International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **07** Issue **02 | 2023**

Sentiment Analysis on E-commerce website in Machine Learning

Shubham Kambli[1], Hitisha Patil[2], Rohit Jage [3], Shweta Shelke[4], Prof. M. Velladurai [5]
[1,2,3,4,5]Department of Computer EngineeringARMIET, Shahapur, India

ABSTRACT

Opinion information is very important for businesses and manufacturers. They often want to know in time what consumers and the public think of their products and services. However, it is not realistic to manually read every post on the website and extract useful viewpoint information from it. If you do it manually, there is too much data. Sentiment analysis allows large-scale processing of data in an efficient and cost-effective manner. In order to know more about sentiment analysis, this author explores the application of sentiment analysis on business to understand its strengths and limitations.
This paper used dataset of the Amazon Alexa reviews, and then built a model to predict the sentiment of the comment given the comment declaration by using Python and machine leaning algorithm- Naïve Bayes and logistic regression. SMOTE is used to cope with the unbalanced dataset and AUC/ROC are used to evaluate which method is best. Sentiment Analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors.

Keywords: Python , Pycharm, Django , Nodejs , Angular15

1. INTRODUCTION

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities. Internet is a resourceful place with respect to sentiment information. From a user's perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. For instance, Twitter currently has three different versions of APIs available, namely the REST API, the Search API, and the Streaming API. With the REST API, developers are able to gather status data and user information; the Search API allows developers to query specific Twitter content, whereas the Streaming API is able to collect Twitter content in real time. Moreover, developers can mix those APIs to create their own applications. Hence, sentiment analysis seems having a strong fundament with the support of massive online data.
Multiple studies about sentiment analysis on Amazon.com reviews have been done. These studies used conventional Machine Learning (ML) like Naive Bayesian (NB), Sup- port Vector Machine (SVM), decision trees or logistic regression, which resulted in relatively good performances ($accuracy > 0.90$).
A sentiment analysis of reviews of Amazon beauty products has been conducted in 2018 by a student from KTH [2] and he got accuracies that could reach more than 90% with the SVM and NB classifiers. He found that SVM was performing better than NB for a large amount of data.
He also focused on summaries of the reviews which are more informative and got higher accuracy than with the complete reviews. Xing Fang and Justin Zahn analyzed different categories of Amazon products (beauty, book, electronic, and home) with 3 different classifiers: NB, SVM and Random Forest. They reached the conclusion that Random Forest usually provided them with more accurate results. They also found that SVM was performing better than NB for larger data sets.

2. LITERATURE SURVEY

One fundamental problem in sentiment analysis is categorization of sentiment polarity. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral). Based on the scope of the text, there are three levels of sentiment polarity categorization, namely the document level, the sentence level, and the entity and aspect level. The document level concerns whether a document, as a whole, expresses negative or positive sentiment, while the sentence level deals with each sentence's sentiment categorization; The entity and aspect level then targets on what exactly people like or dislike from their opinions. Since reviews of much work on sentiment analysis have already been included in, in this section, we will only review some previous work, upon which our research is essentially based. Both lists also include some misspelled words that are frequently present in social media content. Sentiment categorization is essentially a classification problem, where features that contain opinions or sentiment information should be identified before the classification. For feature selection, Pang and Lee suggested to remove objective sentences by extracting subjective ones. They proposed a text-categorization technique that is able to identify subjective content using minimum cut. Gann et al. selected 6,799 tokens based on Twitter data, where each token is assigned a sentiment score, namely TSI (Total Sentiment Index), featuring itself as a positive token or a negative token. The aim of this project is to investigate if sentimental analysis is feasible for the classification of product reviews from Amazon.com. Therefore, we will compare the performance of different classification algorithms on the binary classification (positive vs. negative) of product reviews from Amazon.com.

Thereby, we want to investigate whether the category of products the reviews come from influence the performance of this classification. Once found the best performing classifier, it will be applied on new Amazon.com datasets containing reviews of different product categories and these results will be compared. This paper selects dataset "Amazon Alexa Reviews" from Amazon and would like to do a sentimental analysis about the Amazon Alexa reviews by Python. The author will detect the positive and negative feedback and visualize the contents by word cloud and predict the ratings from the reviews by Naïve Bayes and Logistic regression model, and then compare the accuracy of different machine learning algorithms.

- We have created deep learning model with RNN method and used LSTM layer for it.
- We have used total 20k reviews (10k positive, 10k negative) and there are total 39k+ unique words in it.
- Converted word vector by using glove.6B.100d.txt file.
- Training done on 70:30 rule (70% training, 30% testing).

