Supplementary: A Multi-task Model for Emotion and Offensive Aided Stance Detection of Climate Change Tweets

1 EXPERIMENTAL SET-UP

For various hyper-parameters, we experimented with different values. Following are the details:

- Bi-LSTM memory cells (d_f): We experimented with (100, 200, and 300) neurons.
- fully connected layers: We experimented with (100, 200, and 300) neurons for modality attention layer [MA] (d_a) and stance-specific shared attention layer [SAM] (d_s).
- Hidden Activations for fully connected layers: ReLU [1]
- Output Activations: softmax for stance detection (SD) and sigmoid for emotion recognition (ER).
- For Epoch: We used Keras's "ModelCheckpoint" function for saving the best-attained model weights for a particular epoch. We initially set the limit to 50 epochs and observed after multiple runs of our experiments that the results did not improve after the 20th epoch.
- Adam optimizer [2]: with default learning rate of 0.001 is used.

The best parameters mentioned in the main paper are then selected using TPE in the Hyperopt[1] python library that minimises loss functions.

REFERENCES

[1] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 315–323.
[2] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[1]http://hyperopt.github.io/hyperopt/