A sentiment analysis on bullet screen using machine learning bag of words algorithm

Yiyuan Wang[1] and *Jie* Li[2,*]

[1]School of Media and Communication, Taylor's University, Subang Jaya, Selangor
[2]School of Computer Science, Taylor's University, Subang Jaya, Selangor, Malaysia

Abstract. Bullet Screen is a novel innovation allowing viewers to watch internet films and engage in real-time comment reading and sharing opinions at the same time. This paper aims to adopt machine learning algorithms to develop a model based on Bag of Words to conduct sentiment analysis on the reviewer's comments on the bullet screen. This paper adopted an experimental design with a dataset of 6,300 valid comment datasets retrieved from bilibili.tv. This research used accuracy and loss as performance metrics and compared the performance of the BOW model with the performance of the CNN and RNN models. The result suggested that the accuracy of the BOW model is the highest among all the three compared models, and it has low losses in its performance. This paper contributes new findings to Natural Language Processing (NLP) in social media, enriches the analytical approaches for sentiment analysis on audience expression online, integrates AI in social media, and drives further business innovation in the online video sector. Finally, a methodological limitation of this research is that the measuring metrics used did not incorporate precision and recall. Furthermore, the utilisation of BOW disregards both the arrangement of words and the surrounding context. Furthermore, our scope constraint fails to take into account the utilisation of emoji expressions. In the coming years, the authors intend to train and evaluate the precision of the model using various segments of social media sites. Additionally, authors endeavour to develop a model utilising the CNN architecture to accurately detect emotional expression through emotion classification, based on the findings of current research.

1 Introduction

Audience interaction is a popular concept widely applied in TV drama platforms for audiences to express their feelings along with their watching engagement. In East Asia, a novel social viewing experience, known as bullet screen, allows viewers to simultaneously watch internet films and engage in real-time comment reading and sharing [Cao, 2021]. The comments are superimposed over the videos, appearing on the screen at the same time as the video material.

* Corresponding author: Lijiesat@hotmail.com

Natural Language Processing (NPL) is recognized the most frequently applied technique to process textual data so as to facilitates user experience in the recent decade. Sentiment Classification Analysis is believed to be one of the NPL approaches to understanding the feeling of human expression using Machine Learning (ML) algorithms [2, 3]. The Bag-of-Words model is a straightforward approach of classification from textual input. The concept involves representing each sentence as a collection of words without considering syntax or linguistic patterns. The arrangement of words within a sentence is the only factor in determining its meaning for the model. In recent years, there has been a lot of research focusing on emotion classification using sentiment analysis approaches in the bullet screen application environment [4-7]. However, a limited number of them utilised the BOW model as the analysis approach to determine audience emotion in bullet screen context. The recent research emphasised the precision of their classification results using various machine learning algorithms. Thus, there is still potential for improvement in the classification result by applying hybrid models. This paper aims to adopt the BOW with the machine learning model in sentiment analysis to classify the emotions of the audience. The objectives of the paper is to design, develop, and examine a BOW hybrid model for sentiment analysis on the emotions of the audience from a bullet-screen dataset.

2 Literature Review

2.1 Text Mining and Natural Language Processing

Due to the increasing popularity and extensive use of natural language processing technology, research on short text mining has advanced significantly over the past decade [8]. Research has been ongoing for several years and has focused on various social media platforms, including Twitter and Facebook [9,10]. Bullet screen is a novel kind of commenting that involves the use of small text messages that are closely tied to video content and coordinated with specific time points. Currently, the existing research on bullet-screen reviews primarily concentrates on the analysis of video content [11]. For example, Li et al. [7] introduced a time-depth structured semantic model (T-DSSM) that segments a video using clustering and allows for the extraction of theme distribution from video clips [12]. The extraction of video highlights is accomplished by utilising supervised learning and finding semantic vectors. Liu et al. [12] established a video popularity prediction algorithm that evaluates bullet screens from several perspectives. The study by Gao et al. [13] and Li et al. [7] combines the criteria of uploader, video quality, and herd effect to forecast the popularity of bullet-screen videos, as these aspects can have an impact on their popularity. Kuo [14] proposed a handbook of NLP with Gensim to investigate the pattern, theme, and valuable insights of textual data.

