

Transformer-based Sense Embeddings with Deep Learning Large Language Models for Low-Resource Language Disambiguation

Hlaudi Daniel Masethe, TUT.

Determining a word's accurate meaning in each context is known as Word Sense Disambiguation (WSD)[1], and it is one of the most significant problems in Natural Language Processing (NLP) [2]. This undertaking is particularly challenging for low-resource languages like Sesotho sa Leboa since there are few annotated corpora and linguistic resources available for them. This study explores the application of many transformer-based and deep learning models for WSD in Sesotho sa Leboa, with good results despite the language's resource constraints. This study employs a variety of deep learning architectures, including transformer-based models such as Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM), Bidirectional Gated Recurrent Units (BiGRU), and an LSTM-based Language Model (LSTMLM), as well as models like DistilBERT with Naive Bayes (DistilBERT & NB), DeBERTa, T5, and ALBERT[3][4]. The study makes use of the unique hardware characteristics of the T4 GPU to improve and optimize the runtime of deep learning language models, especially big transformers. The purpose of the NVIDIA T4 Tensor Core GPU is to speed up deep learning and machine learning operations. It works especially well for training and inferring big language models. Every phase entail making efficient use of software optimizations in addition to comprehending and utilizing hardware features. The BiGRU model outperformed other deep learning language models with an accuracy of 79%, demonstrating the effectiveness of bidirectional processing effectively capturing contextual information. With an accuracy of 70%, DeBERTa beat the other transformer-based large language models to enhance pre-training techniques that prioritize spatial and contextual embeddings.

Keywords: Word Sense Disambiguation, Sesotho sa Leboa, Low-Resourced Languages, Deep Learning, Transformer Models, RNN-LSTM, BiGRU, DeBERTa, NLP

References

- [1] Q. Yang, R. Li, Y. Li, and X. Gu, "Word Sense Disambiguation based on Sequence Topic Model using sense dependency," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2021-July, pp. 1–8, 2021, doi: 10.1109/IJCNN52387.2021.9533571.
- [2] B. Li, S. Lu, K. Xie, and Z. Wang, "Accelerating NLP Tasks on FPGA with Compressed BERT and a Hardware-Oriented Early Exit Method," in *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, Nicosia, Cyprus: IEEE, 2022, pp. 410–413. doi: 10.1109/ISVLSI54635.2022.00092.
- [3] H. Choi, J. Kim, S. Joe, and Y. Gwon, "Evaluation of BERT and Albert sentence embedding performance on downstream NLP tasks," in *Proceedings - International Conference on Pattern Recognition*, Milan, Italy: IEEE, 2020, pp. 5482–5487. doi: 10.1109/ICPR48806.2021.9412102.
- [4] A. Aziz Sharfuddin, M. Nafis Tihami, and M. Saiful Islam, "A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification," in *2018 International Conference on Bangla Speech and Language Processing, ICBSLP 2018*, Sylhet, Bangladesh: IEEE, 2018, pp. 1–4. doi: 10.1109/ICBSLP.2018.8554396.