

MANTIS at SMM4H'2022: Pre-Trained Language Models Meet a Suite of Psycholinguistic Features for the Detection of Self-Reported Chronic Stress

Sourabh Zanwar

RWTH Aachen University
sourabh.zanwar@rwth-aachen.de

Yu Qiao

RWTH Aachen University
yu.qiao@rwth-aachen.de

Daniel Wiechmann

University of Amsterdam
d.wiechmann@uva.nl

Elma Kerz

RWTH Aachen University
elma.kerz@ifaar.rwth-aachen.de

Abstract

This paper describes our submission to the Social Media Mining for Health (SMM4H) 2022 Shared Task 8, aimed at detecting self-reported chronic stress on Twitter. Our approach leverages a pre-trained transformer model (RoBERTa) in combination with a Bidirectional Long Short-Term Memory (BiLSTM) network trained on a diverse set of psycholinguistic features. We handle the class imbalance issue in the training dataset by augmenting it by another dataset used for stress classification in social media.

1 Introduction

The global increase in social media use over the past decade has afforded researchers new opportunities to mine health-related information that can ultimately be used to improve public health. The Social Media Mining for Health Applications (SMM4H) Shared Task involved ten natural language processing challenges of using social media data for health research (Weissenbacher et al., 2022). In our submission to the task targeting the classification of self-reported chronic stress on Twitter (Task 8), we built hybrid models that combine pre-trained transformer language models with Bidirectional Long Short-Term Memory (BiLSTM) networks trained on a diverse set of psycholinguistic features.

2 Data

The Twitter data provided by the organizers of Task 8 comprised of a total of 4,195 tweets whose distributions over training, development and testing sets are shown in Table 3 of supplementary material¹. About 37% of the tweets are positive (self-disclosure of chronic stress, Pos) and 63% are negative (non-self-disclosure of chronic stress, Neg). The only preprocessing step that was applied was

¹Supplementary material

the removal of HTML and links from the text. To address the class imbalance in the data, we augmented the data using 1000 items with positive labels and 200 ones with negative labels from the Dreddit dataset (Turcan and McKeown, 2019).

2.1 Measurement of Psycholinguistic Features

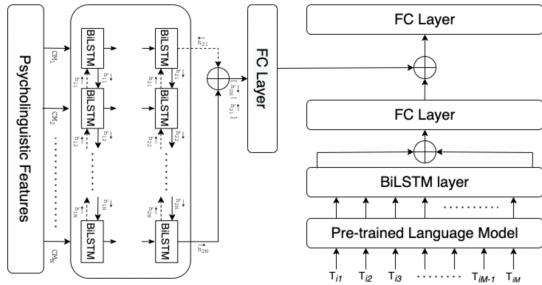
A set of 435 psycholinguistic features used in our approach fall into the following four categories: (1) features of morpho-syntactic complexity (N=19), (2) features of lexical richness, diversity and sophistication (N=77), (3) readability features (N=14), and (4) lexicon features designed to detect sentiment, emotion and/or affect (N=325). Measurements of these features were obtained using an automated text analysis system (for its recent applications, see e.g. Wiechmann et al. (2022) for predicting eye-moving patterns during reading and Kerz et al. (2022) for detection of Big Five personality traits and Myers-Briggs types). Tokenization, sentence splitting, part-of-speech tagging, lemmatization and syntactic PCFG parsing were performed using Stanford CoreNLP (Manning et al., 2014).

3 Description of System Architecture

We conducted experiments with a total of five models: (1) a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018), (2) a fine-tuned RoBERTa model (Robustly Optimized BERT pre-training Approach) (Liu et al., 2019), (3) a bidirectional neural network classifiers trained on measurements of psycholinguistic features described in Section 2.2, and (4) and (5) two hybrid models integrating BERT and RoBERTa predictions with the psycholinguistic features. For each model, we performed experiments with and without data augmentation.

For (1) and (2) we used the pretrained ‘bert-base-uncased’ and ‘roberta-base’ models from the Huggingface Transformers library (Wolf et al., 2020) each with an intermediate BiLSTM layer with 256

