PersianEmo: Enhancing Farsi-Dari Emotion Analysis with a Hybrid Transformer and Recurrent Neural Network Model

Mohammad Ali Hussiny, Mohammad Arif Payenda, lilja Øvrelid
University of Oslo (UiO), University of Agder (UiA), University of Oslo (UiO)
arif.payenda@uia.no, {mohamahu, liljao}@ifi.uio.no

Abstract

Emotion analysis is a critical research domain within the field of natural language processing (NLP). While substantial progress has been made in this area for the Persian language, there is still a need for more precise models and larger datasets specifically focusing on the Farsi and Dari dialects. In this research, we introduce "LearnArmanEmo" as a new dataset and a superior ensemble approach for Persian text emotion classification. Our proposed model, which combines XLM-RoBERTa-large and BiGRU, undergoes evaluation on LetHerLearn for the Dari dialect, ARMANEMO for the Farsi dialect, and LearnArmanEmo for both Dari and Farsi dialects. The empirical results substantiate the efficacy of our approach with the combined model demonstrating superior performance. Specifically, our model achieves an F1 score of 72.9% on LetHerLearn, an F1 score of 77.1% on ARMANEMO, and an F1 score of 78.8% on the LearnArmanEmo dataset, establishing it as a better ensemble model for these datasets. These findings underscore the potential of this hybrid model as a useful tool for enhancing the performance of emotion analysis in Persian language processing.

Keywords: Emotion, Farsi-Dari, Transformer, Recurrent Neural Network, LearnArmanEmo

1. Introduction

Humans express their feelings using various methods such as writing text, audio, video, images, etc. However, one of the most common methods is still writing text. With the increasing use of social networks and the advancement of technology, the expression of emotions through text has risen. Analyzing emotions becomes challenging when people express multiple emotions within a single text Sailunaz and Alhajj (2019). Numerous studies have been conducted across various languages, delving into the intricacies of emotion analysis. Nevertheless, there is an ongoing need for more advanced and accurate approaches. Emotion analysis holds immense potential not only for understanding human behavior but also for enhancing the efficiency of various applications, such as content recommendation systems, mental health monitoring, and customer experience enhancement Kim and Klinger (2018).

Within the Indo-Iranian language family, the Persian (Farsi in Iran, Dari in Afghanistan, and Tajik or Tajiki in Tajikistan) (Spooner, 2012) languages stand out with their unique linguistic structure and cultural context, presenting distinctive challenges and opportunities in the realm of sentiment and emotion analysis. Persian text is enriched with cultural idioms, poetic expressions, and subtle nuances, demanding specialized techniques for accurate emotion categorization. The surge in Persian content on the Internet and social media platforms underscores the pressing need for robust emotion analysis tools tailored to this language. Persian language dialects differ from each other. This discrepancy is regarded as a fundamental challenge in text analysis, especially from an emotional perspective. Further research is required to address this issue. To tackle this problem, we have combined two datasets: LetHerLearn Hussiny and Øvrelid (2023), which focuses on the Dari language, and ARMANEMO Mirzaee et al. (2022), which concentrates on the Farsi language. This approach aids in expanding the research area of emotion analysis in the Persian language. We merged the two mentioned datasets and released them as the "LearnArmanEmo" dataset for the Farsi-Dari dialect of the Persian language. In addition, we introduce a new and more accurate model for emotion analysis of the Persian language.

In section 2, we review relevant literature. Section 3 explains the dataset. Section 5 explains the implemented model and our proposed model. Section 6 presents the experimental setup and result. Finally, Section 7 summarizes our findings and conclusions.

