Learning Emotion-Aware Representations Using Contrastive BERT for Tweet Clustering

1st Sad Yeamin Sayem
BRAC University
Dhaka,Bangladesh
sad.yeamin.sayem@g.bracu.ac.bd

I. Introduction

Emotion recognition in short text, such as tweets, has gained significant attention due to its wide-ranging applications in social media analysis, customer feedback, and mental health monitoring. Tweets, however, present unique challenges for emotion detection because of their brevity, informal language, and frequent use of slang or abbreviations. These factors make it difficult for traditional text representation methods to capture the underlying emotional meaning effectively.

This project proposes fine-tuning a BERT-based encoder with a contrastive loss on pairs of tweets labeled by emotion. The goal is to learn embeddings that better separate different emotions, thus enhancing clustering quality compared to conventional approaches like TF-IDF and GloVe embeddings. Experiments on the TweetEval emotion dataset demonstrate that this method leads to improved clustering metrics and more distinct emotional groupings.

II. Model Architecture

The model architecture combines a pretarined bert encoder with a lightweight projection head. The bert model generates a 768 dimensional representation of the text and then the projection head reduces it to 128 dimension . The projection head contains two layers and a ReLu activation gate and a dropout layer. A constructive loss function is used during training that creates sentence pairs to make better embedding of the similar texts.

III. Data Analysis

For this project the TweetEval dataset's emotion subset was used. This subset is for emotion analysis for classification tasks . The dataset consists of tweets labeled with 4 kinds of emotions , those categories are anger , sadness , joy and optimism. But the labels were not used during training as a clustering algorithm is used in this project. It was only used during evaluation to figure out the clustering quality.The first 100000 instances were taken for this task for computational feasibility. Each of the instances of the dataset contains two columns - text and label . Text contains the tweets and the labels . The dataset was a bit unbalanced. To solve this problem positive and negative pairs from the dataset were formed . Positive pair will contain two tweets containing the same emotion and the negative pair will contain two tweets containing tweets of different emotion labels . This pairwise

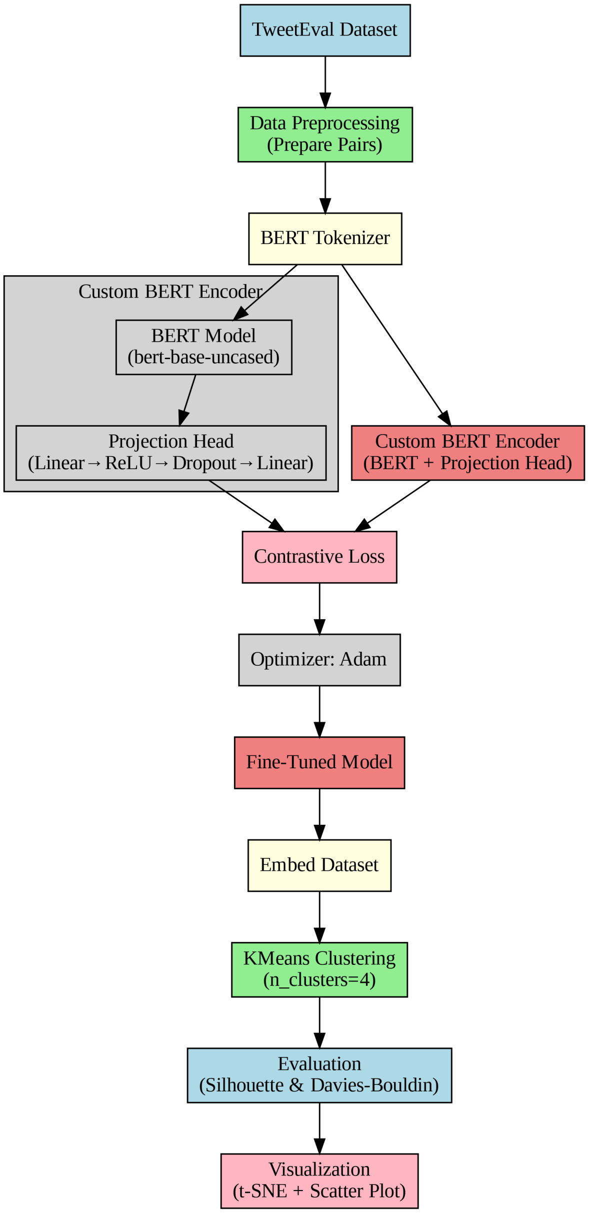
TweetEval Dataset
Data Preprocessing (Prepare Pairs)
BERT Tokenizer
Custom BERT Encoder
BERT Model (bert-base-uncased)
Projection Head (Linear → ReLU → Dropout → Linear)
Custom BERT Encoder (BERT + Projection Head)
Contrastive Loss
Optimizer: Adam
Fine-Tuned Model
Embed Dataset
KMeans Clustering (n_clusters=4)
Evaluation (Silhouette & Davies-Bouldin)
Visualization (t-SNE + Scatter Plot)

Fig. 1. Model block diagram

construction allows the contrastive loss to effectively learn emotion-aware representations by minimizing the embedding distance between semantically similar samples and maximizing it between dissimilar ones.

