



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Emoji Based Opinion Mining Using Deep Learning Techniques

C.GAYATHIRI, S.SIVAPRAVEEN, S.SUDHARSAN, M.VIGNESHWARAN, S.VISVESVARAN

Department of Computer Science and Engineering, Mahendra Institute of Technology, Thiruchengode, Namakkal,  
Tamil Nadu, India

## ABSTRACT

Informal media, such as Twitter, are more relevant today than ever before. Twitter is still a valuable tool for friends to communicate, but it has evolved into a public bulletin board where ordinary people, companies, and even big personalities like politicians and sports routinely publish their thoughts and participate in conversations. Since, Twitter is so extensively utilized throughout the world, the ability to do reliable opinion mining and gauge public opinion and perception on a variety of issues is more vital than ever. Emojis are frequently used to communicate feelings or sentiments that are difficult to express succinctly in language. For emoji-based opinion mining, a deep learning framework is presented to model the influence of emojis on text sentiment polarity. The emojis and words in microblog posts are combined to create emoji representations that include contextual information. This is due to the nature of microblogs, which allow users to publish real-time messages about their thoughts on a wide range of topics, debate current events, complain, and express a wide range of emotions. Data from microblogging platforms and social networks may be used in opinion mining, sentiment analysis, and emotion analysis applications since the audience of these platforms and networks expands every day. Sentiment analysis and opinion mining studies opinions and associated concepts such as feelings and emotions. The field's emergence and quick expansion parallel social media on the Internet, including reviews, forum discussions, blogs, microblogs, Twitter, and social networks.

## I.INTRODUCTION

The polarity of like/dislike feelings is generally the focus of current sentiment analysis approaches. In the middle, neutral feelings can be sensed. Human emotions are considerably more complex and varied than these simple measures. As a result, polarity analysis provides only a limited amount of information about the author's real meaning. It is relatively easy to define simply positive or negative emotions, but it is far more challenging to define a complete and distinct collection of emotions. As a result, researchers have developed a wide range of study techniques for identifying fundamental emotions. Because of the features of computer-mediated communication, which may lack many of the auditory and visual signals that are typically linked with emotional elements of interaction, emotions in virtual communication differ in several ways from those in face-to-face contacts [2]. While text-based communication may not provide auditory or visual signals, alternative ways can convey emotion. Emoticons, or emotional icons, are symbols that may express a wide range of emotions.

Emojis and emoticons are extensively used in social media to communicate emotions, moods, and thoughts, particularly when individuals find it difficult to express their feelings in plain text or when they want to demonstrate sarcasm [7-9]. These emojis and emoticons are extremely valuable since they significantly influence communicating a sentence's sentiment polarity. Emojis and emoticons are thus valuable tools for sentiment analysis. The widespread usage of emojis has piqued the interest of academics [10, 11] because they provide essential semantically and sentimental information to visually complement textual data, which is crucial for deciphering the inherent emotional signals in texts. Emoji embedding's, for example, have been proposed to better comprehend the semantics of emojis [12, 13], and embedding vectors may be used to display and forecast emoji usage based on their contexts. Retraining a deep neural network on an emoji prediction task with pre-trained emoji embedding's to understand the emotional signals of emojis for other tasks such as sentiment, emotion, and sarcasm prediction has also been shown to be beneficial.

## II.LITERATURE SURVEY

### 2.1. OVERVIEW OF EMOJI-BASED OPINION MINING

Although many people confuse emoticons and emojis and often have the same meanings, the two are not the same. Emoticons are a graphical representation of a facial expression made up entirely of characters. Emojis are graphic icons that

express ideas such as “Face with tears of joy” emoji show. Each emoji is a Unicode character that cannot be entered using standard keyboards. Emojis and emoticons are used interchangeably, and emojis are commonly thought to be the new generation of emoticons.

## **2.2. THE POSTINGS WERE EXTRACTED USING AN ATTENTION METHOD**

Proposed HEMOS, a humour detecting deep learning model, in the literature [21]. Positive, negative, optimistic humour for jokes, self-mockery, jocosity, etc., and pessimistic humour for sarcasm and irony on Chinese Web postings were all classified using the attention based Bi-LSTM model. Web’s humorous postings featuring online lingo and emojis were fed into the model, which was then projected into a vector representation. Important words that contributed to the meaning of the postings were extracted using an attention method. Elbow et al. used a data consisting.

## **2.3. MACHINE LEARNING-BASED METHODS**

Tan and Zhang conducted an empirical study of sentiment categorization on Chinese documents. Their results showed that the information gain and SVM features provided the best performances for sentiment classification coupled with domain or topic dependent classifiers. There are also researchers who have combined the machine learning approach with the lexicon-based approach. Chen et al. proposed a novel sentiment classification method which incorporated existing Chinese sentiment lexicon and convolutional neural network.

