Fake News Challenge: Stance Detection with Attention and Conditional Encoding

Stephen Pfohl (spfohl@stanford.edu), Oskar Triebe (oskar2@stanford.edu), Ferdinand Legros (flegros@stanford.edu)

ABSTRACT

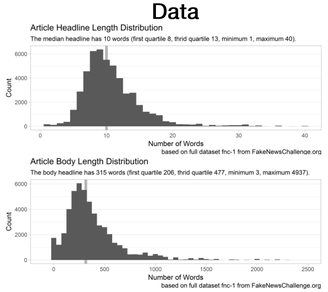
The in-progress Fake News Challenge is a public challenge tasking competitors to develop a stance detection tool that could ultimately be incorporated into a larger automatic fact-checking pipeline.

The challenge pairs 1,648 unique headlines and 1,669 unique article bodies to produce 49,972 body-headline pairs. Each of the body-headline pairs are labeled with either "Unrelated", "Discusses", "Agrees", or "Disagrees“. It is the goal of the stance detection task to predict these labels.

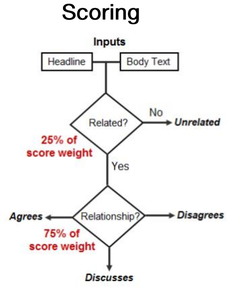
We applied the concepts of neural attention and conditional encoding to long short-term memory networks (LSTM). Our best model achieves a score of 0.808 improving over the current best competition score of 0.795.

