Twitter Sentiment Analysis Using Enhanced BERT

Chapter *in* Lecture Notes in Electrical Engineering · January 2023

DOI: 10.1007/978-981-19-6581-4_21

Suman Mann
Galgotias University

Jyoti Arora
Maharaja Surajmal Institute Of Technology

Twitter Sentiment Analysis Using Enhanced BERT

Suman Mann, Jyoti Arora, Mudita Bhatia, Ritika Sharma, and Rewangi Taragi

Abstract With the fast improvement of Internet social platforms, consumer evaluations have emerged as a critical foundation for clients to recognize merchandise and make decisions. New means of communication, such as microblogging, have evolved in the last decade, and a lot of information is provided by tweets and short messages, that are used to predict feelings of the users and their perception of what is going on around the globe. The most prominent sentiment in the tweet can be recognized by sentiment analysis techniques. Sentiment analysis is the method of extracting and recognizing the user's evaluations of products and models and has various approaches using machine learning algorithms to classify the emotion behind that text. This paper investigates the usage of Enhanced BERT models to recognize the sentiment behind the tweet. BERT or bi-directional encoder representations from transformers was designed to help computers understand the meaning of ambiguous language in text, by using the surrounding text to understand the context in which that text could have been written. For a successful evaluation using Enhanced BERT, the Kaggle SMILE dataset is considered and will be tested for emotions such as happiness, sadness, etc., and categorized as such. Experiments display that this version of the model achieves an accuracy of 0.96.

Keywords Sentiment analysis · BERT · Tweet sentiment analysis · Precision · Recall

S. Mann · J. Arora · M. Bhatia (✉) · R. Sharma · R. Taragi
Department of Information Technology, Maharaja Surajmal Institute of Technology, Affiliated to GGSIPU, New Delhi, India
e-mail: muditabhatia25@gmail.com

S. Mann
e-mail: sumanmann@msit.in

A. J. Kulkarni et al. (eds.), *Intelligent Systems and Applications*, Lecture Notes in Electrical Engineering 959, https://doi.org/10.1007/978-981-19-6581-4_21

1 Introduction

In recent times, the exceptional boom of social networks, private blogs and assessment apps has made the availability of user-generated content material very freely available. Taking this into consideration, it can be said that human beings are now confident enough to express their opinions publicly on such platforms. So, the analysis of this content material presents essential statistics regarding the thoughts of people about different topics.

Twitter is a well-known Web site where users can express themselves through their posts, which are short tweets, with a limit of 140 characters, and express their opinion on important topics and share their day-to-day activities. As a result, people use slang, acronyms, emoticons and abbreviated forms to condense their statements. People also use sarcasm and polysemy to express their viewpoints. With over 100 million users and over 500 million tweets sent every day, Twitter has established itself as a prominent microblogging platform [1].

It is also very useful for categorizing reviews of various markets, which is why Twitter is considered in the proposed model to predict people's doubts and feelings about the different issues going on in the world. Covering almost the entire urban population, Twitter has become an indispensable part of various sentiment checking techniques, and the proposed version takes advantage of the large viewership and usership of the site, to recognize and classify the emotion of its users regarding the daily news, be it the Coronavirus, lockdown, closure of institutions and organizations or just the daily political news. The proposed Enhanced BERT model works to analyze and categorize sentiments of users using their tweets, and the performance of this model is evaluated against the emotion parameters and compared with other existing models, to check whether the desired improved results have been achieved or not.

For this evaluation, the dataset used for the proposed Enhanced BERT model is the Kaggle SMILE dataset, which covers tweets from May 2013 to June 2015, from Twitter handles affiliated to British Museums, with emotions such as happy, sad, disgust, surprise and irrelevant. Thus, this version of the Enhanced BERT model gives better results than its predecessor approaches, because it follows comparative analysis on performance metrics. Unlike the pre-vailing NLP models, which examine the textual content in only one direction, from left to right or vice-versa, Enhanced BERT scans the total series of words all at once; therefore, it is a faster and much more feasible method. The Enhanced BERT model is a steppingstone toward a newer and better approach, to analyze and categorize sentiments of the users even on short datasets, and thus can be deployed for larger twitter datasets, to analyze the emotion with which the user is blogging his or her opinion on the Web site. On this idea of Enhanced BERT, this paper proposes a powerful sentiment analysis approach.

