Persian Text-Based Emotion Detection

AmirMohammd Azadi - 99521046

Sina Zamani - 99521325

Abstraction:

This research project focuses on the challenging task of emotion detection in Persian text, aiming to enhance natural language understanding capabilities for applications in sentiment analysis, social media monitoring, and user engagement analysis in Persian-speaking communities. Emotion detection plays a crucial role in understanding the affective aspects of text data, providing valuable insights into users' sentiments and emotional states.

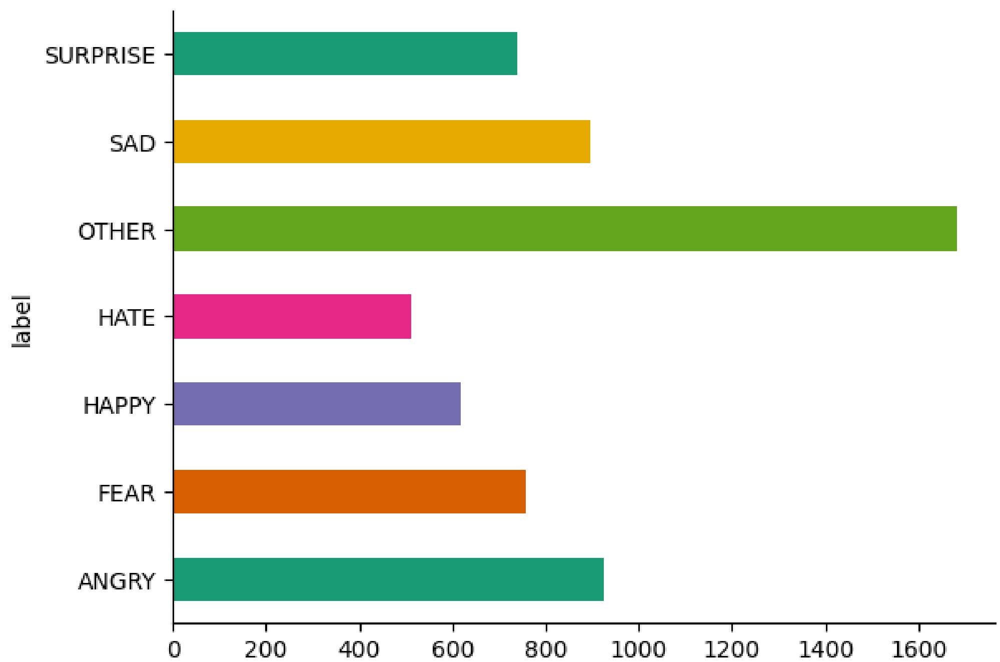
Introduction:

Emotion detection has become a critical component in natural language processing, offering deeper insights into the emotional context of textual data. While extensive work has been done in emotion detection for major languages, there is a noticeable gap in the literature when it comes to languages with non-Latin scripts, such as Persian. This paper addresses this gap by proposing and implementing an emotion detection model specifically tailored for Persian text.

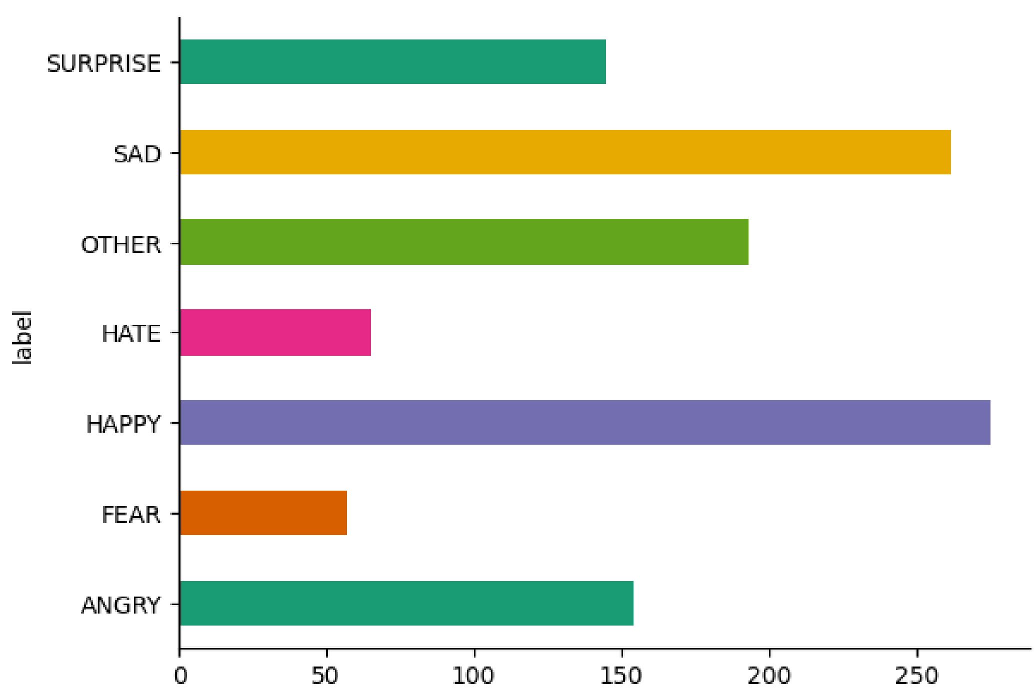
Datasets:

