**Review Sentiment Analysis Using BERT with Mixed Datasets**

**Urim Aliji**

**161538**

***Abstract***

Sentiment analysis plays a crucial role in decision making for organizations that have a broad user base. Some such organizations may include companies selling products, non-governmental organizations operating around the world, or even sports teams aiming to please their fans. People who interact with these organizations in whatever capacity may have vastly different approaches to text-based communication, and with the continuously-expanding reach of the Internet, these text-based interactions between companies and consumers may be noted as incredibly varied. This paper aims to replicate an environment that contains different kinds of user reviews positioned within varied contexts in order to see 1) the outcome of the results of studying these user reviews, and 2) the manner in which BERT models react to such data.

**1 Introduction**

The use of Sentiment Analysis (SA) as a tool for decision making among various organizations has seen a rapid increase, which creates a need to develop more robust systems in security but also, more importantly, in the accuracy of the models so that sentiment of large groups may be correctly analyzed and predicted.

Bidirectional Encoder Representations from Transformers (BERT) is a Google-made transformer-based technique of creating models specifically for Natural Language Processing (NLP). BERT models are fine-tuning models that show great robustness in many different fields of NLP. The vast array of pre-trained models that BERT provides has made the creation of competent models highly accessible.

The prevalence of datasets that are free for use in NLP and specifically in Sentiment Analysis has created an unparalleled environment of innovation. Such datasets may be used in various ways and the data may be parsed and selected at the user’s discretion to create specialized models.

The specific models used in the paper are discussed in the subsequent section. In the sections following that, the datasets and the results of running mixed models with BERT are further expanded upon.

**2 Model**

The classification model for encoding the reviews used a pre-trained small BERT model that uses 8 hidden layers with a hidden size of 512 and 8 attention heads. These models have shown great results and are of great use for systems that are limited in processing power. For pre-processing the data for both the classification and SA models, the English pre-processor by BERT [sic] was used, which is a pre-trained state of the art model for preparing the data for training. The specific pre-trained model uses transformer blocks similar to the bird base structure, which has 12 layers with a hidden size of 768 and with self-attention heads of 12.

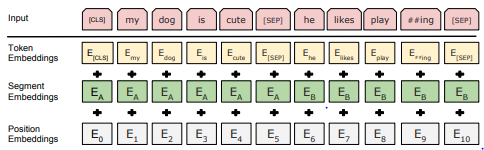


Figure 1: BERT Input Representation

The loss function for both models is Binary Cross Entropy (BCE,) which is fitting for these kinds of models, given that they only have two classes. Furthermore, the structure of the classification model is relatively simple, consisting of the aforementioned layers (viz. the pre-trained and the embedding layers,) one Dense layer that uses a Sigmoid activation function, and a Dropout layer.

Additionally, the structure of the SA model uses the same encoding method and the same BERT pre-trained model as a base. The exact structure of the model differs in that it uses four Dense layers[[1]](#footnote-0) and one Dropout layer.

**3 Dataset**

The classification model utilizes two separate datasets: 1) Stanford’s Large Movie Review Dataset (LMRD) and 2) Amazon’s Gourmet Food Reviews (GFR.) The LMRD consists of 25.000 highly polar and incredibly diverse movie reviews used for training, and the same amount used for testing. The data is pre-labelled with 1, representing a positive review, and a 0, which represents a negative. The GFR dataset, on the other hand, contains 500.000 reviews, and contrastive to LMRD, uses values 0-5 to represent customer satisfaction rather than positives and negatives. The latter dataset was normalized to have the same labels as LMRD for training purposes. During training, the data was uniformly selected from both. This was done in order to avoid muddling the results in the case of the model training more with one dataset rather than the other. Additionally, the SA model used the third dataset, OpinRank Data - Reviews from TripAdvisor and Edmunds, henceforth referred to as OCD, as a supplementary dataset. The specific car reviews collected from 2007, which were 19.000 in total, were labelled by the classification model trained on the previously stated datasets.

**4 Results**

Chart, line chart

Description automatically generated

Figure 2: Graph of loss and validation loss for the sentiment model.

Both models show promising results even when running for small amounts of epochs. The gradual decline of loss value in each epoch supports the claim that by letting the models run longer (by a sensible amount), they will deliver satisfactory results.

The binary accuracy of the model has a value of 0.51, which may be drastically improved with the use of more data during training and, as aforementioned, training for longer periods of time.

**References**

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). *BERT: Pre-training of deep bidirectional Transformers for language understanding*. arXiv.org. Retrieved October 22, 2021, from <https://arxiv.org/abs/1810.04805?source=post_page>.

Google-Research. (n.d.). *Google-Research/BERT: Tensorflow code and pre-trained models for Bert*. GitHub. Retrieved October 22, 2021, from <https://github.com/google-research/bert?fbclid=IwAR2IH8vndGLEa1lWJQSyOXQxEkkIFMD_kCVPyC8ARCYEHB28S4Sn2bVb9F4>.

Li, X., Bing, L., Zhang, W., & Lam, W. (2019, October 4). *Exploiting BERT for end-to-end aspect-based sentiment analysis*. arXiv.org. Retrieved October 22, 2021, from <https://arxiv.org/abs/1910.00883>.

*OpinRank data - reviews from TripAdvisor & Edmunds*. Kavita Ganesan, Ph.D. (2019, May 31). Retrieved October 22, 2021, from <https://kavita-ganesan.com/entity-ranking-data/?fbclid=IwAR3c-aRmrkO0Nbnhp2u3pMzgxRIC5_eAZfb3Tz73v0Xg6zOd-rdXIoFuwMs#.YXLe0e1R0uV>.

Project, S. N. A. (2017, May 1). *Amazon Fine Food Reviews*. Kaggle. Retrieved October 22, 2021, from <https://www.kaggle.com/snap/amazon-fine-food-reviews?fbclid=IwAR35qLwRYzcDb4XjX5QPZIR2V_YPKGj0CtIG-S-flHoH3bXVXnXDwdrvejA>.

*Sentiment classification on large movie review*. Kaggle. (n.d.). Retrieved October 22, 2021, from <https://www.kaggle.com/c/sentiment-classification-on-large-movie-review?fbclid=IwAR05-9fC9_faGSS-eS4JgGet5mW5PysaC-fa1zz8kQZi2F8taQ6SRN1XKsc>.

Sun, C., Huang, L., & Qiu, X. (n.d.). *Utilizing bert for aspect-based sentiment analysis via ...* Retrieved October 22, 2021, from <https://aclanthology.org/N19-1035.pdf>.

Tensorflow hub. (n.d.). Retrieved October 22, 2021, from <https://tfhub.dev/google/collections/bert/1?fbclid=IwAR1sWQJWoptSYzWT3e3zQqpZYiYHHEZAl8MEjlR4x4twwRQAD6ljqYfDqAE>.

Tensorflow hub. (n.d.). Retrieved October 22, 2021, from <https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/4>.

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019, May 4). *BERT post-training for review reading comprehension and aspect-based sentiment analysis*. arXiv.org. Retrieved October 22, 2021, from <https://arxiv.org/abs/1904.02232v2>.

1. The first three Dense layers have a pyramid shape that goes from 128 hidden connections to 256 and afterwards back to 128, ending with a layer that consists only of 1. [↑](#footnote-ref-0)