A Hybrid Deep Learning Model For Sentiment Analysis In Bangla From Bangla-English Code-Mixed Data

Md. Shahriyar Hossain, Mohammed Julfikar Ali Mahbub, G M Sohanur Rahman, Shihab Sharar and Annajiat Alim Rasel

Dept. of Computer Science and Engineering, BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh
{*md.shahriyar.hossain, mohammed.julfikar.ali.mahbub, g.m.sohanur.rahman, shihab.sharar*}*@g.bracu.ac.bd,*
annajiat@bracu.ac.bd

Keywords: Sentiment Analysis, Code-Mixed, Deep Learning, NLP, BNLP, Bangla, Text processing

Abstract: Sentiment analysis research from Bangla text has had a surge in recent years. The barrier of having very little data for proper sentiment analysis is slowly lifting. Thus newer deep learning models are being approached by researchers to perform better sentiment analysis. However, Bangla being a more complex language, researchers cannot take advantage of the already well established English sentiment analysis models. On the other hand Hindi, a language similar to Bangla, already has a lot of successful research in its bucket regarding sentiment analysis. Quite a lot of it has utilized code-mixed data between Hindi-English languages. Since most people use English keyboards to express their emotion/sentiment in Bangla, the motivation for using code-mixed data for Bangla Language comes from here. This paper aims to introduce a new hybrid sentiment analysis model that can perform sentiment analysis from Bengali-English code-mixed text data. In addition to that, we have also implemented some of the unorthodox deep learning models with transfer models for Bangla text processing. We followed strategies used to detect Hindi code-mixed language and chose the optimal one based on research data. Then we further use this approach in similar fashion to perform Bangla sentiment analysis. In the end, we show a comparison of the new model performance in Bangla language as well English and Hindi. Additionally, we compare it with previous BNLP studies for sentiment analysis and find that our model has noticeable improvement.

REFERENCES

Aimal, M., Bakhtyar, M., Baber, J., Lakho, S., Mohammad, U., Ahmed, W., and Karim, J. (2021). Identifying negativity factors from social media text corpus using sentiment analysis method. *arXiv preprint arXiv:2107.02175*.

Alam, F., Hasan, A., Alam, T., Khan, A., Tajrin, J., Khan, N., and Chowdhury, S. A. (2021). A review of bangla natural language processing tasks and the utility of transformer models. *arXiv preprint arXiv:2107.03844*.

Dang, N. C., Moreno-García, M. N., and De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3):483.

Dutta, S., Saha, T., Banerjee, S., and Naskar, S. K. (2015). Text normalization in code-mixed social media text. In *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, pages 378–382. IEEE.

Garain, A., Mahata, S. K., and Das, D. (2020). Junlp@ semeval-2020 task 9: Sentiment analysis of hindi-english code mixed data using grid search cross validation. *arXiv preprint arXiv:2007.12561*.

Hassan, A., Amin, M. R., Mohammed, N., and Azad, A. (2016). Sentiment analysis on bangla and romanized bangla text (brbt) using deep recurrent models. *arXiv preprint arXiv:1610.00369*.

Jahan, M., Ahamed, I., Bishwas, M. R., and Shatabda, S. (2019). Abusive comments detection in bangla-english code-mixed and transliterated text. In *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, pages 1–6. IEEE.

Mai, S., Zeng, Y., Zheng, S., and Hu, H. (2021). Hybrid contrastive learning of tri-modal representation for multimodal sentiment analysis. *arXiv preprint arXiv:2109.01797*.