Pipelined model for classification of the processed tweets

B. M. Bandgar

JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune-411033, India
*Corresponding author E-mail: baapuraob@yahoo.com

Abstract

The The extraction and processing of the real-time tweets is one of the challenging tasks. The processed tweets are classified using unstructured models. These showed no. of neutral tweets after the classification the tweets. Therefore, here pipelined unstructured model is developed the for the classification of tweets and explored to reduce the number of neutral tweets. It showed more reduction in the number of neutral tweets as compared to other unstructured model such as the SWNC model results.

Keywords: *Unstructured models, Classification of Tweets, Opinion Mining.*

1. Introduction

This The extraction and processing of the real-time tweets is one of the challenging tasks. Then further the classification of obtained meaning full messages can be done by different methods. The numbers of researchers have carried out related works. Machedon, R., Rand, W. and Joshi, Y.(4) classified social tweet messages into three categories; informative, persuasive and transformative. Farhan Hassan Khan et. al.(5) tried to improve the accuracy of text classification and resolve the data sparsity issues. Further the classification of the obtained meaningful messages is done by using the unstructured models such as Enhanced Emoticon Classification (EEC) model, Improved Polarity Classification (IPC) model and Sentiworldnet Classification (SWNC). These models result showed no. of the neutral tweets.

Therefore, here we tried to develop the pipelined Unstructured model for the classification of the tweets and explored to reduce number of neutral tweets.

2. Method Details

Initially, the real time tweets are extracted from the twitter website (3) on 18th May 2017 at 10.15am for the keyword Arvind Kejriwal; the extracted tweets were processed to obtain the meaningful messages. Further the obtained meaningful tweets are classified using EEC model. The output of positive and negative tweets is collected in separate file. The output of the neutral tweet is given as input to the IPC model and classified the neutral tweets. Again the outputs of positive and negative classified tweets by IPC model are added in the file of EEC model's resulted positive and negative tweets. Further, reaming neutral tweets from the IPC model are given as input to the SWNC model and classified the neutral tweets again. Finally, collected reaming neutral tweets, positive and negative tweets obtained by SWNC, IPC and EEC models as combine resulted outcome file. The outcome of above said data is shown graphically in Fig 1. Also the outcome of the SWNC model

is shown in Fig 3 Individually; SWNC model showed better result as compared to EEC, IPC models (Bandgar, 2016).
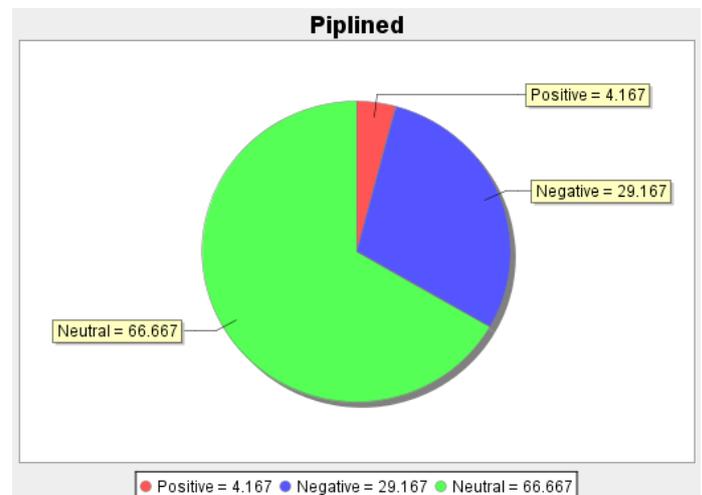

