AMS 691.03: Deep Learning

Term Project Report

Vu Huynh - 115317761

Stony Brook University

Department of Applied Mathematics and Statistics

Academic Year: 2023-2024

Contents

Abstract

This project presents a focused study on the application of the Transformer Encoder architecture for sentiment analysis of Vietnamese text reviews. Due to the limited deep learning research on Vietnamese language processing, this work emphasizes model optimization and architectural efficacy. By employing a standard preprocessing procedure to prepare the dataset, the emphasis was placed on fine-tuning a Transformer Encoder model to effectively classify sentiments. The systematic hyperparameter tuning identified an optimal configuration, which demonstrated the capability of the Transformer architecture to handle the subtleties of sentiment classification, achieving an accuracy of 85.88% on the validation set and 84.53% on the test set. These results highlight the Transformer Encoder model's adaptability and strength in language tasks, even within the context of languages that have typically been underrepresented in natural language processing research.

Chapter 1

Introduction

Deep learning has emerged as a transformative force in the domain of natural language processing (NLP), particularly in sentiment analysis. This field, pivotal in interpreting and categorizing emotions in textual data, has greatly benefited from the advanced capabilities of deep learning techniques. These techniques enable a nuanced understanding of language semantics and context, essential in accurately deciphering sentiment from text.

In the context of the Vietnamese language, sentiment analysis poses unique challenges. Vietnamese, with its distinct syntax and semantic structure, often lacks the extensive datasets and language processing tools readily available for languages like English. This gap presents both a challenge and an opportunity for deep learning applications in Vietnamese sentiment analysis.

At the heart of this project is the transformer encoder architecture, renowned for its effectiveness in text understanding. The architecture's self-attention mechanism is adept at capturing contextual nuances, making it particularly suitable for sentiment analysis tasks. The application of this architecture to Vietnamese text aims to not only address the challenges posed by the language but also to leverage the model's strengths in capturing the subtleties of sentiment expression.

The primary objective of this project is to harness the transformer encoder architecture for classifying Vietnamese reviews into positive or negative sentiments. This endeavor is not merely a technical exercise but a step towards bridging the language processing gap in Vietnamese NLP research. It involves custom adaptations of the transformer model to suit the intricacies of the Vietnamese language and aims to enhance the accuracy and reliability of sentiment classification.

This report follows in a structured manner, beginning with a review of related work in the field, followed by a detailed exposition of the methods employed. Subsequent sections will present the experiments conducted and discuss their outcomes, culminating in a conclusion that encapsulates the key findings and implications of this research.

Chapter 2

Related Work

Sentiment analysis has evolved significantly with the advancement of machine learning techniques. Initially, methods like support vector machines (SVM) and Naive Bayes classifiers were predominant [3]. However, the complexity and subtlety of human language necessitated more advanced approaches.

The emergence of deep learning opened new avenues in sentiment analysis. Particularly, recurrent neural networks (RNN) and long short-term memory (LSTM) networks showed promising results in understanding sequential data [1]. Despite their success, these models faced challenges in processing long-range dependencies within text.

A pivotal development in natural language processing was the introduction of the Transformer model by Vaswani et al. [4]. This model, particularly its encoder component, revolutionized the understanding of text by employing a self-attention mechanism. This mechanism allows the model to weigh the importance of different parts of the input text, enabling a deeper understanding of context and semantics. The Transformer Encoder's ability to process sequences in parallel also marked an improvement over earlier sequence-based models like RNNs and LSTMs in terms of efficiency and performance.

The applicability of the Transformer Encoder architecture extends beyond English and has been explored in various languages, including Vietnamese. The Vietnamese language, with its unique syntactic and semantic characteristics, poses specific challenges for NLP models [2]. The adaptation of Transformer models for

Vietnamese text, therefore, represents a significant step in language-specific NLP research.

This project is inspired by these advancements and focuses specifically on employing the Transformer Encoder architecture for sentiment analysis of Vietnamese reviews. By leveraging the model's capacity for understanding complex language patterns, the project aims to accurately classify sentiments in Vietnamese text, contributing to the growing body of research in language-specific NLP applications.

Chapter 3

Methodology

This project employs the Transformer Encoder architecture to perform sentiment analysis on Vietnamese reviews. The Transformer Encoder, a key component of the Transformer model introduced by Vaswani et al. [4], is particularly well-suited for this task due to its ability to capture context and relationships within text.

3.1 Data Processing

The data processing pipeline for this project is to preparing Vietnamese reviews for sentiment analysis using the Transformer Encoder architecture. The dataset comprises Vietnamese reviews, each labeled as positive or negative.