## **III. PROJECT DESCRIPTION**

### **EXISTING SYSTEM**

Internet slang is informal language used to express ideas on the Chinese Internet in response to events, to mass media and foreign cultures. It also expresses a natural human desire to simplify and update language. Slang that first appears on-line is often adopted to become widely used in everyday life. It includes content relating to all aspects of social life, mass media, economic, political situation etc. Internet slang is arguably the fastest-changing aspect of a language, created by a number of different influences, technology, mass media and foreign culture amongst others. Because Internet slang is not easy to extract automatically, it can cause a significant difficulty in sentiment detecting task. For improving the performance of Chinese social media sentiment analysis, we created a Chinese Internet slang lexicon (examples shown in Table 1). We manually extracted 448 frequent Internet slang terms from the Internet.

### **DISADVANTAGES**

- A number of Emoji characters were added to represent the feelings of sadness, grief and anger.
- Emoji characters were also being used by people in the event with negative feelings.
- In this part we took into consideration the existence of Emoji characters in the text during the sentiment scoring process.

## **IV. PROPOSED SYSTEM**

Proposed a rule-based approach with two phases: a) the sentiment of each sentence is first decided based on word dependency to aggregate the sentences sentiments and then b) the sentiment of each document is calculated. A method that does not require any annotated corpus training data and only requires information on commonly occurring negations and adverbials. Li et al. stated that polarities and strengths judgment of sentiment words comply with a Gaussian distribution, and thus proposed a Normal distribution-based sentiment computation method which allows quantitative analysis of semantic fuzziness of sentiment. The fuzzy semantic model by using an emotion degree lexicon and a fuzzy semantic model. Their model includes text pre-processing, syntactic analysis, and emotion word processing. They empirically showed the advantages of using, the newly-built slang sentiment word dictionary for sentiment classification.

### **ADVANTAGES**

- We collected 448 frequent Chinese Internet slang expressions as a Chinese slang lexicon.
- We converted the emoticons into textual features creating Chinese emoticon lexicon.
- We empirically confirmed implicit humour characteristic to Chinese culture visible on utilized both lexicons with several machine learning approaches for detecting humorous expressions on and confirmed that using both slang and emoticons improves previously proposed method.



**MODULE DESCRIPTION**

**Data Acquisition and Pre-processing**

The dataset contains various data related to the tweets and the user’s profile, while maintaining the structure of a connected network. This data includes the ID of the tweet, the ID of the user, the name of the user, the content of the tweet, the language of the tweet, the number of likes, the number of replies, the number of retweets, the origin of the retweet, the time, the followers, the followers, the link, the source of the video, the cover of the video, and the images.

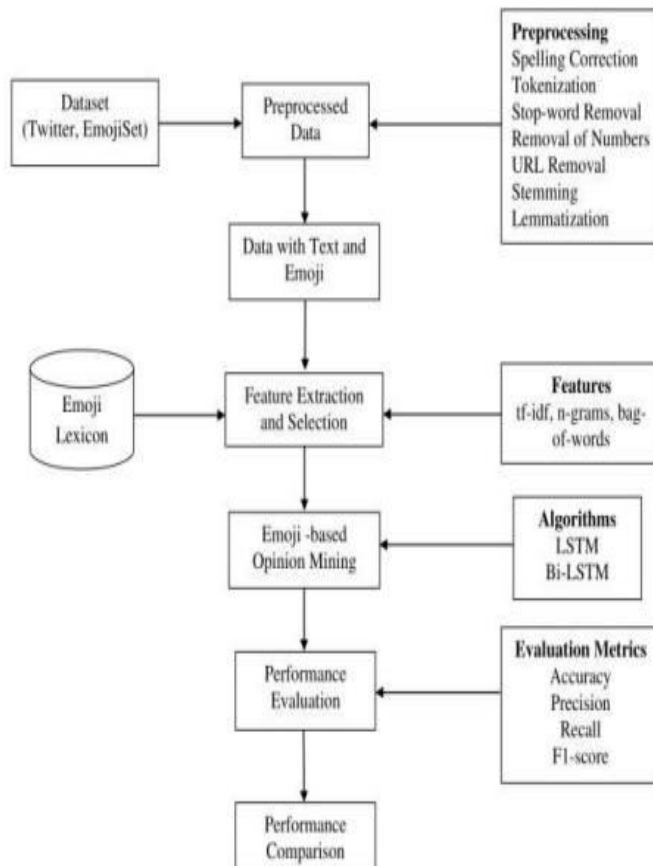
**CNN for Opinion Mining**

CNN is extensively used in image processing and image recognition. It is also known for handling challenges related to NLP tasks such as sentiment analysis. In the CNN architecture, the convolutional layers are made up of neurons that search for patterns in their input. Pooling layers are frequently added after convolutional layers, primarily to minimize the dimensionality of the feature map.

**LSTM for Opinion Mining**

LSTM is effective at dealing with sequential data because it allows for long-term dependencies to remain throughout the network. That explains their broad use in text mining applications. LSTM suppose a relation between data items in sequence. Our objective is to examine the impact of emojis and topics as non-sequential data on the performance of LSTM for opinion polarity. We propose an LSTM architecture in which two input layers are concatenated using a concatenation layer. The first input to the LSTM layer consists of two sequential features, namely the polarity of words and the Part. Of Speech tag of each word. The second input includes the non-sequential features such as the associated topic and emojis.

