Stance Detection for the Fake News Challenge with Attention and Conditional Encoding
MSCI 641 Text Analytics Course Project

Yuqing Zhao
University of Waterloo

Abstract

The Fake News Challenge is a public challenge, which encourages competitors to develop a stance detection tool that can effectively deal with fake news problem and could be incorporated into an AI-assisted fact-checking pipeline.
The MSCI641 Fake News Challenge used the idea of the Fake News Challenge Stage one (FNC-1), and the dataset is provided on the CodaLab challenge website.
I developed four neural network models to deal with this problem: Baseline LSTM, Bidirectional LSTM, LSTM with attention and conditional encoding LSTM with attention (CELA). The final results show that CELA outperformed than other three models, with validation accuracy of 96.60% and the competition weighted score of 1819.50.

Models

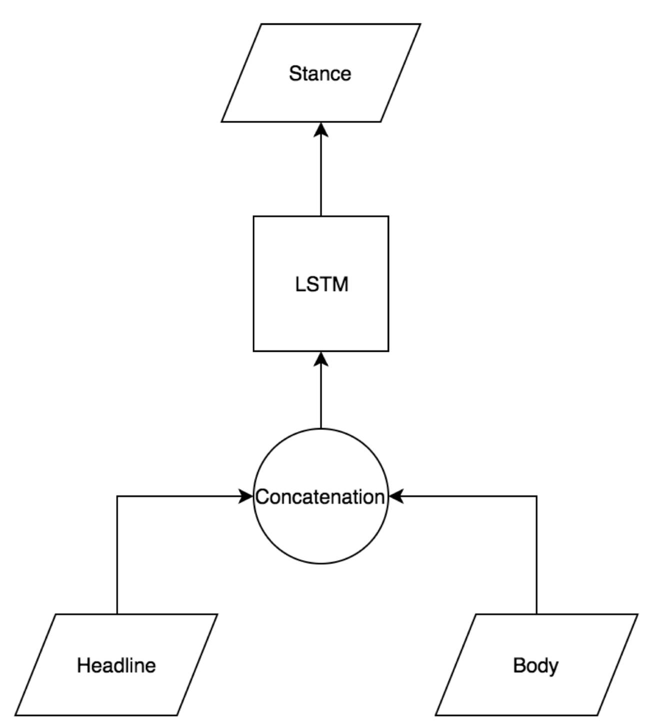

Figure 1. LSTM Model

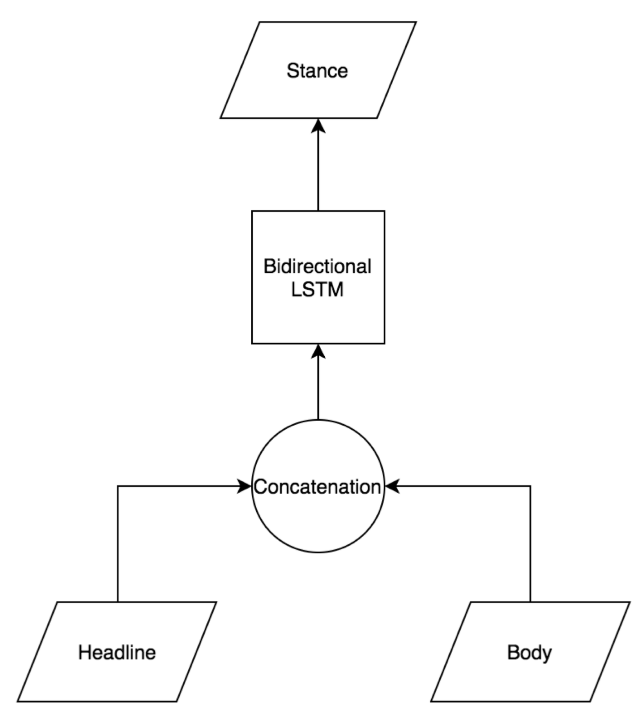

Figure 2. Bidirectional Model

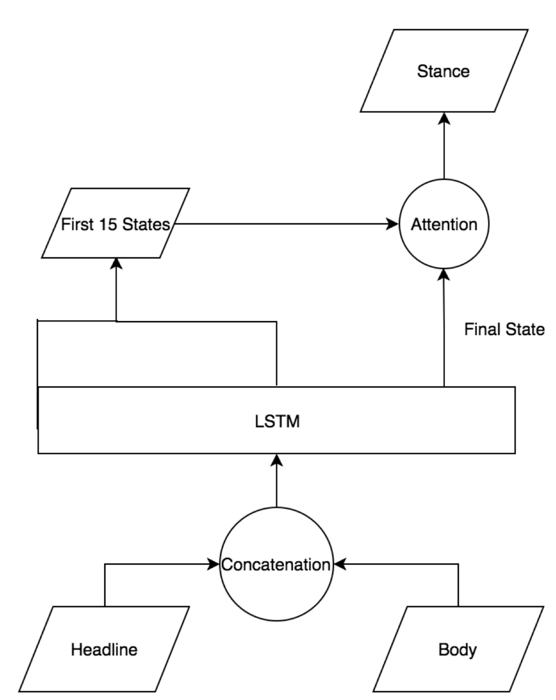

Figure 3. LSTM with Attention Model

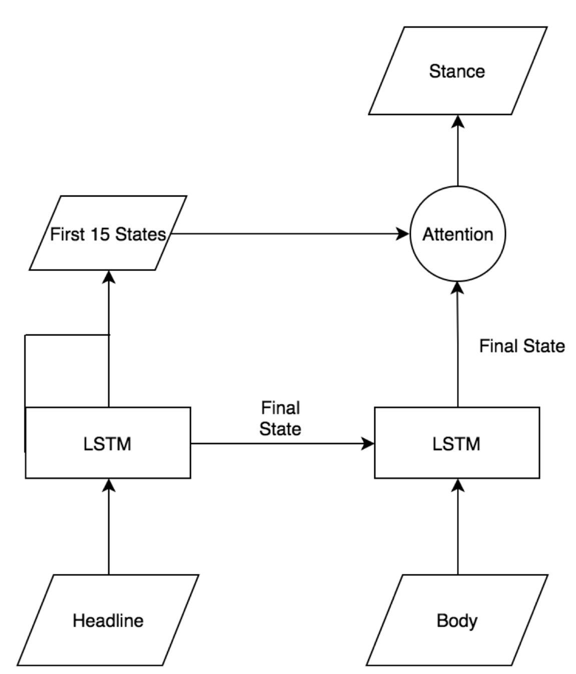

Figure 4. Conditional Encoding LSTM with Attention Model

Results

The provided training dataset was split randomly into training dataset (90%) and validation dataset (10%). Models with parameters which led to the highest validation accuracy was used to predict unlabeled test dataset. (Fitted on training and evaluated on validation).

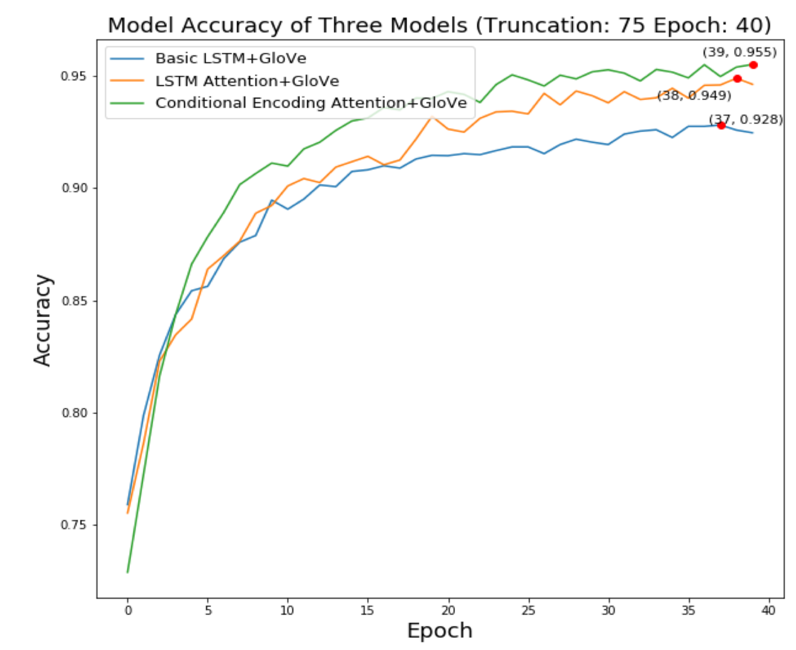

Figure 5. Model Accuracy (Truncation 75)

