden Globe Awards” (GGA), and “British Academy of Film and Television Arts Awards” (BAFTA). In this work, we define movies nominated for either original screenplay or adapted screenplay nominees as “high (award) quality” and the rest as “low (award) quality.”

3 Features

Inspired by (Eliashberg et al., 2014), we construct our domain specific feature sets by starting from collecting text statistics, metadata and content variables. In addition, we explore features that cover diversity-related, emotion-related, and writing style-related angle.

Diversity. In this category, we will use two types of features, type-token ratio and scene-level autoencoder clustering. (Kao and Jurafsky, 2012) justified that type-token ratio is helpful for decerning creative poetry works from mediocre ones, and we suppose the effect would be the same for screenplays. The concept of scene-level autoencoder clustering is similar to topic models – we feed all texts on a scene level into an autoencoder model and perform clustering on the intermediate representations, and finally histogram the clustering result for each movie. We belive the histogram of clusters can represent components of movie scripts.

Emotion. In affective computing domain, Valence-Arousal-Dominance (VAD) and Basic Emotions are two types of measures for emotion representation. We plan to use (Mohammad, 2018a) and (Mohammad, 2018b) lexicons for measuring VAD and and word affect intensity respectively.

Writing Style. Our definition for writing style is in measuring consistency in characters’ linguistic portrays and again in their emotion changes from act to act. We belive that main characters’ language use and sentiment level would change accordingly in the progression of the story. There is a long history in statistical modeling for behavioral activity curve. We will use a certain type (Dawadi et al., 2016) for modeling activity curve and use (Fast et al., 2016) for a more comprehensive emotional consistency modeling.

4 Models

In this work, we will implement traditional machine learning classifiers like SVM, XGBoost, and MLP for predictionto see how each feature we propose perform..

On top of traditional algorithms, we plan to utilize the recent advancements of NN architecture for document classification. First, we are going to implement several widely used architectures such as Hierarchical Attention Network (HAN) (Yang et al., 2016), Bow-CNN (Johnson and Zhang, 2014), XML-CNN (Liu et al., 2017), and Bi-LSTM (Adhikari et al., 2019). Second, we plan to modify the NN architectures based on screenplay writing theory (Weiland) to better capture the special structure of best practices in screenplay writing. Finally, aside from plain NN models, we will be incorporating our domain-specific features into the NN models and get a better look into the topic we proposed.

References

Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. 2019. Rethinking complex neural network architectures for document classification. In *NAACL-HLT*.

BAFTA. British academy of film and television arts award.

R.E. Banchs. 2012. Movie-dic: a movie dialogue corpus for research and development. In *In Proceeding of the 50th Annual Meeting of the ACL*.

Michael Ray Brown. Screenplay structure checklist - story sense.

Prafulla N. Dawadi, Diane J. Cook, and Maureen Schmitter-Edgecombe. 2016. Modeling patterns of activities using activity curves. *Pervasive Mob. Comput.*, 28(C):51–68.

J. Eliashberg, S. K. Hui, and Z. John Zhang. 2014. Assessing box office performance using movie scripts: A kernel-based approach. *IEEE Transactions on Knowledge and Data Engineering*, 26(11):2639–2648.

Jehoshua Eliashberg, Sam K. Hui, and Z. John Zhang. 2006. 1 from storyline to box office : A new approach for greenlighting movie scripts.

Ethan Fast, Bin Bin Chen, and Michael S. Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *CHI*.

GGA. Golden globe award.

Philip John Gorinski and Mirella Lapata. 2018. What's this movie about? a joint neural network architecture for movie content analysis. In *NAACL-HLT*.

Barrie Gunter. 2018. *How Important Is the Screenplay to Box Office Performance?*, pages 145–160. Springer International Publishing, Cham.

Rie Johnson and Tong Zhang. 2014. Effective use of word order for text categorization with convolutional neural networks. In *HLT-NAACL*.

Justine Kao and Daniel Jurafsky. 2012. A computational analysis of style, affect, and imagery in contemporary poetry. In *CLfL@NAACL-HLT*.

Michael T. Lash and Kang Zhao. 2016. Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, 33(3):874–903.

Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. 2017. Deep learning for extreme multi-label text classification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 115–124, New York, NY, USA. ACM.

Saif M. Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Saif M. Mohammad. 2018b. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.

Oscar. Oscar, academy award.

Anil Ramakrishna, Victor R. Martinez, Nikos Malandrakis, Karan Singla, and Shrikanth Narayanan. 2017. Linguistic analysis of differences in portrayal of movie characters. In *ACL*.

Harold L. Vogel. 2014. *Entertainment Industry Economics: A Guide for Financial Analysis*, 9 edition. Cambridge University Press.

K.M. Weiland. Story structure qa: 6 outstanding questions about structure.

WGA. Writers guild of america award.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J. Smola, and Eduard H. Hovy. 2016. Hierarchical attention networks for document classification. In *HLT-NAACL*.