IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

Sentiment Analysis on Movie Reviews Using BERT

Drishti Engineer, Dennis Shah, Dhawal Mange, Vaishali Hirlekar

Department of Computer, Shah and Anchor Engineering College, Mumbai, India

ABSTRACT: Sentiment analysis highlights the distinction between emotions and thoughts and their content. Sentiment analysis is a technique for identifying and validating a person's feelings regarding a certain source of content. When picking which movies to see or buy next, moviegoers usually visit IMDb or Amazon for ratings and reviews. Moviegoers currently choose which films to see by looking at movie ratings and reading some of the reviews on sites such as IMDb and Amazon. Tweets, online diaries, and status updates, among other things, make up a large portion of the inclination data in electronic life. Sentiment analysis of the items under consideration made data extremely useful in revealing people's opinions. Opinion mining is a method of systematic mining. Via Natural language Processing Feelings and thoughts obtained from text, audio, and databank sources can be classified (NLP). The study's major purpose is to apply the BERT model. to evaluate the concept of human emotions. The dataset used consisted of IMDB movie reviews generated by user data. The dataset used is not limited to one language but has a mix of several languages. The experimental methodology is adequate and reliable, which defines the sentiment analysis output.

KEYWORDS: Sentiment analysis, Opinion mining, IMDb reviews, BERT, Adam's learning rate.

I. INTRODUCTION

Opinions are feelings that one person assumes. These announcements often convey feelings about articles or events. Opinion mining, also known as sentiment analysis, is a task that entails determining a client's attitude toward purchasing a particular item or service. Opinion mining is a method for identifying and separating abstract data in content archives. In web-based life, opinion plays a significant role in obtaining client feedback. It makes use of online feedback systems, notifications, and various person-to-person contact locales, such as Twitter, Facebook, YouTube, Blogs, and LinkedIn. Sentences, archives, and viewpoint dimensions are all popular places to study sentiments.

In recent years, sentiment analysis of Social Media messages has snatched critical thinking. Bad content can be separated from first-rate material with the aid of assessment mining. It has substance. The site's state dies are assembled in this document. The lexical philosophy is used to determine the motion picture overviews' general extremity.

The project aims to categorize to establish if movie reviews are positive or negative by classifying their polarity The issue could be approached in two ways: as a multi-label classification task or as a binary classification task, where the polarity can be easily identified as good, poor, or average. As BERT stands for Bidirectional Encoder Representations from Transformers, Google's network architecture has been transformed so that we can see what the state of the art for NLP tasks is.

BERT distinguishes itself from many other models by employing a transformer and displaying the contextual relationship between terms and subwords. Here, the transformer comprises two processes in the form of an encoder and a decoder, with the encoder reading the text input and the decoder predicting the mission.

Feature-based and fine-tuning are two well-established methods for applying pre-trained language representations to subsequent tasks. The task-specific architectures used in the feature-based method contain as additional features the pre-trained representations. The fine-tuning introduces only a few downstream tasks that are taught by simply fine-tuning all pre-learned parameters Both approaches learn general language representations using unidirectional language models during pre-training with the same objective function.

In this paper, BERT is an acronym for "Bidirectional Encoder Representations from Transformers," and it is proposed that BERT stands for Used to "Bidirectional Encoder Representations from Transformers" aims to improve fine-tuning-based techniques. Pre-training objective using a "masked language model" influenced by Cloze (MLM). BERT overcomes the previously mentioned unidirectionality requirement (Taylor, 1953).