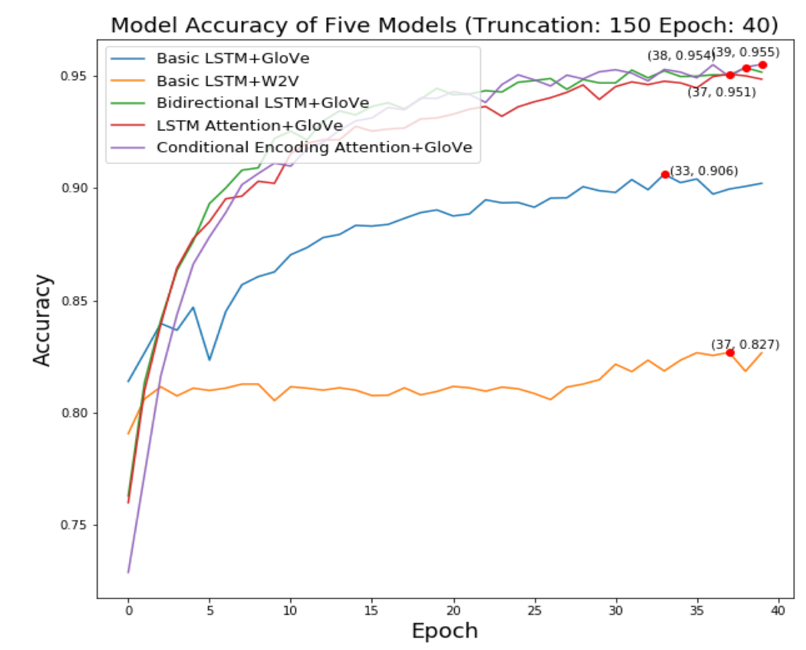

Figure 6. Model Accuracy (Truncation 150)

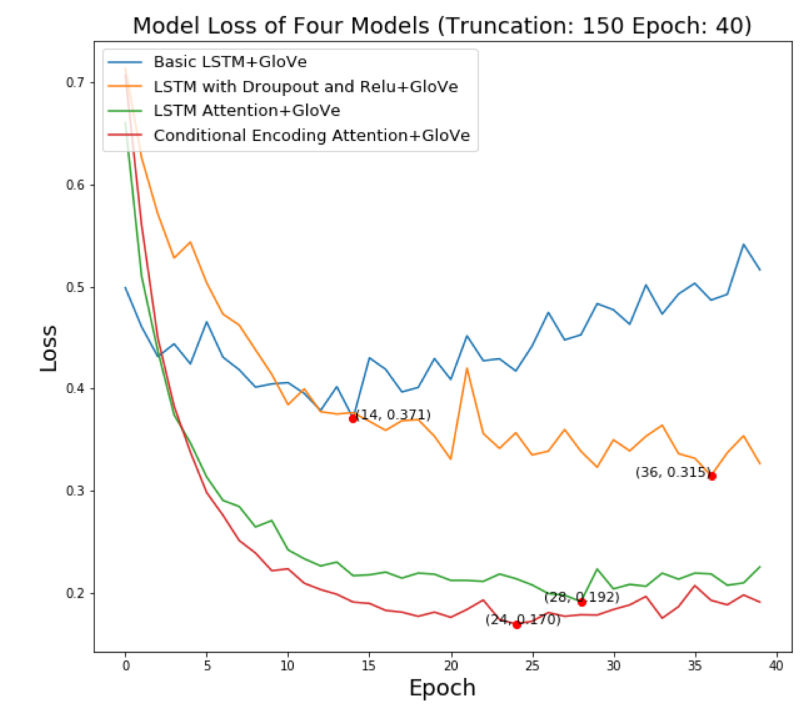

Figure 7. Model Loss (Truncation 150)