2.2 Sentiment Classification

The analysis technique typically involves extracting text, preprocessing, extracting sentiment feature, and determining sentiment polarity [15]. Sentiment analysis is the computational examination of digital text to ascertain if the emotional disposition of message is favourable, negative, or neutral [16, 17]. The dictionary-based methods are widely used in today's world. For example, a lexicon-based algorithm employs adjectives and adverbs to ascertain the semantic orientation of the text. To determine the orientation of any text, a lexicon-based approach extracts combinations of adjectives and adverbs along with their corresponding sentiment values [18]. Catelli et al. [19] did an experiment where they compared the results of using a lexicon-based algorithm with Bert-based sentiment analysis (BERT). On the other hand, lexicon-based methods are improving. Sallam et al. [20] used lexicon-based sentiment analysis for an Arabic literature recommendation system to make it better at filtering and guessing what people would like more accurately.

One of the popular supervised methods is Support Vector Machine (SVM), which extracts features of data and uses models to map them to labels with sentiment definition. CA-SVM is proposed by Cyril et al. [10] their research into sentiment analysis of tweet data from Twitter. The result shows the accuracy of the model is higher than the other counterparts (i.e., ANN and K-Means) at 90% or above. For the unsupervised methods, K-Mean was widely used to group similar texts and infer sentiment according to predominant sentiment in clusters. Rezazadeh Kalehbasti et al., [21] used the R-squared score as the most important metric to determine the representativeness of the K-Mean model has its R-Squared equals 0.6748, which is higher than Linear Regression models and Neural Nets.

Liu et al. [12] come up with an algorithm that primarily estimates the likelihood that a danmaku sample belongs to a specific sentiment category based on the prior probability distribution. It then identifies the sentiment category with the highest probability as the forecast sentiment category. Ye et al. [22] conducted a study where they examined the effectiveness of NB, SVM, and N-gram models in sentiment classification of text comments. They discovered that the accuracy of the SVM and NB algorithms is significantly superior to that of the N-gram model. The figure above demonstrates a CNN-based sentiment classification process named SentiCNN. This method used sentence word vectors in a multidimensional embedding matrix to get features for classification. The results show that SentiCNN is the most accurate model compared to Lexicon-Based, CNN-Rand, LSTM, and NCSL, with an accuracy level of 80% or higher [6].

2.3 Bag of Words Models

The Bag-of-Words (BOW) model is a simple method for extracting features from text data. The idea is to represent each sentence as a bag of words, disregarding grammar and paradigms. Just the occurrence of words in a sentence defines the meaning of the sentence for the model [4]. In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity. The BOW model is a way to extract features from the text so programs could use the text input in machine learning algorithms like neural networks. There are 4 steps for BOW model design and development, including tokenization, vocabulary building, indexing, and vector representation [23]. The study by Hasan and Matin [4] adopted the BOW model to detect the sentiment of tweets, and their result shows the trained model has a 0.8 above recall and 87% accuracy in feature extraction.

Abudabar et al. [24] compared the BOW model with the TF-IDF model based on book review comments, and the result shows the BOW has advantages in precision and recall performance. Prior studies n sentiment analysis have employed word/n-gram bags, word2vec, and doc2vec models to represent sentiment in written native languages. Each of these 40 strategies emphasises different levels of textual granularity, which are represented by abstracting certain concepts from lower to higher levels of granularity [24]. Consequently, the feature vectors generated by each technique represent the abstract level of information. Bag-of-words and n-grams employ local representations to emphasise individual words and brief sequences of phrases. While doc2vec emphasises words that convey their context within the entire document, word2vec emphasises the overall representation of words and phrases in text.

3 Methods

3.1 Data Collection and Preprocessing

Regression statistically identifies the beta coefficient between independent and depend variables for understanding the internal linkage of the input and output of any contructed linear model. Comparing with other sentiment analysis approaches, BOW has advantages in text classification in terms of simplicity (i.e. only considering text data rather than word order and grammar), flexibility (i.e. easily combined with other TF-IDF), compatibility with ML algorithm (e.g. NB and SVM), and higher efficiency in execution (efficient computation and scalability in processing big data). Therefore, in this study, the authors used BOW with regression apporach to process the bullet screen data. There were 8,000 bullet screen comments collected for both training and test purposes using Barriage UI from the 2023 New Year Concert of Bilibili.TV. Authors used Jupyter Notebook to construct the model and examined the performance of the model. Since the nature of bullet-screen accentuates the use of language online, it matters to the accuracy of the analysis result in the end. The categorization of the online language based on emotion classification is therefore important in data preparation. Therefore, the word vocabulary NetLex word segmentation method was used after translation, and part of the result is shown in Table 1 below.

