VIVA-Tech International Journal for Research and Innovation
ISSN(Online): 2581-7280

Volume 1, Issue 1 (2018)
Article No. 3
PP 1-5

VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

Extraction of Emoticons with Sentimental Bar

Shreyas Wankhede[1], Ranjit Patil[1], Sagar Sonawane[1], Prof. Ashwini Save[2]
[1](*Computer Engineering, Viva Institute of Technology/ Mumbai University, India)*
[2](*HOD, Computer Engineering, Viva Institute of Technology/ Mumbai University, India)*

Abstract :*The latest generation of emoticons which are called as emojis that is largely being used in mobile communications as well as in social media. In past few years, more than ten billion emojis were used on Twitter. Emojis which are known as the Unicode graphic symbols, which are basically used as shorthand to express the concepts and ideas of the people. For smaller number of well-known emoticons, their meanings or sentiments are well known but there are thousands of emojis so extracting their sentiments is difficult. The Emoji Sentiment Ranking method which is used to evaluate a sentiment mapping of emojis by using sentiment polarity such as negative, neutral, or positive. The sentimental classification of tweets with and without emoticons are very much different.Finally, the method also gives representation of sentiments and a better visualization in the form of a sentimental Bar.*

Keywords-*Classification of Emoticons, Emoji Sentiment Ranking, Sentiment Bar, Sentiment labels, Sentiment score.*

1. INTRODUCTION

As the use of social media is increasing day by day, emoticons plays a essential role in communication through technology, and many other devices have provided different forms of pictures that do not use type punctuations. They provide another range of expressions and feelings through texting that conveys specific emotions through facial gestures. Nowadays emoticons on smartphones, in chatting, and in many different applications, have become tremendously popular worldwide. For example, Twitter has become very active in sharing content with comments. According to statistics, around 500 million of tweets are dispensed per day. Each tweet expresses different form of emotions.

An emoticon, such as ☺, ☹ and many others are used for facial expression. It also allows the peoples to express their feelings, moods, and emotions, which replaces and enriches a written message with non-verbal elements. It allows user to understand the feelings of their friends and colleagues in better manner. Some social network sites and microblogging tools such as Twitter allows individuals to express their feelings or opinions to specific results. These short messages which are also known as tweets that includes emotions such as happiness, sadness, anger etc in it. Classification of emoticons is basically done in two categories such as positive emoticons and negative emotions. Positive emoticons consistof love and joy whereas negative emoticons consist of sadness and anger.

The simplest forms of representation are the generally denoted as 'emoticon' or 'smiley'. People from Japan popularized a kind of emoticon called kaomoji where ((kao)=face and (moji)=characters) [6]. Sentiment analysis of text is being done by many researchers but for emoticons still it's in developing stage so it's a need to research more on emoticons and give it a limelight to know all about emoji for future users.

Volume 1, Issue 1 (2018)
Article No. 3
PP 1-5

2. RELATED WORK

There are six reported works related to Emoticons found in the literature review.

Georgios Solakidis [1]. In this paper the objective is to evaluate sentiment analysis on multilingual data, also the paper focuses on study and draw conclusion about subjectivity, polarity and the feelings that is expressed in user generated content, which mainly consist of free text document. The approach involves detection and use of self-defining features that available within the data that take accounts in two emotionally rich features: -a] emoticons b] lists of specific keywords. There is machine learning approach on collection of training data using evaluating and comparing the result of two separate elements that is emoticons & keywords. There is graphical comparison between keyword and emoticons on subjectivity level, polarity level. This system integrates and automates all tasks associated with semi-supervised emotion detection.

Nasiya Najeeb [2]. This paper proposes different opinion mining techniques for various type of natural language processing for tracking mood of public about particular product or topic. Many peoples write opinions in forum, microblogging or review websites. This is useful for analysing the data for companies, governments, and individuals for tracking automatically feelings and attitudes. Social networks allow users to express their feeling and convey their emotions via text as well as emoticons. Information in terms of text are extracted ad clustered into emotions and then classified into positive, negative and neutral. An emoticon basically affects the sentence when it occurs because it also provides better sentiment expression of a web user. Text extraction is done manually or automatically after extraction the next step is filtration which includes emoticon replacements. After filtering the data, various classifiers are used to classify text based on emotions. The algorithm called as Word emotion technique is used to extract emotion form each word. Types of emoticons: Textual emoticons: - ":", "=","_", "," Graphical emoticons: - this provides better sentiments e.g. I feel very happy. (using phrase), feel very happy ☺. (using phrase as well as emoticon).

