Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/48/20241057

A new branch of fake review detection research -- A review of fake review detection in the Chinese film industry in the post-epidemic era

Chen Zhaowei

North China Electric Power University, Baoding, 071003, China

williamchen119@163.com

Abstract. In the post-pandemic era, Chinese moviegoers increasingly rely on online movie reviews, but fake reviews by spreaders can mislead moviegoers to make wrong decisions. Fake review detection has been developed to a certain extent in China. However, there is a lack of application research in the film industry. This paper summarizes some of the more advanced fake review detection methods in China in the post-epidemic era from the perspectives of review text detection and reviewer detection, introduces their indicators, feature selection methods, and training methods, and further discusses the specific steps of these methods in the detection of fake movie reviews combined with the characteristics of fake movie reviews. The research of this paper can bring guidance for the future detection of fake movie reviews, and provide a decision-making basis for consumers and investors.

Keywords: Fake Reviews, Fake Review Detection, Movie Reviews, Text Feature, Behavior Features.

1. Introduction

The Chinese film market has ushered in a recovery after the COVID-19 pandemic [1], and the leading platform for film promotion has also switched from offline to online [2]. People usually look at online movie reviews before deciding whether to watch a movie. However, an increasing number of movie companies employ spreaders to fake large amounts of positive or negative reviews online in order to boost their films' box office or hurt the box office of rival films [3]. fake review detection is a method to identify fake reviews from a large number of reviews. The existing research results include text detection, reviewer detection, and comprehensive detection [4]. However, the application of this technology in the Chinese film field is lacking. This paper summarizes the more advanced fake review detection technologies in China and proposes application-specific fake review detection methods for different movie review platforms. Furthermore, this paper envisages the steps of fake review detection when applied to movie reviews, which provides guidance for researching fake movie reviews.

2. An overview of fake review research

Fake reviews are reviews that don't match the actual user experience [4]. Reviews are an important reference for consumers to purchase products, through which they can learn other users' understanding of product features and user experience. Movie movie sales are also largely influenced by reviews. Zhao

DOI: 10.54254/2755-2721/48/20241057

[5] collected the high-box office Chinese film reviews and box office data from 2010 to 2020, calculated the correlation between reviews and box office, and concluded that film reviews would affect the box office. Taking the Chinese film Detective Chinatown 3 as an example, he gave some evidence that this film suffered from malicious reviews. Before watching a movie, it is difficult for users to identify fake reviews only by intuitive feeling. Even if a single review can be judged to be true or false, it is also difficult to see from a large number of data what proportion of reviews are untrue. The two directions of fake review detection are text detection and reviewer behavior detection [4], and the main tools are text analysis, data mining, and statistics. Fake review detection was first proposed by Jindal et al. [6], but only at the text level of reviews. In 2010, Lim [7] proposed a new direction for study reviewers, but most of the research still focused on text research. In 2011, Ott3 et al. proposed a new method to extract lexical features and psychological features, which provided a new research idea. Mukherjee [8] pointed out the group characteristics of fake reviewers in 2013 and initiated research on fake reviewer group detection. While the theory is gradually enriched, a large number of fake review detection studies have appeared in China since 2013. However, there is still a lack of relevant literature on the application and research of this method in the Chinese film field. Therefore, some of the existing methods can be applied to the film field to broaden the application of fake review detection and help consumers better judge the actual word-of-mouth of the movie.

3. The possible application of fake review detection in the Chinese film field

Fake review detection refers to the use of data mining technology to identify the text, the reviewer, or the combination of the two for comprehensive detection to determine fake reviews. In the current film field, the main channels of publicity are social media, short video platforms, and movie apps [2]. The forms of reviews on these platforms have different characteristics. Therefore, different fake review detection methods can be applied to data from different platforms. The general process of fake review detection is data acquisition, data processing, dimension selection, modeling, and effect evaluation. [4] The reviews that have the most significant impact on the box office of the movie are the reviews in the early stage. Different from the reviews of physical goods, the water army of the movie starts to warm up the box office before the movie is released and is mainly active in the early stage of the movie release period. Therefore, the sampling time should be extended to the movie before the release, and the early reviews should be the focus of the study.

3.1. Detection of text fake reviews

Text fake review detection mainly screens fake reviews through text lexical features and emotional features.

