



Sentimental Analysis Using Deep Learning

Mahesh Mishra and Amol Patil

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 3, 2024

Sentimental Analysis Using Deep Learning

Mahesh Mishra^{1,*} and Amol Patil²

^{*1,2}Department of Electronics and Telecommunication, Kharghar

^{*1,2}A C Patil College of Engineering, Kharghar

E-mail address: mahesh070901@gmail.com, agpatil@acpce.ac.in

Abstract

A great amount of data is being evaluated in the newly-emerging field of sentiment analysis. In this approach text emotion recognition plays a crucial role. In order to produce insightful findings on a certain data. It is a powerful methodology that can benefit various domains like health care, ecommerce enterprise solution, government organisation and so on. To implement the procedure with the maximum level of accuracy researchers in the domain of natural language processing (NLP) and machine learning (ML) have investigated a number of techniques. The Recurrent Neural Network (RNN) algorithm serves as its foundation. In this study, the feelings of the Internet Movie Database (IMDB) movie reviews are analysed using the Long Short-Term Memory (LSTM) classifier. The data has already been divided into 25,000 reviews for testing the classifier and 25,000 reviews for training. This model gives a accuracy of 89.9%.

Keywords: Sentimental Analysis; Transformer; LSTM; NLP; BERT

1. Introduction

Sentiment analysis can be performed on a document-level [1], a sentence level [2], or an aspect level (sentiment regarding a particular aspect of an entity) [3]. Effective preprocessing and partitioning of the data improves post-classification performance. The correctness of the categorization performance is examined. Findings indicate the best classification. precision of 89.9% for LSTM approach.. It demonstrates the possibility of incorporating the created remedy in contemporary text-based sentiment analyzers. A data collection that Andrew Maas built using 50k IMDb movie reviews is used in this paper [4]. Additionally, each set has 12.5k reviews, both favourable and unfavourable. It allows users to rate on a scale from 1 to 10, with everything below 4 stars being tagged as bad and anything over 7 stars being recognised as positive, according to the dataset's developer. Not included are reviews with ratings outside of the aforementioned ranges. Each movie has a maximum of 30 reviews. With a standard deviation of 172.91 words, reviews typically include 234.76 words. The collection has 88585 unique words in total. With reference to the IMDb rating system, the reviews are divided into good and negative categories. In this work, 10-fold cross-validation is used to avoid bias in the categorization results. The adaptive moment estimation (Adam) optimizer is used with the LSTM. It adjusts the learning rate for each weight in the neural network by estimating the first and second moments of gradient [4]. There are three layers used. In this instance, 50 nodes are used in the first layer to convey the intended words. The final layer produces two outputs that correspond to the classes being considered, and the second layer is an LSTM with 101 memory units.

Further section 1 explains about overall description of model and the ways data which were divided for training and testing algorithm, section 2 explains about Methodology used for

sentimental analysis, section 2.1 explains about Lexicon and ML based model descriptions, section 2.2 explains overall model description of LSTM and confusion which were generated from this model, section 3 explains about results obtained from model and accuracy and loss graph obtained from model. The section 4 shows a conclusion.

2. Methodology

1. Dataset
2. Data division
3. Preprocessing
4. Vectorization
5. LSTM Classifier
6. Classification Accuracy

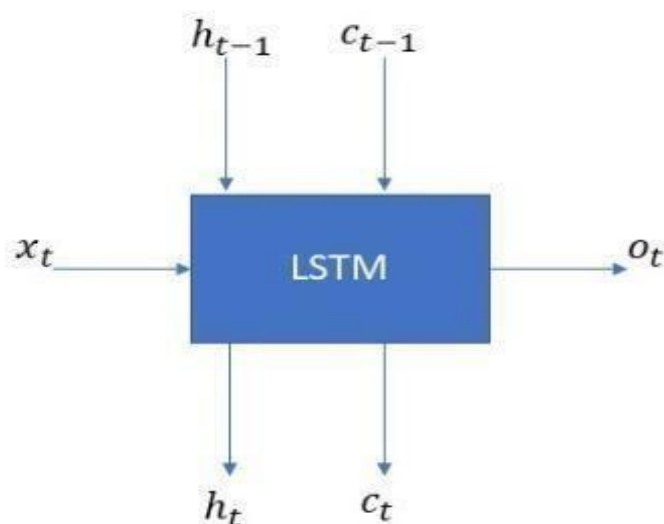
2.1. Lexicon and ML based: -

1-Lexicon-based [5] sentiment analysis and sentiment analysis powered by machine learning [6] The Lexicon Approach relies on tokenizing the text. (tokenization), counting how many times each occurs word and researching each term's subjective meaning from a current lexicon. This model's validation accuracy is 86.67%, while its training accuracy is 91.9%.

