Sentiment Analysis An Intelligent User Interface

Shubham Girdhar(1323003)
shubham.girdhar@stud.fra-uas.de
Frankfurt university of applied sciences
Frankfurt, Hessen

Gaurav Kapadiya(1319237)
gaurav.kapadiya@stud.fra-uas.de
Frankfurt university of applied sciences
Frankfurt, Hessen

kshitij Yelpale(1322509)
kshitij.yelpale@stud.fra-uas.de
Frankfurt university of applied sciences
Frankfurt, Hessen

Pranay Raman(1321759)
pranay.raman@stud.fra-uas.de
Frankfurt university of applied sciences
Frankfurt, Hessen

Shaikh Safir Mohammad(1322554)
safir.shaikh@stud.fra-uas.de
Frankfurt university of applied sciences
Frankfurt, Hessen

Abstract

In this era of technology sentiment classification is an interesting and major topic in the field of Natural Language Processing (NLP). Data-driven methods, including machine learning and deep learning techniques provide effective solutions for the sentiment classification problem. While there has been a lot of research on various machine learning models to improve the accuracy of the task, there has been no comparison, based on statistical measures, between the best of models, particularly for longer texts. Therefore, we reconstruct the classification task for long reviews as a comparison problem and compare two models known to work well, Long Short-Term Memory (LSTM) and Naive Bayes. The study is constructed as a Human-Computer-Interaction (HCI) experiment for movie reviews and uses user feedback to measure the performance of both models. Analyzing the data collected, we show that there is no notable difference in performance of the LSTM and Naïve Bayes models for shorter reviews. However, for longer reviews LSTMs give significantly better results.

Keywords: sentiment analysis, LSTM, Naive Bayes

ACM Reference Format:
Shubham Girdhar(1323003), Gaurav Kapadiya(1319237), kshitij Yelpale(1322509), Pranay Raman(1321759), and Shaikh Safir Mohammad(1322554). 2020. Sentiment Analysis An Intelligent User Interface. In *Proceedings of* . ACM, New York, NY, USA, 2 pages.

1 Introduction

In the new time period, Internet innovation has incredibly impacted human lives . There are audit remarks accessible for each service or item that exists in our every day life on the Internet. The opinions remembered for the audit remarks become progressively significant for vendors to break down and extend their business . On one hand, client audits mirror the quality and issues of items, which is significant for the vendors. One the other hand, the number review reports can be enormous and from alternate points of view, it is hard to finish up the survey reports physically. As a traditional research subject in the field of natural language processing (NLP), sentiments analysis targets mining clients' mentalities and observations (for example, positive, negative or neutral) from text descriptions created by client awareness . In this manner, sentiment analysis is additionally called opinion mining and perspective investigation, which has been generally concentrated in web based business sites, social networks and different fields [1]. Sentiment analysis is an examination field to analyse people's subjective sentiments. for example, feelings, assessments, conclusions and mentalities towards items, administrations, associations, occasions, subjects and their characteristics. Text sentiment analysis task is one of the most significant assignments in the field of natural language processing. It is discovered that the neural system model which depends on Long Short-Term Memory (LSTM) can effectively extract the context relations of sentences[2]. The classification exactness of the model can be improved by presenting Attention Mechanism with filter irrelevant data.

we know from the previous work that Since opinion analysis are a piece of the data mining that can watch public educe about different services and items[3]. It is likewise the stem of natural language processing, text analysis, computational linguistics, bio -measurements, AI strategies. Twitter has a basic length of the tweets as 140 characters. Tweets have positive, negative, neutral (neither positive nor negative) which helps customers to choose better service airlines .The Airline industry is critical in customer reviews. Traditional methods provide inconsistent information about data to the airline industry.[4] The solution could be the simple in the term of perfect result. For example we can easily got to know that what we want to choose either positive , negative or