We have used ArmanEmo, a human-labeled emotion dataset of more than 7000 persian sentences labeled for seven categories. The dataset has been collected from different resources, including Twitter, Instagram, and Digikala comments. Labels are based on Ekman's six basic emotions (Anger, Fear, Happiness, Hatred, Sadness, Wonder) and another category (Other) to consider any other emotion not included in Ekman's model. The data is splited into train and datasets which have about 6000 and 1000 word, respectively.

Here is the number of samples of each class for train data:



And test data:



Preprocessing:

The columns of data didn't have a title. We assign 'text' as the title of the column of sentences and 'label' as the title of the column of labels. The label column in the data are string. We need to change it to numbers. So we map the labels to numbers from 0 to 6 and then change the 'label' column to appropriate number. We define a custom class dataset to prepare data for training, i.e. tokenizing, truncating and padding, while limiting the maximum length for sentences to 128.

Modeling:

As one of our baseline models, we take advantage of a pre-trained language model for Persian, known as ParsBERT. ParsBERT is a monolingual language model based on Bidirectional Encoder Representation Transformer (BERT) architecture. Farahani et al. have shown that the ParsBERT model outperforms the multilingual BERT and previous models in several Persian NLP downstream tasks, including text classification and sentiment analysis. Lighter than the multilingual BERT, ParsBERT has been trained on a larger and more diverse (in terms of the range of topics and style of writing) set of pre-trained Persian datasets.

We also use two variations of a model known as XLM-RoBERTa as our other baseline models. XLM-RoBERTa is another transformer-based masked language model pre-trained on text in 100 languages. This multilingual language model has led to state-of-the-art performance on cross-lingual classification, sequence labeling, and question answering, outperforming multilingual BERT (mBERT) on various cross-lingual benchmarks. Although it is known that ParsBERT as a monolingual language model outperforms multilingual BERT on various tasks in Persian language, we decided to compare the performance of XLM-RoBERTa variations (namely XLM-RoBERTa-base and XLM-RoBERTa-large) against ParsBERT on emotion detection from Persian text.

The initial phase of the research involves utilizing zero-shot learning, a methodology that allows the models to generalize to new and unseen classes without explicit training. This approach is particularly valuable in the context of emotion detection, as it enables the models to understand and categorize emotions in Persian texts without the need for extensive labeled data in the target language.

Following the zero-shot learning phase, the study proceeds to fine-tuning the transformer models on a labeled dataset of Persian emotional texts, i.e. ArmanEmo. Fine-tuning involves adapting the pre-trained models to the specific linguistic nuances and emotional expressions present in Persian, thereby improving their performance on the targeted emotion detection task. We used the default hyperparameters for models except for learning rate (1e-5) and We train the models for 3 epochs.

The paper provides a comprehensive overview of the experimental setup, including data preprocessing, model architecture details, and training procedures. Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the performance of the models. Comparative analyses are conducted to highlight the strengths and weaknesses of each transformer model in the context of Persian emotion detection.