2-The machine learning approach depends on training various classifiers with a data set known as the training set in order to make the system more complex while also adding sophistication.

3-LSTM

Figure 1:LSTM



The Figure 1 shows a inputs and outputs of an LSTM for a single timestep. This represents the input, output, and equations for a single timestep of a time-unrolled representation. CNN output or the input sequence itself can both be used as the LSTM's input $x(t)$. The inputs from the prior timestep LSTM are $h(t-1)$ and $c(t-1)$. The LSTM's output for this timestep is $o(t)$. Additionally, the LSTM generates the $c(t)$ and $h(t)$ for use by the subsequent time step LSTM.

$$f_t = \sigma_g(W_f \times x_t + U_f \times h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i \times x_t + U_i \times h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o \times x_t + U_o \times h_{t-1} + b_o)$$

$$c'_t = \sigma_c(W_c \times x_t + U_c \times h_{t-1} + b_c)$$

$h_t = o_t \cdot \sigma_c(c_t)$, $c_t = f_t \times c_{t-1} + i_t \cdot c'_t$ here f_t is forget gate, i_t is input gate, O_t is output gate,

C_t is cell state and h_t represents hidden state.

2.2.End to end sentimental analysis using LSTM

Figure 2 displays the suggested machine learning (ML) model for sentiment analysis of text. In this study, 10-fold cross-validation is used to prevent bias in categorization results. Gates are added to loops in LSTM to maintain a level of relevancy [7]

Figure 2: LSTM Algorithm

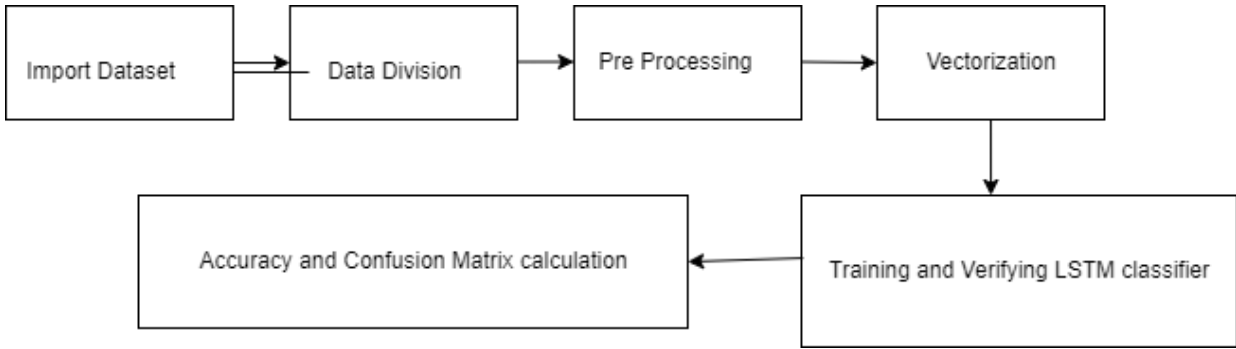


figure 3:Confusion Matrix

	Positive	Negative
Positive	4553	447
Negative	4436	564

Figure 3 summarises and presents the confusion matrix results together with the accuracy score, which is found to be around 89.9%. These findings demonstrate that the classifier achieves an appropriate precision for the desired dataset while taking the classifier's accuracy score into account.

Figure 4: Sentiments

Review	Movie	Sentiment
"Making a brilliant, original horror film is pretty hard these days. It was a great movie, and I suggest that you go see it."		Positive
"The opening ten minutes are admittedly great. Drew Barrymore reaching for the phone, only to find herself in conversation with a stalker is inspired. What a pity then about the rest of the film."		Negative

Fig 4 provides an overview of the dataset's good and negative assessments. Above Fig 4 shows a sentiments of people for Movie Scream.

3.Results And Discussions

Figure 5: Output Prediction

Sentiment	Actual (Positive = 1) (Negative = 0)	Predicted
<i>I love the movie</i>	1	1
<i>It was entertaining</i>	1	1
<i>It was a waste of time</i>	0	0
<i>It was uninteresting</i>	0	0

Figure 5 provides a summary of four movie review samples to show the classification results. As can be seen, four out of the four sentiments proved to be true and legitimate.

