International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Sentiment Analysis using RNN

Yashas Hirethota[1], Shreesha HK[2], Sharanbasappa JB[3], Dr. Radhika K R[4]

Department of Information Science and Engineering, B.M.S College of Engineering, Bengaluru, Karnataka, India

Abstract: ***Sentiment analysis is a critical component of natural language comprehension with numerous real-world applications. Typical sentiment analysis attempts to predict whether a text is positive or negative. The given text has only one aspect and polarity, so this task works. Predicting the elements mentioned in a phrase, as well as the emotions connected with each of them, is a more generic and difficult assignment. Aspect-based sentiment analysis is a generic term for this activity***.

Keywords: RNN, Word2Vec, Sentiment Analysis, Data Pre-processing, LSTM, Layers, Test, Train, Score

1. Introduction

Sentiment analysis is a set of computational and natural language processing-based approaches for identifying, extracting, and characterizing subjective information represented in a text, such as opinions. The fundamental goal of sentiment analysis is to categorize a writer's attitude toward diverse issues into three categories: positive, negative, and neutral. Sentiment analysis has a wide range of applications in a variety of fields, including corporate intelligence, politics, sociology, and more.

In recent years, however, social networking websites, microblogs, wikis, and Web applications have emerged, resulting in an unprecedented increase in user-generated data that is ready for sentiment mining. Web postings, Tweets, videos, and other data that all reflect opinions on a variety of topics and events provide a wealth of opportunity to study and analyse human sentiment and opinions.

Sentiment analysis is a set of computational and natural language processing-based approaches for identifying, extracting, and characterizing subjective information represented in a text, such as opinions. The fundamental goal of sentiment analysis is to categorize a writer's attitude toward diverse issues into three categories: positive, negative, and neutral. Sentiment analysis has a wide range of applications in a variety of fields, including corporate intelligence, politics, sociology, and more.

In recent years, however, social networking websites, microblogs, wikis, and Web applications have emerged, resulting in an unprecedented increase in user-generated data that is ready for sentiment mining. Web postings, Tweets, videos, and other data that all reflect opinions on a variety of topics and events provide a wealth of opportunity to study and analyse human sentiment and opinions.

Human language is a complicated beast. It's tough to teach a machine to recognise the different linguistic nuances, cultural variances, slang, and misspellings that appear in internet discussions. It's considerably more difficult to teach a machine to recognise how context affects tone.

Sentiment analysis is particularly valuable in social media monitoring since it helps us to see how the general population feels about a given issue. Thanks to real-time monitoring features, social media monitoring platforms like Brand watch Analytics make this process faster and easier than ever before.

2. Motivation

Sentiments of users that are expressed on the web have great influence on the readers, product vendors. The advent of social networks has opened the possibility of having access to massive blogs, recommendations and reviews. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking one's friends and family for opinions because there are many user reviews and discussions about the product in public forums on the Web.

For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available. The availability of accurate data and tools for cleaning data, has made sure a lot of progress in the sentimental analysis field, making this one of the most researched fields.

3. Aim and Objective

This project proposes a deep learning-based model for sentiment analysis using twitter dataset. The model is able to detect sentiments from the content of the tweets.

- Develop an analysis method that helps in capturing the emotion of tweet,
- The data is cleaned and unnecessary information is removed using various kinds of methods.
- Design the solution using Machine Learning or Deep learning algorithms so that we achieve approximately 90% accuracy using a good dataset.
- It is proposed to be implemented using the RNN algorithm.
- Provide connection to the NLP and the ML / DL algorithm.

Scope:

We see a lot of scope in this field as Machine Learning is used in almost every industry these days. Some of the recent and famous applications of sentiment analysis:

- Analysing sentiments on social media.
- Spam detection in social media.
- Influence on social media.
- Reviews of new products launched, and many more.

4. Existing System

A. Lexicon-based sentiment analysis

One of the two primary ways to sentiment analysis is the use of a lexicon, which involves estimating sentiment based on the semantic orientation of words or phrases in a text. This method necessitates the creation of a dictionary of positive and negative words, each having a positive or negative emotion value ascribed to it. There have been various techniques to building dictionaries, including manual and automatic approaches. A chunk of text communication is typically represented as a bag of words in lexicon-based techniques. Following this representation of the message, all positive and negative words or phrases inside the message are assigned sentiment values from the lexicon. To make the final prediction about the overall sentiment for the message, a combining function, such as sum or average, is used. Aspects of a word's local context, such as negation or intensification, are frequently taken into account in addition to its sentiment value.