3.1.1 Text Preprocessing

Once loaded, the text data undergoes a series of preprocessing steps implemented in the ‘preprocess_text‘ function. This function performs various cleaning tasks:

- Removal of URLs and HTML tags to eliminate irrelevant content.
- Stripping of punctuation and digits, focusing on textual content.
- Elimination of emojis, which are not crucial for sentiment analysis in this context.

These preprocessing steps are essential for reducing noise in the data and ensuring that the model focuses on linguistically relevant features.

3.1.2 Tokenization and Vocabulary Building

The preprocessed text data is then tokenized using a basic English tokenizer obtained from the ‘torchtext.data.utils‘ module. This tokenizer splits the text into individual tokens (words), a necessary step for text representation in the Transformer model.

Following tokenization, we build a vocabulary from the tokenized text data using the ‘build_vocab_from_iterator‘ function. This vocabulary serves as a mapping between tokens and their corresponding indices, a critical component for converting text into a numerical format that can be processed by the model.

3.1.3 Data Collation and Batch Preparation

Finally, the tokenized text data is collated into batches using the ‘collate_batch‘ function. This function pads each sequence in a batch to match the length of the longest sequence, ensuring uniformity in input size. The prepared batches, comprising tokenized and padded text along with their labels, are then ready for input into the Transformer Encoder model for training and evaluation.

The implementation of these data processing steps lays the foundation for effective sentiment analysis, ensuring that the input data is in an optimal format for the Transformer Encoder to analyze and classify.

3.2 Model Architecture

The Transformer Encoder model is central to this project’s approach to sentiment analysis. The architecture of the model, as implemented in the ‘model.py‘ file, is designed to effectively process and classify Vietnamese text. The model comprises several key components:

3.2.1 Positional Encoding

The model begins with a 'PositionalEncoding' module, which is crucial for capturing the sequence information in the text data. This module adds a positional encoding to the input embeddings, providing the model with information about the relative or absolute position of the tokens in the sequence. The positional encoding uses sine and cosine functions of different frequencies.

3.2.2 Embedding Layer

The input text is first passed through an embedding layer, which converts token indices into dense vectors of a fixed size. This layer is started with the vocabulary size and the embedding dimension, transforming the input tokens into continuous representations.

3.2.3 Transformer Encoder Layers

Following the embedding layer, the model employs a series of Transformer Encoder layers. These layers are built using the 'nn.TransformerEncoderLayer' class in PyTorch, configured with parameters such as the embedding dimension ('d_model'), the number of attention heads ('nhead'), the dimension of the feedforward network ('dim_feedforward'), and the dropout rate. The Transformer Encoder uses self-attention mechanisms to weigh the influence of different parts of the input sequence, allowing the model to capture complex dependencies in the text.

3.2.4 Adaptive Average Pooling

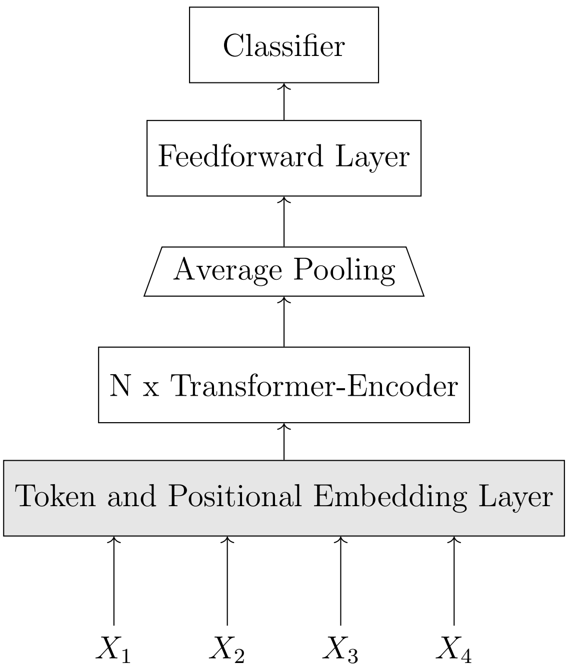
After the Transformer Encoder layers, the model applies an adaptive average pooling layer. This layer averages the encoder outputs, reducing their dimensionality and preparing them for the final classification layer. This step is crucial for distilling the essential features from the text.

3.2.5 Linear Layers and Classification

The model concludes with two linear layers for classification. The first linear layer reduces the dimension from ‘d_model‘ to a third of its size, followed by a ReLU activation function. The second linear layer maps this representation to two output units, corresponding to the positive and negative sentiment classes. The final output of the model is the sentiment classification for each input review.

This architecture, combining positional encoding, a Transformer Encoder, and classification layers, is specifically designed to handle the nuances of Vietnamese text and effectively classify sentiments in reviews.

