Emotion detection of Arabic tweets using transformers

Habiba Reda ,Manar Mohamed ,Mohamed Abdel Mohsen, Mahmoud Khalifa ,Omnya Ramadan
Supervised by Dr Shaimaa Lazem[3], Dr. Amira Youssef[1], Eng. Ahmed El-Sayed[2]
(1) Faculty of Computer and Data Science, Alexandria University. (2) Faculty of Engineering, Alexandria University.
(3) Informatics Research Institute, City of Scientific Research and Technological Applications.

PROBLEM STATEMENT

Social media platforms have become an essential means for communicating feelings to the entire world due to rapid expansion in the Internet era and Nowadays people can express their feelings and thoughts using pictures, videos, audio, or textual content and this led to a massive amount of unstructured data generated from this Social media platforms and The question is **what if we can use this data to detect the emotions and feelings of the users?**

There is a growing interest in using emotion analysis on the Arabic text to understand public reactions and this could be useful in many fields such as monitoring people's mental health, especially after crises and pandemics such as covid-19 where Surveys by such organizations as YouGov show that, during the pandemic, workers were more stressed than ever.

The Arabic language is now facing many problems due to the increase in verb conjugations, the formation, and the presence of more than one word that leads to the same meaning, and many other problems that we should put into considerations when we work in Arabic text

The purpose of this study is to Compare state-of-the-art Arabic transformer models in **Nile University Dataset, which comprised of 10,065 tweets and categorized into 8 emotions (sadness, anger, fear, sympathy, joy, surprise, love, and none)** and our aim is to improve the models' results using Different combinations of prepossessing techniques so that, we can use them in emotion detection

RELATED WORK

a brief review on the previous work
that have been done before to the emotion classification of Arabic Twitter data related to different topics.

The top results on the Nile University data were on Ahmed ElSayed, Shaimaa Lazem and Mohamed Abougabal , work with 75.88% accuracy, 75.6% precision, 75.7% recall and 75.5% F1 score achieved by AraBert word embedding for feature extraction and the weighted voting ensemble of Logistic Regression, GRU and AraBERT.

Al-Khatib and El-Beltagy best results obtained using a Complement Naive Bayes algorithm with accuracy 68.12%

Essam et al. performed emotion classification in the dataset collected in Al-Khatib and El-Beltagy and they used preprocessing techniques onArabic text as shown in Table I and tested them with multiple classic machine learning
and deep learning models and the best accuracy achieved was 72% using LSTM and 69.41% using Logistic Regression.

Methodology

Dataset:

The dataset was collected by a research group at Nile University (NU). It is a balanced dataset consists of 10,065 Arabic tweets mostly using Egyptian dialect, and was manually annotated using eight emotions (sadness, anger, joy, surprise, love, sympathy, fear and none). It was collected using the" Olympics" hashtags from Egypt in the period between Jul 2016 and Aug 2016. The dataset was split into 80% for training and 20% for testing.

Preprocessing techniques:

Data pre-processing is important to minimize the noise and get a better classification accuracy and we applied a lot of techniques in preprocessing the Arabic text such as removing stop words, non-Arabic letters and spaces, Diacritics, Punctuation, Links, mentions, retweet indicators and suffixes and prefixes also, we reduced the Arabic words to their roots using different stemmers and we handled emojis and emoticons in the data in different ways

Experimental design:

We have done 3 experiments on this dataset where each of them started by using the same pre-processing techniques in with some small changes, in the first experiment we removed the emojis and emoticons from the dataset and we used this data in Arabic Transformer models to measure the performance and in the second experiment we replaced emojis and emoticons with an equivalent Arabic emotion to see that if this step will give us better accuracy, and in the last experiment we replaced each emoticon with it's equivalent emoji and we kept the emojis in the dataset as it is and we passed this data in AraBERT-Twitter model as this model have had emojis added to their vocabulary and the model is expected to handle the emojis in the dataset.

Performance metrics:

Accuracy, precision, recall, and F-score were used. For each classifier, the error metrics were calculated for each individual label (joy, sadness, sympathy, anger, fear, surprise, love, none). The final error metrics were calculated as the average of all the metrics for all the labels.

Results & Discussion

Classical machine learning models:

We used mainly SVM, Logistic Regression, Decision Tree and Random Forest models and the best results was in experiment 2 and the best performance was logistic regression model with 67% accuracy, 76.8% Precision, 76.2% Recall and 76.2% F-score Where the difference in performance in experiment 2 is not more than 1% increase than experiment

Arabic Transformer models:

We used AraBERT-Base, AraBERT-Twitter, MarBERT, ArBERT and Qarib and best results were when we used MarBERT model in experiment 2 as expected we got 77% accuracy, 76.8% Precision, 76.2% Recall and 76.2% F- score and these results are the best among all the previous work on this dataset It was expected that the results from experiment 2 will be the best one because of the mapping of the emojis we did and also, MarBERT according to it's Arciheture we expected that it will have the best performance and In case of AraBERT-Twitter it performs better in experiment 1 than experiment 3 where we kept the emojis
the results from experiment 1 is shown in the table below

Model	AC	PR	RE	F1
AraBERT-Base	73%	74%	74.9%	74.8%
AraBERT-Twitter	75%	74.8%	74.3%	74.3%
MarBERT	75%	74.9%	74.9%	74.8%
ArBERT	71%	71.4%	71%	71.1%
Qarib	74%	74.6%	74.1%	74.3%

Arabic Transformer models (Experiment 1)

CONCLUSIONS & FUTURE WORK

Our results show that the eight emotions could be detected in Arabic text with an overall accuracy of 77% in MARBERT model.
In the future, we want to make an application analyzes the feedback of the customers in products to improve the quality of it and other an application talking about analyzing the most common emojis that people use on Facebook to predict their psychological health and we want to use different transformers like ULMfits and try different preprocessing techniques to improve the accuracy

REFERENCES

A. Essam, H. Raafat, N. Hazem, S. Samer," Towards The Early Detection of Depression using Sentiment Analysis of Arabic Tweets", Graduation Project Report, Computer and Systems Engineering Depart- ment, Alexandria University, 2019.

A.Gruzd, (2016). "Netlytic: Software for Automated Text and Social Network Analysis" [Online]. Available: https://netlytic.org/.

Ahmed El-Sayed, Shaimaa Lazem and Mohamed Abougabal, An Improved Emotion-based Analysis of Arabic Twitter Data using Deep Learning 2021