Machine learning algorithms have a major drawback in that they rely on tagged data. It's exceedingly tough to ensure that enough data is collected and labelled appropriately. Aside from that, the fact that a lexicon-based approach is easier for a human to understand and modify is seen as a key benefit for our job. We found that creating an adequate lexicon was easier than collecting and labelling relevant corpora. Given that the data gathered from social media is generated by users from all over the world, the algorithm's ability to handle solely the English language is limited. As a result, the sentiment analysis technique should be easier to translate into many languages.

B. Machine learning Techniques

Social networking sites make their data available on the internet in an easy and unrestricted manner. This abundance of data piques the attention of young researchers who want to pursue a career in sentiment analysis. On social media discussion boards, people express their emotions and perspectives [6]. Researchers are hired by businesses to look into the facts that aren't widely known about their products and services. Multinational corporations are most concerned with the automatic and spontaneous determination of sentiments from reviews. Machine learning approaches have increased sentiment analysis accuracy and speed up autonomous data evaluation in recent years.

C. *Naïve Bayes*

The positive or negative orientation of a text author determines the sentiment dichotomy. The supervised classifier Naive Bayes provides a mechanism to communicate positive, negative, and neutral feelings in web content. To categorize words into their appropriate categories, the Naive Bayes classifier uses conditional probability. When it comes to text classification, the advantage of utilizing Naive Bayes is that it only requires a minimal dataset for training. Pre-processing of online data removes numeric, foreign words, html elements, and special symbols, resulting in a set of words. Human specialists manually tag words with classifications like good, negative, and neutral. As a result of this pre-processing, word-category pairings are generated. For the training set, this pre-processing generates word-category pairings. Consider the word 'y' from the test set (unlabelled word set) and a document window of n-words (x1, x2,.....xn). Given a data point 'y,' the conditional probability that it belongs to the category of n-words from the training set is:

$$P(y/x1,x2,\cdots\cdots xn)=P(y)\times\prod i=1nP(xi/y)$$

Consider the following example of a movie review for the film "Exposed." The following are the outcomes of the Naive Bayes experiment. Experts who are human.

D. RECURRENT NEURAL NETWORK (RNN)

RNN has proven popular in SC. When contrasted to the CNN, the RNN model offers two distinct advantages. For starters, each layer of CNN has different parameters, but each step of RNN has the same parameters (i.e., it reduces the number of parameters needed to learn). Second, because the output of one step is dependent on the output of the previous stage, RNN requires a large amount of memory. As a result, when compared to CNN, RNN is better at processing sequential data.

As a result, RNN is a reliable network architecture for sequential data processing. It supports cyclical connections and reuses weights over multiple instances of neurons, each with its own set of time steps. This concept can explicitly assist the network in learning the whole history (i.e. current states) of past states. RNN converts an arbitrary length sequence to a fixed length vector using this attribute.

The average of the hidden states of all words is utilized as a feature for classification in LSTM, which treats the entire document as a single sequence. Because the LSTM cannot extract the aspect information, it performs poorly. The LSTM model outperforms the CNN model in most cases. Wang et al proposed using LSTM to encode complete tweets, with the hidden state utilized to predict sentiment polarity. With sentiment lexicons, negation words, and intensity words, Qian et al presented linguistically regularized LSTMs for SA

5. Proposed System

We aim to develop, train, and analyse a deep learning model which consists of RNNs and/or LSTMs which is capable of analysing and predicting Sentiment, with high accuracy and precision. Specifically, we intend to build a model that is able to detect the actual Sentiment of the sentence.

We also aim to make the model robust, and tolerable to a fair amount of noise, if any, present in the user input.

a) *Long Short-Term Memory (LSTM)*

LSTM is a special category of RNN that possesses the capability to capture long-term dependencies and their selective remembering property which enables them to focus only on the important parts for prediction.

A very basic LSTM module consists of a cell state and three gates that enable selective remembering by determining whether information to learn, unlearn, or retain. The cell state is employed to ensure that the information flow is

uninterrupted by only a few linear exchanges. A forget gate, an input gate, and an output gate are all included in each unit.

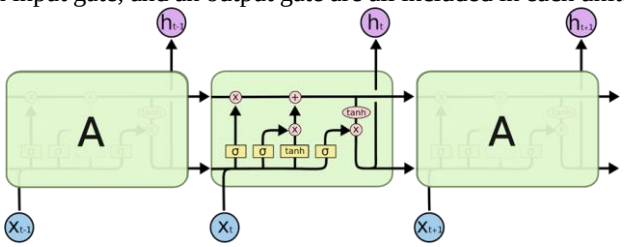

Figure 1: LSTM Cell

LSTM has memory and can store the information from previous timesteps which is how it efficiently learns the network With a few modifications, the model can be made bi-directional to capture the future and past context for each word which better helps understand the importance of each information unit. It captures the long-term dependency in any given information. Exploits the sequential nature of data such as speech which means that no two words are randomly placed next to each other, they occurring together define some relationship between them which might be important for context extraction.