, ,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

neutral service or item. Solution will be perfect to helping us to take decision about service or item. LSTM is a special type of recurrent neural network , has achieved great success in solving many problems. LSTM network introduces a self-cycling mechanism that makes it easier to learn long-term dependent information than a simple cycling structure.[2] A basic task in sentiment analysis is classifying the polarity of a given text whether the expressed opinion is positive, negative, or neutral. More Advanced, “beyond polarity” sentiment classification looks, for instance, at emotional states such as “angry”, “sad”, and “happy”. We always like when YouTube recommends our favourite videos automatically, or when the Restaurant works on our reviews and facilitates with better results next time we visit. Until and Unless your customers are happy and interestingly interacting, your business is up and running. The User Comments/Reviews contain crucial information which cannot be ignored. By analysing the User Sentiments, We can Improve the User Experience by providing better Results. This makes users happy and there is more human machine interaction. Firstly we want to Build a generic Machine Learning model for text classification of English language user sentiments. We can also give one name to that model for example “emotion model” The Model accurately extracts people’s opinions from a large number of unstructured review texts and classifies them into sentiment classes, positive, neutral or negative! (Model Training).We would like to use top Classification techniques such as Deep Neural Network using CNN, LSTM etc. and compare the results. LSTM is better in analyzing emotion of long sentences and can produce better accuracy rate and recall rate. It can get complete sequence and information effectively.so, we could like to use LSTM and Bayes for our classification problem. This Model can then be further used in multiple specific applications like Recommender Systems, Hotel/Restaurant Review Analysis etc.

2 Related work

2.1 Multi task classification

As a significant subject in the NLP field, sentiment classification consistently is an intresting and vital exploration point. With the expanding popularity of deep learning, an expanding number of NLP assignments depend on word embeddings. In 2013, Google opened word2vec, a tool for computing word vectors. It can map words to low-dimensional vector space, reduce the expense of figuring, and make words to be identified with one another as opposed to free. It is a decent distributed representation strategy [1], which defeats the defect of a one-hot encoding discrete representation. The motivation behind the sentiment analysis task is to acquire the sentiment polarity contained in the sentence, while the sentence data contained in the word is not finished, so the sentence encoder is expected to remove the highlights of the sentence, in order to create the vector presentation of the sentence. Convolutional neural system (CNN) was initially applied in the field of the picture, yet it has additionally been broadly applied in the field of NLP as of late and has accomplished great outcomes[1]. The system structure of deficient association and weight sharing lessens the multifaceted nature of the model. The presence of a convolutional layer makes CNN great at catching spatial nearby relationship, and the pooling layer in the system can extraordinarily decrease the calculation[1]. The presence of word2vec makes it workable for CNN to apply to the content, empowering CNN to acquire nearby data in the content and direct feeling arrangement. Comparative works use CNN extensions to characterize sentiment and finished multi-class classification tasks with great outcomes . In theory, the RNN language model could cover the time request structure of the entire content, and manage long-term dependence issue. in the practice, however, RNN could not get familiar with the information effectively[5]. At the point when the interval between the overall data of writings and the current location to be anticipated turns out to be huge, a few issues will come out. As there are too many unfold layers in the back propagation through time optimization algorithm(BPTT), which prompts history data misfortune and angle weakening while at the same time preparing[5]. To defeat this trouble, a few researchers set forward a procedure named Long Short-Term Memory (LSTM), which prompts better exploratory outcomes in some application situations[5].

2.2 Word embeddings

Great word installing is the basis of the characteristic language handling task. As far as improving the quality of word vectors, Mikolov et al. [6] and Pennington [7] first found that the word vectors learned through a RNN have intresting direct substructers conducted an exhaustive preparing of the worldwide word-word co-occurrence of measurable information from the corpus and the subsequent worldwide vector (GloVe) shows intresting straight base with regards to word vector space like [8]. Maas et al. [9] took in the word embeddings with conclusion dependent on the old style neuro-probabilistic language model [9]. Tang et al. [10] proposed three models, consid-ering the passionate propensity of text, and learning the word embeddings with feeling. In their analyses, they likewise utilize the GloVe to get introductory word embeddings. they first use assumption vocabulary as the additional data to pretrain a word sentiment classifier and afterward get the sentiment embending of each word including the words not in sentiment dictionary. Concatenating the word inserting and its sentiment embedding can improve the nature of word embeddings.

References