3.2.6 Model Architecture Diagram



Chapter 4

Experimental Results

4.1 Experimental Setup

4.1.1 System Configuration

The experiments were conducted on a robust computational platform to ensure efficiency and performance. The system was configured with the following specifications:

```
| GPU | Name                     | Memory |
| 0   | NVIDIA A100-PCIe-40GB    | 40960MiB |
```

The utilization of NVIDIA A100 GPU, equipped with 40GB of memory each, facilitated the parallel processing of the deep learning model, providing an optimal environment for extensive computation and large model training.

4.1.2 Learning Rate Schedule

The training employed a learning rate decay strategy to optimize convergence. The initial learning rate was set at 0.001, with a scheduled decay to one-tenth of its value every 500 steps. This approach allowed the model to make large updates to the weights initially, for faster learning, and smaller updates as training progressed, for fine-tuning.

This setup provided a balance between rapid learning and the stabilization of the model's weights as it approached optimal performance, which is often necessary for training deep neural networks effectively.

4.2 Dataset and Preprocessing

The dataset comprised Vietnamese text reviews, intended for sentiment classification. Preprocessing steps were applied to prepare the data for training, including tokenization and sequence length truncation. The characteristics of the dataset after preprocessing are as follows:

- Vocabulary size: 33,337 unique tokens
- Maximum sequence length: 2,314 tokens
- Batch size: 64

These parameters were critical in shaping the input data for the model, ensuring the uniformity of sequence lengths and the comprehensive representation of the Vietnamese language's lexical diversity.

4.3 Hyperparameter Tuning

4.3.1 Hyperparameter Space

The hyperparameter search space was comprehensively defined to identify the most effective model configuration. The parameters were varied as follows:

- Learning rate (`lr`): {0.01, 0.001}
- Batch size: {64}
- Number of encoder layers: {2, 3}
- Dimension of the feedforward network (`dim_feedforward`): {128, 216, 512}

- Number of attention heads (`nhead`): {4, 8}

The search was conducted across all the combinations of these options, resulting in a comprehensive exploration of the defined hyperparameter space.

4.3.2 Best Hyperparameter Configuration

The search for the optimal hyperparameter configuration yielded the following best-performing parameters:

Hyperparameter	Value
Learning rate	0.001
Batch size	64
Number of encoder layers	3
Dimension of feedforward network	512
Number of attention heads	8

Table 4.1: Best hyperparameter configuration based on validation accuracy.

The model with these parameters achieved the highest validation accuracy of 85.88%, indicating its superior performance over other configurations within the search space.

4.4 Model Performance on Test Set

Utilizing the optimal hyperparameters from the tuning phase, the model's performance on the test dataset achieved a final accuracy of **84.53%**. The training epochs consistently showed progressive enhancements in training accuracy while decreasing the loss metric.

The training employed a batch size of 64 and a learning rate of 0.001. The model's architecture consisted of 3 encoder layers, a feedforward dimension of 512, and 8 heads in the multi-head attention mechanism. These parameters were identified as the superior combination within the hyperparameter space explored during the tuning process.

The test loss was reported as 0.3643, indicating the model's generalization capability on unseen data. The incremental training process, demonstrated over successive epochs, confirms the model's learning efficiency and stability.

Chapter 5

Conclusion

This study presented a deep learning approach to sentiment analysis of Vietnamese reviews using the Transformer Encoder architecture. The key contributions and findings of this work are as follows:

- We developed a robust preprocessing pipeline that normalized the Vietnamese text data and prepared it for deep learning applications, considering the unique linguistic features of the Vietnamese language.
- The Transformer Encoder model was successfully implemented and adapted to the sentiment classification task, demonstrating the model's capability to understand and classify sentiment with high accuracy.
- A comprehensive hyperparameter tuning process was undertaken, revealing that a deeper model with more attention heads and a larger feedforward network dimension leads to better performance, with the best configuration achieving an 85.88% accuracy on the validation set.
- The model's performance on the test set was robust, with a test accuracy of 84.53%, indicating good generalization from the model despite the complexity and nuances of Vietnamese sentiment analysis.

In conclusion, the Transformer Encoder model has proven to be a powerful tool for sentiment analysis, adapting well to the Vietnamese language. The findings of this study not only contribute to the growing body of research in sentiment analysis for low-resource languages but also pave the way for future work, which may include the exploration of different Transformer-based architectures, the inclusion of more granular sentiment categories, and the application of the model to other domains of Vietnamese text.

Bibliography

[1] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[2] Dat Quoc Nguyen, Thanh Vu, and Dai Quoc Nguyen. Improving vietnamese pos tagging and dependency parsing using bert embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4879–4885, 2019.

[3] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10:79–86, 2002.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.