

Research on Sentiment Analysis and Visualization of Barrage Text

Guangping He, Yirong Wang and Liping Wang

Chengdu University of Technology, Chengdu 610059, China.

Abstract

With the rise of the barrage video and live broadcast industry, barrage is not only a commentary text, but also an important tool for people to communicate in real time. The bullet screen has the characteristics of strong interaction, fast language update, and close connection with video content. It has important applications for assisting business decision-making, video highlight segment detection, and network video surveillance. Starting from the sentiment analysis of barrage texts, this article will focus on the introduction of automated analysis techniques and visualization methods of barrage texts, and provide new perspectives and theoretical support for future sentiment analysis of barrage texts.

Keywords

Barrage, Sentiment Analysis, Visualization.

1. Introduction

In recent years, the number of users in the online video industry has been increasing. A large number of online video users have generated a large number of online video comments. "Barrage" comments are a popular video commenting method in recent years. The comment content is directly displayed on the video interface. When there are many comments on a video, the video interface will have an effect like countless bullets flying by. The term barrage comes from Japan, and originally meant intensive shelling of targets. Barrage originated from Niconico, Japan. AcFun was the first to introduce it in China. As the barrage system becomes more and more complete, domestic mainstream video platforms such as Bilibili, Tencent Video, Youku, Iqiyi, etc., and live broadcast platforms such as Douyu, Huya, etc. The barrage service was opened, and barrage sprang up like bamboo shoots.

Barrage comment data contains the comment text and the corresponding video time point information. Therefore, compared with general comment data, the barrage data can more accurately and specifically reflect the user's instant emotions and praise and criticism when watching the video. These sentiment and evaluation information can provide references for other users when choosing videos[1]. Using sentiment analysis technology to extract sentiment information from online video bulletins and display it in combination with visualization methods can help users obtain the overall sentiment tendency of the video review text and the changes in the review sentiment over time. With the promotion of the barrage function in major mainstream video websites, the opinions and emotional expressions in the barrage comments will be more universal and reference.

Based on the above analysis, the application of sentiment analysis and visualization methods in network video barrage data analysis is tentatively explored.

2. Relevant technology realization of barrage sentiment analysis

Sentiment analysis, also known as Opinion mining, is mainly to analyze and study the texts published by people that contain subjective content such as comments and opinions. This information expresses

people's various emotional colors and emotional tendencies. Sex, such as joy, anger, sadness, joy and criticism, praise, etc. Emotions and emotions are an individual's private state, which cannot be directly observed and verified by others[2]. We can only analyze the text results of their subjective expressions. Through sentiment analysis, we can understand the hidden information of the user's attitude and emotion contained in the text. The analysis and mining of these hidden information helps to understand the user's preferences, and the user's attitude towards a certain topic, and then assist us in making decisions[3].

2.1 Granularity Classification of Barrage Sentiment Analysis

According to the analysis granularity of sentiment analysis tasks, text sentiment analysis is divided into text sentiment analysis at the text level, sentence level and aspect level. The task of text-level sentiment analysis is to define the sentiment tendency of the entire document, and the sentiment analysis of sentence-level text is to determine the overall sentiment orientation of the text in sentence units[4]. The former two are both coarse-grained sentiment analysis, while aspect-level text sentiment analysis is a more fine-grained sentiment analysis. It not only analyzes the overall subjective sentiment tendency of the corpus, but also analyzes the specific subjective sentiment targeted Aspect of emotional tendencies. This can dig out richer and more valuable information. For example, the corpus "The special effects of this movie are good, but the plot is too bad" is a positive praise for the subjective emotions of the movie's special effects, and a negative derogatory meaning for the movie's plot.

2.2 Computer Technology in Barrage Sentiment Analysis

At the same time, there are three main types of tasks related to sentiment analysis in terms of technical implementation:

- 1) Rule-based sentiment analysis of this article
- 2) Text sentiment analysis based on machine learning
- 3) Text sentiment analysis based on deep learning technology

2.2.1 Method based on sentiment dictionary and rules

The method based on sentiment rules is a relatively simple method of sentiment analysis, that is, through the establishment of an sentiment dictionary, the number of positive and negative sentiment words in the text or the sentiment value of the sentiment words are counted, and finally the sentiment score is calculated according to a series of artificial rules. Determine the sentiment category of the text. Rule-based text sentiment analysis methods rely heavily on the coverage of sentiment dictionaries and the completeness of manual rules. They tend to have better results in specific domains, but they cannot be widely used in other domains, and the sentiment words in the dictionary need to be continuously expanded and expanded. Update[5].