2. Related Work

One of the approaches to emotion recognition was the use of lexicons such as Word-Net Affect and SentiwordNet, which apply linguistic rules and sentence structures Shivhare et al. (2015); Rahman et al. (2017). Some researchers used emotion detection methodologies based on corpora employ supervised learning techniques to extract sources of information, which are categorized from textual datasets containing a predefined set of emotions derived from theories like

SIGUL2024 Workshop, pages 257–263

Ekman, Parrot, and others Sailunaz and Alhajj (2019); Rachman et al. (2016); Wang and Pal (2015). Bandhakavi et al. (2017) illustrate how the use of a generative Unigram Mixture Model (UMM) can facilitate the simultaneous modeling of the emotional and neutral attributes of terms within labeled. Del Arco et al. (2020) constructed a multilingual dataset based on Twitter called Emo-Event, encompassing both English and Spanish languages, this dataset comprised 8409 labeled instances in Spanish and 7303 labeled instances in English. This research presented linguistic analyses and employed machine learning methods to discern emotions, achieving an accuracy of 0.64 for Spanish and 0.55 for English. The other approach to emotion detection is the use of machine learning algorithms that can learn to identify patterns in data and predict emotions expressed in text such as Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR), etc Pang and Lee (2004); Suhasini and Srinivasu (2020); Hasan et al. (2019). Jayakrishnan et al. (2018) developed machine learning algorithms to measure the magnitude of emotions in the Twitter dataset by distinguishing the intensity levels of four different emotional categories: ”Happiness”, ”Sadness”, ”Anger”, and ”Terror”. In many research papers, deep learning approaches such as LSTM, BiLSTM, GRU, CNN, and BERT models are addressed. Chatterjee et al. (2019) developed a model known as SS-BED for the detection of contextual emotions from textual dialogues and classified four emotion classes as ”Happy”, ”Sad”, ”Angry”, and ”Others” using two LSTM layers utilizing distinct word embedding matrices. Cortiz (2022) conducted an experiment that demonstrated the effectiveness of various transformer models for the task of emotion recognition. The authors implemented several Transformer language models, including BERT, DistilBERT, RoBERTa, XL-Net, and ELECTRA. These models were fine-tuned using a fine-grained emotion dataset that included 28 different emotion classes. Recently many researchers used hybrid approaches based on combined various methods, enhancing the likelihood of surpassing individual methods by leveraging their strengths while mitigating their respective Tzacheva et al. (2020); Ochsner and Gross (2005); Khanpour and Caragea (2018). (Ramalingam et al., 2018) a hybrid model incorporating both keyword-based and learning-based methods was developed, resulting in a remarkably high accuracy score for emotion recognition. Liu et al. (2019) has been widely used for different classification tasks, including emotion analysis, and allows modification in terms of the languages, amount of data, learning rates, and batch size.

While there has not been extensive prior research on emotion detection models and approaches in the Persian language, but there have been efforts focused on developing of emotion datasets. The ARMANEMO Mirzaee et al. (2022) dataset constituted an important step in this direction. It is based on the 7500 comments from social media, and the dataset was annotated using a mixture of manual and automatic steps into 7 classes.

The LetHerLearn dataset, as presented in Hussiny and Øvrelid (2023), comprises 7,600 emotional tweets gathered from Twitter using specific keywords related to the ban on education in Afghanistan. This dataset was manually annotated into 7 classes.

3. Datasets

In this study, we used two Persian datasets. One is called LetHerLearn, and it comes from Twitter. This set is about supporting the right to education for girls in Afghanistan. The LetHerLearn set we used has 7600 tweets. The authors considered seven different classes: ”Anger”, ”Disgust”, ”Fear”, ”Happiness”, ”Sadness”, ”Surprise”, and ”Other”. The other set is called ARMANEMO, and it was gathered from Twitter, Instagram, and comments on DigiKala. In ARMANEMO, there are also seven classes, but they have slightly different names: ”Anger”, ”Fear”, ”Happiness”, ”Hatred”, ”Sadness”, ”Wonder”, and ”Other”. The authors of ARMANEMO mentioned in their paper that the main dataset had 7500 sentences, but the available dataset only has 7274 instances. To make sure our new method is evaluated correctly, we kept the same number of emotion classes as in both sets. Both datasets have been annotated with Ekman’s Ekman (1992) method with seven distinct classes. The only distinction between the two datasets lies in the classification labels ”Disgust” and ”Surprise” used in LetHerLearn, which correspond to ”Hatred” and ”Wonder” in ARMANEMO, respectively. In the combined dataset, the label ”Hatred” is replaced with ”Disgust,” and ”Wonder” is replaced with ”Surprise.” All other classes remain consistent across LetHerLearn, ARMANEMO, and the LearnArmanEmo dataset. Table 1 presents the statistical report for LetHerLearn, ARMANEMO and LearnArmanEmo datasets.