The raw review data is cleaned for different elements, which could worsen the performance of the classifier. This is done by:

- Removal of HTML tags.
- Filtering every symbol except for letters (a-z) and numbers (0-9).
- Filtering out every word with length of 3 symbols or lower.

In Deep Learning, Recurrent Neural Networks (RNN) are a family of neural networks that excels in learning from sequential data. A class of RNN that has found practical applications is Long Short-Term Memory (LSTM) because it is robust against the problems of long-term dependency.

Data which contains 20k mixed reviews including 39k+ unique words.In LSTM Words and punctuation marks are both considered symbols.

Function for building the dictionary and reverse dictionary. The model with a 512-unit LSTM cell

Constants and training parameters, Symbols to vector of int as input, One-hot vector as label, Training step optimization, Sample prediction and accuracy data per training sub-session, Loss and optimizer

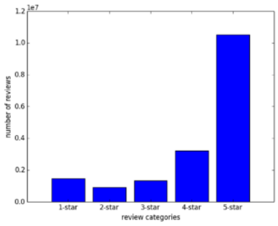
Improved LSTM

For Testing:

- We gave a sample review and tested with accuracy of 85.36.
- We enhanced the algorithm to take a single line review as well.

Data collection Data used in this paper is a set of product reviews collected from amazon.com. From February to April 2014, we collected, in total, over 5.1 millions of product reviewsb in which the products belong to 4 major categories: beauty, book, electronic, and home (Figure 3(a)). Those online reviews were posted by over 3.2 millions of reviewers (customers) towards 20,062 products. Each review includes the following information: 1) reviewer ID; 2) product ID; 3) rating; 4) time of the review; 5) helpfulness; 6) review text.

Every rating is based on a 5-star scale, resulting all the ratings to be ranged from 1-star to 5-star with no existence of a half-star or a quarter-star. Sentiment sentences extraction and POS tagging It is suggested by Pang and Lee that all objective content should be removed for sentiment analysis. Instead of removing objective content, in our study, all subjective content was extracted for future analysis. The subjective content consists of all sentiment sentences. A sentiment sentence is the one that contains, at least, one positive or negative word. All of the sentences were firstly tokenized into separated English words. Every word of a sentence has its syntactic role that defines how the word is used. The syntactic roles are also known as the parts of speech. There are 8 parts of speech in English: the verb, the noun, the pronoun, the adjective, the adverb, the preposition, the conjunction, and the interjection. In natural language processing, part-of-speech (POS) taggers [have been developed to classify words based.



3. PROPOSED WORK

A sentiment token is a word or a phrase that conveys sentiment. Given those sentiment words proposed in [27], a word token consists of a positive (negative) word and its part of-speech tag. In total, we selected 11,478 word tokens with each of them that occurs at least 30 times throughout the dataset. For phrase tokens, 3,023 phrases were selected of the 21,586 identified sentiment phrases, which each of the 3,023 phrases also has an occurrence that is no less than 30. sentiment score should fall into the interval of For positive word tokens, we expect that the median of their sentiment scores should exceed 3, which is the point of being neutral . For negative word tokens, it is to expect that the median should be less than the sentiment score information for positive word The histogram chart describes the distribution of scores while the box-plot chart shows that the median is Similarly, the box-plot chart in shows that the median of sentiment scores for negative word tokens is lower than 3 In fact, both the mean and the median of positive word tokens do exceed 3, and both values are lower than 3, for negative word tokens . The ground truth labels The process of sentiment polarity categorization is twofold: sentence-level categorization and review-level categorization. Given a sentence, the goal of sentence-level categorization is to classify it as positive or negative in terms of the sentiment that it conveys. Training data for this categorization process require ground truth tags, indicating the positiveness or negativeness of a given sentence. However, ground truth tagging becomes a really challenging problem, due to the amount of data that we have. Since manually tagging each sentence is infeasible, a machine tagging approach is then adopted as a solution. The approach implements a bag-of-word model that simply counts the appearance of positive or negative (word) tokens for every sentence. If there are more positive tokens than negative ones, the sentence will be tagged as positive, and vice versa. This approach is similar to the one used for tagging the Sentiment 140 Tweet Corpus. This approach is similar to the one used for tagging the Sentiment 140 Tweet Corpus. Training data for review-level categorization already have ground truth tags, which are the star-scaled ratings. Table 3 Statistical information for word tokens Token Type Mean Media