The following section explains the previous works that have been done in these field, which have laid the foundation for successfully carrying out the research done for the proposed model, and the later sections discuss the architecture and techniques of Enhanced BERT, explaining the various preprocessing tactics that are used in the two-step pipeline to convert the text into machine-readable text, by removing the emojis, symbol, hashtags and other twitter jargon. Also, the experimental analysis and results performed using the Enhanced BERT model based on various classifiers such as precision, recall and F1 score, which are performance metrics are explained in detail. These metrics help us to calculate the accuracy with which that emotion was recognized. The results achieved by experimental analysis of this model were comparatively analyzed based on their F1 scores. The final section concludes that the Enhanced BERT model which provides a new powerful approach in the field of sentiment analysis.

2 Literature Survey

Many ML and NLP approaches are used today to help us understand the text and take our valuable information from that. Dhruv Rathee and Suman Mann have used ML and deep learning technologies to detect phishing attacks which requires the ML model to identify the language of such mails and mark it as suspected attack [1].

Few sentiment analyses are done by characterizing posts about electric products like mobile phones using ML approach. Many ML approaches are used in such cases to recognize the sentiment with which the text has been written [2]. Newer research tried to preprocess the dataset, then extracted an adj from the dataset that had important meaning (feature vector), used this list and subjected ML algorithms such as Naïve Bayes, Maximum Entropy and SVM. At last, they measured the performance in terms of recall, precision and accuracy [3].

Another approach was applied where the tweet was preprocessed and classified based on its content as happy, sad and no change; and compared the performance of the algorithm based on precision and recall [4]. But recognizing emotion with which tweet has been written and separating the categories is also an important part, which these approaches failed to do.

The technique applied could only recognize various possibilities of emotions with which the tweet could have been writ-ten. Andrea Chiorrini and Alex Mircoli investigated a similar model and deployed it with different classifiers, achieving an accuracy of 92% [5]. They tested their model for emotion recognition and sentiment classification. This model achieved a high accuracy and formed the basis for many further research grounds on this aspect.

Duyu Tang, Nan Yang, Furu Wei proposed a model to check the sentiment-specific embedding. Their work highlighted the use of NLP for learning task word embedding and was highly useful to predict the sentiment of particular words in the tweet, whether they were expressed with anger, grief, sorrow, etc. [6]. Walaa Medhat, Ahmed Hassan, Hoda Korashy gave a sophisticated update on previous approaches to sentiment analysis. They tried to group together many datasets and then distinguish the different emotions and tabulate them so [7].

Marco Pota, along-with his associates, tested state-of-the-art techniques to check existing BERT models for newer datasets, including emojis and the Italian language.

They gave a methodological approach to build newer models for languages other than English. The research also tried opinion mining to recognize people's opinions. The research also tried opinion mining to recognize people's opinions and attitudes for day-to-day situations. It includes extracting the sentiment from the user's tweet and then classifying its polarity [8].

Vasu Negi, Suman Mann and Vivek Chauhan proposed an approach to recognize the Devanagari character recognition using artificial neural network which can be further be improved if a similar BERT model is developed with all the resources and books written in that language [9]. The approach followed by Suman Mann, Anish Batra and Guneet Singh Sethi for their research for personalized automation using artificial intelligence was based on the communication of data from one end to radio via wireless media and was implemented for the micro-processing stage for this research [10].

Suman Mann and Sakshi Hooda observed that recent trends of incorporating computing technology into the medical fields, keeping track of patients, diseases, and their treatments, has led to accumulation of large amounts of data which can essentially be successfully replaced as classifiers for predicting the severity of the patient's condition, by application of data analysis and data mining techniques [11]. This also formed the basis for the research proposed, as it proved to be an important step in realizing that data analysis could be done for unique and peculiar datasets as well. Similar approaches proposed by Deepa Gupta and Suman Mann introduced intelligent devices for communicating data with other environments using sensor networks [12], while others proposed an effective and improved method for securing the preprocessing techniques for enhanced results [13].

While many researchers have worked toward enhancing the performance of their techniques in this field, to test out the sentiment behind the text and to classify that correctly, the area where the research faced a pit stop was the accuracy of their model. The proposed Enhanced BERT model proves to be a steppingstone toward a newer and better approach to analyze and categorize sentiments of the users, based on the accuracy achieved for various emotions amidst the large dataset they were tested for.