4. Preprocessing

During the preprocessing stage, the ARMANEMO dataset underwent several cleaning and normalization steps, which involved removing irrelevant information such as URLs, links, hashtags, mentions, and HTML tags. Each record was normalized using the Persian text preprocessing tool called Hazm, and punctuation and digits were also

Dataset	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Other
ARMANEMO	1077	575	813	892	1158	884	1874
LetHerLearn	1727	569	606	1597	1280	490	1338
LearnArmanEmo	2804	1144	1419	2489	2438	1374	3212

Table 1: Statistical report of ARMANEMO,LetHerLearn, and LearnArmanEmo datasets

removed. The LetHerLearn dataset was already cleaned and did not need to be cleaned.

5. Proposed Approach

In our proposed model, we employed the XLM-RoBERTa-large model as an encoder to tokenize the input data and generate contextual word embeddings for each token. To regularize word embeddings, the result is fed into a spatial dropout layer in the dense vectors, which represent the contextual embeddings of each token. The Bi-GRU component accepts word embeddings and processes the long-range dependencies within the word embedding sequence. Subsequently, a deep attention mechanism scores the different parts of the sequence, highlighting informative regions. The attention layer's output is passed to a dense layer to extract complex relationships and significant patterns from processed embeddings. An additional dropout layer is added to ensure regularization. Finally, a classification layer with softmax activation is used to estimate the probability distributions of the different emotional classes. Figure 1 illustrates the overall workflow of our proposed approach.

XLM-RoBERTa-large: is a multilingual transformer model pre-trained on a vast corpus of text from multiple languages.

BiGRU: is a recurrent neural network model that is particularly adept at capturing sequential dependencies in textual data.

Dense Layer: the proposed model uses two dense layers. The first layer that has functionality is to capture the connection between the hidden state produced by the BiGRU layer and the class labels to facilitate feature extraction and representation. The second layer has functionality for the final classification process. The softmax activation function is used in this layer to transform the output values into a probability distribution.

Deep Attention layer: we incorporate a deep attention layer to improve the model's ability to focus on significant parts of the input data and to improve overall performance by effectively capturing relationships between data and class labels. This layer contains weights and biases that are initialized by the model during construction. It allows the model to compute attention scores and dynamically weight input features dynamically.

6. Experimental setup and Results

This section describes the experimental setup and results of our proposed approach for Persian text emotion analysis. We tested our models on LetHerLearn, ARMANEMO, and LearnArmanEmo datasets, considering the ultimate goal of accurately analyzing our proposed methods. Finally, the developed models are compared with the existing approaches to examine the proposed model's predictive performance.

6.1. Experiments

We implemented various models, including LSTM, BiLSTM, BiGRU, ParsBert, ParseBert + BiGRU, XLM-Roberta-Large, and XLM-Roberta-Large + BiGRU models. All neural network models made use of fastText (Grave et al., 2018) word embedding with 300 dimensions for the Persian language.

Neural Network Models: the neural network model has 128 neurons. Both dropout and recurrent dropout rates were set to 0.25. An additional layer of 64 neurons with the same dropout rates was added to each model. This was followed by another layer of 32 neurons using the Adam optimizer with a learning rate of 0.001.

ParsBERT: the hyperparameters were set for five epochs with a batch size of 32, using the Adam optimizer with β_1 = 0.9, β_2 = 0.999, and a learning rate of 2e-5.

ParsBERT + BiGRU: the hyperparameters for ParsBERT + BiGRU is two Bidirectional GRU layers with 256 and 128 units, with dropout values of 0.2 respectively. The model has 32 units of deep attention layer with 32 units, an added dense layer with 64 units with ReLU activation, and dropout layers with a rate of 0.2 are used to prevent overfitting, followed by another dense layer with the softmax activation function.

XLM-RoBERTa-large: the XLM-RoBERTa-large has 5 epochs, batch size of 32, learning_rate of 0.00001, and optimizer of AdamW.