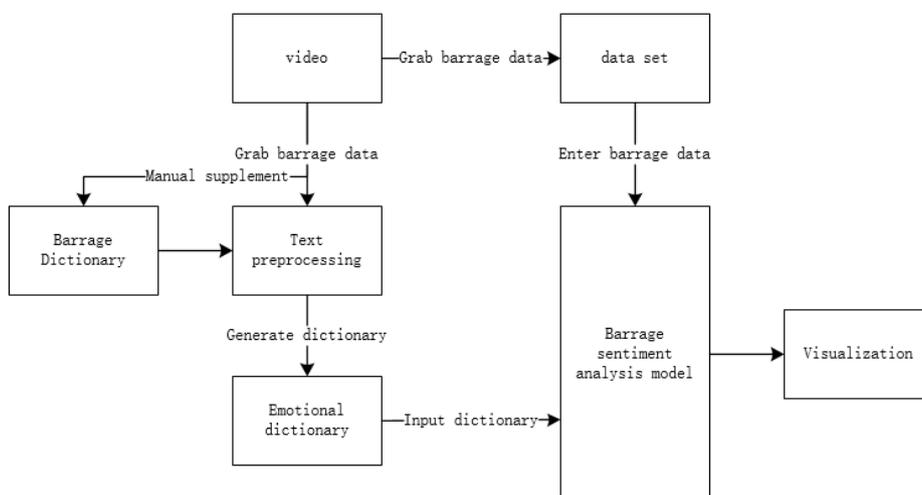


Figure 2.1. Rule-based sentiment analysis process

2.2.2 Method based on shallow machine learning

Based on the method of machine learning for sentiment analysis, the sentiment analysis task is regarded as a classification problem, the text in the sentiment dictionary is vectorized as a feature, and the sentiment tendency expressed by the sentiment dictionary text is used as a label, a training data set is constructed, and then used Machine learning classification algorithms are trained to classify texts representing different emotional tendencies into different categories[6]. Commonly used are Naive Bayesian (NB), support vector machines (SVM), and Maximum Entropy (ME). It essentially extracts features of different categories from the labeled data set, and then uses related methods for sentiment analysis[7].

Although machine learning methods perform well in some experiments, they are prone to problems such as sparse feature vectors and difficulty in feature extraction during feature extraction. For this reason, machine learning-based methods require a lot of feature engineering, such as data preprocessing, stop word processing, and text vectorized representation (one-hot (also called one-hot encoding), TF-IDF, n-gram).

2.2.3 Deep learning-based methods

Deep learning is a multi-layer representation learning algorithm. Compared with traditional machine learning algorithms, it has a deeper network structure and stronger expression ability, and can extract deeper features from data[8]. Therefore, the deep learning method is also the mainstream method currently used for sentiment analysis. Its classic basic network models include: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and cyclic neural Network (Recurrent Neural Network, RNN). In order to mine text sequence information more comprehensively on these basic network structures, bidirectional network structures such as BiLSTM and BiGRU are derived. At the same time, in order to make the model pay more attention to the important information of the text sequence, the attention mechanism (Attention Mechanism) in the field of computer vision is introduced[9].

It is worth mentioning that with the in-depth mining of the semantic information represented by the text, deep semantic analysis techniques, such as the combination of Semantic Dependency Parsing (SDP) and deep learning, open up a new kind of sentiment analysis. Ideas.

3. Visualization of barrage sentiment analysis results

The barrage text contains the review content and its corresponding video time point information. Therefore, compared with general product reviews, the barrage text can more accurately and specifically reflect the user's instant emotions and praise and criticism when watching a video or live broadcast [10]. Using sentiment analysis technology to extract sentiment information from videos or live broadcasts and display them in combination with visualization methods can help research and analyze the overall sentiment tendency of the bullet screen text and the changes in the sentiment of the bullet screen text over time.

The visualization of barrage sentiment analysis is mainly the time change graph of the number of barrage, sentiment word tag cloud and sentiment trend graph. In order to show the visualization effect, this article captured the one-hour live video barrage of a game on the Huya live broadcast platform, and after generalized sentiment analysis, combined with visualization for specific analysis.

3.1 Time change graph of the number of bullet screens per unit time.

The time attribute of the barrage is the main difference between it and the general text. It is the relationship between the number of barrage and time, which reflects the activity of the barrage and the emotional intensity of the participants from the side.

4. Conclusion

The barrage platform allows online video users to send and watch comments while watching the video. The barrage comments of some videos even contain more and more popular information than the video itself. The barrage comments have become a second time for users to make videos. A new way of editing. With the popularity of the barrage function and the increase in the number of users, the opinions and emotions in the barrage comments will become more universal and referential.

Starting from the sentiment analysis of the barrage text, this article focus on the introduction of the automated analysis technology and visualization methods of the barrage text. Part of the barrage data from the chicken-eating game on the Huya live broadcast platform is used as an experimental data set. It uses sentiment analysis technology and visually displays the sentiment data in order to fully tap the barrage content resources.

References

- [1] Pang B, Lee L. Thumbs up?: Sentiment Classification Machine Learning Techniques [C]. In: Conference on Empirical Methods in NLP Proceedings Morristown 2002:79-86
- [2] Yu H, Hatzivassiloglou V Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences [C]. In: Proceedings of the Conference on Empirical Methods in NLP. Morristown: ACL, 2003:129-136
- [3] Hu M, Liu B. Mining and Summarizing Customer Reviews [C]. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM_ 2004: 168—177.
- [4] Kim S M, Hovy E. Determining the Sentiment of Opinions [C]. In: Proceedings of the 20th International Conference on Computational Linguistics. Morristown: ACL, 2004:1367—1373
- [5] Yang S, Li S, Zheng L, et al. Emotion Microblog [C]. In: Proceedings of the on Web Society (SWS'09). 2009: 71-75 Mining Research on 1st IEEE Symposium
- [6] Rao D, Ravichandran D. Semi—Supervised Polarity Lexicon Induction [C]. In: Proceedings of the 12th Conference of the European Linguistics Chapter of the Association for Computational Morristown: ACL, 2009: 675-682
- [7] Turney P D, Littman M L. Measuring praise and criticism: Inference of semantic orientation from association [J]. ACM Transactions on Information Systems (TOIS), 2003, 21(4): 315—346.
- [8] Riloff E, Wiebe J, Wilson T. Learning subjective nouns using extraction pattern bootstrapping [J]. 2003, 4: 25—32.
- [9] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis [C]// Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Vancouver: ACM Press, 2005:347—354.
- [10] Tang D, Qin B, Wei F, et al. A joint segmentation and classification framework for sentence level sentiment classification [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 2015, 23(11):1750-1761.