6. Implementation

A. Dataset:

Dataset used for Training and testing the model is sourced from the Twitter data of 1.6 million tweets, available from Kaggle.

B. Main Data Structures used:

1) Arrays, such as NumPy arrays, to store the converted images.
2) Pandas dataframe to read write and create CSV file for the images
3) Dictionaries, to store the key-value pairs.

Data Processing- Preprocessing the data is the procedure for preparing raw data for use in deep learning model. It's the first and most important age in building a machine learning model. In this step we will create a CSV file with the location of the files for training and test photos as well as their associate class if any, so that they can be readily traced

C. Model architecture

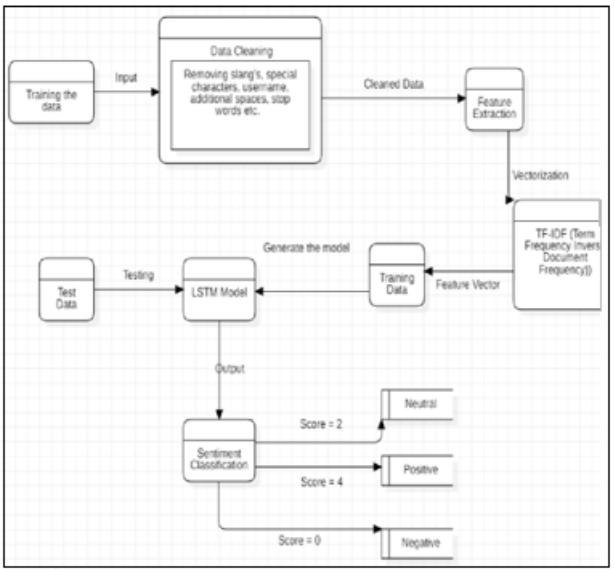

Figure 2: System Design

In this system design, we have shown the architecture of our implementation. From the dataset, we will split the train and test values (in our case 80:20) and the Data preprocessing takes place where the data cleaning happens (removing slangs, special character etc.) and the cleaned data goes to the feature extraction (using TF-IDF) and we get the Feature vector to the training data to pass to the LSTM Model where the training and testing happens to get the classification / Prediction model to get the sentiment scores.

7. Results

The model was trained using the above-mentioned dataset, and was evaluated rigorously. The variation of the training and validation accuracy can be seen below:

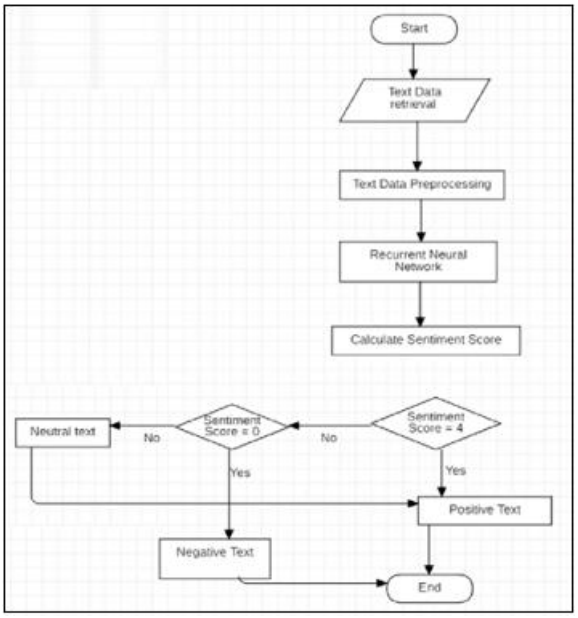

Figure 3: Module Design

In this module design, we see the model in depth. Where after the model is run and text data is received and preprocessing (cleaning) is done and when the sentences are taken to the RNN Model to get the sentiment scores.

When the sentiment score is 4, we take it as a positive text and if the score is not 4 then we check if the score is 0. If the score is 0 then we take it as a negative text, and the score is not 4 nor 0 we take it as a neutral text where we consider it as a positive text as there are no major negative words.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 300, 300)	87195900
dropout (Dropout)	(None, 300, 300)	0
lstm (LSTM)	(None, 100)	160400
dense (Dense)	(None, 1)	101

Total params: 87,356,401
Trainable params: 160,501
Non-trainable params: 87,195,900

Figure 4: Model Architecture

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.80	0.80	0.80	160046
1	0.80	0.80	0.80	159954
accuracy			0.80	320000
macro avg	0.80	0.80	0.80	320000
weighted avg	0.80	0.80	0.80	320000

Figure 5: Classification Report

8. Conclusion

Sentiment analysis is an incredibly valuable technology for businesses because it allows getting realistic feedback from your customers in an unbiased (or less biased) way.