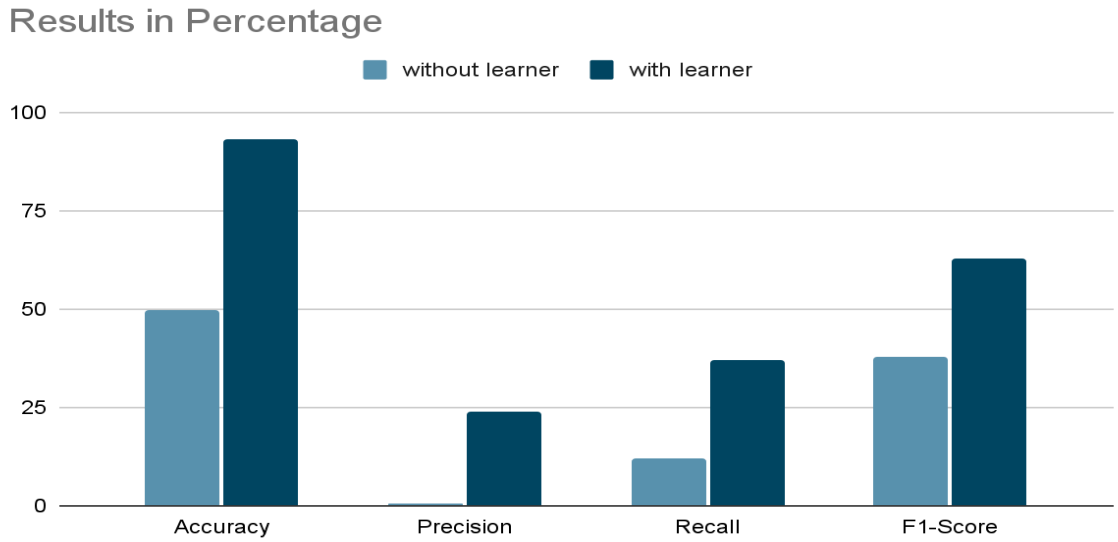
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

II. LITERATURE SURVEY

In sentiment analysis, many alternative approaches to text processing are applied. Constructing lexical chains, machine learning, and a variety of other methodologies are all beneficial for this goal. Statistical methods, as well as domain knowledge-driven analysis, are employed. For sentiment analysis, such approaches have proven to be beneficial.



Researchers have worked in a variety of languages, including Thai, Nepali, Bengali, and Malayalam. The results of sentiment analysis processing are both time-saving and accurate. The first draught was written in Hindi, Marathi, and Bengali. As a consequence of the numerous surveys, the necessity for the same has been identified.

C. Nanda, M. Dua, and G. Nanda [1] proposed Reviews and opinions are important in determining a user's level of satisfaction with a specific organization. The polarity, i.e. positive, negative, and neutral, is then determined using these. This paper discusses a method for conducting Sentiment Analysis on Hindi movie reviews.

R. I. Permatasari, M. A. Fauzi, P. P. Adikara, and E. D. L. Sari [2] Instead of relying solely on Bag of Words Features, this study used Nave Bayes with Ensemble Features. This ensemble used a variety of features, including Twitter-specific features, textual features, part of speech features, lexicon-based features, and Bag of Words.

Malini R; Sunitha M.R [3] Sentiment analysis identifies and validates an individual's feelings about a particular source of matter. Tweets, online diaries, status updates, posts, and other forms of electronic life contain a significant amount of inclination data. Sentiment analysis of the items under consideration made data extremely useful in revealing people's opinions.

K. Topal and G. Ozsoyoglu [4] Moviegoers often consult IMDb or Amazon for movie ratings and feedback when deciding Which film should I watch or buy next? Currently, viewers make decisions on which films to see by reading reviews and looking at movie ratings on sites like IMDb and Amazon.

Saad R. et al [11] Via Natural Language Processing (NLP), this computerized system assists in identifying a human emotion derived from textual data, voice, and other database (DB) sources (Natural Language Processing). The text is divided into three sections: positive, negative, or neutral by SA.

To acquire generally hyperparameters Words can be represented in a variety of techniques that have been employed for decades. Modern NLP systems use pre-trained word embeddings, which offer substantial advantages over learning embeddings from scratch (Turian et al., 2010). Objectives for left-t language modeling and objectives for distinguishing pre-train word embedding vectors, correct from wrong words in the left and right contexts were used. (Mnih and Hinton, 2009).

Traditional word embedding research is generalized in a different dimension by ELMo minimizing to and its forerunner (Peters et al., 2017, 2018a). To extract context-sensitive characteristics, a left-to-right and right-to-left language model is used. The sum of each token's contextual representations from left to right and right to left is its contextual representation. In a range of significant NLP benchmarks, In addition to question responding (Rajpurkar et al., 2016), sentiment analysis is an important part of the research process. (Socher et al., 2013), and named entity recognition (Peters et al., 2018a), ELMo advances the state of the art (Tjong Kim Sang and De Meulder, 2003).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

BERT

The bidirectional training of Transformer, a prominent attention model, to language modeling is an example of BERT's primary technical achievement. This study looked at a combination of left-to-right and right-to-left training, unlike other studies that from left to right or right to left, I examined a text sequence. This research looked at a left-to-right text sequence. The findings of the paper suggest that bi-directionally trained language models and single-direction language models may have a better sense of language context and flow. Masked LM is a new methodology described by the researchers (MLM) in previously impossible models that were previously impossible to train in.

