Nepali COVID-19 Tweets Sentiment Analysis Using Transformer-based Models

Jeevan Chapagain

University of Memphis, Memphis, TN, USA
jchpgain@memphis.edu

Abstract

With the increase in COVID-19 cases during the pandemic in 2019, millions of people lost their lives not only due to the pandemic but also because of the mental states and sentiments of individuals evoked by the fear of the virus. The availability of mobile phones and the ease of posting information on Twitter enable users to generate an enormous volume of tweet data on a wide range of topics and events. People expressed their feelings and emotions during the pandemic through tweets which can be studied to identify how's the person feeling and what they are going through. In our study, we specifically focused on tweets related to COVID-19 in the Nepali language which were tweeted during the time of the pandemic. We evaluated several state-of-the-art transformers models on NepCOV19Tweets datasets. Our best-performing models achieved F_1 score of **0.73** in classifying the sentiments of Nepali COVID-19 tweets.

Introduction

Twitter serves as a platform for individuals to freely express and share their viewpoints on a wide range of topics, encompassing social injustice, political parties and their actions, governmental policies, and the evaluation of company products. Given the extensive user base and the diverse languages used in tweets, such as English, Spanish, German, Hindi, Nepali, and others, the application of natural language processing (NLP) techniques becomes imperative for conducting various analyses. Throughout the COVID-19 pandemic, social media became a crucial outlet for countless individuals to express their emotions, as millions of lives were lost. Consequently, people posted diverse tweets reflecting their sentiments and experiences during this time, offering an opportunity to explore the psychological and emotional wellbeing of individuals through a comprehensive study.

Sentiment Analysis involves employing Natural Language Processing (NLP) techniques to evaluate human emotions pertaining to various subjects. Sentiments are typically categorized as positive, negative, or neutral. Businesses heavily rely on people's opinions and sentiments to shape their product development and marketing strategies. Likewise, analyzing individuals' tweets enables the exploration of their behavior, psychology, and more. Sentiment Analysis has been widely conducted on diverse subjects, utilizing textual data from sources such as news articles, movie reviews, and tweets.

Several studies have (Rustam et al. 2021; Nemes and Kiss 2021; Naseem et al. 2021; Garcia and Berton 2021) been conducted to analyze COVID-19 tweets in both English and other languages. However, there is a lack of research in resource-poor languages like Nepali, which will be elaborated upon in the relevant literature section. Recent decades have witnessed progress in NLP for Nepali languages, with studies focusing on various NLP tasks, such as news classification (Shahi and Pant 2018), detection of offensive language in Nepali social media (Niraula, Dulal, and Koirala 2021) and named entity recognition in Nepali (Niraula and Chapagain 2022). The limited research in the Nepali language can be attributed to several key factors. Firstly, as a resource-poor language, traditional approaches like TF-IDF and bag-of-words may not yield optimal results. Secondly, insufficient data availability poses a challenge in training models for the Nepali language. Lastly, tokenization proves challenging in morphologically rich languages like Nepali, as words can exhibit diverse forms with varying grammatical features such as number, gender, honorifics, and tense. Furthermore, the attachment of different suffixes to the same word introduces additional variability in word forms.

With the recent advancement and state-of-the-art performance, transformers have been widely used and showed great results even in resource-poor languages like Nepali. Due to the promising results shown by Transformers for the Nepali language, we studied the transformer-based models in analyzing the Nepali COVID-19 tweets. The models that we used in this study are namely: DB-BERT, NepBERT, NepaliBERT, and BERT-bbmu.

The remainder of this paper is organized as follows: Related Work reviews related work, providing context and highlighting the contributions of previous research in sentiment analysis and focusing on the Nepali language. The methodology section describes the methodology, detailing our approach to selecting transformer models. The Experiment and Results section presents the results of the experiments and metrics used to evaluate the performance of the pre-trained models followed by the conclusion with a summary of our contributions and suggestions for future research directions.