Results:

The performance of the models are as follows:

Model	Precision (Macro)	Recall (Macro)	F1 (Macro)	Accuracy
ParsBERT (zero-shot)	0.13	0.15	0.08	0.15
XLM-Roberta-base (zero-shot)	0.03	0.14	0.05	0.23
XLM-Roberta-large (zero-shot)	0.02	0.14	0.03	0.13
ParsBERT (trained)	0.68	0.64	0.64	0.65
XLM-Roberta-base (trained)	0.65	0.66	0.65	0.66
XLM-Roberta-large (trained)	**0.76**	**0.73**	**0.74**	**0.75**

As shown in above table, the fine-tuned XML-RoBERTa-large model significantly outperforms other models on our test set in terms of average macro F1 score, precision, recall, and Accuracy.

The performance of the best model among our baseline models (XLM-RoBERTa-large) is summarized in below Table. The model achieves a macro average F1 score of 75.39 on the test set. It presents the best performance on emotions like Happiness, Fear, and Sadness. On the other hand, it obtains the lowest F1 score on emotions like Anger, Other, and Hatred.

Emotion	Precision	Recall	F1	Support (No. of Test Examples)
Anger	0.69	0.66	0.68	154
Fear	0.72	0.82	0.77	57
Happiness	0.90	0.74	0.81	275
Hatred	0.80	0.60	0.68	65
Other	0.60	0.80	0.69	193
Sadness	0.77	0.87	0.82	262
Wonder	0.83	0.61	0.71	145
Accuracy			0.75	1151
Macro Average	0.76	0.73	0.74	1151
Weighted Average	0.77	0.75	0.75	1151

We manually investigated the mislabeled sentences to better analyze the situations where the best model is performing poorly. Table 4 presents some randomly selected samples of these mislabeled sentences generated by the best model. Going through some of these examples, what can be inferred is that the model outputs the wrong label whenever a given sentence carries mixed emotions. In such situations, the assignment of one and only one exact emotion to the sentence might be challenging even for humans. That is why multi-label classifiers are used against multi-class classifiers to include the presence of more than one emotion in a given sentence.

Sentences	Ground Truth Label	Model Prediction
دیشب بعد از ارسال تویت مربوط به آثار باستانی تویت دیگری نوشتم ولی هرچه منتظر شدم ارسال نشد، از همون موقع تا الان تویتر نداشتم، ناراحت بودم که نکنه پیامی داده باشین ومن نبینم که الحمدالله خبری نیست خوب، چه خبر؟ من نبودم خوش گذشته؟	Happy	SAD
واقعا حال به هم زنه این حجم از داستان سرايي درباره تجاوز يا چيزهاي شبيه به اون براى جذب لايك و توجه	Hate	Angry
درود بر شرف آریایی ات آقای هومن سیدی. عجب فیلمی بود این مغزهای کوچک زنگ زده	Happy	Surprise
یه زمانی دخترخالهم میگفت یاد گرفتن زبان آلمانی زیادم سخت نیست. انگار کن انگلیسی رو تر زدن توش. اما الان که قصد یادگیریش رو دارم، حقیقتاً گرخیدم. نمیدونم سهماهه میشه به اون حد لازمش برسم یا نه	Fear	Sad
دیدم حمیدرضا عارف در مصاحبه با انصاف نیوز گله کرده که انتقادها به حرفهایش حاصل شنیدن حرفهای تقطیع شده بوده. دیدم راست میگوید. کل مصاحبه را نگاه کردم. راست میگفت. حرفهایش انتقاد برانگیز نیست. حال به هم زننده و تهوع آور است.	Hate	Angry
امشب گفت نامزدی دوستش که ادم روشنفکری است بهم خورده و دختر بشدت نگران حرف مردم گفتم وای به مردم روشنفکرمون که نگران مردمند پس تقصیری بر مادر حاشیه من نیست که بعد از یکسال و درگوشی درمورد تجاوز به دخترش حرف میزنه و ترس از ابرو داره	Sad	Fear

References:

1. Hossein Mirzaee, Javad Peymanfard, Hamid Habibzadeh Moshtaghin, Hossein Zeinali, ArmanEmo: A Persian Dataset for Text-based Emotion Detection
2. Mehrdad Farahani, Mohammad Gharachorloo, Marzieh Farahani, and Mohammad Manthouri. Parsbert: Transformer-based model for persian language understanding. Neural Processing Letters, 53(6):3831–3847, 2021.