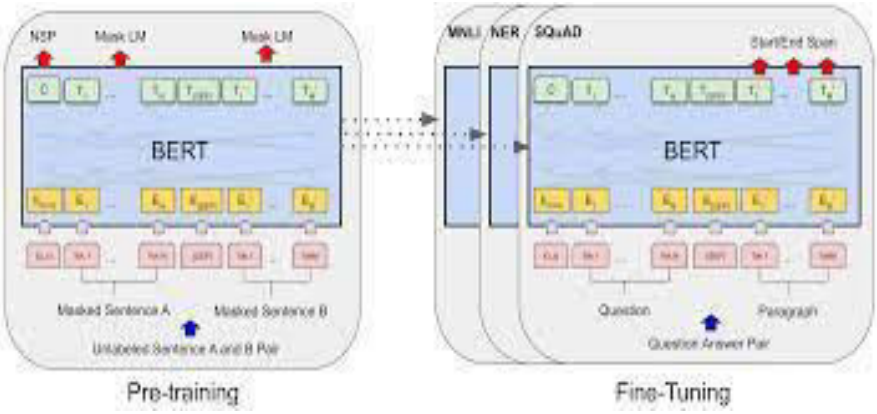


Fig 2: BERT Model

Transfer learning is frequently employed in the field of computer vision by pre-training a neural network model on a known task, such as ImageNet, and then fine-tuning which the trained neural network is utilized as the foundation for a new purpose-specific model.

A. Input Data

The surveys from the movie review dataset are used as input data. The particular film is selected from the dataset, and audits for that film are shown on the website page. Following the release of a new motion picture, the survey results for that film are added to the dataset.

The data used here is IMDB Movie Reviews, which is batched data. This data includes 5,000 movie reviews for preparation and 5,000 for research.

Processing

The text pre-processing techniques are:

Tokenization:

Tokens are a square of characters that appear in the content report's detail. These substance reports are separated into tokens and used for data processing.

Removal of Stop Words:

A web search tool or another popular language planning application can provide a set of stop-records or a single stop-list. "an", "an", "of", "the", "you", and "and" are some of the
most frequently used stop words in English. As a result, all terms that appear repeatedly but provide no data for the assignment are eliminated.

C. Classification

Everything done is to turn the data into a format that our model can understand. A lot of things have been done in terms of preprocessing. Now after preprocessing the data is ready to be fed to the model. The model then analyses the data being fed and classifies it into the categories mentioned by the programmer.

D. Prediction

The artificial neural networks methodology is used to predict movie review outcomes. Machine learning has been transformed by neural frameworks, which are a specialized set of calculations. Neural Networks are general limit approximations in and of themselves, which is why they can be used to solve almost any machine learning problem involving an unconventional mapping from the commitment to the yield space.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

One of the reasons to use Python is this. Lower case text should be present since BERT lower case model is in use. The model being used is a fine-tuning model that can perform token and sentence level tasks over a wide set of tasks. Fine-tuning is performed except for batch size, learning rate, and several training epochs, the hyperparameters are the same in most of the models. After the results are prepared, the next step is to create a model called "create a model." First, use the tf hub to name module BERT. It will only make one sheet. This will aid in the training and adaptation of BERT for sentimental responsibilities (which are, to find the polarity of movie reviews). Two approaches are usually used by BERT:

1. Masked Language Modeling (MLM)

2. Next Sentence Prediction (NSP)

Masked LM (MLM)

15% of the total Each word in each word sequence is replaced with a [MASK] token before being sent into BERT. The model then attempts to forecast the original value of the masked words based on the context supplied by the sequence's other non-masked phrases. To put it another way, output word prediction necessitates: On top of the encoder output, a classification layer is applied.

- The output vectors are translated into the vocabulary dimension by multiplying them by the embedding matrix.
- Softmax is used to calculate the likelihood of each word in the lexicon.