Fei Jiang [3]. A decision tree based user's context classifier and prediction model is designed to classify tweets according to emoticons expressed through the emoticons. The emoticons which are proposed by are used for mapping different emotions such as Love, Happiness, Pity, Furious, Heroic, Fearful, Disgust, Wonder and Peace [3]. These emotions are mandatory part of human nature that can be considered. The methodology for user's context personalization based on emoticons involves two major phases such as: a) Training Phase b) Testing Phase. Along with words, emoticons also extracted. Emoticons that are extracted from standard library. A decision tree is generated which is based on classifier and prediction model for performing emotion classification.

Alexander Hogenboom [4]. Nowadays people increasingly use emoticons to express their feelings or sentiments. People uses emoticons for products, services organizations, individual issues, events topics and their attributes through social media (Twitter). As twitter have text message limitations to 140 characters. So, people uses emoji's instead of big texts to express their sentiments. Emoticons are ASCII art emoticons are also called as smileys. Emoticons always adds essence for plain text and convey joy ☺, sadness ☹, laughter etc. To exploit or to understand emoji's in automated system first need to analyse that emoticons can typically relate to sentiments of the data in which they occur. They affect sentences or paragraphs. Some paragraphs contain only one emoticon which shows different sentiments, but in other paragraphs there are multiple use of emoticons so it will affect the sentence in which it occurs. Till now textual based sentiments were used in twitter. But now people uses emoticon to express their feelings so now it will be based on lexicon based sentiment analysis for emoticons.

Anthoniraj Amalanathan [5]. They proposed Emoticons Space Model (ESM) for sentiment analysis. In this paper the ESM technique treats each emoticons differently and also integrates that do not have clear emotional meaning. ESM simplifies emoticon signals and consistently performs previous state operations. ESM consist of two phase: -a) Projection Phase:-obtain co-ordinates of the posts are obtained based on coordinate of words. b) Classification Phase: -Use co-ordinate of the posts as features for supervised sentiment classification task.

Maryam Hasan [6]. In social media there many tools are widely used by personally to express their feeling and comments in the form of text message. Detection of emotions in plain text has a wide range of applications which includes human individual emotions and also public emotions of other people. They propose new approach in which classifying text messages automatically according their emotional states. There is one of the model studied that Circumflex Emotional model. This model characterized along two dimensions a) Valence b) Arousal. The Twitter messages are selected as input data set to the system and in that data set they provide a very large amount of available group of emotions. Main thing is they used

Volume 1, Issue 1 (2018)
Article No. 3
PP 1-5

supervised classifier for to detect classes of multiple emotions. In detection of emotions from the text messages there are some problem such as sparse and high dimensional feature vectors of messages. For the tackle of those problem they utilize Lexicon of emotions which uses these steps Designing and implementing a method that automatically label twitter messages based on the emotions of their authors, then Resolving the problem of high dimensional feature space in twitter dataset and lastly Achieving highest accuracy for classifying twitter messages based on their emotional states. The accuracy is compared with several machine learning algorithms and methods such as SVM, KNN, Decision Tree and Naive Bayes.

3. PROBLEM DEFINITION

The new generation of emoticons which are known as emoji's that is increasingly being used in mobile communications as well as in social media. For smaller number of popular emoticons, their sentiments are well known but there are thousands of emoji's so extracting their sentiments is challenging. The existing system for exploiting emoticons, classification of emoticons was based only on sentiment score and polarity but did not use Sentiment Rank and Bar which will provide better understanding of emoticons.

4. METHODOLOGY

The method proposed in this paper aims for automation of sentiment analysis for emoticons. It uses two main approaches first one is the emoji sentiment lexicon which calculates the sentiment score and the second one is emoji sentiment ranking which considers ranking and positions of the emoticons for sentiment analysis.Proposed system also gives graphical representation on sentiments after extracting emoticons. Fig 4.1: shows System flow diagram for proposed system.

