Emotion Analysis of Tweets Banning Education in Afghanistan

Mohammad Ali Hussiny and Lilja Øvrelid

Department of informatics, University of Oslo

March 26, 2024

Introduction

- Emotion Detection: identifying expression and recognition of feelings has crucial role for human communication.
- LetHerLearn dataset: a Persian Dari corpus of emotion-annotated Twitter data based on the collection and analysis of tweets.
- LetHerLearn dataset: contains 7,600 tweets posted in reaction to the Talibans ban of womens rights to education in 2022 and has been manually annotated according to Ekmans emotion categories.

Demand and Importance

- There is a lack of research and resources for different Persian language varieties.
- All the available datasets are developed based on Iranian social media and speakers.
- This work presents the first emotion annotated dataset for the Dari variant of Persian language.
- The goal of this work is to provide insights into peoples real-time perspectives, attitudes, concerns and reactions in the face of this oppression.

Data Collection

- We collected data from Twitters API using relevant Hashtags (LetHerLearn, AllOrNone, LetHerwork, LetAfghanistanGirlLearn and letAfghangirllearn).
- The included tweets were all posted from December 20, 2022 up to March 10, 2023.
- Collected around fifty thousand tweets and selected 7600 tweets for manual labeling.

Data Preprocessing

- Remove irrelevant or noisy data.
- Remove duplicates tweets.
- Annotators were instructed to remove tweets in languages such as Pashto, Uzbek, and others written in Persian script.

Data Annotation

- Data were labeled by two Dari native speakers based on Ekmans set of fundamental emotions.
- A1 Labeled 4000 tweets and A2 Labeled 3600 tweets.
- 100 twets were provided to both annotators, we measured the consistency of annotations using Cohens Kappa and the agreement achieved was 0.80.
- The guidelines provided to the annotators contain detailed descriptions of the six emotions with example words typically associated with the different emotions.

Data Annotation

Tweet	Label
خدا لعنت کند کسانیکه را که کاشانه و ماوایم خراب کرد May God condemn those who have destroyed our home and shelter	Disgust
دختر یعنی ریشه دواندن در دل خاک همیشه در حال پیشرفت و توسعه میباشد کسی دختران حذف کرده نمیتوانند A girl is like a tree, she keeps growing strong, impenetrable roots deep in the ground	Happiness
از ترس طالبان کسی صدایش بلند کرده نمیتواند No one can raise their voice due to the dread of the Taliban	Fear
احساس سوختن به تماشا نمی شود آتش بگیر تا که بدانی چه می کشم Watching someone on fire doesn't truly convey any feeling, however once you experiencing the torment of being on fire, you will grasp the real pain	Anger

Figure 1: LetHerLearn example tweets with emotion label

Dataset Statistics

- The dataset has 88,875 words, and 16,276 words are unique.
- We can see that Anger is the most observed emotion, followed by Happiness, and Surprise is the least observed emotion with only 503 occurrences.