Table 1. Language Translation, Interpretation, and Classification.

Original Words	Translation	Remarks	Result
做的不错，下次别做了	Well done, but do not do it again.	Dispirited	Negative (0)
满级大佬	Top Boss	Compliment	Positive (1)
我膨胀了	I'm bloated	Self-mockery	Negative (0)

Moreover, in the data cleaning process of this experiment, the authors removed the data with an obscure definition to be understood so as to the data with an incomplete expression. There were 6,300 valid barriage data to be used for both test and training purposes. The split of the dataset follows the 25-75 rule, meaning 25% of the total samples should be used as training samples. In this experiment, tokenization and numericalization were discussed in the next section.

3.2 Algorithm Development

In this paper, the authors design the proposed model using BOW. As shown in Figure 1 above, there are 5 layers, including the input layer, embedding vector layer, pooling layer, linearity layer, and the final output layer, implemented in the proposed model. The algorithm 1 shown below interprets how the sentiment analysis is conducted from the data input to final result using the BOW model proposed. In this model, first of all, the embedding layer processes the input data from the dataset. Then, the embedding later takes the tokens (T) in sequence, and n represents the total number of the input sentences (see Equation 1 below).

Algorithm 1 Sentiment Analysis - BOW-NReg
1. Input: x; Tokenization: T(); Vocabulary: V; Occurence: O; Numericalization: N Result Prediction: R 2. Import Numpy, Torch, and Matplot Lab
3. Retrieve data → Training dataset (‘content‘, ‘label’) || Performance dataset (‘content‘, ‘label’) For each (x ∈ Training dataset) T(x);

```
Count (T(x)) = C;
V = Ascending C ∈ [T(x1) ... T(xn); O1 ... On] ;
For each N(X) Matches V
    For each Xn ∈ N(X)
        if  labelled 0:
        Xn → negative;
        else labelled 1:
        Xn → positive;
    return Xn;
R = reg(∑_0^n Xn | r);
Return R;
```

Tokenization is the process of utilising a tokenizer to handle the strings in our collection. A tokenizer is a function that converts a string into a list of individual strings.

$$\mathrm{T} = (\mathrm{T0},\ \mathrm{T1}\ \ldots\ \mathrm{Tn}) \in \mathrm{Z} \tag{1}$$

To accomplish the research objective, researchers must first decompose the string into distinct tokens and subsequently transform these tokens into numerical values. The taken token is marked as X to be put in a dimensional map. The dimension ID is noted as d, and the t stands for the number of tokens from the input layer. In Equation 2 below, the tensor matrix should be constructed based on the token and number of dimensions.

$$X = (X0,\ X1\ \ldots\ Xn) \in Rdt \tag{2}$$

Also, in order to improve the accuracy of the vocabulary, we took 30% of the total sample dataset as the training set to set the vocabulary based on the most occurrences in the dataset using the torchtext.vocab.build_vocab_from_iterator() function. After marking the most frequent token, the less frequent token remains unidentified. Afterwards, as shown in Table 2 below, all the tokens (string values) were numericized by giving the indexing ID in ascending order based on the number of occurrences.

Table 2. Numericalization Example.

I	like	the	film	,	but	not	this	actor
1	17	4	55	4	14	12	30	102

4 Experiment Result and Discussion

First of all, the loss using the BOW-NReg model was recorded during the 12 epochs. As shown in the diagram in Figure 1, the training loss is marked by a blue-coloured line. From each epoch, the loss gradually went down from above 0.7 to 0.280 on the 12th epoch. Also, the trend of the train loss has become less sloping in the last few epochs. The orange-coloured line represented the variation of the loss in actual performance after the training process. The loss started at 0.652 and ended at 0.314 after the 12th epoch.

For the accuracy, first of all, the accuracy using the BOW-NReg model was recorded during the 12 epochs. As shown in the diagram in Figure 3 above, the training accuracy is marked in a blue-coloured line. From each epoch, the accuracy gradually increases from 0.631 to 0.874 on the 12th epoch. Also, the trend of the train's accuracy became less sloping in the last few epochs. The orange-coloured line represented the variation in accuracy in actual performance after the training process. The accuracy started at 0.723 and ended at 0.851 after the 12th epoch.

