

Review of Effective Models for Sentiment Analysis

Mehul Awasthi

Department of Computer Science and Engineering, SRM University, Chennai, Tamil Nadu, India-603203

Email address: mehulawasthi7[AT]gmail.com

Abstract— The sentiment analysis of data has become rather important in today’s world where people share their opinions and views freely on social media platforms such as Twitter. Companies use various sentiment analysis models to analyse the data available on social media. However some of these models are unable to discern the contextual meanings of words in a sentence and can end up classifying erroneously. This paper looks to suggest ways to improve the sentiment analysis models and their accuracy of classification of tweets as negative or positive. One way is by combining different single models to overcome shortcomings of a single model [1]. The other is classification on the basis of features such as negation and conjunctions [2].

Keywords— Classification, Sentiment Analysis, Social Media, Twitter.

I. INTRODUCTION

Twitter is extremely popular amongst its users as it allows them to freely express their views and opinions on a public platform [3]. Millions of web users connected through social networks like Twitter, take to social networks to share their opinions on different aspects of life every day, thus making tweets rich sources of data containing diverse views and perspectives of a wide number of people from different walks of life. Topics of tweets can range from movies to politics to sports to stock markets and so on and so forth. This presents a goldmine for data analysts to work upon for valuable market insight or very simply, public sentiment. But the humungous volume of twitter feeds poses a problem for data analysts, who may be looking to gauge the general opinion (via sentiment analysis) of the public on a particular topic. Companies can use this insight to take important decisions regarding their product launches, advertisement campaigns, or future strategies. This analysis of people’s sentiments from textual data is called sentiment analysis and it has become very prevalent amongst major companies [4]. Seeing as to how many companies base their decisions on the analysis of such data, it is pertinent that the analysis of such data be accurate. To correctly gauge the sentiment of the public is what an ideal sentiment analysis model should do. The problem with the sentiment analysis models today is that most of them are not able to correctly identify the contextual meaning of words in a tweet. For example the sentence, ‘The country had a low poverty rate.’ may be classified as a negative sentence by some sentiment analysis models because of the presence of words such as ‘low’ and ‘poverty’, but in reality the sentence is a positive one.

This paper will first look at how a hybrid model could overcome this flaw, after which classification of a tweet on the basis of features such as negation and conjunction will be taken into consideration.

II. A HYBRID MODEL WITH THREE LAYERS

Usually, the conventional sentiment analysis methods make use of a sentiment analysis dictionary.

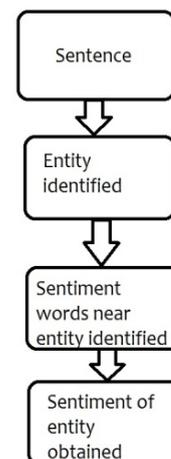


Fig. 1. Working of a traditional sentiment analysis model

The way this works is simple. First the method finds out the entities in the sentence. Next the sentiment words in the vicinity of these entities are found out. Finally, by the application of the sentiment dictionary, the sentiment of the entity is obtained. There is an inherent flaw with this method of sentiment analysis. It is highly possible that the sentiment word has no direct relation to the entities close by. On the other hand, they may be related to certain aspects of the entity. Take for example the sentence, ‘This house has large rooms and a low rent’. The words ‘large’ and ‘low’ are related to the words, ‘rooms’ and ‘rent’ respectively and not the entity ‘house’.

What the Three Layer model proposes to do is that, it will first extract the entities out and then the aspects related to that entity. Next, the sentiment words that describe the sentiment of the aspect will be found and finally, the sentiment of the entities will be found by combining sentiments of their aspects. This model can go a long way in solving problem of correctly identifying the contextual meaning of words in a sentence. For example, the model will be able to identify that ‘low’ in ‘low poverty’ expresses a positive fact, whereas in ‘low survival’ it expresses a negative fact. This is so because the model considers the sentiment words attached to the respective aspects.

This model could go a long way in providing a more accurate and precise sentiment analysis model, and must be given due consideration.

III. FEATURE BASED SENTIMENT ANALYSIS

Another model we can consider for more effective sentiment analysis is to create a model which deals with features such as negation and conjunctions. Let us take a look at how each one of these features can be dealt with, one by one.

A. Negation

First let us take up the case of negation handling. Take for example the sentences, 'The food is good' and 'The food is not good.' Traditional sentiment analysis models would classify both as positive opinions because of the presence of the word 'good' in both the sentences. However the second sentence is a negative opinion. This illustrates the problem of the traditional models very clearly. In order for a model to be accurate, it must consider the negation (if any) before classifying a tweet.

B. Conjunctions

It is normal for a single sentence to contain more than one opinion on something. For example the sentence, 'The food was good but the service was bad', expresses two opinions. One opinion is positive about the food, and the other is a negative opinion about the service. Traditional sentiment analysis models struggle to classify such sentences. In order to correctly classify such sentences for their sentiment, the sentences must be split before and after the conjunction and judged separately for their sentiment. Such a model would process the above example sentence as, 1)'The food is good.' and 2)'The service is bad'. The proposed model would thereby correctly classify the sentiment of such a sentence.

A model implementing the above would be more effective in correctly classifying sentences than the traditional sentiment

analysis models. If implemented well enough, the model could be a stepping stone towards more sophisticated models in the future.

IV. CONCLUSION

In this paper, two different models have been reviewed for sentiment analysis that are more effective and efficient than the traditional sentiment analysis models present today. One method is a hybrid three layer model that can overcome the problem of traditional sentiment analysis model-wherein they struggle to correctly classify sentences based on the contextual meaning of words. The other model suggested takes into consideration features such as negation and conjunctions for more effective and accurate classification and sentiment analysis. Both these methods have an immense scope for further improvement and in the future can be built upon to have an ideal sentiment analysis model which is accurate, precise and consistent

REFERENCES

- [1] K. Yang, Y. Cai, D. Huang, J. Li, Z. Zhou, and X. Lei, "An effective hybrid model for opinion mining and sentiment analysis," *IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017.
- [2] P. Sharma and N. Mishra, "Feature level sentiment analysis on movie reviews," *2nd International Conference on Next Generation Computing Technologies (NGCT)*, Dehradun, India, 14-16 October 2016.
- [3] H. Halpin and M. Tuffield, "A Standards-based, Open and Privacy-aware Social Web," W3C Incubator Group Report, 2010.
- [4] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analysing and comparing opinions on the web," in *Proceedings of the 14th International Conference on World Wide Web*, ACM, pp. 342-351, 2005.