In movie reviews, lexical features are mainly the collocation between words. For example, when a movie is deliberately badly reviewed, there may be a collocation of the movie character name plus a certain character plot plus offensive words. The sentiment feature refers to the positive and negative of the words used in the movie. For example, praise represents positive reviews, and disparagement represents negative reviews. Different words have different emotional intensities, which can be defined as different values. From the perspective of the purpose of improving the box office or hitting the box office, fake reviews generally have a strong emotional tendency, that is, praise or disparage. Fake reviews can also be found through the detection of emotional features.

3.1.1. Detection based on lexical features. Lexical features are continuous words and sentences with similar word collocations can be grouped into one group, which plays a great role in identifying fake reviews. Lexical feature detection has a high requirement for collocation detection. n-grams is an early method for storing words, which is suitable for storing low-dimensional words, but not for complex languages. However, the Word2vec model used in the Amazon dataset can only recognize single words, and it is difficult to identify groups of words that express specific meanings together. In order to solve this problem, Zhang [9] proposed a Bigram-Word2vec model based on the Word2vec model. This model uses word pairs to create a dictionary, which has a better effect on recognizing continuous text. The

DOI: 10.54254/2755-2721/48/20241057

results are better than the Word2vec model under different sentence lengths. In terms of the choice of integration method, they proposed the "binary weighted hard voting method" to deal with the problem of an equal number of votes of classifiers, and the accuracy of the new model reached 81.27%, which had better integration effect. When applied to movie reviews, this method is suitable for movie reviews on movie apps, because the reviews on movie apps have the characteristics of long text, so there may be more phrases, so it is easier to extract consecutive word pairs. We can first use the Bigram model to process a large amount of text, cut each sentence into a number of word pairs, then count and record the number of word pairs, select high-frequency word pairs, build a dictionary, and then look for word pairs in movie reviews, replace the word pairs in movie reviews with the values of the same word pairs in the dictionary, and then use Word2vec model to train. Then, the binary classification hard voting method is used for integration.

The detection of sentiment features is mainly realized by extracting and analyzing words, and the positive and negative directions of words constitute the positive and negative directions of sentences. The comments with subjective tendency are the comments with emotional features. Fake movie reviews usually have absolute subjective tendencies, such as extreme dissatisfaction or satisfaction. SVM finds the optimal margin to obtain language classification, and LSTM uses the recurrent network to store longer word models. Cao [10] et al. added the Transformer model to the SVM model to extract emotional factors and verified it with the Amazon data set. After adding emotional factors, the effectiveness reached 66.96%, which was higher than the SVM and LSTM models and made progress. When applied to the field of movie reviews, this method is suitable for the movie reviews of short video platforms, because the reviews of short video platforms have a word limit and are generally short. Only using strong emotions within a short word number is more likely to achieve the effect of false propaganda. The NLTK library can be used to label the reviews, and the adjectives, adverbs, and superlations with strong emotions can be screened out. Then the pyltp library is used to quantify, and the Transformer model is used to extract the feature vector of the review.

3.1.2. Detection based on lexical features and sentiment features. The detection of lexical and emotional features is not always done in isolation. Fake movie reviews are likely to be characterized by strong emotions and high-frequency words at the same time. Zhang [11] used the Roberta sentiment pre-training model to train lexical features and sentiment features and then integrated the training to achieve a cross-domain fake review detection accuracy of 88.00%. However, this method also has a defect, that is, in the collected data, there are very few negative data, which is the imbalance of positive and negative samples. If this method is used to identify fake movie reviews, it is likely that the vast majority of the collected samples are real reviews, which cannot achieve the purpose of effectively identifying fake reviews. Recognizing the imbalance of positive and negative samples, Dushan [12] et al. adopted manual labeling and voting methods to process samples and proposed a SMOTE-RE model with multiple dimensions. SMOTE oversampling can solve the over-fitting problem in random oversampling, and this model can make the number of positive and negative comments balanced by reconstruction. They selected nine dimensions of reviewer features and comment text features, such as anonymity, additionality, response, usefulness, time interval, emotion, score deviation, and text length, and used the optimized decision tree algorithm of random forest to train. Finally, the prediction accuracy reached 87.71%, and the positive and negative comments were balanced. This method can be applied to the comment detection of social media commenters with a certain number of followers because the number of this type of comment is small, which is conducive to manual annotation. Firstly, the method of voting and marking is used to obtain the data classification, and the sentiment polarity, score deviation, and text length are used as multi-dimensional features to establish the feature vector of the comment. Then the SMOTE algorithm is used to balance the positive and negative samples.

