Supplemental Material - Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm

Bjarke Felbo[1]**, Alan Mislove**[2]**, Anders Søgaard**[3]**, Iyad Rahwan**[1]**, Sune Lehmann**[4]

[1]Media Lab, Massachusetts Institute of Technology
[2]College of Computer and Information Science, Northeastern University
[3]Department of Computer Science, University of Copenhagen
[4]DTU Compute, Technical University of Denmark

A Supplemental Material

A.1 Preprocessing Emotion Datasets

In the Olympic Games dataset by Sintsova et al. each tweet can be assigned multiple emotions out of 20 possible emotions, making evaluation difficult. To counter this difficulty, we have chosen to convert the labels to 4 classes of low/high valence and low/high arousal based on the Geneva Emotion Wheel that the study used. A tweet is deemed as having emotions within the valence/arousal class if the average evaluation by raters for that class is 2.0 or higher, where ‘Low’ = 1, ‘Medium’ = 2 and ‘High’ = 3.

We also evaluate on the ISEAR databank (Wallbott and Scherer, 1986), which was created over many years by a large group of psychologists that interviewed respondents in 37 countries. Each observation in the dataset is a self-reported experience mapped to 1 of 7 possible emotions, making for an interesting benchmark dataset.

A.2 Pretraining as Regularization

Figure 1 shows an example of how the pretraining helps to regularize the target task model, which otherwise quickly overfits. The chain-thaw transfer learning approach further increases this regularization by fine-tuning the model layer wise, thereby adding additional regularization.

A.3 Emoticon to Emoji mapping

To analyze the effect of using a diverse emoji set we create a subset of our pretraining data containing tweets with one of 8 emojis that are similar to the positive/negative emoticons used by Tang et al. (2014) and Hu et al. (2013). The positive emoticons are :) :) :-) :D =) and the negative emoticons are :(: (:-(. We find the 8 similar emojis in our dataset seen in Figure 2 as use these for creating the reduced subset.

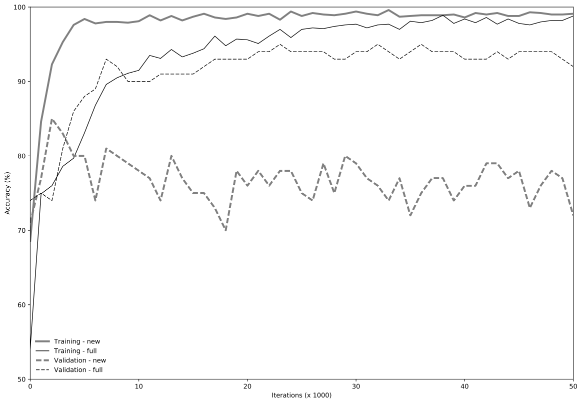

Figure 1: Training statistics on the SS-Youtube dataset with a pretrained model vs. a untrained model. The architecture and all hyperparameters are identical for the two models. All layers are unfrozen.

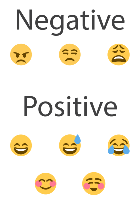

Figure 2: Emojis used for the experiment on the importance of a diverse noisy label set.

A.4 Emoji Clustering

We compute the predictions of the DeepMoji model on the pretraining test set containing 640K tweets and compute the correlation matrix of the predicted probabilities seen in Figure 3. Then we use hierarchical clustering with average linkage on the correlation matrix to generate the dendrogram seen in Figure 4. We visualized dendrograms for various versions of our model and the overall structure is very stable with only a few emojis changing places in the hierarchy.

References

Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. 2013. Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, pages 607–618.

Valentina Sintsova, Claudiu-Cristian Musat, and Pearl Pu. 2013. Fine-grained emotion recognition in olympic tweets based on human computation. In *4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*.

Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*. pages 1555–1565.

Harald G Wallbott and Klaus R Scherer. 1986. How universal and specific is emotional experience? evidence from 27 countries on five continents. *International Social Science Council* 25(4):763–795.