Figure 6:Loss graph

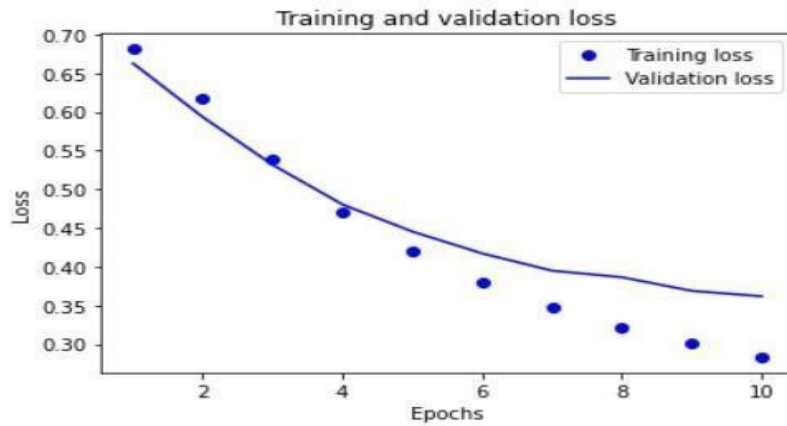


Figure 6 shows a Graph of Loss vs Epochs in which blue dotted and Line shows a Training and validation loss.

Figure 7:Accuracy graph

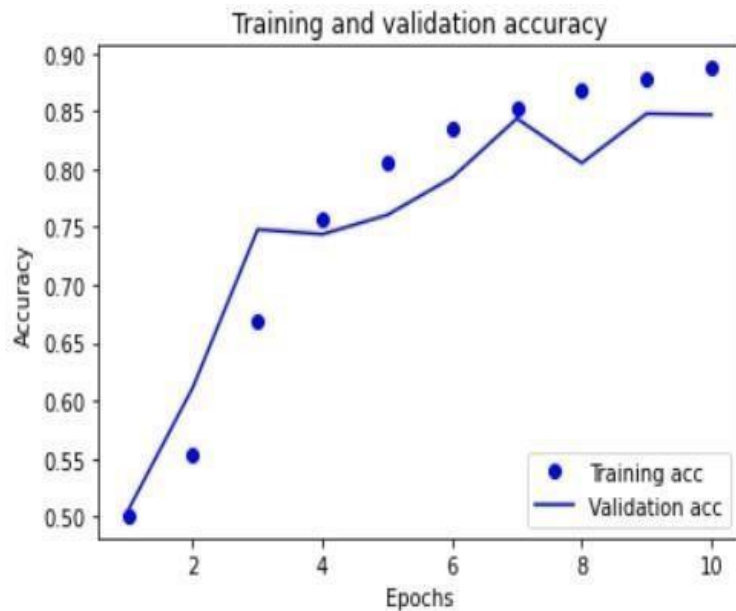


Figure 7 shows a Graph of Accuracy vs Epochs in which blue dotted and Line shows a Training and validation accuracy.

4-Conclusion

The technique of extracting views from text data and categorising them as positive, negative, or neutral is known as sentiment analysis, sometimes known as opinion mining. This paper uses a Long Adam optimizer and Short-Term Memory classifier are used to automatically classify the IMDb movie that has been processed reviews. 10k reviews in all, 5k of which are positive and 5k for unfavourable feelings. The model is giving a accuracy of 89.9%. By utilising additional data preconditioning techniques, a higher level of accuracy can be achieved. Additionally, using ensemble classifiers or deep learning techniques will increase classification accuracy. In future going to be using hybrid approach as LSTM and Transformer Based.

References:

- [1] P. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, 2002, ACL -
- [2] M. Hu and B. Lui, Mining and summarizing customer reviews, 2004, KKD-
- [3] T. Wilson, J. Wiebe and P. Hoffman, Recognizing contextual polarity in phrase level sentiment analysis, 2005, ACL.
- [4] Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, 8(1), 28 Sailunaz, K., Dhaliwal, M., Rokne, J., & Alhajj, R. (2018-Jiang, S., & Chen, Y. (2017, September). Hand Gesture Recognition by Using 3DCNN and LSTM with Adam Optimizer. In *Pacific Rim Conference on Multimedia* (pp. 743-753). Springer, Cham.
- [5] *Expert Systems with Applications*, 62, 1-16 Ansah J, Kang W, Liu L, Liu J, Li J. Information Propagation Trees for Protest Event Prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining 2018 Jun 3* (pp. 777-789). Springer, Cham
- [6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, Lexicon based methods for sentiment analysis, *Computational linguistics* Vol.37, No.2, 2011, pp 267-307
- [7] IEEE. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016, August). Attentionbased bidirectional long short-term memory networks for relation classification.