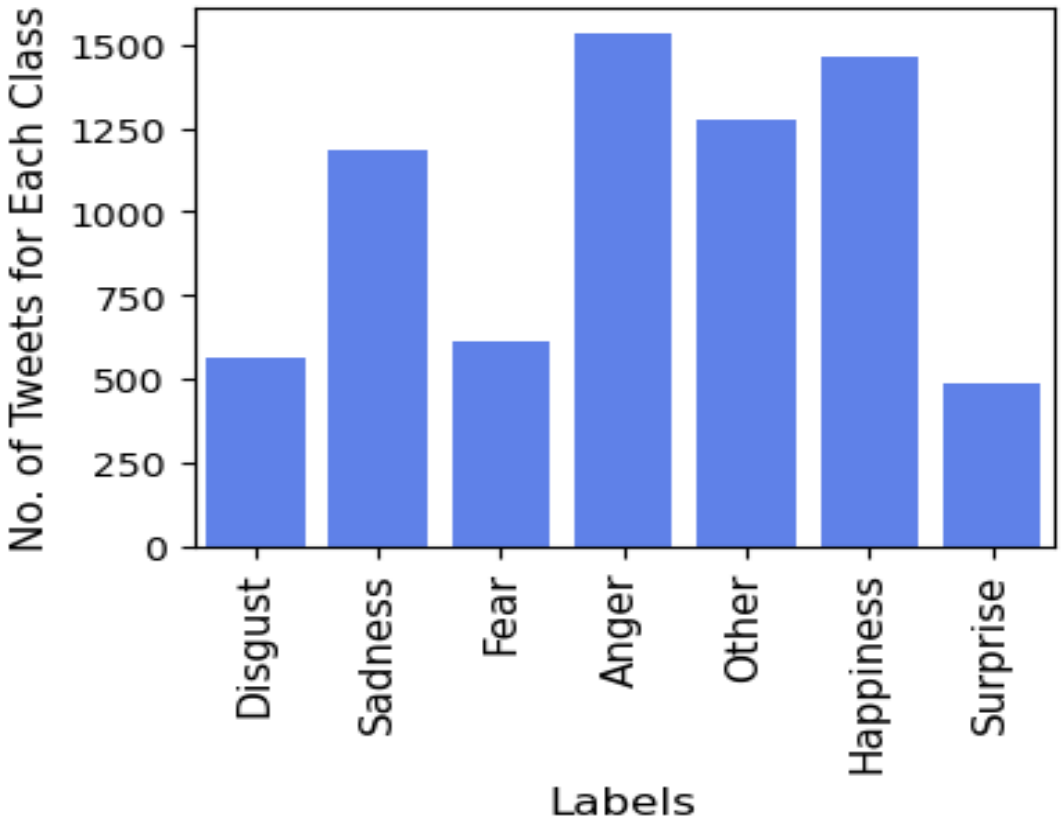

Figure 2: Number of tweets for each emotion class

Dataset Statistics

- We split the dataset into train, dev and test set using 80:10:10 for experimentation accordingly.

Type	Train	Dev	Test
Anger	1366	174	187
Disgust	462	50	57
Fear	483	64	59
Happiness	1266	179	152
Sadness	1032	120	128
Surprise	394	46	50
Other	1082	128	128
Total	6085	761	761

Table 1: Data distribution for experiments

Words Occurrence

- We observe that some words frequently occur in all classes such as Taliban, Afghanistan, girls, women, everyone.

Class	Words
Anger	'work', 'should', 'islam', 'society'
Disgust	'curse', 'tribe', 'damnation', 'filthy'
Fear	'fear', 'explosion', 'escape', 'arrest'
Happiness	'justice', 'hope', 'fight', 'rights'
Sadness	'pain', 'close', 'forgot', 'tired'
Surprise	'again', 'wish', 'someday', 'men'
Other	'life','world' 'further', 'iran'

Table 2: Frequent words in each class following removal of stopwords.

Models

- We evaluate a number of deep neural models on our dataset:
  1. Long Short-Term Memory Network(LSTM).
  2. Bi-directional Long Short-Term Memory(Bi-LSTM).
  3. Gated Recurrent Unit (GRU).
  4. Convolutional Neural Network (CNN).
- All models made use of fastText word embeddings with 300 dimensions for Persian language.

Models

- We used a pre-trained language model for Persian, ParsBERT which is a monolingual BERT model and XLM-RoBERTa large a multilingual transformer-based language model.
  1. seven epochs with a batch size of 32.
  2. the maximum length of encoded sequence is 128.
  3. Adam optimizer with learning rate scheduler of (2e-5).
- XLM-RoBERTa large
  1. seven epochs with a batch size of 32.
  2. the maximum length of encoded sequence is 128.
  3. AdamW optimizer with $\beta 1$ = 0.9, $\beta 2$ =0.999, learning rate scheduler is (0.00001).

Result

- The results of our experiments are summarized in below table, which shows the evaluation result of the different models described in the previous slides.

Model	Precision	Recall	F1
LSTM	0.67	0.63	0.65
BiLSTM	0.66	0.63	0.64
GRU	0.65	0.62	0.60
CNN	0.66	0.60	0.62
ParsBERT	0.65	0.65	0.65
XML-RoBERTa	**0.70**	**0.70**	**0.70**

Table 3: Macro Average Precision, Recall and F1 Result.

Result

- The results for the individual classes show that the highest scores were obtained for the Disgust and Fear classes, and the most difficult classes were the Other class and the Anger class 4.

Class	Precision	Recall	F1_Score
Anger	0.52	0.57	0.54
Disgust	**0.86**	0.84	**0.85**
Fear	0.84	**0.86**	**0.85**
Happiness	0.67	0.71	0.69
Sadness	0.58	0.61	0.59
Surprise	0.82	0.85	0.84
Other	0.62	0.44	0.52
Macro Average	0.70	0.70	0.70

Table 4: Individual class performance using XLM-RoBERTa-large model.

Error Analysis

- Error analysis shows that there is no direct correlation between class frequency and performance.
- Assigning a maximum of one emotion to each tweet is problematic for some tweets that have more than one have more than one emotion.
- Overlapping words between classes.

Error Analysis

- Some examples of misclassified predictions.

Tweet	True Label	Predicted Label
چه زمین لرزه وحشت ناکی آن هم در این فصل سرد What a terrifying earthquake, in this cold season	Anger	Happiness
سنگدل سیری گرسنه های را نصیحت میکند درد گرسنگی تحمل کند A hard heart satiety advise the hungry to endure the pain of hunger	Anger	Sadness
طالبان زیر فشار خارجی ها شریعت را فراموش میکنند The Taliban forget the Shariah under foreign pressure	Fear	Anger
وای چی دردهای جانسوزی Oh,What tragic and painful situation	Surprise	Happiness

Figure 3: Examples of misclassified tweet.

Conclusion

- We have presented the first Dari emotion-annotated dataset of tweets collected following the Talibans ban of womens education in afghanistan.
- In future work we would like to experiment with cross-variant Persian emotion detection as well as multitask learning of sentiment and emotion.

Thank you