Related Works

Several studies (Phan et al. 2020; Goel, Gautam, and Kumar 2016; Joyce and Deng 2017; Mehta et al. 2020) have extensively worked on sentiment analysis of tweets across various topics. The outbreak of COVID-19 in 2019 led to a significant surge in COVID-19-related tweets on Twitter, prompting numerous studies in the field of sentiment analysis. Most of these studies have focused on resource-rich languages like English and German, while only a limited number of works have addressed sentiment analysis of COVID-19 tweets in resource-poor languages such as Nepali. In our review, we examined existing research on sentiment analysis of COVID-19 tweets in both English and Nepali, as well as other resource-poor languages like Arabic and Brazilian.

Nemes et al. (Nemes and Kiss 2021) proposed an RNN model to classify COVID-19-related tweets into Positive or Negative categories, using a dataset of self-created tweets. They conducted a performance comparison with TextBlob (Loria et al. 2018) and observed that their method outperformed TextBlob. Similarly, Rustam et al. (Rustam et al. 2021) conducted a study where they employed five machine learning methods (Extra Tree classifier, Decision Tree, Random Forest, XGBoost classifier, and Long Short-Term Memory) to categorize COVID-19 tweets into Positive, Negative, and Neutral sentiments. They utilized two commonly used text representation techniques, namely Bag-of-Words (BoW) and Term-Frequency and Inverse Document Frequency (TF-IDF).

Basiri et al. (Basiri et al. 2021) developed an ensemble deep learning model, combining Convolutional Neural Network (CNN), Bidirectional Gated Recurrent Unit (BiGRU), Support Vector Machine (SVM), and Naive Bayes (NB), to classify the sentiment of COVID-19 tweets. In a study by Aljameel et al. (Aljameel et al. 2021), unigram and bigram coupled with a TF-IDF approach were employed to represent Arabic COVID-19 tweets, which were subsequently classified using machine learning algorithms including SVM, K-Nearest Neighbor (KNN), and NB. De et al. (de Melo, Figueiredo et al. 2021) conducted sentiment analysis on Brazilian news articles and COVID-19-related tweets, revealing similar sentiment patterns across both types of content.

Sitaula et al. (Sitaula et al. 2021) presented a sentiment analysis approach for Nepali COVID-19 tweets, employing deep learning techniques. They introduced the NepCov19Tweets dataset, a benchmark dataset of Nepali COVID-19 tweets. The researchers trained three convolutional neural network models using different text representations (FastText, domain-specific, and domain-agnostic), combining them to form an ensemble CNN for tweet sentiment classification. In another study, Shahi et al. (Shahi, Sitaula, and Paudel 2022) proposed a hybrid feature approach, combining FastText and TF-IDF, to represent Nepali COVID-19 tweets for sentiment classification. They evaluated the performance of nine machine learning algorithms on these hybrid features. Additionally, Sitaula et al. (Sitaula and Shahi 2022) suggested the utilization of a multi-channel convolutional neural network (MCNN) to classify Nepali COVID-19 tweets into positive, neutral, and negative sentiment classes.

Although there are few studies done for Nepali COVID-19 tweets, all of the studies focus on traditional machine learning methods or deep learning methods using feature extraction methods but none of them have studied using the state-of-the-art transformers model.

Methodology

Transformers have been widely used in text classification tasks and have shown good performance in Nepali language tasks like NER (Niraula and Chapagain 2022, 2023). Because of their state-of-the-art performance, the transformers-based models for the Nepali language have been widely available. Monolingual Nepali transformer models are developed by training from scratch, utilizing only Nepali text data, whereas multilingual models are created by training the models to incorporate multiple languages. For this experiment, we used the following variations of transformer models that comprise monolingual Nepali transformer models as well as multilingual models. All of the models used in this experiment are available on Huggingface.

DB-Bert: Sakonii/distilbert-base-nepali model was trained on a Nepalitext language modeling dataset which was created by compiling the OSCAR, cc100, and a collection of Nepali articles scraped from Wikipedia. To train the model, the texts in the training set were segmented into blocks of 512 tokens. This model used the same configuration as the original distilbert-base-uncased model.

NepBERT: amitness/nepbert model was pre-trained on the Nepali CC-100 dataset, which includes 12 million sentences, using a Tesla V100 GPU from Google Colab. This model contains 83.5 million parameters, and it uses a Byte-level BPE tokenizer with a vocabulary of 52,000 tokens.

