BEmoLexBERT: A Hybrid Model for Multilabel Textual Emotion Classification in Bangla by Combining Transformers with Lexicon Features

Ahasan Kabir, Animesh Chandra Roy, Zaima Sartaj Taheri
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chattogram-4349, Bangladesh
{ahasankabir146, animeshroycse, zstaheri1999}@gmail.com

Abstract

Multilevel textual emotion classification involves the extraction of emotions from text data, a task that has seen significant progress in high-resource languages. However, resource-constrained languages like Bangla have received comparatively less attention in the field of emotion classification. Furthermore, the availability of a comprehensive and accurate emotion lexicon specifically designed for the Bangla language is limited. In this paper, we present a hybrid model that combines lexicon features with transformers for multilabel emotion classification in the Bangla language. We have developed a comprehensive Bangla emotion lexicon consisting of 5336 carefully curated lexicons across nine emotion categories. We experimented with pre-trained transformers including mBERT, XLM-R, BanglishBERT, and BanglaBERT on the EmoNaBa (Islam et al., 2022) dataset. By integrating lexicon features from our emotion lexicon, we evaluate the performance of these transformers in emotion detection tasks. The results demonstrate that incorporating lexicon features significantly improves the performance of transformers. Among the evaluated models, our hybrid approach achieves the highest performance using BanglaBERT(large) (Bhattacharjee et al., 2022) as the pre-trained transformer along with our emotion lexicon, achieving an impressive weighted F_1 score of 82.73%. The emotion lexicon is publicly available at https://github.com/Ahasannn/BEmoLex-Bangla_Emotion_Lexicon

1 Introduction

Multilabel emotion classification involves the assignment of several emotion labels to a provided text. This allows for a more comprehensive representation to understand underlying emotional content. However, achieving accurate multilabel emotion classification in resource-constrained languages presents a unique challenge. Limited availability of annotated data, linguistic diversity, and cultural variations pose significant hurdles. Additionally, the lack of comprehensive emotion lexicons specific to these languages further complicates the task. To tackle these challenges, we present an innovative approach that combines the power of transformers with emotion lexicon features for multilabel emotion classification in Bangla. By leveraging pretrained transformer models and developing an extensive Bangla emotion lexicon, we aim to enhance the accuracy and effectiveness of emotion classification.

2 Related Work

Emotion detections from textual content have gained considerable attention in recent years. While extensive research has been carried out in high-resource languages such as English, Chinese, and Arabic, there remains a scarcity of studies specifically targeting emotion detection and classification in the Bangla language. Irtiza Tripto and Eunus Ali (2018) presented an LSTM-based method for emotion classification in Bangla YouTube comments by achieving 59.23% accuracy. Pal and Karn (2020) used logistic regression for detecting emotions (joy, anger, sorrow, suspense) from Bangla short stories. Rayhan et al. (2020) predicted six emotions from 7214 Bangla texts with the CNN-BiLSTM model outperforming BiGRU, achieving 66.62% accuracy. Das et al. (2021) developed s corpus of 6,523 texts for classifying six emotion categories employing various transformer models, among which XLM-R showed the best results with an F_1-score of 69.73%. Parvin et al. (2022) developed an emotion corpus comprising 9,000 Bangla texts in six emotion categories and proposed a weighted ensemble of CNN and BiLSTM. Islam et al. (2022) introduced a manually annotated Bangla noisy dataset comprising of 22,698 Bangla texts from various domains, labeled for six fine-grained emotion categories.

Researchers found that combining transformers with handcrafted features has enhanced performance in various tasks. For abusive language detection Koufakou et al. (2020) combined lexicon features with BERT and found improved results in four different datasets. Similarly, De Bruyne et al. (2021) combined Dutch transformer-based BERT models named BERTje and RobBERT with lexicon-based methods, resulting in a marginal performance improvement. There are many emotion lexicons have been developed in English but not that much available in Bangla. Mohammad and Turney (2013a) has developed a crowdsourced English lexicon named NRC EmoLex consists of 14,000 lemmas labeled in Plutchik (1980) eight basic emotions and two sentiments. Abdaoui et al. (2017) used a semi-automatic translation method to construct French expanded emotion lexicon named FEEL, from the NRC Emolex.

