UFAL-ULD at BLP-2023 Task 2 Sentiment Classification in Bangla Text

Sourabrata Mukherjee[1], Atul Kr. Ojha[2], Ondřej Dušek[1]
[1]Charles University, Faculty of Mathematics and Physics, Prague, Czechia
[2]Insight SFI Centre for Data Analytics, DSI, University of Galway, Ireland
{mukherjee,odusek}@ufal.mff.cuni.cz
atulkumar.ojha@insight-centre.org

Abstract

In this paper, we present the UFAL-ULD team's system for the BLP Shared Task 2: Sentiment Analysis of Bangla Social Media Posts. The Task 2 involves classifying text into Positive, Negative, or Neutral sentiments. As a part of this task, we conducted a series of experiments with several pre-trained sequence classification models – XLM-RoBERTa, BanglaBERT, Bangla BERT Base and Multilingual BERT. Among these, our best-performing model was based on the XLM-RoBERTa-base architecture, which outperforms baseline models. Our system was ranked 19th among the 30 teams that participated in the task.

1 Introduction

Sentiment analysis, the task of determining the sentiment expressed in textual data, is a critical component of natural language processing (NLP). It plays a vital role in understanding public opinion, social media trends, and user sentiments in various languages. In the context of the Bangla language, sentiment analysis poses unique challenges due to the language's specific characteristics such as complex morphology, making it an intriguing area of research. Sentiment analysis classification in Bangla remains less explored in comparison to English for reasons ranging from the non-availability of the datasets to lack of development of models or shared tasks (Rosenthal et al., 2017; Patwa et al., 2020; Barnes et al., 2022).

This paper presents our system developed for the BLP Shared Task 2 (Hasan et al., 2023a,b). Our objective is to provide a comprehensive description of our approach and results. We aim to contribute valuable insights and techniques to the field of Bangla sentiment analysis. Our team conducted a series of experiments utilising several pre-trained sequence classification models where we contributed to fine-tuning and hyper-parameter tuning to optimize the performance of these models.[1] We further employed focal loss to counter class imbalance in the data. The resulting system placed 19th out of 30 submissions in the shared task.

2 Related Works

Several works on Bangla sentiment were released prior to the present shared task challenge. Ali et al. (2020) introduces "BanglaSenti", a dataset of 61,582 Bangla words for sentiment analysis, originally developed for social media, with potential uses in emotion detection and opinion mining. This dataset's polarity categorizations are crucial for understanding sentence sentiment. Kabir et al. (2023) provides "BanglaBook," a dataset of 158,065 Bangla book reviews categorized as positive, negative, or neutral. Pre-trained models outperform manually crafted features, underscoring the need for more training resources in Bangla sentiment analysis, and an error analysis reveals common classification mistakes in under-resourced languages. Patra et al. (2015) focuses on Sentiment Analysis in Twitter for Indian languages, including Bangla, Hindi, and Tamil. The paper presents the first sentiment analysis attempt for these languages and ranks participating teams based on accuracy, achieving a maximum accuracy of 55.67% for Hindi. Bhowmick and Jana (2021) explores sentiment analysis in Bangla using pre-trained transformer models, achieving state-of-the-art performance with a maximum accuracy of 95% for a two-class sentiment classification task, setting a benchmark for Bangla sentiment analysis.

3 Dataset

The data for the BLP Shared Task 2 (Hasan et al., 2023a) is curated from two primary sources (Islam et al., 2021; Hasan et al., 2023a): MUltiplatform BAngla SEntiment (MUBASE) and SentNob

[1]Our code is available at https://github.com/souro/classification_tasks_bangla

Proceedings of the First Workshop on Bangla Language Processing (BLP-2023), pages 336–339

datasets. These datasets collectively contribute diverse textual data for sentiment analysis in the Bangla language. Each item of the data is annotated as having a positive, neutral or negative sentiment. Automating such annotation is the subject of BLP Shared Task 2.

The MUBASE dataset offers a substantial multiplatform collection. It includes a range of textual data, such as Tweets and Facebook posts, each meticulously annotated with their respective sentiment polarity labels. The SentNob dataset comprises public comments extracted from various social media platforms, specifically associated with news and video content. This dataset encompasses a broad spectrum of 13 distinct domains, spanning subjects such as politics, education, and agriculture.