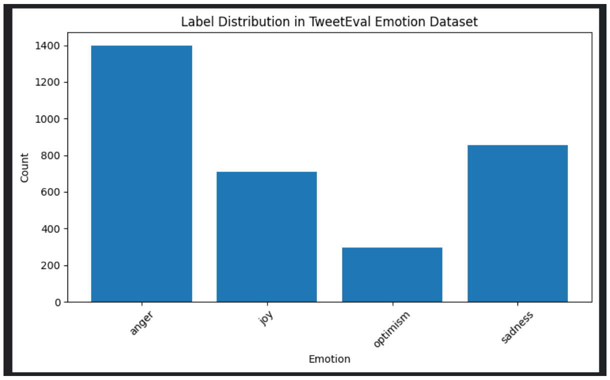

Fig. 2. Data visualization

IV. Methodology

A. Hyperparameter tuning and optimization

In this project, several key hyperparameters were optimized to ensure that the model achieved optimal performance for emotion-based clustering. Those hyperparameters are learning rates, batch size, the architecture of the projection head, margin in the contrastive loss function, and the number of clusters for K-Means. For this task, Adam optimizer was used and the learning was set to 1e-5. This value of the learning rate was found to be optimal after finding it most efficient through multiple trials and runs. The batch size was set to 32 for limited capacity of my gpu, ram and efficient performance. Moreover , To map BERT's 768-dimensional output into a lower-dimensional embedding space, a projection head was added. This component included two fully connected layers: the first reduced the vector size to 512 with a ReLU activation and a dropout of 0.3 to reduce overfitting. The second layer further reduced the size to 128 dimensions. This smaller embedding size made the clustering process more efficient and helped the model focus on key semantic differences between inputs. Finally , constructive loss margin is used to define how far apart dissimilar examples should be in the embedding space. After several trials and runs, the margin 1.5 gave the best separation between positive and negative pairs. As the dataset has four labels, the value of the number of clusters was set to 4. In the tuning process, instead of relying on automated hyperparameter tuning , the values were adjusted manually after trials and runs, observing the model's performance .

B. Model perameter, normalization and dropoput

The model is built on top of bert-base-uncased transformer and this model has 110 million parameters. The projection head may add few thousands more perameter but the main complexity of the model comes from the bert-base-uncased model . To reduce overfitting, a dropout layer (p=0.3) was added after the first layer of the projection head . This helped the model to do better generalization by randomly disabling some neurons during each training iterations. Additionally, the contrastive loss acts as a form of regularization — by helping the model to push dissimilar away from each other and pulling similar texts together , improving the models performence.

V. Results and Analysis

The Bert model outperformed traditional method of clustering like tf-idf and glove embedding. It achieved the highest Silhouette Score (0.1528) and the lowest Davies-Bouldin Score (1.6933), indicating better cluster quality and separation. In comparison, TF-IDF with PCA and GloVe averaging showed weaker performance, confirming the effectiveness of fine-tuned contextual embeddings for emotion clustering.

TABLE I
Clustering Performance Comparison on TweetEval Emotion Dataset

Method	Silhouette Score	Davies-Bouldin Score
Contrastive BERT	0.1528	1.6933
TF-IDF + PCA	0.0392	3.8790
GloVe Average	0.0664	2.7207

To better understand the clustering results, a two-dimensional visualization of the learned embeddings was created using t-SNE. This technique reduces the high-dimensional embeddings to two dimensions while preserving the local structure, allowing a visual inspection of how well the model separates different emotion clusters.

The plot below illustrates the distribution of the clustered data points, where each color represents a distinct cluster formed by the k-means algorithm.

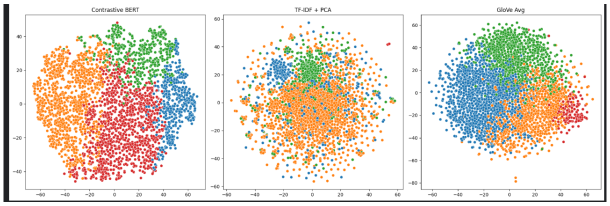

Fig. 3. T-SNE plot

VI. Limitations, challenges and Future work

One of the big challenges was that the dataset was noisy and had unmeaningful data. Moreover , the texts were short , which made it difficult to understand the similarities for the model . Also , as these are tweets , they mostly don't follow strict grammatical rules and it can affect berts understanding of the data. Another limitation was the computational power. It requires expensive computational power to run this model for a longer epoch. While the current approach demonstrates strong potential for emotion-based clustering using contrastively trained embeddings, there are several scopes for future research. One clear area for improvement is the scaling up of training. Due to hardware constraints, a subset of the available data was used. Training on the full dataset could help the model learn richer and more generalized emotional representations. Another promising work would be is to explore multi-view or multi-modal learning. Since social media posts often include images, videos, or audio clips alongside text, using these other forms of data in a multi-modal neural network could significantly enhance emotion detection, especially in cases where textual representation is ambiguous.