3 Methodology

3.1 Text Preprocessing

In this paper, we have experimented with both raw and preprocessed texts. We have used (Islam et al., 2022) dataset for fine tuning our transformers which is built from user comments from various social media sites. To retain relevent features and to eliminate unnecessary complexities, we have preprocessed the texts. The preprocessing steps are shown in Figure 1. As social media data's contains many hyperlinks and user mentions , we have removed them from our texts. Emojis and Emoticons convey rich emotional information. To standardize their representation, we replaced emoticons and emojis with their corresponding word formats. Punctuation marks, such as question mark and exclamation marks carry significant emotional information, we have replaced them with special keyword tokens and removed other insignificant punctuation. We have removed special symbols and stopwords in the preprocessing stage.

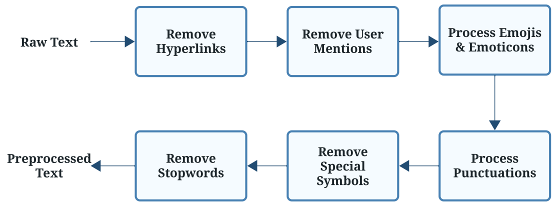

Figure 1: Text Preprocessing Steps

3.2 Development of Emotion Lexicon

We have developed a Bangla Emotion Lexicon named BEmoLex, consisting of 5336 lexicons across 9 emotion classes: Love, Joy, Surprise, Anger, Sadness, Fear, Disgust, Trust, and Anticipation. A semi-automatic translation (Abdaoui et al., 2017) method was followed to generate the Bangla lexicon by leveraging existing English lexical resources, especially the NRC Emolex (Mohammad and Turney, 2013b). The initial automated translations done by Google Translate[1] were subsequently reviewed and validated by three human translators who are proficient in both English and Bangla. The translators can add a new term, remove an existing term, or make necessary adjustments to ensure an accurate representation of emotions considering the cultural and linguistic nuances of the Bangla language. To enhance the coverage and diversity of the lexicon, we have manually incorporated handpicked strong emotive words, and expanded terms with Bangla synonyms.

Table 1 provides an overview of the emotion lexicon, detailing the count of lexicons distributed across various emotion categories. The data highlights substantial coverage in the Anger, Sadness, Fear, and Anticipation classes, each containing an extensive lexicon count, surpassing 700 entries. In contrast, the Love and Surprise categories exhibit relatively lower lexicon counts, with 356 and 301 entries, respectively. The Trust, Joy, and Disgust categories are relatively balanced, each contributing approximately 10% of the lexicon entries, ensuring a comprehensive representation of emotions within the dataset.

3.3 Development of Hybrid Model

In this section, we presents our hybrid model called BEmoLexBERT by integrating lexicon features with transformer-based models. We have used pre-trained BERT models and fine-tuned them for multilevel emotion classification. We preprocessed each raw text and tokenized them. The tokens were given as input into the BERT layer as shown in Figure 2.

For each target text, a lexicon vector is created, and each vector is appended to the lexicon encoding. The dimensions of the vector align with the

[1] https://translate.google.com/

Love	Joy	Surprise	Anger	Sadness	Fear	Disgust	Trust	Anticipation
356	603	301	756	788	752	513	543	724