XLM-RoBERTa-large + BiGRU: the XLM-RoBERTa-large + BiGRU model uses two Bidirectional GRU layers with 256 and 128 units, with dropout values of 0.2 respectively. The model has 32 units of deep attention layer with 32 units, an added dense layer with 64 units with tanh activation, and dropout layers with a rate of

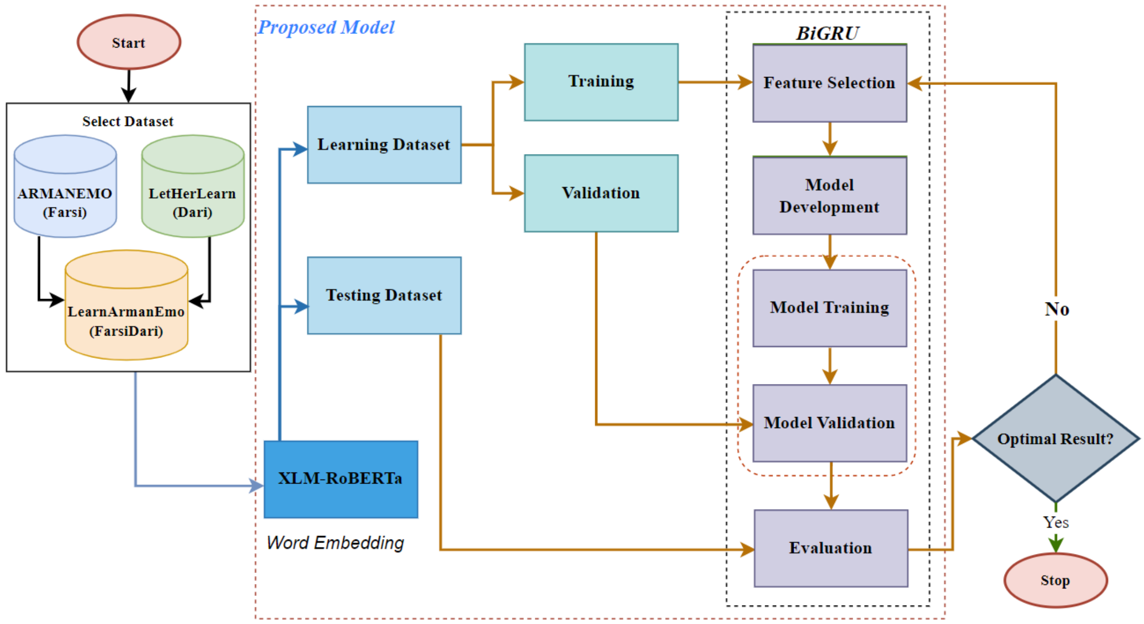

Figure 1: Workflow proposed model (XLM-RoBERTa-large + BiGRU) for the LearnArmanEmo dataset analysis.

0.2 are used to prevent overfitting, followed by another dense layer with the softmax activation function. The proposed model uses the AdamW optimizer with a batch size of 32 and a learning rate of 0.00001.

6.1.1. LetHerLearn results

The results of various deep learning and BERT models applied to the LetHerLearn dataset show that the proposed XLM-RoBERTa-large + BiGRU model achieves the highest precision of 0.735, recall of 0.724, and F1 score of 0.729. This suggests that the XLM-RoBERTa-large + BiGRU model has a better performance in predicting the emotion of Persian text compared to other methods. Data splitting is based on the main article of LetHerLearn. Table 2 offers a comprehensive overview of the proposed models, emphasizing significant differences.

Model	Precision	Recall	F1
LSTM	0.673	0.632	0.652
BiLSTM	0.664	0.633	0.648
BiGRU	0.653	0.624	0.638
ParsBERT	0.65	0.65	0.65
ParsBERT + BiGRU	0.681	0.683	0.682
XLM-RoBERTa-L	0.70	0.70	0.70
Proposed Model	**0.735**	**0.724**	**0.729**

Table 2: The comparison results on the LetHerLearn dataset, we used the results of ParsBERT & XLM-RoBERTa-large from the original paper Hussiny and Øvrelid (2023)

6.1.2. ARMANEMO results

Our implementation shows that the proposed XLM-RoBERTa-large + BiGRU model achieves the highest precision of 0.773, recall of 0.770, and F1 score of 0.771. Our results indicate that the XLM-RoBERTa-large + BiGRU model has better performance in predicting the emotion of Persian text compared to other models. The data partitioning is based on the main article of ARMANEMO. Table 3 provides comprehensive results of the proposed models, emphasizing significant differences.