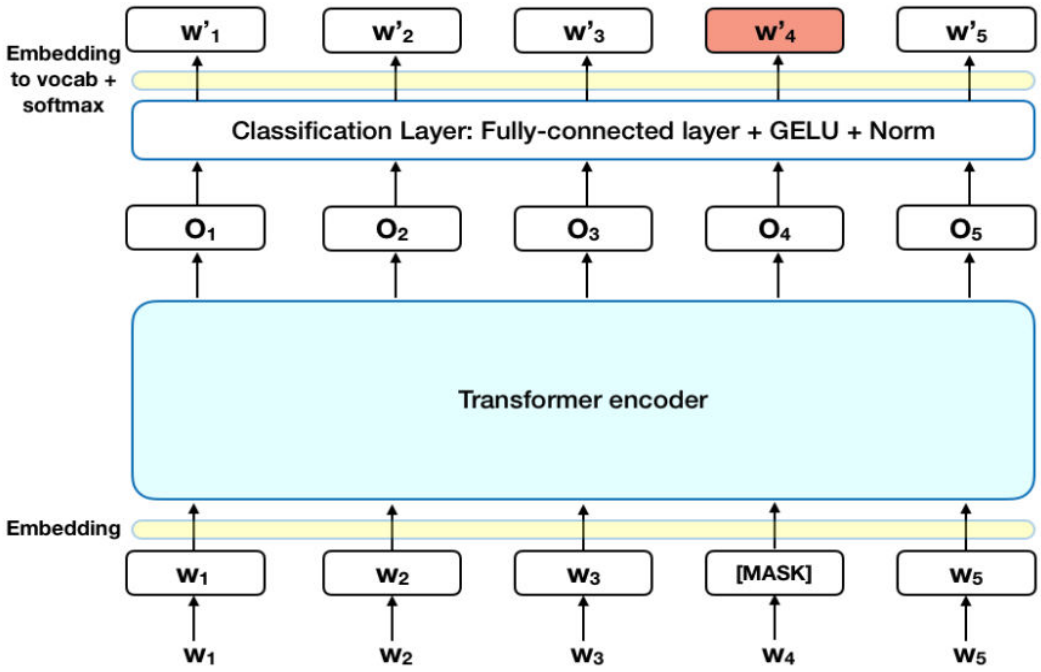


Fig 4: Masked LM (MLM) [14]

The BERT loss function solely examines masked value prediction and ignores non-masked word prediction. As a result, the model converges slower than directed models, although this is compensated for by its enhanced context-awareness.

III. METHODOLOGY

This segment outlines the steps involved in mining a movie dataset for sentiment analysis. The following polarity labels are issued as Social Media (Twitter) and IMDb: positive, negative, positive-positive, positive-negative, negative-negative, negative-negative, and negative-positive.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

Fig 1: Block diagram for Sentiment Analysis

Data Collection:- Data collection is the manner of accumulating and measuring statistics from infinite exceptional sources. To apply the information we gather to broaden realistic synthetic intelligence (AI) and system mastering solutions, it needs to be accumulated and saved in a manner that makes feel for the commercial enterprise hassle at hand.

Text Preprocessing:- Pre-processing of the gathered raw data is required to remove unwanted data and avoid words that are completed by standardization procedures. Using a simple classification layer pre-processed data is broken down using the feeling investigation to classify the tweets into positive, negative, or unbiased categories.

Sentiment Detection:- The sentiment detection of texts has witnessed a booming interest in recent years, due to the increased availability of online reviews in digital form and the ensuing need to organize them. Till now, there are mainly four different problems predominating in this research community, namely, subjectivity classification, word sentiment classification, document sentiment classification, and opinion extraction.

Sentiment Classification:- Sentiment classification is the automated process of identifying opinions in a text and labeling them as positive, negative, or neutral, based on the emotions customers express within them. Using NLP to interpret subjective data, sentiment classification can help you understand how customers feel about your products, services, or brand.

Presentation of Output:- This step will show the output regarding whether the sentence is positive, negative, or neutral. The results of the analysis are used to predict movie ratings using neural networks. There are both positive and negative terms in the dictionary. The prediction model is constructed by obtaining all of these features in their entirety.

IV. IMPLEMENTATION

The gradient pre-normalization phase ensures that all gradients (individual layers/weight matrices) have unit L2 norms throughout the model. Because the gradient direction rather than the amplitude affects updates, pre-normalization is critical. This is especially beneficial in circumstances where the gradient's orientation is strongly conserved, such as big batches. The larger the batch size, the closer the (stochastic) gradient approximation to the real (full-batch) gradient precision and the less common noisy gradients are.

Fine-tuning is another term for this technique, which will change three parameters: preparation, assessment, and adaptability. The fine-tuning hyperparameter used here is the learners. The learning rate used is Adams's learning rate 2e-5. Using the learning rate affected the accuracy of the model drastically.

Text representations that have been pre-trained Textual representations that have been pre-trained are commonly utilized to increase performance on NLP tasks. These representations are fed as features to downstream models after being trained on huge corpora (typically unsupervised). These properties can also be fine-tuned on the downstream task in deep networks.