NepaliBERT: Ranjan/NepaliBERT is a BERT-based language model designed for Nepali. This model was pre-trained using 6.7 million text lines from the Large Scale Nepali Corpus and the OSCAR Nepali corpus, resulting in 82 million parameters. A word-piece tokenizer with a vocabulary size of 50,000 tokens was utilized in the model's training process.

BERT-bbmu: The transformers model bert-base-multilingual-uncased has undergone pre-training through a self-supervised method on a vast multilingual dataset (Devlin et al. 2018). During this process, the text was converted to lowercase and split into tokens using WordPiece, and a shared vocabulary of 110,000 was used. This model is trained in 102 different languages.

Experiments and Results

DataSet and Experiment Settings

For our experiments, we used the NepCOV19Tweets dataset (Sitaula et al. 2021) which consists of tweets related to COVID-19 dated from 11 Feb 2020 to 10 Jan 2021 using the geo-location of Nepal where tweets were searched using the keyword #COVID19 (in the Nepali language). The datasets consist of 33,247 clean tweets which consist of three labels: POSITIVE, NEGATIVE, and NEUTRAL. We randomly split the dataset using an 80-20 split strategy to create training and testing sets. We used the following hyperparameters for training all of the transformer's models.

- Learning rate: 0.0001
- Optimizer: AdamW
- Batch size: 16
- No. of epochs: 10

Results

All transformer models are trained and evaluated on NepCOV19Tweets. We used precision, recall, and F1 scores to measure the performance of our models. Table 1 shows the performance of the different models along with their precision, recall, and F1 scores.

Model	**Pre.**	**Rec.**	$\mathbf{F_1}$
DB-BERT	0.73	0.73	0.73
NepBERT	0.70	0.71	0.70
NepaliBERT	0.31	0.45	0.28
BERT-bbmu	0.70	0.68	0.69

Table 1: Model comparison using F_1 score

The DB-BERT model, which is a distilled version of BERT pre-trained on Nepali data, achieved the highest F1-score of 0.73, indicating a balanced performance in terms of precision and recall. The NepBERT model closely followed with an F1-score of 0.70, demonstrating its effectiveness in capturing the nuances of the Nepali language for sentiment analysis tasks.

In contrast, the NepaliBERT model exhibited relatively lower performance with an F1-score of 0.28. This could be attributed to differences in pre-training data or fine-tuning strategies employed for this model. The BERT-bbmu model, a multilingual variant, achieved an F1-score of 0.69, which is comparable to NepBERT but slightly lower than DB-BERT. This performance gap could be attributed to the model's ability to handle multiple languages, potentially compromising its performance on Nepali-specific tasks.

The results highlight the importance of language-specific pre-training and fine-tuning for achieving optimal performance in sentiment analysis tasks. The DB-BERT model emerges as the top performer, closely followed by NepBERT, indicating their suitability for Nepali sentiment analysis applications. However, further investigation into the pre-training data, fine-tuning strategies, and model architectures may be necessary to understand the performance differences observed among the evaluated models.

We also provided the detailed performance of the best-performing models in Table 2. The results show the model is performing well for all three labels.

The DB-BERT model demonstrated strong performance in classifying neutral and positive sentiments, achieving F1 scores of 0.85 and 0.79, respectively. However, its performance in classifying negative sentiments was relatively lower, with an F1-score of 0.31. This could be attributed to the imbalanced distribution of negative sentiment instances

Sentiment	**Pre.**	**Rec.**	$\mathbf{F_1}$	**Support**
Positive	0.78	0.80	0.79	279
Negative	0.46	0.23	0.31	26
Neutral	0.85	0.84	0.85	380

Table 2: Performance evaluation of the best-performing model per sentiment label

in the dataset, as indicated by the low support value of 26. Despite the lower performance for negative sentiment detection, the overall results highlight the effectiveness of the DB-BERT model in capturing the nuances of Nepali language for sentiment analysis tasks. The high precision and recall values for neutral and positive sentiments suggest that the model can reliably identify these sentiment classes

Conclusions

The study evaluated the performance of various BERT models for sentiment analysis of Nepali tweets. The results demonstrated that the DB-BERT model, a distilled version of BERT pre-trained on Nepali data, outperformed other models, achieving an overall F1-score of 0.73. The NepBERT model, pre-trained on Nepali data, closely followed with an F1-score of 0.70, indicating its effectiveness in capturing the nuances of the Nepali language.