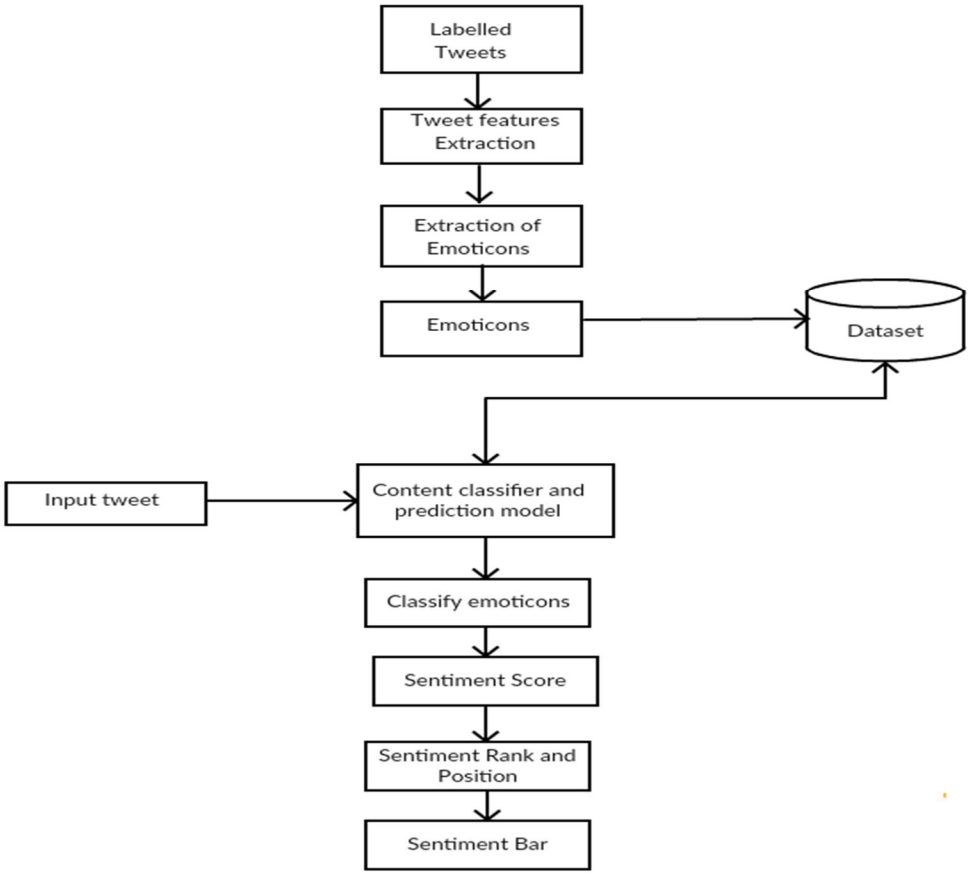

Fig 4.1: System flow diagram of proposed model.

Volume 1, Issue 1 (2018)
Article No. 3
PP 1-5

The sentiment of emoticons is evaluated from the sentiment of tweets. In the training phase initially labelling of the tweets will be done. Sentiment labels can take one of the three values which are negative, neutral and positive. A label 'c' is discrete three valued variables {-1, 0, +1} [5]. After this extraction of emoticons are done from the tweets by separating the text from emoticons. Then the emoticons are stored in the dataset for classification and further process. When the user enters the tweet in the text field of Twitter the content classifier is used for classifying the text and emoticons in that tweet with having access to the dataset. In classification, Emoji Sentiment Ranking uses the overall mapping of emojis. The position of an emoji is determined by sentiment score $\overline{S}$and neutrality p0. The sentiment score will be in the range of (-1, +1) and computed as p+ to p-. The positive emojis will be towards right hand side with green colour, negative ones towards left hand side with red colour and the neutral emojis with yellow colour.

The frequently used emoticons are given a higher rank and less frequently emoticons are given as lower rank while others are given as mid-point rank. Emojis, on the other hand, can appear in groups and also at the end of the tweets. In the sentiment distribution for the set of relevant tweets the system will find the number of occurrences of emoticons in the tweets, and also the sentiment label c by using discrete probability distribution formula. After finding Sentiment score and Rank based on position the next step is to form Sentiment Bar. The sentimental bar is a useful for proper visualization of the sentiments attached to an emoji. The sentimental bar will include all the properties of emoticons such as p-, p0, p+ and $\overline{S}$. The coloured bar extends from −1 to +1, which is the range of the sentiment score (red, yellow, green). The grey bar is centred at $\overline{S}$ and is extended, but never beyond the range of $\overline{S}$and gives the Sentiment Bar for analysis of different emoticons whether they are having positive, neutral or negative sentiment.