The datasets were separated into sections for training, development, development-test and test (for final evaluation), consisting of 35,266, 3,934, 3,426 and 6,707 comments respectively.[2]

4 Experiments

This section provides a detailed description of our system's design, methodology, and the steps taken to achieve competitive results in the BLP Shared Task 2.

4.1 Data Preprocessing and Cleaning

Our system began with data preprocessing and cleaning using techniques provided by BNLP Shared Task 2 organisers (Hasan et al., 2023a). In addition to basic processing, we paid special attention to Unicode handling: We used the Bangla NLP toolkit to fix and normalize all Unicode characters into the NFKC normalization form.[3]

4.2 Model Selection

To develop our sentiment analysis system, we experimented with several pre-trained sequence classification models in a constrained setting.

We fine-tuned these models using various pretrained masked language models derived from the BERT (Devlin et al., 2019) architecture: XLM-RoBERTa (base and large versions) (Conneau et al., 2019), BanglaBERT '' (Bhattacharjee et al., 2022), Bangla BERT Base (Sarker, 2020) and BERT-base-multilingual-cased (Devlin et al., 2018). Among these, our best-performing model was based on the XLM-Roberta-base (Conneau et al., 2019) architecture. [4]

4.3 Hyperparameter Tuning

We conducted hyperparameter tuning on the development data to optimize our model's performance. The best hyperparameter settings we utilized are as follows: batch size 5, learning rate 1e-5, AdamW optimizer (Loshchilov and Hutter, 2019), 15 epochs gradient clipping (`clip_grad_norm`) 1.0, weight decay 0.01, dropout rate 0.1.

4.4 Handling Class Imbalance: Focal Loss

To address class imbalance issues, we experimented with oversampling and undersampling techniques. Although we obtained promising results using these methods, we eventually discovered an alternative approach that improved our results even further – focal loss.

Focal loss (Lin et al., 2017) is a specialized loss function designed to address class imbalance and focus on hard-to-classify examples. Specifically, we used the following parameters for the focal loss: $\alpha = 1, \gamma = 2$. Focal loss provided superior results compared to the simple cross-entropy loss, leading us to integrate it into our best-performing model.

5 Results

The official evaluation metric for the BLP Shared Task 2 is micro-F1 (Hasan et al., 2023a). This metric is commonly used in multi-class classification tasks, including sentiment analysis. It combines precision and recall to provide an overall assessment of the system's performance across all sentiment classes. The performance of our sentiment analysis system in the BLP Shared Task 2 is a testament to the effectiveness of our approach. We achieved a micro-F1 score of 0.6768 (see Table 1) on the evaluation dataset, substantially outperforming baselines, but placing 19th in the shared task out of 30 entries.

6 Conclusion

In this paper, we detailed the approach and methodologies we used for the BLP Shared Task 2: Senti-

[2] https://github.com/blp-workshop/blp_task2/tree/main/data

[3] We use the `fix_unicode=True`, `unicode_norm=True` and `unicode_norm_form="NFKC"` parameters.

[4] We use the models from HuggingFace: https://huggingface.co/xlm-roberta-base, https://huggingface.co/xlm-roberta-large, https://huggingface.co/csebuetnlp/banglabert, https://huggingface.co/sagorsarker/bangla-bert-base, https://huggingface.co/bert-base-multilingual-cased.

Model	micro-F1
Random Baseline	0.3356
Majority Baseline	0.4977
n-gram Baseline	0.5514
BLP Shared Task 2 winning system	0.7370
Our system	0.6768

Table 1: UFAL-ULD team and baseline systems results

ment Analysis of Bangla Social Media Posts. Our system demonstrated a strong performance, achieving a micro-F1 score of 0.6768, signifying its proficiency in classifying Bangla social media posts into Positive, Negative, and Neutral sentiments. Our team's system was ranked 19th among the 30th teams that participated in the task.

Through data preprocessing, model selection, hyperparameter tuning, and the incorporation of advanced techniques such as Focal Loss, we optimized our system to excel in sentiment analysis tasks for the Bangla language. Our results substantially outperformed baseline models, underscoring the effectiveness of our strategies.

Our work contributes to the advancement of sentiment analysis in Bangla social media, enabling a deeper understanding of user sentiment and trends in the Bangla-speaking community. We believe that our system's success marks a significant step toward improving sentiment analysis in underrepresented languages, and we look forward to further advancements in this field.

