Sentiment Analysis for Marketing and Product Review using a Hybrid Approach

Yash Sandansing[1], Tanay Rathod[2], Debarati Ghoshal[3]

[1]Studying BE INFT
[2]Studying BE INFT
[3]Asst. Prof. VIT

---***---

Abstract - *Sentiment Analysis has become a growing field with the advancement in technology. It is a powerful and uniquely humane technique but a primitive technology which is being applied through machines. Sentiment Analysis is primarily and majorly used for product review and/or marketing purposes but these technologies show very limited insight with very little accuracy. This is due to the use of a single technology at a time.i.e. A neural network or else an nltk. This paper aims to merge the two majorly used technologies i.e token based analysis and neural networks to not only classify this data but also gain further insight into the classified data and thus help a brand to analyse its data much better.*

Key Words: Machine Learning, RNN, LSTM, Sentiment, Lexicon, token.

1. INTRODUCTION

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.

[1] Since the birth of this coincides with the rise in use of social media, it became an ever evolving and impressive technology with a consistent data flow to help it grow[3]. This data flow came in the form of social media posts, forums, discussions, blogs, reviews etc. Sentiment Analysis has grown to be a major part studied in Natural Language Processing[4].

In recent years, emotion recognition in text has become more popular due to its vast potential applications in marketing, security, psychology, human-computer interaction, artificial intelligence, etc[4]. As humans we are driven by emotions to almost everything around us. We like/dislike something or someone because of how it makes us or used to make us feel at one point in our life.

[2] mentions that 87% of internet users are influenced in their purchase and decision by customer's review. So that, if an organization can catch up faster on what their customers think, it would be more beneficial to organize to react on time and come up with a good strategy to compete with their competitors. Not only that but if the organization has a deeper insight into it's customer's mind

and knows the extent to which the customer is satisfied, then it would help them get ahead of their competitors. Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches[5].

Knowledge based techniques classify text based on a predefined category on the basis of unambiguous affect words present in the phrase[6]. These emotions categorised as happy, sad, bored, afraid can be condensed into positive, negative and neutral for our purpose. Statistical methods include assigning a probable affinity to particular emotions[8].

According to [7], it uses some of the methods like SVMs, latent semantic analysis, neural networks, etc.

Hybrid approaches involve using a combination of techniques from both of these domains to try and achieve better results. The hybrid approach we aim to use is a Long Short Term Memory(LSTM)of a Recurrent Neural Network(RNN).i.e.LSTM-RNN alongside a powerful natural language toolkit(NLTK).

1.1 Theoretical Background

In this section, we provide an introduction to SA, followed by a brief review of the literature on information sharing, including the distinction between RNN and knowledge based methods examined in previous studies. We then identify research gaps we intend to fill in this study. Drawing on the theoretical background, we go on to build our proposed model.

There are primarily two basic approaches for the automatic extraction of sentiment, which are lexicon-based approach and machine-learning-based approach[11-15].

1.Lexicon-based approach

A lexicon based approach also known as token based analysis makes use of a predefined list of words i.e. a predefined dictionary where each word is assigned a predefined sentiment[13].The lexicon methods vary according to the context in which they were created and involve calculating orientation for a document from the semantic orientation of texts or phrases in the documents

[11]. Besides, a lexicon sentiment is to detect word-carrying opinion in the corpus and then to predict opinion expressed in the text.[12] shows the lexicon methods along with the basic paradigm which are:

i.Preprocess a phrase by removing punctuation.

ii.Initialize a total polarity score(s) as 0.

iii.Match individual tokens with words in the dictionary.

iv. If token is present and:

 a.positive, increase the value of s with a positive value, making s positive.

 b.negative, decrease the value of s with a negative value, making s negative.

v.Check the total polarity of post:

 a.if s>threshold, then phrase is positive.

 b.if s<threshold, then phrase is negative.