The Fig 4.2 shows the Sentimental Bar from extraction of emoticons. When user will enter the tweet in the text field of twitter, this tweet may consist of text and emoticons. It is necessary to extract the emoticons from tweet. The content classifier will extract the emoticons from the tweet and separates emoticons from text. For example, if user enters emoji with smile face as shown in Fig 4.2 the labelling of this emoji is done in the training phase with positive value. Then the sentiment score is assigned to it which is in the range of -1 and +1. Based on the occurrence and number of counts of this emoji, the ranking of this emoji is done by using probability distribution. Finally, a Sentimental Bar is generated for the emoji as shown below.

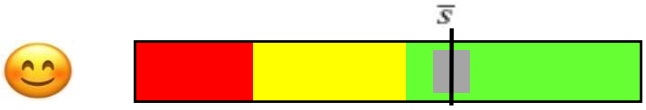

Fig 4.2: Sentimental Bar

5. EXPECTED RESULT

The occurrence and position of emoticons matters a lot for the prediction of sentiment analysis and hence the proposed system includes Sentiment Rank and Position Approach for better sentiment analysis. A graphical representation for sentiments of tweets in a form of Sentimental Bar for easy analysis of sentiments which represents in the form of red-negative, yellow neutral & green-positive. The Emoji Sentiment Ranking will be an important resource for helping humans during the representation process, or even for the automatically labelling of tweets with emojis for sentiment analysis.

6. CONCLUSIONS

This model describes the construction of an Emoji Sentiment Lexicon and the Emoji Sentiment Ranking for different emoticons in tweets based on their occurrence. The Emoji lexicon method can also be used for grouping the emoticons with a sentiment including the text. The Emoji Sentiment Lexicon and Emoji Sentiment Ranking approaches will be constructed for different emoticons used in the tweets based on their occurrence for predicting better sentiments. This approach has analysed and used the sentiment properties of the emojis in depth and also gives some interesting facts regarding emoticons. In future, it will be interesting to monitor and analyse how fast the usage of emojis are growing in communication, and whether textual communication will be replaced with different technique. Till now many researchers have focused on text based sentiment analysis but have not given much priority for emoticon sentiments so the proposed system focuses on emoticon sentiments and thus generates sentimental bar for it.

Volume 1, Issue 1 (2018)
Article No. 3
PP 1-5

REFERENCES

[1] S. Georgios, K. N. Vavliakis, and P. A. Mitkas, "Multilingual Sentiment analysis using emoticons and keywords", *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on. Vol. 2. IEEE*, 2014.

[2] N. Najeeb, and S. Jamal. "A Survey on Mining User Opinion from Texts and Emoticons."

[3] J. Fei, "Microblog sentiment analysis with emoticon space model", *Journal of Computer Science and Technology 30.5 (2015)*, pp. 1120-1129.

[4] A. Hogenboom, "Exploiting emoticons in sentiment analysis." *Proceedings of the 28th Annual ACM Symposium on Applied Computing, ACM*, 2013.

[5] Amalanathan, Anthoniraj, and S. Margret Anouncia. "Social network user's content personalization based on emoticons." *Indian Journal of Science and Technology 8.23 (2015)*.

[6] M. Hasan, E. Rundensteiner, and E. Agu. "Emotex: Detecting emotions in twitter messages." (2014).

[7] B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong. "Polarity Analysis of Texts using Discourse Structure". *20th ACM Conference on Information and Knowledge Management (CIKM 2014), pp. 1061-1070.Association for Computing Machinery, 2014*.

[8] Ip Amy. "The impact of emoticons on affect interpretation in instant messaging." *Retrieved April 26 (2002): 2014*.

[9] Zhou, Rui, Jasmine Hentschel, and Neha Kumar. "Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 2017*.

[10] Tchokni, SimoEditha, Diarmuid O. Séaghdha, and Daniele Quercia. "Emoticons and Phrases: Status Symbols in Social Media." ICWSM. 2014.

[11] N. N. Bazarova "Social sharing of emotions on Facebook: Channel differences, satisfaction, and replies." *Proceedings of the 18th ACM Conference on Computer Supported cooperative Work & Social Computing. ACM, 2015*.

[12] A. Sarlan, C. Nadam, and S. Basri. "Twitter sentiment analysis." *Information Technology and Multimedia (ICIMU), 2014 International Conference on. IEEE, 2014*.