Fig. 1: Pipelined unstructured model classification % result in pie graph

The flow chart of Pipelined Unstructured model for tweets classification is shown in Fig 2.
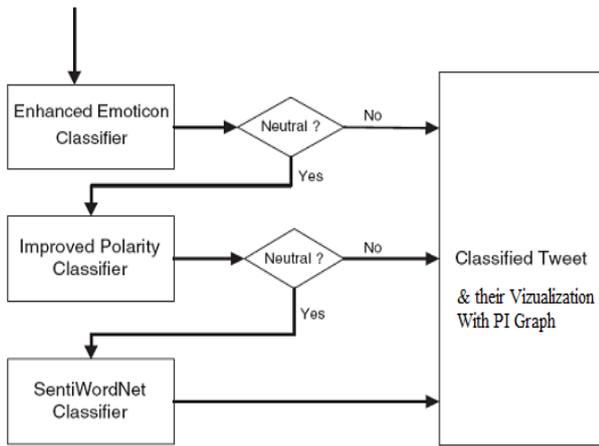
Fig. 2: Flow chart of Pipelined Unstructured model for classification tweets

3. Results and Discussion

The classification of the processed tweets is carried out using the Pipelined Unstructured model, which constituted of EEC, IPC and SWNC models. The results are shown in the % form of classification of Positive, Negative and Neutral tweets in pie graph. The result of the Pipelined Unstructured model is compared with the SWNC model obtained results because the Individually the SWNC model showed more better classification results over other models such as EEC, and IPC. The results were verified for the different real-time extracted tweets. One of the resulted data is shown in the Fig 1 and Fig 3.

By comparing the Fig 1 and Fig 3, it is observed that the no. of neutral tweet % is reduced by 20 to 30 % by Pipelined Unstructured model. Thus, the Pipelined Unstructured model gives the better results of the classification over the other unstructured model results.
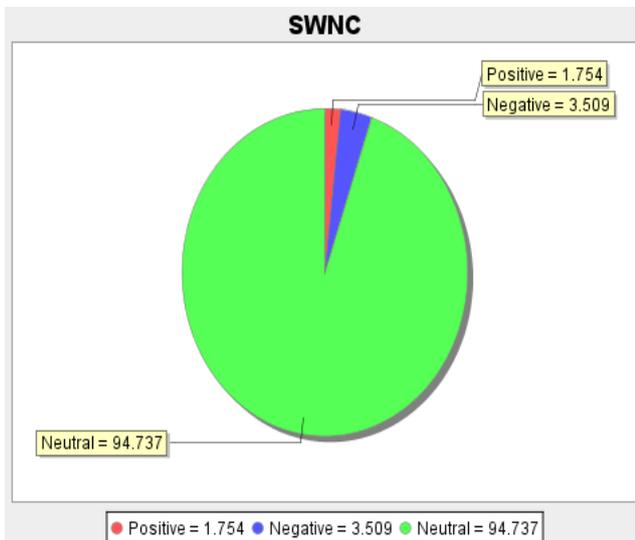
4. Conclusion

The processed tweets are classified using Pipelined Unstructured model constitutes of three unstructured models EEC, IPC and SWNC. It showed more accuracy of classification over the SWNC unstructured model results. The number of neutral tweets in classification of the tweets is reduced to 20-30 % by Pipelined Unstructured model. Thus, we classified tweets with more accurate % over the other machine learning algorithms(Nagy, 2012) and obtained the real time opinions of the people on the particular events/products etc. For this purpose we developed an indigenous windows based user friendly application in Java.

Acknowledgment

References

[1] Bandgar B. M. and Binod Kumar, *Real time extraction and processing of social tweets*, International Journal of Computer Science and Engineering, E-ISSN No.-2347-2693,, Vol.3 No.3, (2015) , pp. 1- 6.

[2] Bandgar B. M. and Dr. Sheeja Azije, *Analysis of real time social tweets for opinion mining*, International Journal of Applied Engineering Research ISSN No. 0973-4562 Vol. 11 No. 2, (2016), pp. 1404 - 1407.

[3] https://www.twitter.com

[4] R. Machedon, W. Rand, Y. Joshi, *Automatic Classification of Social Media Messaging using Multi-Dimensional Sentiment Analysis and Crowd sourcing,* http://dx.doi.org/10.2139/ssrn.2244353 (Available at SSRN: http://ssrn.com/abstract=2244353), 2013.

[5] Farhan Hassan Khan et. al. , *TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems*, (2013), http://dx.doi.org/10.1016/j.dss.2013.09.004