Limitations

While our system achieved competitive results, it is essential to acknowledge its limitations. First, our system's performance may vary depending on the specific characteristics of the social media posts and the domains they pertain to. Further fine-tuning and adaptation may be required for specialized applications.

Second, the dataset used for this shared task, although diverse, may not encompass the full breadth of Bangla social media discourse. As such, our system's performance may be influenced by potential biases in the training data.

Finally, while we have strived to optimize our system's performance, there may be room for further improvements through the exploration of alternative models, techniques, or additional linguistic resources.

Ethics Statement

We adhere to ethical guidelines in conducting our research and participating in the BLP Shared Task 2. We have respected privacy and data protection principles throughout our work, ensuring that any data used in our experiments adheres to appropriate consent and privacy regulations.

Furthermore, our system's output is intended for research and analysis purposes only. We emphasize responsible use and interpretation of sentiment analysis results, recognizing that automated sentiment analysis tools can influence decision-making and public perception.

We are committed to transparency and open collaboration in the field of natural language processing and sentiment analysis. We encourage ethical research practices and advocate for fairness, accountability, and transparency in AI and NLP technologies.

Acknowledgements

This research was supported by the European Research Council (Grant agreement No. 101039303 NG-NLG) and by Charles University projects GAUK 392221 and SVV 260575. We acknowledge of the use of resources provided by the LINDAT/CLARIAH-CZ Research Infrastructure (Czech Ministry of Education, Youth, and Sports project No. LM2018101).
Atul Kr. Ojha would like to acknowledge the support of the Science Foundation Ireland (SFI) as part of Grant Number SFI/12/RC/2289_P2 Insight_2, Insight SFI Centre for Data Analytics.

References

Hasmot Ali, Md. Fahad Hossain, Shaon Bhatta Shuvo, and Ahmed Al Marouf. 2020. Banglasenti: A dataset of bangla words for sentiment analysis. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–4.

Jeremy Barnes, Laura Oberlaender, Enrica Troiano, Andrey Kutuzov, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, and Erik Velldal. 2022. SemEval 2022 task 10: Structured sentiment analysis. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1280–1295, Seattle, United States. Association for Computational Linguistics.

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.

Anirban Bhowmick and Abhik Jana. 2021. Sentiment analysis for Bengali using transformer based models. In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, pages 481–486, National Institute of Technology Silchar, Silchar, India. NLP Association of India (NLPAI).

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.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*, Minneapolis, MN, USA. ArXiv: 1810.04805.

Md. Arid Hasan, Firoj Alam, Anika Anjum, Shudipta Das, and Afiyat Anjum. 2023a. Blp-2023 task 2: Sentiment analysis. In *Proceedings of the 1st International Workshop on Bangla Language Processing (BLP-2023)*, Singapore. Association for Computational Linguistics.

Md. Arid Hasan, Shudipta Das, Afiyat Anjum, Firoj Alam, Anika Anjum, Avijit Sarker, and Sheak Rashed Haider Noori. 2023b. Zero- and few-shot prompting with llms: A comparative study with fine-tuned models for bangla sentiment analysis.

Khondoker Ittehadul Islam, Sudipta Kar, Md Saiful Islam, and Mohammad Ruhul Amin. 2021. SentNoB: A dataset for analysing sentiment on noisy Bangla texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3265–3271, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Mohsinul Kabir, Obayed Bin Mahfuz, Syed Rifat Raiyan, Hasan Mahmud, and Md Kamrul Hasan. 2023. BanglaBook: A large-scale Bangla dataset for sentiment analysis from book reviews. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1237–1247, Toronto, Canada. Association for Computational Linguistics.

Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2999–3007. IEEE Computer Society.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *ICLR*, New Orleans, LA, USA. arXiv. ArXiv:1711.05101 [cs, math].

Braja Gopal Patra, Dipankar Das, Amitava Das, and Rajendra Prasath. 2015. Shared task on sentiment analysis in indian languages (SAIL) tweets - an overview. In *Mining Intelligence and Knowledge Exploration - Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings*, volume 9468 of *Lecture Notes in Computer Science*, pages 650–655. Springer.

Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 774–790, Barcelona (online). International Committee for Computational Linguistics.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

Sagor Sarker. 2020. Banglabert: Bengali mask language model for bengali language understanding.