3 Proposed Enhanced BERT Model

In the proposed version of sentiment analysis, the model ex-tracts tweets from the Twitter site, and the accumulated tweets are then be subjected to preprocessing under a supervised set of rules at the saved data as is shown in Fig. 1. This approach follows the previously proposed models for categorizing into six groups of varying ages, using certain facial recognition features, involving the same procedural pipeline, namely preprocessing, feature extraction and classification, as is used in the proposed model [14]. The steps for analyzing the tweet make use of preprocessing tactics that transform the Twitter text into plain text, such as emoji, and classifying this text on the pre-trained model [15].

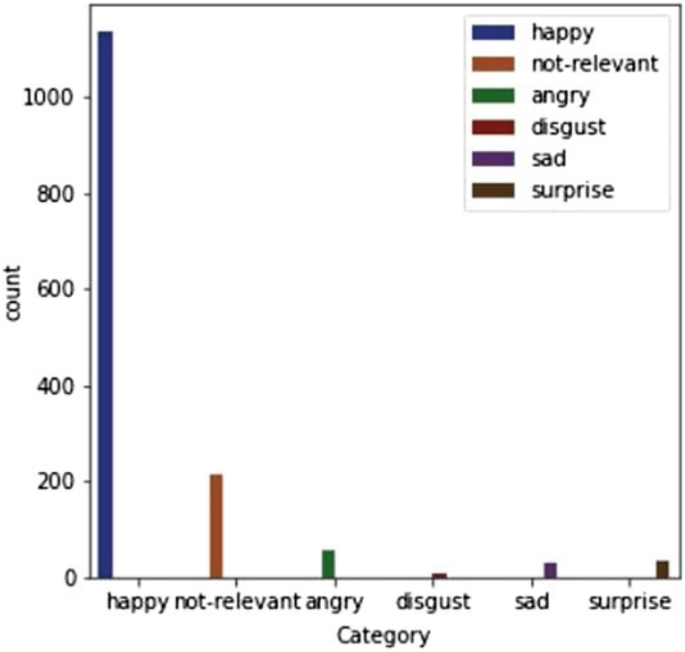

Fig. 1 Emotions on tweet intensity dataset: distribution of class of the test set. Categories: anger, disgust, happiness, surprise and sadness

Preprocessing strategies are used to transform the textual content so that it may be easily subjected to required set of rules under the proposed model. Tweets have several unique features, i.e. emoji, user mentions, hashtags as well as regular Internet constructs and different noisy sources, random numbers, etc., which must be preprocessed so that the machine can understand the underlying sentiment behind these tweets [16].

BERT preprocessing modules includes several preprocessing techniques that helps to save a lot of time. This technique is called BERT to Analyze Twitter Data [17].

The BERT architecture is recognized as one of the most popular models among the modern language modeling architectures. Its generalization functionality is such that it could be tailored to specific tasks in line with specific needs, be it relation extraction or Named Entity Recognition (NER), query answering or sentiment analysis [17].

Following are some insights into the BERT architecture:

There are two inputs, the first from word tokens and the second from segment layer.

After adding these two, they result in a position embedding, followed a layer normalization and dropout.

The next step is Multi-head Self-Attention layers which are 12 in number, and each is having 9 steps each.

After this, the two outputs are—one for Next Sentence Prediction (NSP) and one for Masked Language Modeling (MLM) [18].

In emotion recognition and classification, precision and recall are performance metrics that are applied to the data retrieved from a collection of text, and precision and recall are the two building precision and recall metrics into a single metric. At the same time, the F1 score is a measure of the test's accuracy [7]. The precision and recall are calculated and the F1 score is evaluated. Based on blocks of the F1 score. The goal of the F1 score is to combine the F1 score of the emotions, the comparison table is formulated, and the performance of the model can be easily evaluated. A good improvement in the F1 score shows good achieving accuracy for that emotion, whereas a decline or a slow improvement in the score shows that that emotion was not very well evaluated or that emotion could not be recognized as well as the others in the dataset.