3.2. Fake review detection based on a combination of text and reviewers.

The detection of fake reviewers is also a way of fake review detection. Fake movie reviewers may have some unusual characteristics compared with other users. For example, a large proportion of them have

DOI: 10.54254/2755-2721/48/20241057

dedicated accounts, may have imperfect account information, or are concentrated in certain regions. In addition, their behavior is different from other users, they usually post a large number of fake reviews at a time, and they are not very active when they are not on task, in terms of personal information and behavior. Reviewer behavior research is used to identify fake reviews based on account information or abnormal behavior of these users. The active period of the spreader is matched with the release period of movies.

Accordingly, it can be analyzed that users who are significantly more active during the release period of movies than at other times are more likely to be spreaders. In addition, in the review control behavior of fans of stars, fans are highly correlated because they have similar follow lists and retweet blog posts, which is the similarity of information between reviewers.

Fake reviews may have both textual and reviewer behavior characteristics, so text detection and reviewer detection can be combined. Yang et al [13] pointed out that most of the previous fake review detection equated fake review detection with binary classification, without considering reviewer behavior. Therefore, they proposed to use dual convolutional neural networks to fuse text information and reviewer behavior information, and the accuracy was 3.84% higher than that of the binary classification method. Zhang [14] et al pointed out that the ability of various convolutional neural network models to recognize the behavior of reviewers is not good enough, so they proposed a hierarchical heterogeneous graph attention network for fake review detection. This method can identify the text structure hierarchy, and can also identify the features of text information and behavior information from heterogeneous graphs, and obtain the weights of different features. After fitting and training the help data, the F1 value of this algorithm can reach 94.20%, which is 21.6% higher than that of convolutional neural network detection. However, this method also does not consider the imbalance of the sample. In the sample, the proportion of fake reviews is very small, so the error rate in the large category detection is very low, but this is misleading. These studies usually use overfitting or simple sampling, which is not very appropriate. In order to solve this problem, Tao [15] et al. proposed to use EasyEnsemble to process samples, take a subset from the large class, and merge the small samples to train the XGBoost base classifier to form the most reasonable classifier. This method can avoid data loss and overfitting, and it is faster. The results show that the AUC value is 3.32% higher than that of the simple under-sampling method. Using this method to identify fake movie reviews is suitable for social media, because the user information on social media is the easiest to obtain, and social media users usually need to fill in more detailed user information, which is different from other platforms. Because the comments of movie spreaders may self-replicate, they can be used as a polarity. Then, the emotional polarity of the review, the historical text-similarity of the reviewer, whether or not to comment for a long time, and so on are used as features. The under-sampling method is used to obtain a subset, and then the large sample subset and the small sample subset are merged. With XGBoost ensemble, more perturbations can be considered and the training effect is better.

Fake reviewers do not always act alone, and sometimes a group of fake reviewers will cooperate with each other to fake real reviews. Such groups can confuse the public and guide social opinion, which is very dangerous. In the film industry, this kind of group is called movie spreader. In this group, superiors are responsible for publishing the task, and a large number of subordinates are responsible for executing it. They will release a large number of fake reviews at a certain time point to distort the public perception, such as at the beginning of the film's release. They also fake realistic interactive comments in the form of dialogues. Smart spreaders even avoid Posting comments in groups to avoid causing data anomalies on servers. However, these groups also have certain characteristics, such as the same time window, similar content to each other, and different ratings from other users. Combining text detection and reviewer group detection can also improve the fitting effect. The traditional methods of mining fake reviewer groups are based on the structure, behavior, and network structure of the group. In order to identify fake reviewer groups more efficiently, Ye [16] et al. proposed a spectral clustering group detection algorithm based on the reviewer similarity matrix. Through their own interactive behavior, the reviewers are divided into several categories. Experimental results show that the classification effect of this method is better than the K-means and hierarchical clustering methods. This method is suitable for

DOI: 10.54254/2755-2721/48/20241057

movie apps and social media because most of the user data of these two platforms are text categories, which are easier to quantify. When using this method, we need to measure the degree of correlation between movie reviewers in several aspects. The first is how often two reviewers comment on the same movie, the similarity of their ratings for the same movie, and their mutual reviews, praise, retweet times, etc. Then, we need to measure the characteristics of users, such as the proportion of strong emotional reviews in the total reviews. It uses a weighted graph composed of reviewers, movies, and reviews, and uses spectral clustering for group detection.