DATA AND SCORING

Data



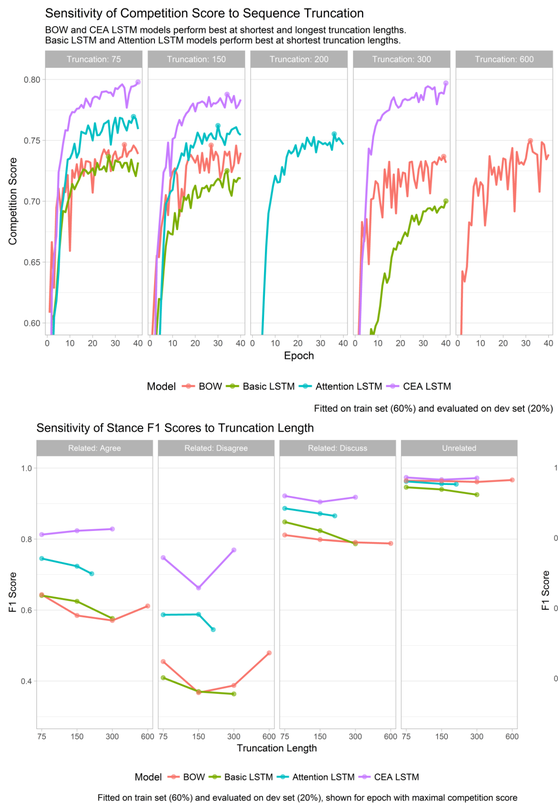
Scoring

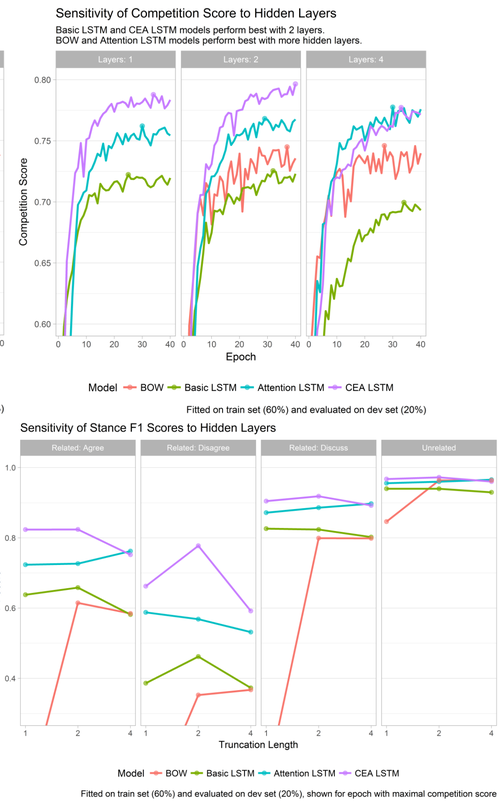


Class	Unrelated	Discusses	Agrees	Disagrees
Samples (%)	73.13	17.82	7.36	1.68

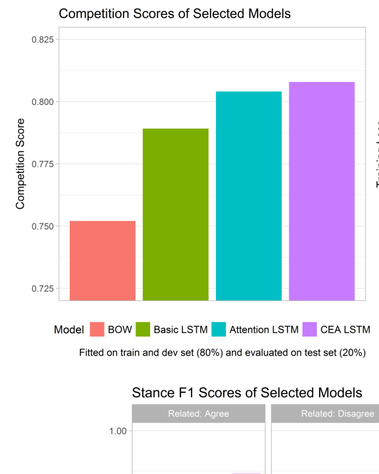
RESULTS

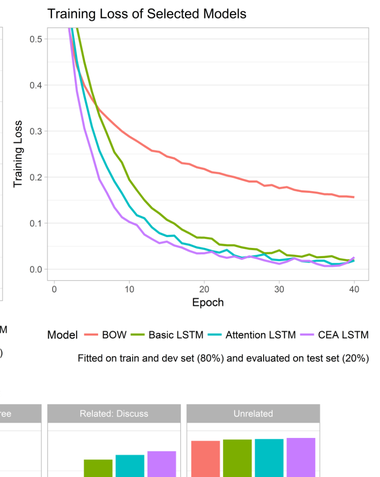
Hyperparameter Selection

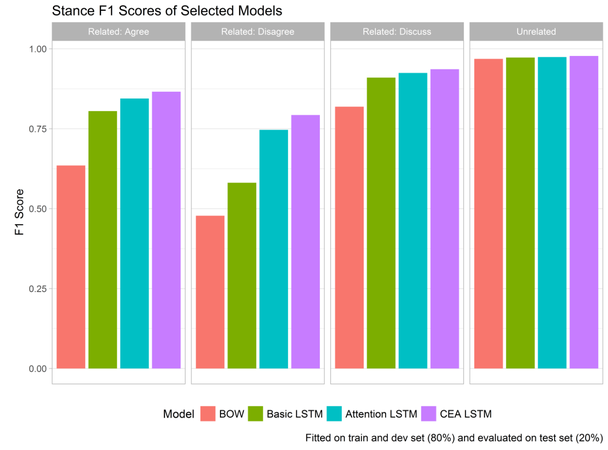




Test Performance of Selected Models

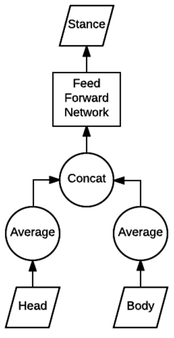




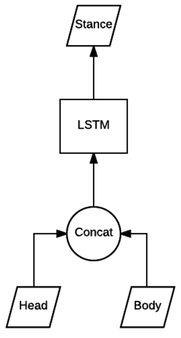


MODELS

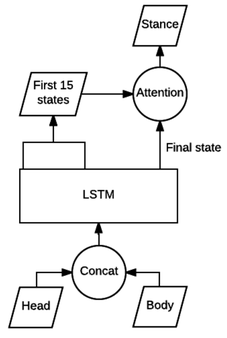
Bag of Words



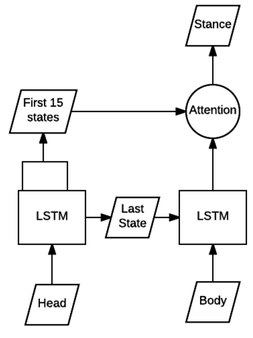
Basic LSTM



LSTM with Attention



Conditional Encoding LSTM with Attention (CEA)



ATTENTION

- Let all vectors be column vectors.
- Define the attention window the be the first L output states produced by the LSTM.
- Let k be the dimension of the hidden state, and N be the total sequence length.
- Let $Y \in \mathbb{R}^{k \times L} = [h_1, \ldots, h_N]$ be a matrix of the output states of the LSTM in the attention window.
- Let $e_L \in \mathbb{R}^L$ be a vector of 1s.
- Let $W^y, W^h, W^p, W^h \in \mathbb{R}^{k \times k}$ and $w \in \mathbb{R}^k$ be trainable matrices.
- A final state h^* is produced as follows

$$M = \tanh(W^y Y + W^h h_N e_L^T)$$

$$\alpha = \text{softmax}(w^T M)$$

$$r = Y \alpha^T$$

$$h^* = \tanh(W^p r + W^x h_N)$$

CONCLUSION

- Attention and conditional encoding allows for the model to utilize information from longer sequences without sacrificing performance
- Approach may extended to bidirectional models and more complex attention mechanisms
- Downsampling and a custom loss function may improve performance

REFERENCES

- Dean Pomerlau and Delip Rao. Post-facto Fake News Challenge
- Sepp Hochreiter and Jrgen Schmidhuber. Long Short-term Memory. Neural Comput., 9(9):1735–1780, 11 1997.
- Tim Rocktaschel, Edward Grefenstette, Karl Moritz Hermann, Tom Kocisky, and Phil Blunsom. Reasoning about Entailment with Neural Attention. 9 2015.