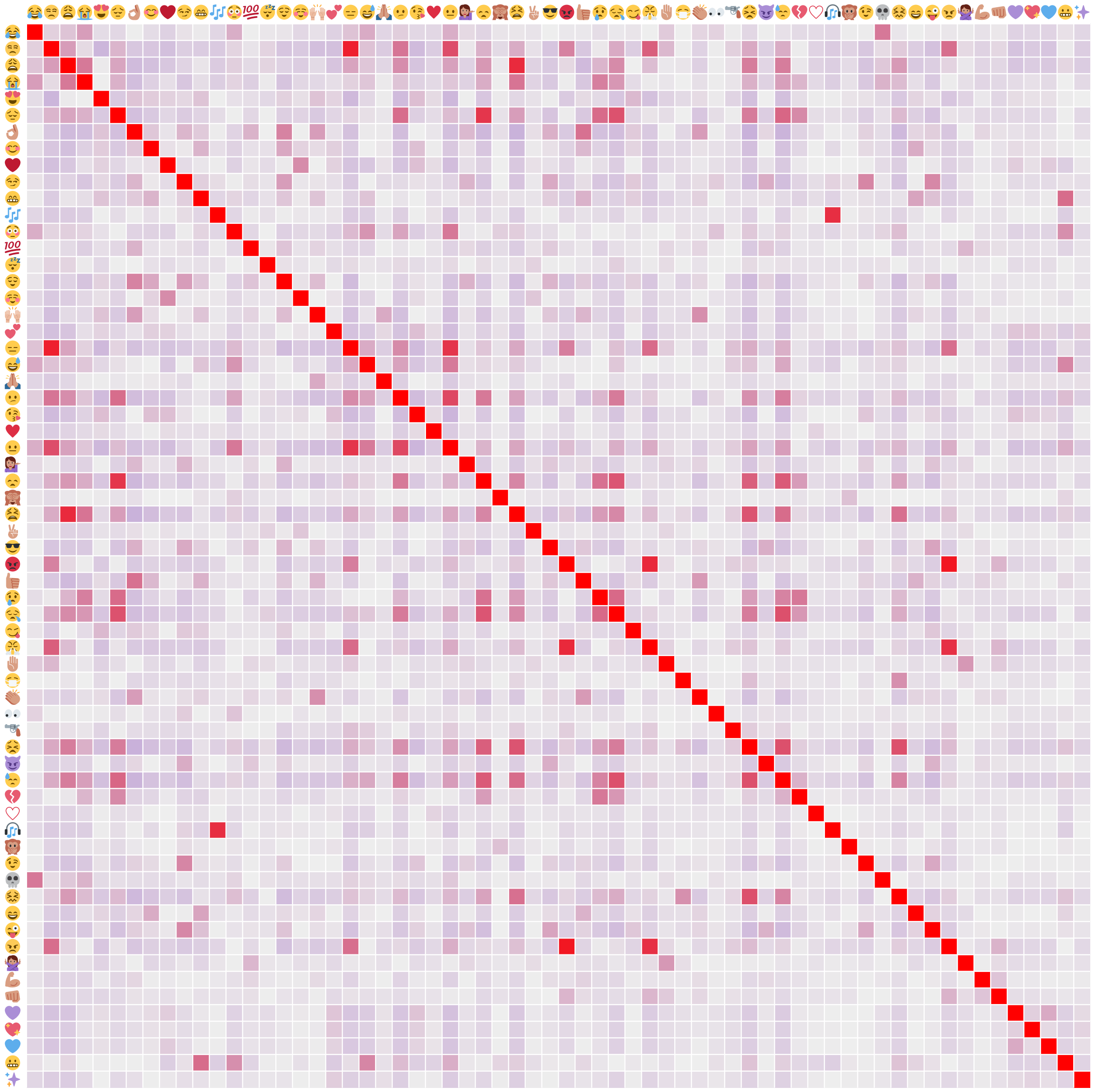

Figure 3: Correlation matrix of the model's predictions on the pretraining test set.

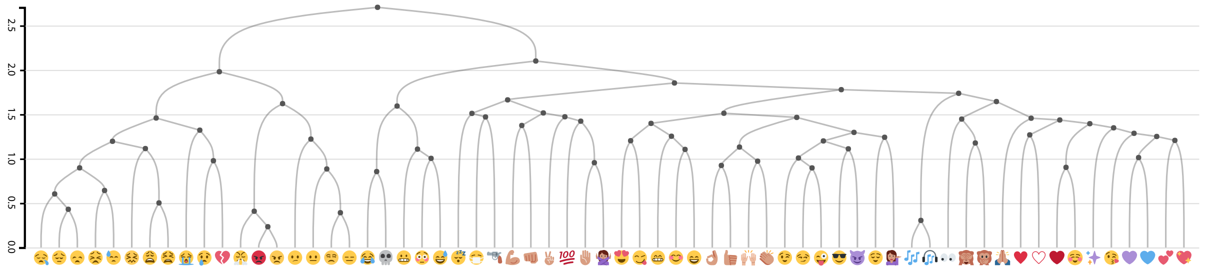

Figure 4: Hierarchical clustering of the DeepMoji model's predictions across categories on the test set. The dendrogram shows how the model learns to group emojis into overall categories and subcategories based on emotional content. The y-axis is the distance on the correlation matrix of the model's predictions measured using average linkage.