Table 1: Number of Lexicons in Each Emotion Category

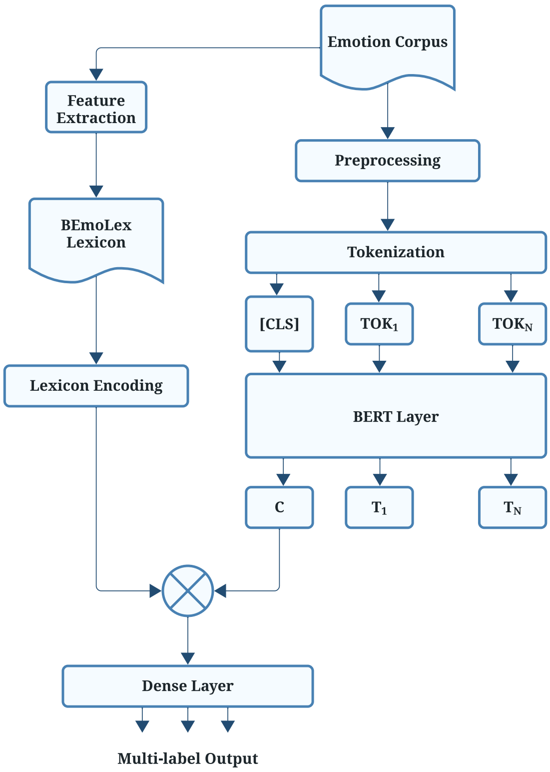

Figure 2: BEmoLexBERT, a hybrid model incorporating lexicon features with transformer

emotion categories in the training data. We perform a search for each term in the target text within the lexicon, incrementing the corresponding lexicon feature value for any matched emotion. The lexicon encoding vector is concatenated with the [CLS] token representation from the BERT layer, creating a comprehensive feature vector that effectively integrates both contextualized information and explicit emotion-related information. Finally, the combined feature vector is passed through a dedicated classification layer for Multilevel emotion classification.

4 Results and Analysis

We conducted a total of 24 experiments to evaluate the model we advocate. We have experimented with six pre-trained transformers as shown in Table 2 on both the plain model and hybrid model for both raw texts and preprocessed texts. The transformers were fetched from the Hugging Face [2] transformer library. Pretrained transformers are fine-tuned with 20,468 instances of the training section and tested with 2,272 instances of the testing section from the EmoNaBa dataset (Islam et al., 2022). Throughout the experiments, a consistent batch size of 8 was maintained. We have trained the models for 20 epochs and the learning rate was 2e-5. The experiments were carried out in a Google Colab [3] environment. The weighted F_1 score was selected as the primary evaluation metric.

The outcomes of our proposed hybrid model are presented in Table 3, showcasing a comparative analysis with a plain model concerning both raw and preprocessed texts. A noteworthy observation is that the preprocessing steps yielded only marginal enhancements over the raw text inputs. Remarkably, m-BERT demonstrated the most substantial improvement, achieving a 0.82% increase in the F1 score following preprocessing in the plain model. To further understand these marginal gains, we conducted a manual inspection of each preprocessing step. The results revealed that the primary improvement stemmed from the conversion of emojis and emoticons into textual forms, while the impact of the other preprocessing steps remained negligible. This observation underscores the proficiency of large language models in effectively handling noisy data, thereby minimizing the necessity of rigorous preprocessing. Furthermore, in the plain model, we can notice a slight improvement in the F_1 score when transitioning from XLM-R to XLM-R (large), showing an improvement of 3.22%. Similarly, the F_1 score increased by 3.42% when transitioning from BanglaBERT to BanglaBERT (large). Among the multilingual models, it is evident that they performed relatively lower compared to the monolingual models.

Across all transformers, the hybrid model demonstrated improved performance compared to the plain model. In the case of BanglaBERT, we observed a notable increase of 2.42% in the F_1 score for the hybrid model. For BanglaBERT

[2] https://huggingface.co/
[3] https://colab.google/

Pretrained Transformers	Parameters	Language	Reference
BanglaBERT	110M	Bangla	(Bhattacharjee et al., 2022)
BanglaBERT (large)	335M		
BanglishBERT	110M	Bangla & English	
mBERT	180M	Multilingual	(Devlin et al., 2018)
XLM-R (base)	270M	Multilingual	(Conneau et al., 2019)
XLM-R (large)	550M		