Model	Precision	Recall	F1
LSTM	0.650	0.623	0.636
BiLSTM	0.631	0.622	0.626
BiGRU	0.654	0.651	0.652
ParsBERT	0.671	0.655	0.667
ParsBERT + BiGRU	0.702	0.691	0.696
XLM-RoBERTa-L	0.759	0.758	0.753
Proposed Model	**0.773**	**0.770**	**0.771**

Table 3: The comparison results on the ARMANEMO dataset, we used the results of ParsBERT & XLM-RoBERTa-large from the original paper Mirzaee et al. (2022)

6.1.3. LearnArmanEmo dataset results

We combined both datasets to specify the results and performance of the proposed algorithm more precisely. Deep learning algorithms exhibit more effective results with larger datasets and we randomly divided the LearnArmanEmo into three dis-

Model	Precision	Recall	F1
LSTM	0.672	0.660	0.666
BiLSTM	0.671	0.670	0.670
BiGRU	0.661	0.673	0.667
ParsBERT	0.713	0.714	0.714
ParsBERT + BiGRU	0.735	0.734	0.735
XLM-RoBERTa-L	0.773	0.774	0.774
Proposed Model	**0.792**	**0.786**	**0.789**

Table 4: The comparison results on the LearnArmanEmo dataset

tinct parts, 80% for training, 10% for validation, and 10% for testing. The results obtained by the XLM-RoBERTa-large + BiGRU algorithm outperform other algorithms, demonstrating a precision of 0.77, a recall of 0.77, and an F1 score of 0.77. Table 4 provides comprehensive results of the proposed models, emphasizing significant differences.

Table 5 presents the scores for each class based on the XLM-RoBERTa-large + BiGRU model. The results indicate that the "Disgust" and "Fear" classes achieved the highest F1 scores, whereas the "Sadness" and "Surprise" classes posed more challenges.

Class	Precision	Recall	F1
Anger	0.750	0.761	0.755
Disgust	**0.942**	0.860	**0.899**
Fear	0.821	**0.871**	0.845
Happiness	0.773	0.812	0.792
Sadness	0.725	0.710	0.717
Surprise	0.743	0.724	0.733
Other	0.792	0.764	0.778

Table 5: Individual class performance based on proposed model

6.2. Evaluation and Result

The results of our experiment and comparisons indicate that the ensemble model XLM-RoBERTa-large with BiGRU is effective and outperforms other models. These models demonstrate higher abilities in recognizing emotions in Persian texts. The combined model not only performs better on individual datasets but also excels when datasets are combined. BERT models, with their transformer architecture, excel at capturing context and semantic understanding in text, while the recurrent neural network adeptly captures sequential nuances. Simultaneously, the performance of the BiGRU model is determined by its results, which exhibit better outcomes due to its forward and backward direction, aiding in improved emotion recognition.

7. Conclusion

In this research, we implemented various models for the nuance of emotion analysis within Persian texts. Additionally, we introduced an improved approach that yields better results for Persian emotion analysis. This model combines the power of a transformer model, namely XLM-RoBERTa-large, with the sequential insights harnessed by a recurrent neural network, BiGRU. Our innovative model underwent rigorous evaluation on two existing datasets, LetHerLearn and ARMANEMO, each representing distinct linguistic nuances and contextual challenges. This model yielded favorable results when merging both datasets into a larger Persian emotion dataset. The outcomes of our experimentation reveal promising results for the proposed model, achieving an F1 score rate of 72.9% for the LetHerLearn dataset, a more commendable F1 score of 77.1% on the ARMANEMO dataset, and an F1 score of 78.8% on the LearnArmanEmo dataset.

LearnArmanEmo[1] is a combination of two datasets in the geographical area of Persian language speakers (Farsi and Dari). Due to the differences in writing and the ways of expressing feelings considering the words, it is necessary to augment the dataset with a larger volume. We aim to broaden the new dataset to include multimodal data, integrating text, images, and audio to better comprehend dialect complexity.