4. Conclusion

The purpose of this review is to extend fake review detection to the Chinese film field. With the recovery of the Chinese film market after the COVID-19 pandemic, there are more and more fake reviews. However, the application of fake review research methods in the film industry is still very little discussed. According to the summary of the methods developed for fake review detection in China, it can be concluded that the possibility and steps of the application of review text detection, reviewer detection, and comprehensive detection methods in film reviews can be obtained. Different detection methods can be applied to different movie review platforms. However, there are more and more pictures and videos in the current movie reviews, and this paper does not extend the detection method for the reviews containing pictures and videos because there are still few ways to recognize pictures and videos. According to this conclusion, researchers should focus on introducing algorithms to identify pictures and videos in reviews, adopt appropriate quantitative methods, and consider the content of pictures and videos as a dimension of review detection, which is helpful to further improve the effectiveness of fake review detection in future research. This paper successfully summarizes several advanced fake review detection methods in China and discusses the method applied in fake movie reviews, which provides guidance for the application of fake review detection in the film field, helps movie-viewers to analyze the authenticity of movie reviews to make decisions, avoid economic losses, and helps platforms to strengthen the censorship mechanism and reduce fake reviews. In order to avoid misleading the public with fake reviews.

References

[1] Li A and Yang Z 2021 The box office of the summer season exceeded 10 billion yuan, and domestic films achieved impressive results *China. Secur. J.* **2021** 5
[2] Ma L 2022 Analysis of film marketing strategies in the context of new media: Taking "Ex 3: Goodbye ex" As an example *West. China. Broadcast. TV.* **43** 80
[3] Wang Y 2020 Research on online ‘water army’ identification for douban network film critics *Intell. Comput. Appl.* **10** 218
[4] Yuan L, Zhu Z and Ren T Survey on Fake Review Recognition *Comput. Sci.* **48** 111
[5] Zhao Y. A Study on the Harm and Its Countermeasures of Chinese Paid Internet Trolls to The Film Industry *Northwest Normal Univ.* 2021
[6] Jindal N and Liu B 2008 Opinion spam and analysis *Proc. 2008 Int. Conf. on Web Search and Data Mining (Palo Alto).* (New York: Association for Computing Machinery ACM) p 219
[7] Lim E P, Nguyen V A, Jindal N, Liu B and Lauw H W 2010 Detecting product review spammers using rating behaviors *Proc. 19th ACM Int. Conf. on Information and Knowledge Management (Toronto).* (New York: Association for Computing Machinery ACM) p 939
[8] Mukherjee A, Venkataraman V, Liu B and Glance N 2013 What yelp fake review filter might be doing? *Proc. Int. AAAI Conf. on Web and Social Media (Cambridge).* vol 7 (Palo Alto: AAAI Press AAAI) p 409
[9] Zhang D, Liu Y, Zhang W, Shen F and Yang J 2020 Fake comment detection based on heterogeneous ensemble learning *J. Shandong. Univ. (Engineering Science)* **50** 1
[10] Cao D, Li S, Chen H, Zhang J and Zhang Q 2021 Fake Review Detection Method Fusing Sentiment Features *J. Inform. Eng. Univ.* **22** 326

DOI: 10.54254/2755-2721/48/20241057

[11] Zhang D, Huang L, Zhang R, Xue H and Lin J 2021 Fake Review Detection Based on Joint Topic and Sentiment Pre-Training Model *J. Comput. Res. Dev.* **58** 1385
[12] Du S, Yang M and Qiu R 2023 Research on Recoanition of Fake Online Product Reviews Based on SMOTE-RF and Multidimensional Feature Vector *J. Intell.* **42** 156
[13] Yang C, Li T, Tan S and Yang X 2020 Spam Review Detection Based on Double Convolutional Neural Network *Comput. Digit. Eng.* **48** 1954
[14] Zhang R and Zhang X 2021 Opinion spam detection based on hierarchical heterogeneous graph attention network *J. Comput. Appl.* **41** 1275
[15] Tao C and Yang J 2021 Detection of spam reviews based on subjectivity and EasyEnsemble algorithm *Appl. Res. Comput.* **38** 1403
[16] Ye Z and Wang 2021 Fake Review Groups Detection Based on Spectral Clustering *Comput. Appl. Softw.* **38** 175