Table 2: Description of Pretrained Transformers Used for Experiments

Transformers	Plain Model				Hyrbrid Model			
	Raw Text		Preprocessed Text		Raw Text		Preprocessed Text	
	Acc	F_1	Acc	F_1	Acc	F_1	Acc	F_1
m-BERT	55.67	68.07	56.03	68.89	57.19	70.17	57.61	70.23
XLM-R	61.11	73.23	61.39	73.37	62.12	75.31	62.33	75.78
XLM-R (large)	62.92	76.66	62.81	76.59	63.15	77.61	63.07	77.35
BanglishBERT	62.35	76.39	62.97	76.86	64.33	77.94	64.59	78.27
BanglaBERT	64.84	77.96	65.03	78.25	66.39	80.07	66.73	80.67
BanglaBERT(large)	67.02	81.23	67.11	81.67	68.05	82.67	68.17	82.73

Table 3: Result comparison of pretrained transformers on plain model & hybrid model, both for raw texts & preprocessed texts. Acc denotes Accuracy in percentage and , F_1 denotes weighted F_1-score.

(large), a substantial 1.06% increase in F_1 score was achieved. Similarly, the performance of XLM-R (large) and XLM-R showed improvements with the hybrid model, presenting gains of 0.76% and 2.41%, respectively. Moreover, m-BERT and BanglishBERT displayed enhanced results, boasting improvements of 1.34% and 1.41% in F_1 score, respectively, when utilizing our hybrid model on preprocessed data. These consistent findings underscore the remarkable effectiveness of our proposed hybrid model, which skillfully incorporates lexicon features to deliver enhanced performance compared to the plain model.

We have manually identified some instances where the plain model failed but the hybrid model succeeded. One such example is the sentence, "ছোট বেলার এতো কাছের বন্ধু এমন মীরজাফর হয়ে যাবে ভাবি নি !" (I never thought that such a close childhood friend would become Mirjafar!). In the Bengali language, the word "মীরজাফর" (Mirjafar) is metaphorically used to describe someone who has deceived or cheated. The plain model detected the emotions of Sadness and Surprise in this sentence solely based on contextual analysis. In contrast, the hybrid model correctly identified it as a combination of Sadness, Anger, and Surprise emotions. The word "মীরজাফর" (Mirjafar) is classified under the Anger category in our emotion lexicon. The hybrid model, in addition to contextual analysis, effectively leveraged the lexical information of "মীরজাফর" (Mirjafar) within the sentence, enhancing its focus on the Anger emotion category. This example highlights the hybrid model's proficiency in capturing nuanced emotional cues by combining lexical and contextual information.

4.1 Comparison with existing works

In order to assess the effectiveness of our hybrid model, a comparison is conducted with existing techniques in the field. The previous methods (Pal and Karn, 2020; Rayhan et al., 2020; Das et al., 2021) are implemented on the EmoNaBa (Islam et al., 2022) dataset, and the outcomes are measured using the weighted F_1-score. To accommodate the multilabel emotion classification, necessary adjustments are made to convert the previous multiclass approaches (Pal and Karn, 2020; Das et al., 2021).

Methods	F_1	Reference
TF-IDF + LR	64.28	(Pal and Karn, 2020)
CNN + BiLSTM	68.57	(Rayhan et al., 2020)
XLM-R	73.23	(Das et al., 2021)
BEmoLexBERT	**82.73**	Proposed

Table 4: Performance comparison with existing works. F_1 denotes weighted F_1-score in percentage. The best score is denoted with bold letters.

In table 4 the results indicate that deep learning-based approaches (Rayhan et al., 2020) outperform machine learning-based approaches (Pal and Karn, 2020) in our test data. However, it is observed that the transformer based models (Das et al., 2021) demonstrating superior results than the deep learning methods and machine learning methods. Our proposed hybrid model, BEmoLexBERT, in combination with BanglaBERT(large) and the BEmoLex emotion lexicon, outperforms existing techniques for multi-label emotion classification in Bangla. It achieves an impressive F_1-score of 82.7%.

5 Conclusion

In this study, we have introduced BEmoLexBERT, a novel hybrid model that integrates transformers with lexicon features to enhance multilabel emotion detection in Bangla texts. A critical component of this work is the development of BEmoLex, a specialized emotion lexicon tailored to the nuances of the Bangla language. This lexicon encompasses a comprehensive repository of 5,336 unique lexicons, thoughtfully categorized into nine distinct emotion classes.

