Name – Uttkarsh Rai
Group – G-43
Roll no. - 2101320100174

Title and Author Information

Title: Depression Classification From Tweets Using Small Deep Transfer Learning Language Models

Authors:

- Muhammad Rizwan
  - Department of Artificial Intelligence, Islamia University of Bahawalpur, Pakistan
  - Department of Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan
- Muhammad Faheem Mushtaq
  - Department of Computer Science and Information Technology, Islamia University of Bahawalpur, Pakistan
- Urooj Akram
  - Department of Computer Science and Information Technology, Islamia University of Bahawalpur, Pakistan
- Arif Mehmood
  - Department of Computer Science and Information Technology, Islamia University of Bahawalpur, Pakistan
- Imran Ashraf
  - Department of Information and Communication Engineering, Yeungnam University, South Korea
- Benjamín Sahelices
  - Department of Informatics, University of Valladolid, Spain

Abstract:

In this paper, the authors explored the classification of depression intensity using a labeled Twitter dataset. They conducted a detailed performance evaluation of four transformer-based pre-trained small language models: Electra Small Generator (ESG), Electra Small Discriminator (ESD), Xtreme Distil-L6 (XDL), and Albert Base V2 (ABV)—for classifying depression intensity using Tweets. The authors fine-tuned the models to achieve the best performance by applying different hyperparameters. They tested the models by classifying the depression intensity of labeled tweets for three label classes: 'severe', 'moderate', and 'mild', using downstream fine-tuning of the parameters. Evaluation metrics such as accuracy, F1 score, precision, recall, and specificity were calculated to assess the models' performance. The authors also performed a comparative analysis of these models with a moderately larger model, DistilBert, which has 67 million tunable parameters, for the same task and experimental setup. The results indicated that ESG outperformed all other models, including DistilBert, achieving the best F1 score of 89% with comparatively less training time. This outcome, according to the authors, is due to ESG's superior capability in capturing deep contextualized text representations. They further suggested optimizing ESG for use in low-powered devices. Overall, the authors concluded that their study contributes to achieving better classification performance in depression detection and aids in selecting the most suitable language model in terms of performance and training efficiency for Twitter-related natural language processing tasks.

Dataset Information

For their experiments, the authors used a dataset of depression-related tweets. They extracted these tweets using the Twitter public APIs, employing various depression-related hashtags as seed words. To annotate the tweets, they utilized Python libraries such as VADER and TextBlob, calculating quantitative sentiment polarity and subjectivity scores. The authors then filtered out tweets with low subjectivity to retain only opinionated tweets in their dataset.

The authors assigned labels to the tweets based on three depression intensity classes: 'mild', 'moderate', and 'severe', aligning with ICD-10 depression diagnostic criteria.

The distribution of the dataset is as follows:

- Mild: 29,931 tweets
- Moderate: 28,106 tweets
- Severe: 15,331 tweets

The authors split the dataset into training (70%), test (15%), and validation (15%) sets.

Parameters Used

The authors employed four small pre-trained language models: Electra Small Generator (ESG), Electra Small Discriminator (ESD), Xtreme Distil-L6 (XDL), and Albert Base V2 (ABV).

The hyperparameters used for fine-tuning these models were:

- Optimizer: Adam
- Loss function: Categorical cross-entropy
- Learning Rate: 2e-5, 5e-5, 8e-5
- Batch Size: 64
- Tokens Length: 64
- Training With: One Cycle Policy

Project Made

In this project, the authors focused on classifying the intensity of depression in tweets using small, transformer-based language models.

Their approach involved:

- Extracting and labeling depression-related tweets.
- Preprocessing the tweets to remove noise (URLs, mentions, hashtags, non-ASCII characters).
- Fine-tuning the language models (ESG, ESD, XDL, ABV) for the specific task of depression intensity classification.
- Evaluating the models' performance using metrics like accuracy, F1 score, precision, recall, and specificity.
- Comparing the performance of the small models with a larger model (DistilBert).

Gap Filled

The authors aimed to address certain gaps in previous research on depression detection using social media data.

- Multi-class classification: They noted that most earlier work focused on binary classification (depressed vs. non-depressed), whereas their study classified depression intensity into 'severe', 'moderate', and 'mild' categories.

Results with a Data Table Showing Accuracy

The paper does present accuracy results, but they are integrated into Table 6, which combines accuracy with other metrics like precision, recall, and F1-score. To give you a clear view, I'll extract the accuracy data from Table 6 and present it separately.

Here's how to interpret the table:

- Model: The specific language model used (ESG, ESD, XDL, ABV, DistilBert).
- Learning Rate: The learning rate that was used to train that model.
- Accuracy: The accuracy score achieved by the model on the test set.

Accuracy of Models on Test Set:

Model	Learning Rate	Accuracy
ESG	2.00E-05	0.88
ESG	5.00E-05	0.89
ESG	8.00E-05	0.9
ESD	2.00E-05	0.88
ESD	5.00E-05	0.89
ESD	8.00E-05	0.9
XDL	2.00E-05	0.84
XDL	5.00E-05	0.87
XDL	8.00E-05	0.88
ABV	2.00E-05	0.88
ABV	5.00E-05	0.89
ABV	8.00E-05	0.89
DistilBert	2.00E-05	0.88
DistilBert	5.00E-05	0.88
DistilBert	8.00E-05	0.88

Problems in This Project and Potential Improvements

The authors do imply that there is a problem that can be improved. Here are a few areas where the authors suggest potential improvements or where limitations of their study suggest room for improvement:

- Further Optimization of ESG for Low-Powered Devices: The authors specifically mention that Electra Small Generator (ESG) shows promise, but they propose "further optimization of ESG to make it suitable for low-powered devices". This suggests that while ESG is effective, it may not be ideal for resource-constrained applications in its current state.
- Generalizability: While the study achieves strong results on their specific Twitter dataset, the generalizability of these models to other social media platforms or different datasets could be explored further. Different social media platforms may have unique language styles and user behaviors, which could impact the models' performance.

- Dataset Bias: The authors used a dataset created from tweets with depression-related hashtags. Datasets created in such a way can introduce bias. Hashtags may not perfectly capture all instances of depression-related expression, and the sentiment analysis tools used for labeling (VADER, TextBlob) might have limitations. Improving the dataset creation process, perhaps with more diverse data sources or more refined labeling techniques, could be beneficial.
- Real-time Application: The study focuses on offline analysis of tweets. There could be further work on how to implement these models in real-time systems for immediate depression detection and intervention.