An Encoder-Only Transformer for Sentiment Classification of Uzbek-Language Reviews

Bakhtiyor Bekmurodov Farkhod ogli
Farkhod Makhmudkhudjayev
Central Asian University
Machine Learning and Natural Language Processing
December 2025

Abstract

Sentiment analysis of user-generated reviews is essential for understanding customer satisfaction and improving digital services. Traditional methods such as bag-of-words and recurrent neural networks often fail to capture long-range dependencies and contextual information. In this work, we propose an encoder-only Transformer model for sentiment classification of Uzbek-language reviews. The model employs multi-head self-attention to capture contextual relationships without recurrence, and a Byte Pair Encoding tokenizer trained from scratch to handle the language's morphological features. The model was trained on **350,000 Uzbek-language user reviews**, demonstrating stable convergence with a final training loss of **0.1891**, and required approximately **22 minutes of GPU training on Google Colab**. In addition, the attention mechanism provides interpretability through token-level importance. These results show that lightweight Transformer models are effective for sentiment analysis in low-resource languages.

1. Introduction

With the rapid growth of e-commerce platforms and digital services, large volumes of user-generated reviews are produced daily. Analyzing these reviews is essential for understanding customer satisfaction, identifying service issues, and supporting data-driven

decision-making. Sentiment analysis, a key task in natural language processing (NLP), aims to automatically determine the emotional polarity of textual data.

Early sentiment analysis approaches relied on frequency-based representations such as bag-of-words (BoW) and TF-IDF, which ignore word order and contextual relationships. Later, neural network models including CNNs and RNNs were introduced to capture sequential patterns; however, RNN-based architectures suffer from limited parallelization and difficulty in modeling long-range dependencies.

The Transformer architecture, introduced by Vaswani et al., addresses these limitations through self-attention mechanisms that enable global contextual modeling and efficient parallel training. Motivated by these advantages, this work investigates an encoder-only Transformer model for sentiment classification of **Uzbek and Russian-language** user reviews, focusing on an efficient and lightweight architecture suitable for low-resource and multilingual settings.

The main contributions of this work are:

1. Implementation of an encoder-only Transformer trained from scratch.
2. Construction of a BPE-based tokenizer tailored to the dataset.
3. Empirical evaluation and analysis of model performance and parameter efficiency.
4. Demonstration of attention-based interpretability for text classification.

2. Related Work

Early sentiment analysis methods relied on manually engineered features and frequency-based representations such as bag-of-words and n-grams, which are limited in capturing semantic meaning and contextual dependencies. Neural approaches later introduced convolutional neural networks (CNNs) for local feature extraction and recurrent neural networks (RNNs), including LSTMs and GRUs, to model sequential information. However, RNN-based models suffer from limited parallelization and increased training complexity.

The Transformer architecture addressed these limitations by replacing recurrence with self-attention mechanisms, enabling efficient parallel computation and improved modeling of long-range dependencies. Encoder-only Transformer models, such as BERT, have since achieved strong performance across various NLP tasks, including sentiment analysis. Motivated by these successes, this work focuses on a lightweight encoder-only Transformer designed for training from scratch on a domain-specific dataset.

3. Dataset and Preprocessing

3.1 Dataset Description

The dataset consists of approximately **350,000 Uzbek-language user reviews** collected from the Uzum marketplace platform. Each review is associated with a user rating, which is mapped into three sentiment classes: **negative, neutral, and positive**. Reviews with a length of **80 characters or more** were removed to maintain consistent input length and reduce computational overhead.

3.2 Text Normalization

Text preprocessing is essential for improving model performance and ensuring consistent tokenization. The following normalization steps were applied:

- Removal of noise and unsupported symbols
- Retention of Latin and Cyrillic characters (Uzbek and Russian), digits, and whitespace
- Normalization of accented and special characters to standardized Uzbek forms

These steps reduce vocabulary fragmentation and produce clean, normalized text.

3.3 Tokenization and Padding

A **Byte Pair Encoding (BPE)** tokenizer was trained from scratch on the cleaned dataset with a vocabulary size of **30,000**. BPE effectively handles rare words and subword units, which is

particularly beneficial for morphologically rich languages. Each review was tokenized and **padded or truncated to a fixed length of 80 tokens**, ensuring uniform input dimensions.

4. Model Architecture

4.1 Encoder Architecture

The model adopts an **encoder-only Transformer architecture** composed of a stack of identical layers. Each input token is first mapped to a continuous vector through a learned embedding layer. To preserve sequence order, **positional embeddings** are added to the token embeddings at the bottom of the encoder.

Each encoder layer consists of two sub-layers: a **multi-head self-attention mechanism** and a **position-wise fully connected feed-forward network**. A residual connection followed by layer normalization is applied around each sub-layer to facilitate optimization and improve training stability.

4.2 Self-Attention Mechanism

Self-attention computes representations of a sequence by relating each token to all other tokens. Given queries **Q**, keys **K**, and values **V**, attention is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

where dkd_kdk denotes the dimensionality of the key vectors. **Multi-head attention** extends this mechanism by allowing the model to jointly attend to information from multiple representation subspaces, improving its ability to capture diverse contextual relationships.

4.3 Output Representation and Classification

The final encoder layer produces contextualized token representations. These representations are aggregated via **mean pooling** over the sequence dimension to obtain a fixed-length sentence embedding. The pooled representation is then fed into a linear projection layer to produce logits corresponding to the three sentiment classes.

4.4 Parameterization

Table 1 summarizes the number of trainable parameters in each component of the model.

Table 1. Model Parameter Breakdown

Parameter	**Value**
Vocabulary size	30,000
Embedding dimension	64
Number of encoder layers	4
Number of attention heads	4
Feed-Forward network dimension	64
Dropout rate	0.2
Maximum sequence length	80 tokens
Batch size	64
Learning rate	3x10^-3
Number of classes	3
Number of training iterations	10
Total trainable parameters	**2,123,459**

5. Experiments and Results

The model was trained for **10 epochs** using **AdamW** with a learning rate of 3×10−33 \times 10^{-3}3×10−3 and a batch size of 64 on a GPU, taking approximately **25 minutes**. Training

showed **stable convergence**, reaching a final loss of **0.1891**. Qualitative inspection indicates that the model reliably classifies reviews into negative, neutral, and positive sentiments.

6. Discussion

A key advantage of the proposed model is the **self-attention mechanism**, which allows interpretability by highlighting the tokens that most influence predictions. However, the use of **mean pooling** may dilute important token-level information in some cases. Future work could explore incorporating a special classification token, using pretrained embeddings, or evaluating the model on additional datasets to further improve performance and interpretability.

7. Conclusion

This work presented an **encoder-only Transformer** for sentiment classification of Uzbek-language reviews. Leveraging **self-attention** and **subword tokenization**, the model captures contextual information effectively and achieves stable convergence. The results demonstrate that **lightweight Transformers trained from scratch** can successfully handle sentiment analysis in low-resource language settings. Future work will focus on improving evaluation metrics and enhancing interpretability through attention visualization.

References

1. Vaswani, A., et al. *Attention Is All You Need*. NeurIPS, 2017.

2. Devlin, J., et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL, 2019.