Our comprehensive evaluation, involving a comparative analysis between our proposed hybrid model and the plain model, underscores the efficacy of our approach in significantly enhancing emotion detection performance. Notably, the monolingual models outperformed their multilingual counterparts, while the examination of preprocessing steps revealed their marginal benefits, suggesting that large language models are proficient in managing noisy data, thereby reducing the necessity for extensive preprocessing.

Furthermore, we conducted a comparative assessment with other existing models, and the results underscored the state-of-the-art performance achieved by BanglaBERT(large) in conjunction with the BEmoLex lexicon. These findings collectively highlight the potential and significance of our approach in advancing multilabel emotion classification in the context of the Bangla language.

Limitations

The success of our approach heavily depends on the comprehensiveness of the emotion lexicon. Words and expressions that are not part of the lexicon may be overlooked, leading to inaccurate results. Lexicons require continuous updates and maintenance as languages evolve, and new words or expressions emerge. While transformers excel at understanding context, there might be cases where lexicon-based features do not align perfectly with the contextual analysis.

Ethics Statement

We acknowledge that bias in emotion classification models is an important ethical concern. We have conducted a meticulous review of our training data and lexicon to ensure that our models do not reinforce stereotypes or biases, taking into account the intricate linguistic and cultural intricacies of the Bangla language. Our lexicon, thoughtfully curated, is a testament to our commitment to respecting the rich cultural diversity and sensitivities of the Bangla-speaking community.

Acknowledgements

We sincerely appreciate the anonymous reviewers and our pre-submission mentor for their valuable insights and constructive suggestions, which contributed to the enhancement of this work.

References

Amine Abdaoui, Jérôme Azé, Sandra Bringay, and Pascal Poncelet. 2017. Feel: a french expanded emotion lexicon. *Language Resources and Evaluation*, 51(3):833–855.

Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1318–1327, Seattle, United States. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Avishek Das, Omar Sharif, Mohammed Moshiul Hoque, and Iqbal H Sarker. 2021. Emotion classification in a resource constrained language using transformer-based approach. *arXiv preprint arXiv:2104.08613*.

Luna De Bruyne, Orphée De Clercq, and Véronique Hoste. 2021. Emotional robbert and insensitive bertje: combining transformers and affect lexica for dutch emotion detection. In *Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), held in conjunction with EACL 2021*, pages 257–263. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Nafis Irtiza Tripto and Mohammed Eunus Ali. 2018. Detecting multilabel sentiment and emotions from bangla youtube comments. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–6.

Khondoker Ittehadul Islam, Tanvir Yuvraz, Md Saiful Islam, and Enamul Hassan. 2022. Emonoba: A dataset for analyzing fine-grained emotions on noisy bangla texts. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 128–134.

Anna Koufakou, Endang Wahyu Pamungkas, Valerio Basile, Viviana Patti, et al. 2020. Hurtbert: Incorporating lexical features with bert for the detection of abusive language. In *Proceedings of the fourth workshop on online abuse and harms*, pages 34–43. Association for Computational Linguistics.

Saif M Mohammad and Peter D Turney. 2013a. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465.

Saif M Mohammad and Peter D Turney. 2013b. Nrc emotion lexicon. *National Research Council, Canada*, 2:234.

Aditya Pal and Bhaskar Karn. 2020. Anubhuti–an annotated dataset for emotional analysis of bengali short stories. *arXiv preprint arXiv:2010.03065*.

Tanzia Parvin, Omar Sharif, and Mohammed Moshiul Hoque. 2022. Multi-class textual emotion categorization using ensemble of convolutional and recurrent neural network. *SN Computer Science*, 3(1):62.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

Md Maruf Rayhan, Taif Al Musabe, and Md Arafatul Islam. 2020. Multilabel emotion detection from bangla text using bigru and cnn-bilstm. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. IEEE.