Further analysis of the DB-BERT model's performance across different sentiment classes revealed its strength in identifying neutral and positive sentiments, with F1-scores of 0.85 and 0.79, respectively. However, the model showed relatively lower performance in detecting negative sentiments, with an F1-score of 0.31, which is potentially due to the imbalanced distribution of negative sentiment instances in the dataset. Overall, our experiment shows that both multilingual or monolingual transformer models can perform well in a task of Sentiment Analysis even for a resource-poor language like Nepali.

One of the future works could focus on addressing the class imbalance issue by employing techniques such as oversampling or data augmentation to increase the representation of negative sentiment instances in the training data. Furthermore, investigating the impact of different fine-tuning approaches, including hyperparameter tuning and regularization techniques, may lead to further performance improvements.

References

Aljameel, S. S.; Alabbad, D. A.; Alzahrani, N. A.; Alqarni, S. M.; Alamoudi, F. A.; Babili, L. M.; Aljaafary, S. K.; and Alshamrani, F. M. 2021. A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *International journal of environmental research and public health*, 18(1): 218.

Basiri, M. E.; Nemati, S.; Abdar, M.; Asadi, S.; and Acharrya, U. R. 2021. A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. *Knowledge-Based Systems*, 228: 107242.

de Melo, T.; Figueiredo, C. M.; et al. 2021. Comparing news articles and tweets about COVID-19 in Brazil: sentiment analysis and topic modeling approach. *JMIR Public Health and Surveillance*, 7(2): e24585.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Garcia, K.; and Berton, L. 2021. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied soft computing*, 101: 107057.

Goel, A.; Gautam, J.; and Kumar, S. 2016. Real-time sentiment analysis of tweets using Naive Bayes. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 257–261. IEEE.

Joyce, B.; and Deng, J. 2017. Sentiment analysis of tweets for the 2016 US presidential election. In *2017 ieee mit undergraduate research technology conference (urtc)*, 1–4. IEEE.

Loria, S.; et al. 2018. textblob Documentation. *Release 0.15*, 2(8).

Mehta, R. P.; Sanghvi, M. A.; Shah, D. K.; and Singh, A. 2020. Sentiment analysis of tweets using supervised learning algorithms. In *First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019*, 323–338. Springer.

Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P. W.; and Kim, J. 2021. COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Transactions on Computational Social Systems*, 8(4): 1003–1015.

Nemes, L.; and Kiss, A. 2021. Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 5(1): 1–15.

Niraula, N.; and Chapagain, J. 2022. Named Entity Recognition for Nepali: Data Sets and Algorithms. In *The International FLAIRS Conference Proceedings*, volume 35.

Niraula, N.; and Chapagain, J. 2023. DanfeNER-Named Entity Recognition in Nepali Tweets. In *The International FLAIRS Conference Proceedings*, volume 36.

Niraula, N. B.; Dulal, S.; and Koirala, D. 2021. Offensive Language Detection in Nepali Social Media. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, 67–75.

Phan, H. T.; Tran, V. C.; Nguyen, N. T.; and Hwang, D. 2020. Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access*, 8: 14630–14641.

Rustam, F.; Khalid, M.; Aslam, W.; Rupapara, V.; Mehmood, A.; and Choi, G. S. 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos one*, 16(2): e0245909.

Shahi, T.; Sitaula, C.; and Paudel, N. 2022. A hybrid feature extraction method for Nepali COVID-19-related tweets classification. *Computational Intelligence and Neuroscience*, 2022.

Shahi, T. B.; and Pant, A. K. 2018. Nepali news classification using Naive Bayes, support vector machines and neural networks. In *2018 International Conference on Communication Information and Computing Technology (ICCICT)*, 1–5. IEEE.

Sitaula, C.; Basnet, A.; Mainali, A.; Shahi, T. B.; et al. 2021. Deep learning-based methods for sentiment analysis on Nepali COVID-19-related tweets. *Computational Intelligence and Neuroscience*, 2021.

Sitaula, C.; and Shahi, T. B. 2022. Multi-channel CNN to classify Nepali covid-19 related tweets using hybrid features. *arXiv preprint arXiv:2203.10286*.