Unlike the prevailing NLP models, which examine the textual content in only one direction, from left to right or vice-versa, Enhanced BERT scans the total series of words all at once. Enhanced BERT employs a transformer that is a device for recognizing relationship among words in a dataset. An Enhanced BERT system basically consists of various encoders and a decoder. The encoders serve the function of taking the input while the decoder generates predictions. Since the essential motive of BERT is to generate a pre-trained model, encode takes priority over the decoder.

The SMILE Twitter Emotion dataset, which is used in this study, is a collection of tweets mentioning 13 Twitter handles affiliated with British museums that was collected between May 2013 and June 2015. It has 3085 tweets, each expressing one of five emotions: rage, disgust, delight, surprise, and sadness [19]. This system is more powerful than the prevailing ones, as it may be feasible to realize how the records decided from the illustration of the result may have an effect in a specific field.

4 Experimental Results and Analysis

The SMILE Twitter Emotion dataset, which contains 3085 tweets labeled regarding the five emotions as shown in Fig. 2, was used for evaluating the performance of the proposed architecture on the job of sentiment. The model was assessed using following metrics: classification of accuracy and the F1 score. Each occurrence in the dataset is linked to a label emotion as well as a metric called intensity, which measures the emotion's intensity [20].

Let x_{ij} be the data number belonging to jth class, classified as ith class [21].

Let C be the classes number and N be the total amount of data. The accuracy achieved by a classifier is computed as [21]:

$$\text{accuracy} = \left(\frac{1}{N}\right)\sum_{i=1}^{C} x_{ii} \tag{1}$$

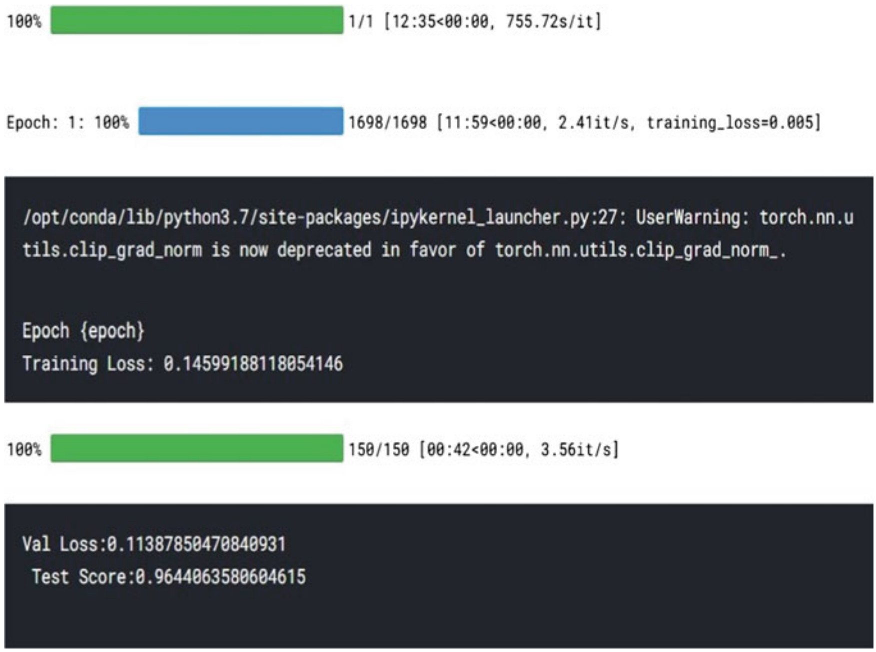

Fig. 2 Accuracy of the model ≈ 96.44%

Precision and recall of ith class are calculated as follows [21]:

$$\text{precision}_i = \frac{x_{ii}}{\sum_{j=1}^{C} x_{ij}} \tag{2}$$

$$\text{recall}_i = \frac{x_{ii}}{\sum_{j=1}^{C} x_{ji}} \tag{3}$$

F1 score of ith class is equal to [20]:

$$F_{1i} = 2 \cdot \frac{\text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} \tag{4}$$

So, the F1 score achieved by the model is defined as the average of F_{1i} [21]:

$$F_{1i} = \frac{1}{C} \sum_{i=1}^{C} F_{1i} \tag{5}$$

The comparative F1 results for the models on the sentiment emotion categories of happy, sad and anger are presented in Table 1.