Adam belongs to a class of AdaGrad-inspired algorithms that normalize the first gradient moment by the second moment's norm. Adam introduces the first two gradient moments: mean and variance, as running averages. Adam is the preferred method for training NLP, reinforcement learning, GANs, and other models. (ref) Adam is the learner in the presence of high noisy gradients, making it resistant to weight initializations and initial learning rate choices. Learning rate has a substantially greater impact on model performance when compared to batch size. So, if you're deciding between different learning rates and batch sizes, it's probably better to spend more time optimizing the learning rate.

Model accuracy has a strong negative association with the learning rate. As a result, lower learning rates are preferable in our search space. Let's look at the accuracy at different learning rates and batch sizes using the Parallel Coordinates plot once more.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004084|

This verifies what the Parameter Importance plot already indicated: low learning rates are desirable, and batch size is unimportant. RTE only has two classes, unlike most entailment classes ("entailment" and "not entailment").
The authors integrate learning rate warmup and bias correction in their algorithm for BERT pretraining. However, according to a subsequent portion of the appendix, bias correction can be eliminated because its behavior is comparable to that of warmup.

V. RESULTS AND DISCUSSION

Without learners, the precision score was 41.82% and the Recall score was minimized to 49.62% while the F1 score was 33.83% Accuracy score was 49.56%. After using the learning rate hyperparameter the results are a precision score of 93.24%, a Recall score of 93.22%, and F1 score of 93.22%, and an Accuracy score of 93.22%. From this, it is inferred that the initial learning rate was low and the new learning rate ensured that the model will be trained faster than the initial speed. The learning rate helped in fine-tuning the model and increased the accuracy.

VI. CONCLUSION

Using the principle of sentiment analysis, this paper examined the polarity of IMDB movie reviews. Here the BERT classifier is used. A model is proposed in which the data set of IMDB movie reviews is used. Analysis of Social Media sentiment is critical; analyzing sentimental argument trend comments is extremely difficult. It's also difficult because there are recurring characters, slang terms, whitespaces, and misspellings. Natural Language Processing (NLP) is used to supervise these issues. Until implementing NLP, individual tweets are pre-processed.
The paper's core ideas are tried to convey without getting bogged down in technical specifics in this overview. It is strongly advised that offsetting those interested in learning more read the complete article and the auxiliary pieces linked to it. The BERT source code and models, which include 103 languages and were graciously provided as open source by the study team, are another helpful resource.
The basic idea of layer-wise adaptive optimizers is shown with a focus on the various published variations of LAMB and our implementation of NVLAMB, as well as how they improve on current optimizers that use a shared global learning rate across all layers. As explained in our previous blog, As a result, training times on modern parallel hardware have substantially decreased, from days to less than an hour. Also, in our BERT repository, give implementations based on PyTorch and Tensorflow.

REFERENCES

[1] C. Nanda, M. Dua, and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning", 2018 International Conference on Communication and Signal Processing
[2]R. I. Permatasari, M. A. Fauzi, P. P. Adikara and E. D. L. Sari, "Twitter Sentiment Analysis of Movie Reviews using Ensemble Features Based Naïve Bayes", 2018 International Conference on Sustainable Information Engineering and Technology.
[3]Malini R; Sunitha M.R, "Opinion Mining on Movie Reviews", 2019 1st International Conference on Advances in Information Technology.
[4]K. Topal and G. Ozsoyoglu, "Movie review analysis: Emotion analysis of IMDb movie reviews", 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
[5] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", LREC.Vol. 10. 2010.
[6] Prabowo, Rudy, and Mike Thelwall, "Sentiment analysis: A combined approach", Journal of Informetrics 3.2 (2009): 143-157.
[7] Ion Smeureanu, Cristian Bucur, "Applying Supervised Opinion Mining Techniques on Online User Reviews", Informatica Economica, 2012.
[8]Singh, V. K., et al. "Sentiment analysis of movie reviews: A new feature based heuristic for aspect-level sentiment classification", Automation, and the Computing, Communication, Control and Compressed Sensing (iMac4s), International Multi-Conference on IEEE,2013.
[9]Hu, X., Zhang, X., Lu, C., Park, E.K. and Zhou, X., 2009, June. Exploiting Wikipedia as external knowledge for document clustering. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 389-396). ACM.
[10]Boubacar, A. and Niu, Z., 2014. Conceptual clustering. In Future Information Technology (pp. 1-8). Springer, Berlin, Heidelberg.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details