[1] The dataset and codes will be made available under a Creative Commons Attribution 4.0 International License.

8. Bibliographical References

Anil Bandhakavi, Nirmalie Wiratunga, Deepak Padmanabhan, and Stewart Massie. 2017. Lexicon based feature extraction for emotion text classification. *Pattern recognition letters*, 93:133–142.

Ankush Chatterjee, Umang Gupta, Manoj Kumar Chinnakotla, Radhakrishnan Srikanth, Michel Galley, and Puneet Agrawal. 2019. Understanding emotions in text using deep learning and big data. *Computers in Human Behavior*, 93:309–317.

Diogo Cortiz. 2022. Exploring transformers models for emotion recognition: a comparision of bert, distilbert, roberta, xlnet and electra. pages 230–234.

Flor Miriam Plaza Del Arco, Carlo Strapparava, L Alfonso Urena Lopez, and M Teresa Martín-Valdivia. 2020. Emoevent: A multilingual emotion corpus based on different events. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1492–1498.

Paul Ekman. 1992. Facial expressions of emotion: an old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 335(1273):63–69.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*.

Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. 2019. Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, 7:35–51.

R Jayakrishnan, Greeshma N Gopal, and MS Santhikrishna. 2018. Multi-class emotion detection and annotation in malayalam novels. In *2018 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5. IEEE.

Hamed Khanpour and Cornelia Caragea. 2018. Fine-grained emotion detection in health-related online posts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1160–1166.

Evgeny Kim and Roman Klinger. 2018. A survey on sentiment and emotion analysis for computational literary studies. *arXiv preprint arXiv:1808.03137*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Kevin N Ochsner and James J Gross. 2005. The cognitive control of emotion. *Trends in cognitive sciences*, 9(5):242–249.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

Fika Hastarita Rachman, Riyanarto Sarno, and Chastine Fatichah. 2016. Cbe: Corpus-based of emotion for emotion detection in text document. In *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pages 331–335. IEEE.

Romana Rahman et al. 2017. Detecting emotion from text and emoticon. *London Journal of Research in Computer Science and Technology*.

VV Ramalingam, A Pandian, Abhijeet Jaiswal, and Nikhar Bhatia. 2018. Emotion detection from text. In *Journal of Physics: Conference Series*, volume 1000, page 012027. IOP Publishing.

Kashfia Sailunaz and Reda Alhajj. 2019. Emotion and sentiment analysis from twitter text. *Journal of Computational Science*, 36:101003.

Shiv Naresh Shivhare, Shakun Garg, and Anitesh Mishra. 2015. Emotionfinder: Detecting emotion from blogs and textual documents. In *International Conference on Computing, Communication & Automation*, pages 52–57. IEEE.

Brian Spooner. 2012. 4. persian, farsi, dari, tajiki: language names and language policies. In *Language policy and language conflict in Afghanistan and its neighbors*, pages 89–117. Brill.

Matla Suhasini and Badugu Srinivasu. 2020. Emotion detection framework for twitter data using supervised classifiers. In *Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19*, pages 565–576. Springer.

Angelina Tzacheva, Jaishree Ranganathan, and Sai Yesawy Mylavarapu. 2020. Actionable pattern discovery for tweet emotions. In *Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2019 International Conference on Human Factors in Artificial Intelligence and Social Computing,* *the AHFE International Conference on Human Factors, Software, Service and Systems Engineering, and the AHFE International Conference of Human Factors in Energy, July 24-28, 2019, Washington DC, USA 10*, pages 46–57. Springer.

Yichen Wang and Aditya Pal. 2015. Detecting emotions in social media: A constrained optimization approach. In *Twenty-fourth international joint conference on artificial intelligence*.

9. Language Resource References

Hussiny, Mohammad Ali and Øvrelid, Lilja. 2023. *Emotion Analysis of Tweets Banning Education in Afghanistan*.

Mirzaee, Hossein and Peymanfard, Javad and Moshtaghin, Hamid Habibzadeh and Zeinali, Hossein. 2022. *ArmanEmo: A Persian Dataset for Text-based Emotion Detection*.