Table 1 Comparison between different techniques based on their F1 score

Technique used	F1 score		
	Anger	Happy	Sad
Baseline-STLM	0.670	0.558	0.599
SS-STLM	0.712	0.597	0.793
BERT	0.736	0.930	0.808

The model achieved best results on the happy emotion and worst results on anger emotion, as can be seen from table below. Thus, it can be proved that the model has made substantial improvements in recognizing the various emotions behind the tweets, than the previous models could.

The results indicated in Fig. 2 show that the accuracy for the BERT model is 96.44% which means the model performed well even on small datasets. Thus, within a short period of time, a BERT model that works on test data with a good score can be built.

5 Conclusion

The proposed model was able to construct a practical sentiment analysis approach of Twitter system based on the Enhanced BERT model. The proposed technique is based on a two-step procedure, with the first step involving a series of preprocessing techniques to convert Twitter jargon, such as emoji into plain text, and the second step involving an updated version of BERT that had been pre-trained on plain text to fine-tune and classify tweets according to their polarity, such as happy, sad, angry, surprised and not relevant or irrelevant tweets, and the second step utilizing a version. Experiments display that this version of the model achieves an accuracy of 0.96. Thus, this version of the Enhanced BERT model gives good results even on short datasets, and thus can be deployed for larger twitter datasets, to analyze the emotion with which the user is blogging his or her opinion on the Web site. On this idea of Enhanced BERT, this paper proposes a powerful sentiment analysis approach.

References

1. Rathee D, Mann S, Detection of e-mail phishing attacks—using machine learning and deep learning. Int J Comput Appl 183(47). https://doi.org/10.5120/ijca2018918026
2. Joselson N, Hellen R (2019) Emotion classification with natural language processing
3. Gautam G, Yadav D (2014) Sentiment analysis of Twitter data using machine learning approaches and semantic analysis. IEEE, pp 103–110
4. Radford A, Narasimhan K (2019) Improving language understanding by generative pre-training, pp 11–35
5. Chiorrini A, Mircoli A, Diamantini C, Potena D (2021) Emotion and sentiment analysis of tweets using BERT, pp 15–32

6. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M (2015) Sentiment embeddings with applications to sentiment analysis. IEEE Trans Knowl Data Eng 28(2):496–509
7. Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5:1093–1113
8. M. Pota, M. Ventura, R. Catelli, and M. Esposito (2020) An effective BERT-based pipeline for Twitter sentiment analysis: a case study in Italian, pp 112–117
9. Negi V, Mann S, Chauhan V (2009) Devanagari character recognition using artificial neural network. Int J Eng Technol 2161–2167
10. Batra A, Sethi GS, Mann S (2019) Personalized automation of electrical and electronic devices using sensors and artificial intelligence—"The Intelligizer System". Computational intelligence: theories, applications and future directions—volume I. AISC, vol 798
11. Hooda S, Mann S (2020) Sepsis-diagnosed patients' in-hospital mortality prediction using machine learning: the use of local big data-driven technique in the emergency department. Int J Grid Distrib Comput 13(1)
12. Mann S, Gupta D, Arora Y, Chugh SP, Gupta A (2021) Smart hospitals using artificial intelligence and internet of things for COVID-19 pandemic. Smart healthcare monitoring using IoT with 5G
13. Gupta D, Jha SK, Mann S (2021) Internet crimes—it's analysis and prevention approaches. In: 2021 9th international conference on reliability, Infocom technologies and optimization (trends and future directions) (ICRITO), 2021, pp 1–4. https://doi.org/10.1109/ICRITO51393.2021.9596396
14. S. Mann et al (2018) Estimation of age groups using facial recognition features. Int J Eng Comput Sci 23945–23951
15. Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv:1810.04805, no 2, pp 18–40
16. Mancini M, Mircoli A, Potena D, Diamantini C, Duca D, Toscano G (2020) Prediction of pellet quality through machine learning techniques and near-infrared spectroscopy. Comput Ind Eng 15–31
17. Mishne G (2005) Experiments with mood classification in blog posts. Live J 1104–1153
18. Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B (2014) Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd annual meeting of the Association for Computational Linguistics, vol 1, pp 1555–1565
19. Kaggle. SMILE Twitter emotion dataset—useful for learning sentiment analysis
20. Chang L-H, Pyysalo S, Kanerva J, Ginter F (2019) Towards fully bilingual deep language modeling. ArXiv:2010.11639, pp 5–16
21. Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv:1810.04805