Figure 1: Proposed hybrid model architecture



hidden units (Al-Omari et al., 2020). For (3) - the model based solely on psycholinguistic features, we constructed a 2-layer BiLSTM with a hidden state dimension of 32. The input to that model is a sequence $CM_1^N = (CM_1, CM_2 \dots, CM_N)$, where CM_i , the output of CoCoGen for the i th sentence of a document, is a 435 dimensional vector and N is the sequence length. To predict the labels of a sequence, we concatenate the last hidden states of the last layer in forward (\vec{h}_n) and backward directions (\overleftarrow{h}_n). The result vector of concatenation $h_n = [\vec{h}_n | \overleftarrow{h}_n]$ is then transformed through a 2-layer feedforward neural network, using the Rectifier linear unit (ReLU) as an activation function. The output of this network is then passed to a Dense Fully Connected (FC) Layer with dropout of 0.2 and is finally passed on to a terminal, fully connected layer. The output is a K dimensional vector, where K is the number of emotion labels. The architecture of the hybrid classification models - models (4) and (5) - consists of two parts: (i) a pre-trained Transformer-based model with a BiLSTM layer and FC layer on top of it and (ii) the psycholinguistic features of the text fed into a BiLSTM layer and a FC layer. We concatenate the outputs of these layers before passing them into a final FC layer with a sigmoid activation function. Since the evaluation is based on the sensitivity of the system - i.e., here the classification of positive labels is more important - , we reduced the threshold of positive labels from 0.5 to 0.3. The model used to generate predictions for the test set was the RoBERTa-PsyLing hybrid model with the following configuration: BiLSTM-PsyLing: 2-layers, hidden size of 512 and dropout 0.2. We trained this model for 12 epochs, saving the model with the best performance from the development set. The optimizer used is AdamW with a learning rate of $2e-5$ and a weight decay of $1e-4$. We trained 5 models with 5 random 80% splits of the training data and superimposed a meta-learner to get the

Table 1: Results on the validation set

Model	Aug	Acc	Prec	Rec	F1
PsyLing	No	69	67	67	67
BERT	No	75	75	77	74
RoBERTa	No	71	77	76	71
BERT-Hybrid	No	78	77	79	78
RoBERTa-Hyb.	No	79	78	79	78
PsyLing	Yes	71	69	67	68
BERT	Yes	79	78	80	78
RoBERTa	Yes	78	79	81	78
BERT-Hybrid	Yes	81	80	81	80
RoBERTa-Hyb.	Yes	82	81	83	81

final predictions.

4 Results

An overview of the performance metrics on the validation set is presented in Table 1. We found that the proposed hybrid models consistently outperformed the standard transformer-based baseline models, with an improvement in F1 of up to 3%. Training the models on the augmented data led to an average increase in performance of 3.8% F1 over the non-augmented data. Highest performance on the validation set (F1 = 81) was achieved by the RoBERTa hybrid model. Error inspection indicated that the the majority of errors are False Positives (see Table 5²). This behavior was intentionally evoked by lowering the threshold for positive labels, as task scoring focused on the F1 assessment of these labels. Manual inspection of the errors also revealed that some predictions were incorrect due to labeling errors in the development set. Examples of such cases can be found in Table 6².

Table 2: Results on the test set (courtesy of challenge organizers)

Model	Acc	Prec	Rec	F1
RoBERTa-Hybrid	79.8	72	76	75

5 Conclusion

We developed hybrid classification systems for the detection of self-reported chronic stress that integrate pre-trained transformer language models with BiLSTM networks trained on a diverse set of psycholinguistic features. Our experiments show that such hybrid models significantly outperform base transformer models for both augmented and non-augmented data.

²Tables 5,6 in supplementary material

References

- Hani Al-Omari, Malak A. Abdullah, and Samira Shaikh. 2020. *Emodet2: Emotion detection in english textual dialogue using bert and bilstm models*. In *2020 11th International Conference on Information and Communication Systems (ICICS)*, pages 226–232.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Elma Kerz, Yu Qiao, Sourabh Zanwar, and Daniel Wiechmann. 2022. Pushing on personality detection from verbal behavior: A transformer meets text contours of psycholinguistic features. *arXiv preprint arXiv:2204.04629*.
- 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*.
- Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.
- Elsbeth Turcan and Kathy McKeown. 2019. Dreaddit: A Reddit dataset for stress analysis in social media. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 97–107. Association for Computational Linguistics.
- Davy Weissenbacher, Ari Z. Klein, Luis Gascó, Darryl Estrada-Zavala, Martin Krallinger, Yuting Guo, Yao Ge, Abeed Sarker, Ana Lucia Schmidt, Raul Rodriguez-Esteban, Mathias Leddin, Arjun Magge, Juan M. Banda, Vera Davydova, Elena Tutubalina, and Graciela Gonzalez-Hernandez. 2022. Overview of the seventh social media mining for health applications smm4h shared tasks at coling 2022. In *Proceedings of the Seventh Social Media Mining for Health (SMM4H) Workshop and Shared Task*, pages 33–40.
- Daniel Wiechmann, Yu Qiao, Elma Kerz, and Justus Mattern. 2022. Measuring the impact of (psycho-) linguistic and readability features and their spill over effects on the prediction of eye movement patterns. *arXiv preprint arXiv:2203.08085*.
- Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.

A Supplementary material

Supplementary material can be found here <https://bit.ly/3xq68bx>.