If Done right, it can be a great value-added to your systems, apps, or web projects.

The model built in in-grained Sentiment Analysis involves determining the polarity of the opinion. It can be a simple binary positive/negative sentiment differentiation.

This type can also go into the higher specification in-grained Sentiment Analysis involves determining the polarity of the opinion.

It can be a simple binary positive/negative sentiment differentiation.

The model built has an accuracy of 80% and has predicted the given sentences into positive, negative and neutral. We have True Negative of 128344 and True positive of 127304 and False Negative of 32650 False positive of 31702.

9. Future Work / Enhancements

The model further can be trained and updated to analyse different types of sentimental analysis like emotion detection where it is the process of identifying and analysing the emotions expressed in textual data is known as emotion analysis or for intent analysis where it is the process of analysing text data to determine the author's intent.

Intentions underpin much of human behaviour and action, and understanding intentions can help you interpret these behaviours.

The model can be used commercially by organization by improving the datasets and models to use them to find out the emotions or opinion on their products and services.

With the help of a stored procedure, we can get the data called in real time to a front end.

References

[1] SNEHA SUKHEJA, SHALU CHOPRA and M.VIJAYALAKSHMI (2020). Sentiment Analysis using Deep Learning – A survey. IEEE. (ICCSEA).
[2] SHIHAB ELBAGIR AND JING YANG (2019). TwitterSentiment Analysis Based on Ordinal Regression. IEEE. Digital Object Identifier 10.1109/ACCESS.2019.2952127
[3] MEHMET UMUT SALUR AND ILHAN AYDIN (2020). A Novel Hybrid Deep Learning Model for Sentiment Classification. IEEE. Digital Object Identifier 10.1109/ACCESS.2020.2982538
[4] ALHASSAN MABROUK , REBECA P. DÍAZ REDONDO AND MOHAMMED KAYED (2020). Deep Learning-Based Sentiment Classification: A Comparative Survey. IEEE. Digital Object Identifier 10.1109/ACCESS.2020.2992013
[5] Natalie Friedrich , Timothy D. Bowman , Wolfgang G.Stock & Stefanie Haustein (2021),Adapting sentiment analysis for tweets linking to scientific papers-Heinrich Heine University Düsseldorf, Institute of Linguistics and Information, Department of Information Science, Düsseldorf (Germany)100.121.12-Cornell University
[6] A.H. Alamoodi a, B.B. Zaidan a,g , A.A. Zaidan a, O.S.Albahri a , K.I. Mohammed a , R.Q. Malik(2019) Sentiment analysis and its applications in fighting COVID-19 and infectious diseases-Sciencedirect-2020.114155
[7] Walaa MedhatAhmed HassanHodaKorashy(2019)-Sentiment analysis algorithms and applications-Sciencedirect-2019.11658155
[8] SHAHID SHAYAA , NOOR ISMAWATI JAAFAR, SHAMSHUL BAHRI , AININ SULAIMAN Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges Berkshire Media Sdn Bhd, Petaling Jaya-2019-10.1007/978-91-33-6822-7, (125-185)
[9] Bilal Abu-Salih, Pornpit Wongthongtham, Dengya Zhu, Kit Yan Chan, Amit Rudra, Bilal Abu-Salih, Pornpit Wongthongtham, Dengya Zhu, Kit Yan Chan, Amit Rudra, Sentiment Analysis on Big News Media Data, Social Big Data Analytics, 10.1007/978-981-33-6652-7, (177-218), (2021).
[10] Enrique Bigne, Carla Ruiz, Antonio Cuenca, CarmenPerez, Aitor Garcia, What drives the helpfulness of online reviews? A deep learning study of sentiment

analysis, pictorial content and reviewer expertise for mature destinations, Journal of Destination Marketing & Management, 10.1016/j.jdmm.2021.100570, 20, (100570), (2021).

[11] Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learningfor sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4 (2018): e1253.

[12] Ain, Qurat Tul, et al. "Sentiment analysis using deeplearning techniques: a review." *Int J Adv Comput Sci Appl* 8.6 (2017): 424.

[13] Mabrouk, Alhassan, Rebeca P. Díaz Redondo, andMohammed Kayed. "Deep learning-based sentiment classification: A comparative survey." *IEEE Access* 8 (2020): 85616-85638.