2. Machine-learning-based-approach

Machine learning has been widely used in various fields of sentiment analysis. ML methods for sentiment analysis can be classified into two types: supervised and unsupervised. [16]Every instance in any dataset used by machine learning algorithms is represented using the same set of features. The features may be continuous, categorical or binary. If instances are given with known labels (the corresponding correct outputs) then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabeled[16]. Although supervised algorithms are a good approach, they are not suitable for larger datasets[16]. A machine learning method involves preprocessing[17], processing, sequencing and vectorization of its data before passing it into the neural network.

Different approaches for sentiment analysis include ANN, SVM, RNN, CNN all with varying applications and accuracy.

1.2 Proposed model

The proposed model is a combination of both token based as well as machine learning method. The model used for token based analysis uses TextBlob and the one for machine learning uses RNN. In order to get data from twitter, we use it's API also known as Tweepy.

The proposed model initially consists of retrieving tweets from twitter using keywords relevant to the search. But since these tweets are in json format, they need to undergo some preprocessing in the form of cleaning out unwanted symbols, emoticons and links. This involves removing the id of the tweet, date, time, place and many more factors attached.

After the tweets are cleaned thoroughly, they are ready to be passed through the models.

The tweets will be passed through both the models separately(lexicon and rnn).

1.Lexicon model: The lexicon model uses a predefined API called TextBlob to classify the tweets and assign it a score.

These scores can be set to either binary or decimal (between -1 and +1) to denote the sentiment of the tweet.

2. RNN: The second model used is an LSTM-RNN model.

For this model to work, data needs to be converted into a sequence.

Then, that sequence is converted into a fixed sized vector.

These vectors are passed through the neural network which gives a binary output.

Here, like in hybrid methods for sentiment analysis, RNN is used to decide whether the tweet is positive, negative or neutral. Whereas, the lexicon method determines the score of the tweet after the classification.

So, this model not only categorizes the tweets but also helps to decide the power of the sentiment.i.e how strong is the individual's sentiment.

Conclusion

The sentiment of the general public towards a product can be known with less computational expense. The output can show progress over a certain time period.

REFERENCES

[1] Liu B. Sentiment analysis: mining opinions, sentiments, and emotions. The Cambridge University Press, 2015.

[2] A. K. Jose, N. Bhatia, and S. Krishna, "TwitterSentimentAnalysis". NationalInstituteof TechnologyCalicut, 2010.

[3]Lei Zhang,Shuai Wang, Bing Liu:"Deep Learning for Sentiment Analysis:A survey",arxiv, Cornell University, 2018.

[4]Vong Anh Ho, Duong Huynh-Cong Nguyen, Danh Hoang Nguyen, Linh Thi-Van Pham, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen: "Emotion Recognition for Vietnamese Social Media Text",arxiv.org, Cornell University, 2019.

[5]Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems.

[6] Ortony, Andrew; Clore, G; Collins, A (1988). The Cognitive Structure of Emotions

[7]Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics.arxiv.org, Cornell University.

[8]Stevenson, Ryan; Mikels, Joseph; James, Thomas (2007). "Characterization of the Affective Norms for English Words by Discrete Emotional Categories"

[9]"Sentiment Analysis of Twitter Data": Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, Department of Computer Science, Columbia University, New York, cs.columbia.

[10]Aliza Sarlan, Shuib Basri:"Twitter Sentiment Analysis",2014 International Conference on Information Technology and Multimedia (ICIMU).

[11] M.Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, " LexiconBased Methods for Sentiment Analysis," Association for Computational Linguistics, 2011.

[12] M. Annett, and G. Kondrak, "A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs," Conference on web search and web data mining (WSDM). University of Alberia: Department of Computing Science, 2009.

[13] P. Goncalves, F. Benevenuto, M. Araujo and M. Cha, "Comparing and Combining Sentiment Analysis Methods", 2013.

[14] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis:The Good the Bad and theOMG!", (Vol.5). International AAAI, 2011.

[15] S. Sharma, "Application of Support Vector Machines for Damage detection in Structure," Journal of Machine Learning Research, 2008.

[16]S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Overview paper.

[17]Ye Yuan, You Zhou, "Twitter Sentiment Analysis with Recursive Neural Networks".