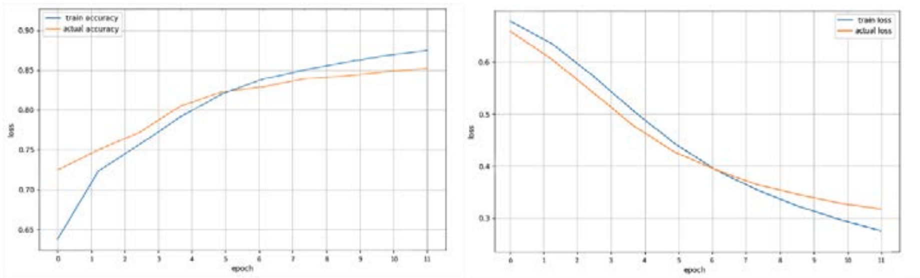

Fig. 1. Variance of the Accuracy and Loss over 12 Epoch.

The Table 3 above shows the comparison result among models from previous research. The current BOW-NReg is based on neural network algorithm performs better accuracy than BERT-SupCL by Xu & Wang [25]. BOW-Reg has similar performance on accuracy to UgramBOW by Hasan & Matin [4], which was also built based on BOW model. Also, the F1 metrics performs higher than BERT-SupCL but lower than UgramBOW.

Table 3. Performance Comparison among BOW and Results from Previous Research.

Model		Accuracy	F1 value	Basic Model
BOW-NReg	Current	0.874	0.831	Regression
UgramBOW	Hasan & Matin [4]	0.850	0.853	BOW
BERT-SupCL	Xu & Wanag [25]	0.773	0.716	BERT

Further understanding of using BOW method in sentiment analysis accentuates the simplicity and effectiveness of the algorithm after adopting the regression method. BOW is a effective method to determine sentiment in a context where long-range dependency is less important [4]. In the context of bullet screen comments analysis, the sequence of single comment is limited and the contextual implication is less complicated than other textual context. In this case, the use of BOW-NReg in bullet screen comment analysis is a effective and concise way to determine the sentiment, sine it maintains the accuracy of the task performance based on the comparison results shown above.

5 Conclusions

As an emerging trend in online watching experiences for online audiences, the screen bullet technique allows users to interact while watching the same video clip. This paper adopted the BOW, which is a classification algorithm, to conduct a regression sentiment analysis to evaluate the audience emotion (positive or negative). The implication of the result indicated, by comparing with its counterparts, the proposed BOW-NReg model has better performance in terms of highest accuracy and less loss in the actual performance stage. To be summarized from the findings, the current BOW-NReg has its advantage to be used in semtiment analysis for social media applications to understand emotion of audience with significant interpretability, effecient large dataset processing capability in the real business environment.

From the academic angle along, this paper, firstly, contributes new findings to Natural Language Processing in social media and enriches the analytical approaches for sentiment analysis on audience expression online. Secondly, the use of BOW-NReg model provides a reliable possibility to understand audience attitude in the context of bullet screen commnets. On top of that, to the social media industry, this paper also contributes to the integration of AI in social media and drives further business innovation in the online video sector (e.g. Robotic Process Automation). Also, this paper contributes to the evolution of social media platforms and accentuates the impact of bullet screens on modern business practices. Lastly, there is no doubt that an accurate model with remarkable prediction can contribute to the improvement of the user experience.

At the meantime, the concerns over the data privacy in information security is considerable for the further application development based on the current developed model. However, there are a few limitations. The methodological limitation of this research is that the measurement metrics adopted did not include precision and recall. Second, the use of BOW-NReg ignores word order and context. Also, our limitation on scope lacks in considering the use of emoji expressions.