**DATA FLOW DIAGRAM**







### V.RESULTS

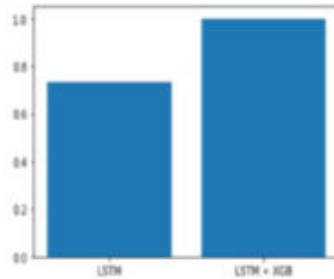


Fig 7. Accuracy graph for LSTM and XGBOOST

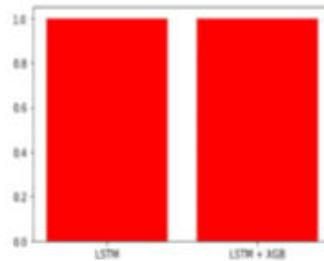


Fig 8. Sensitivity graph for LSTM and XGBOOST

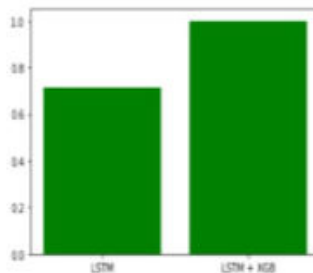
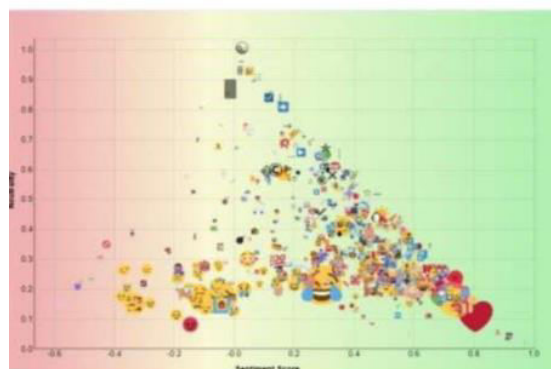


Fig 9. Specificity graph for LSTM and XGBOOST



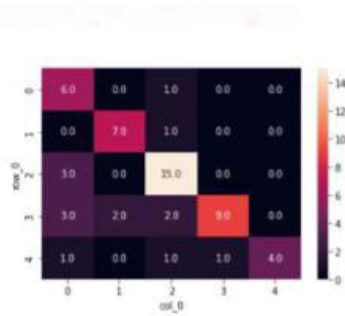


Fig 5.4. Confusion matrix diagram for LSTM



Fig 5.5. Confusion matrix diagram for XGBOOST

Author	Accuracy	Precision	Recall	F1-score
Model-LSTM	0.803	0.834	0.850	0.813
Model-BiLSTM	0.857	0.865	0.902	0.864
Model-XGBOOST	0.956	0.793	0.734	0.745

Table 1: Main results Emojiset Dataset

## VI.CONCLUSION

In this proposed work, the SA is performed on mobile product reviews using various deep learning algorithms. The novelty of this work was to develop an Emoji-based sentiment lexicon based on the text polarity scores. The performance evaluators considered are sensitivity, specificity, F1-score, and accuracy. The mobile reviews considered had both text and emojis. This data was pre-processed and data labelling was done. During data labelling, we used the Emoji2Vec model for calculating the polarity scores. The feature extraction is done using the word embedding models CBOW and SG. Then the LSTM, CNN, Bi-LSTM, and CNN-LSTM models were developed along with their hyper parameters. The experimented results prove that all the models perform well. The CNN-LSTM proved to have high accuracy. The limitation was about the computational requirements for processing large data and few neural network models. Moreover, the SA cannot be limited to only identifying positive and negative labels of text and Emoji data. The aspects of the products also have to verify. So, in the future, this work is extended to complex Deep learning architectures and aspect-based.

## REFERENCES

1. Aldunate, N., Gonzalez-Ibáñez, R.: An integrated review of emoticons in computer-mediated ~ communication. *Frontiers in psychology* 7, 2061 (2017)
2. Chen, Z., Xu, R., Gui, L., Lu, Q.: Combining convolution neural network and word sentiment sequence features for chinese text sentiment analysis. *Journal of Chinese Information Processing* (2015)
3. Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* 20(3), 273–297 (1995)
4. Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S.: Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524* (2017)
5. Guibon, G., Ochs, M., Bellot, P.: From emojis to sentiment analysis. In: *WACAI 2016* (2016)
6. Ho, T.K.: Random decision forests. In: *Document analysis and recognition, 1995., proceedings of the third international conference on*. vol. 1, pp. 278–282. *IEEE* (1995)
7. Kim, Y.: Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014)



8. Kulkarni, V., Wang, W.Y.: Tfw, damngina, juvie, and hotsie-totsie: On the linguistic and social aspects of internet slang. arXiv preprint arXiv:1712.08291 (2017)
9. Li, D., Rzepka, R., Ptaszynski, M., Araki, K.: Emoticon-aware recurrent neural network model for chinese sentiment analysis. In: The Ninth IEEE International Conference on Awareness Science and Technology (iCAST 2018) (2018)
10. Li, R., Shi, S., Huang, H., Su, C., Wang, T.: A method of polarity computation of chinese sentiment words based on gaussian distribution. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 53–61. Springer (2014)



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details