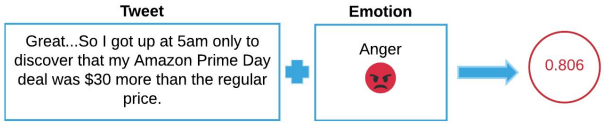
CMPSCI 585 - Natural Language Processing

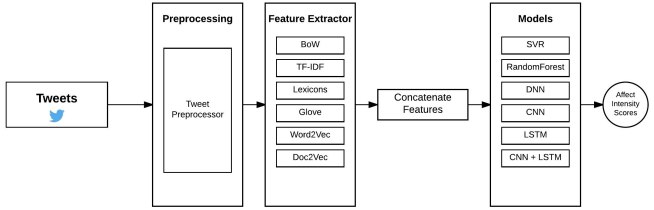
Modeling Affect Intensity In Tweets

Aarsh Patel, Lynn Samson

Task



Approach



- We tried a variety of features from basic sparse features such as tf-idf to distributional representations such as GloVe and word2vec

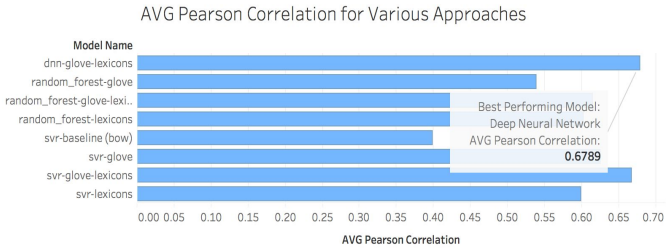
Data

- SemEval 2018 Affect Intensity dataset
- Only considering English dataset
- Each data point:
  - Tweet, emotion, intensity score

Emotion	Train	Dev
Anger	1701	200
Fear	2252	200
Joy	1616	200
Sadness	1533	200
All	7102	800

The number of instances in the Tweet Emotion Intensity dataset.

Experiments & Results



- As expected, dense features (e.g. lexicons) clearly outperform sparse features (e.g. bag-of-words baseline)
- Highest accuracy achieved by a concatenation of the lexicon and GloVe feature vectors
- The best performing model is the deep neural network trained on the GloVe-lexicon features

Future Work

- Implementing more advanced deep learning architectures such as LSTMs and CNNs
- Experimenting with different distributional representations such as doc2vec & emoji embeddings

References

Mohammad, S. M., & Bravo-Marquez, F. (2017). WASSA-2017 Shared Task on Emotion Intensity. *8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

Duppada, Venkatesh, and Sushant Hiray. "Seernet at emoint-2017: Tweet emotion intensity estimator." *arXiv preprint arXiv:1708.06185* (2017).