References

1. Cao, X.: Bullet screens (Danmu): texting, online streaming, and the spectacle of social inequality on Chinese social networks. Theory, Culture & Society, 38(3), 29-49. 2021
2. Zhou, F., Zhou, X.: Across Time and Space, I Am Together with Many, Many Others": Digital Writing and Temporality on Chinese Social Media: Social Media+ Society, 8(3), 20563051221117564. 2022
3. Li, B., Naraine, M. L., Zhao, L., & Li, C. (2023). A Magic "Bullet": Exploring Sport Fan Usage of On-Screen, Ephemeral Posts During Live Stream Sessions. Communication & Sport, 11(2), 334-355. 2023
4. Hasan, T., Matin, A.: Extract Sentiment from Customer Reviews: A Better Approach of TF-IDF and BOW-Based Text Classification Using N-Gram Technique. In Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020 (pp. 231-244). Springer Singapore. 2021
5. Hsieh, Y. H., Zeng, X. P.: Sentiment analysis: An ERNIE-BiLSTM approach to bullet screen comments. Sensors, 22(14), 5223. 2022
6. Huang, M., Xie, H., Rao, Y., Liu, Y., Poon, L. K., Wang, F. L.: Lexicon-based sentiment convolutional neural networks for online review analysis. IEEE Transactions on Affective Computing, 13(3), 1337-1348. 2020
7. Li, Z., Li, R., Jin, G.: Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. Ieee Access, 8, 75073-75084. 2020
8. Amaratunga, T.: NLP Through the Ages. In Understanding Large Language Models: Learning Their Underlying Concepts and Technologies (pp. 9-54). Berkeley, CA: Apress. 2023
9. Stier, S., Bleier, A., Lietz, H., Strohmaier, M.: Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. In Studying Politics Across Media (pp. 50-74). Routledge. 2020
10. Cyril, C. P. D., Beulah, J. R., Subramani, N., Mohan, P., Harshavardhan, A., & Sivabalaselvamani, D.: An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM. Concurrent Engineering, 29(4), 386-395. 2021
11. Liu, J., Zhou, Z., Gao, M., Tang, J., Fan, W.: Aspect sentiment mining of short bullet screen comments from online TV series. Journal of the Association for Information Science and Technology, 74(8), 1026-1045. 2023
12. Lv, G., Xu, T., Chen, E., Liu, Q., Zheng, Y.: Reading the videos: Temporal labeling for crowdsourced time-sync videos based on semantic embedding. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1). 2016
13. Gao, H., Yin, H., Peng, L., Wang, H.: Effectiveness of social video platforms in promoting COVID-19 vaccination among youth: A content-specific analysis of COVID-

19 vaccination topic videos on Bilibili. Risk Management and Healthcare Policy, 1621-1639. 2022

14. Kuo, C.: The Handbook of NLP with Gensim: Leverage topic modeling to uncover hidden patterns, themes, and valuable insights within textual data. Packt Publishing Ltd. 2023.
15. Mehta, P., Pandya, S.: A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research, 9(2), 601-609. 2020
16. Wankhade, M., Rao, A. C. S., Kulkarni, C.: A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731-5780. 2022
17. Haleem, A., Javaid, M., Singh, R. P.: An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. BenchCouncil transactions on benchmarks, standards and evaluations, 2(4), 100089. 2022
18. Machová, K., Mikula, M., Gao, X., & Mach, M.: Lexicon-based sentiment analysis using the particle swarm optimization. Electronics, 9(8), 1317. 2020
19. Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. Electronics, 11(3), 374. 2022
20. Sallam, R. M., Hussein, M., & Mousa, H. M.: Improving collaborative filtering using lexicon-based sentiment analysis. International Journal of Electrical and Computer Engineering, 12(2), 1744. 2022
21. Rezazadeh Kalehbasti, P., Nikolenko, L., Rezaei, H.: Airbnb price prediction using machine learning and sentiment analysis. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 173-184). Cham: Springer International Publishing. 2021
22. Ye, Q., Huang, P., Zhang, Z., Zheng, Y., Fu, L., Yang, W: Multiview learning with robust double-sided twin SVM. IEEE transactions on Cybernetics, 52(12), 12745-12758. 2021
23. Juluru, K., Shih, H. H., Keshava Murthy, K. N., Elnajjar, P.: Bag-of-words technique in natural language processing: a primer for radiologists. RadioGraphics, 41(5), 1420-1426. 2021
24. Abubakar, H. D., Umar, M., & Bakale, M. A. (2022). Sentiment classification: Review of text vectorization methods: Bag of words, Tf-Idf, Word2vec and Doc2vec. SLU Journal of Science and Technology, 4(1 & 2), 27-33. 2022
25. Xu, L., Wang, W.: Improving aspect-based sentiment analysis with contrastive learning. Natural Language Processing Journal, 3, 100009. 2023