4. RESEARCH AND METHODOLOGY

Data collection Data used in this paper is a set of product reviews collected from amazon.com. From February to April 2014, we collected, in total, over 5.1 millions of product reviewsb in which the Data collection Data used in this paper is a set of product reviews collected from amazon.com. From February to April 2014, we collected, in total, over 5.1 millions of product reviewsb in which the products belong to 4 major categories: beauty, book, electronic, and home (Figure 3(a)). Those online reviews were posted by over 3.2 millions of reviewers (customers) towards 20,062 products. Each review includes the following information: 1) reviewer ID; 2) product ID; 3) rating; 4) time of the review; 5) helpfulness; 6) review text. Fang and Zhan Journal of Big Data (2015) 2:5 Page 4 of 14 Figure 3 Data collection (a) Data based on product categories (b) Data based on review categories. Every rating is based on a 5-star scale(Figure 3(b)), resulting all the ratings to be ranged from 1-star to 5-star with no existence of a half-star or a quarter-star. Sentiment sentences extraction and POS tagging It is suggested by Pang and Lee [5] that all objective content should be removed for sentiment analysis. Instead of removing objective content, in our study, all subjective content was extracted for future analysis. The subjective content consists of all sentiment sentences. A sentiment sentence is the one that contains, at least, one positive or negative word. All of the sentences were firstly tokenized into separated English words. Every word of a sentence has its syntactic role that defines how the word is used. The syntactic roles are also known as the parts of speech. There are 8 parts of speech in English: the verb, the noun, the pronoun, the adjective, the adverb, the preposition, the conjunction, and the interjection. In natural language processing, part-of-speech (POS) taggers [29-31] have been developed to classify words based on their parts of speech.

A POS tagger can also be used to distinguish words that can be used in different parts of speech. For instance, as a verb, “enhanced" may conduct different amount of sentiment as being of an adjective. The POS tagger used for this research is a max-entropy POS tagger developed for the Penn Treebank Project [31]. The tagger is able to provide 46 different tags indicating that it can identify more detailed syntactic roles than only 8. As an example, Table 1 is a list of all tags for verbs that has been included in the POS tagger. Each sentence was then tagged using the POS tagger. Given the enormous amount of sentences, a Python program that is able to run in parallel was written in order to improve the speed of tagging. As a result, there are over 25 million adjectives, over 22 million adverbs, and over 56 million verbs tagged out of all the sentiment sentences, because adjectives, adverbs, and verbs.

5. CONCLUSION

As the results on the test data shows, LSTM networks are the most suitable for binary sentiment analysis on Amazon.com product reviews. Based on the results on the evaluation datasets, we can conclude that LSTM performs very well (*accuracy* > 0.90) for binary classification, and that does not depend strongly on the type of product where the reviews come from. As it can be seen clearly from confusion matrices in Figure 2, the LSTM network both performs accurate results for positive and negative classes. Since the training dataset is also balanced, getting balanced results from both classes shows the model's reliability.

In conclusion, LSTM networks are very suitable for classification of the sentiment on product reviews. These results do not depend on the type of product where the reviews are given to. Since we get 90% accuracy with just training 1% of data, we can say that sentimental analysis methods are feasible with Amazon review data.

6. ACKNOWLEDGEMENT

We have immense pleasure in presenting the paper for our project entitled "Congestion Control System Using Machine Learning" (For Road).

We would like to take this opportunity to express our gratitude to a number of people who have been sources of help &encouragement during the course of this project.

We are very grateful and indebted to our project guide and our respected HOD for providing their enduring patience, guidance & invaluable suggestions. They were the one who never let our morale down & always supported us through our thick & thin. They were the constant source of inspiration for us & took utmost interest in our project.

We would also like to thank all the staff members for their invaluable co-operation & permitting us to work in the computer lab.

We are also thankful to all the students for giving us their useful advice & immense co-operation. Their support made the working of this project very pleasant.

7 REFERENCES

[1] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," 2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018, no. May, pp. 1–6, 2018.

[2] S. Paknejad, "Sentiment classification on Amazon reviews using machine learning approaches," 2018.

[3] X. Fang and J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, vol. 2, no. 1, 2015. [Online]. Available: http://dx.doi.org/10.1186/s40537- 015-0015-2

[4] Z. Zhou and L. Xu, "Amazon Food Review Classification using Deep Learning and Recommender System," Stanford University, pp. 1–7, 2009. [Online]. Available: https://cs224d.stanford.edu/reports/ZhouXu.pdf

[5] J. Nowak, A. Taspinar, and R. Scherer, "LSTM recurrent neural networks for short text and sentiment classification," Lecture Notes in Computer Science (including sub- series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10246 LNAI, pp. 553–562, 2017.

[6] "Amazon Reviews for Sentiment Analysis." [Online]. Available: https://www.kaggle.com/bittlingmayer/amazonreviews#train.ft.txt.bz2