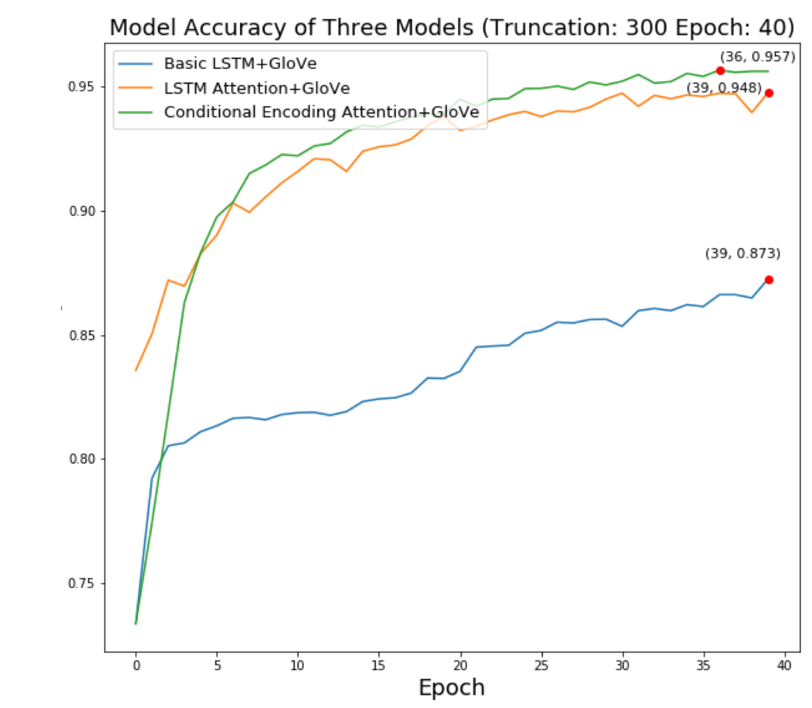

Figure 8. Model Accuracy (Truncation 300)

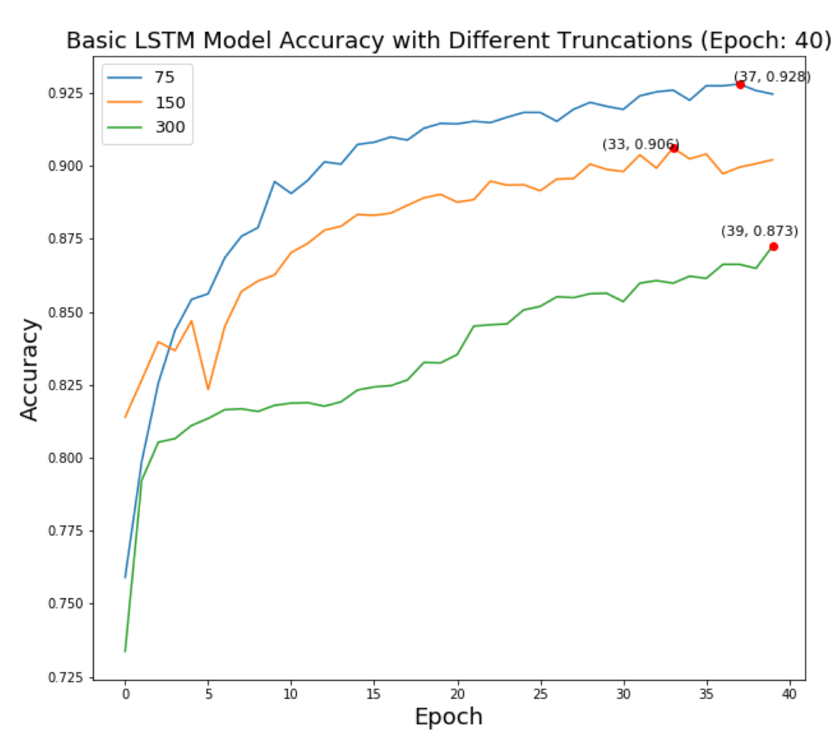

Figure 9. LSTM Model Accuracy with Different Truncations

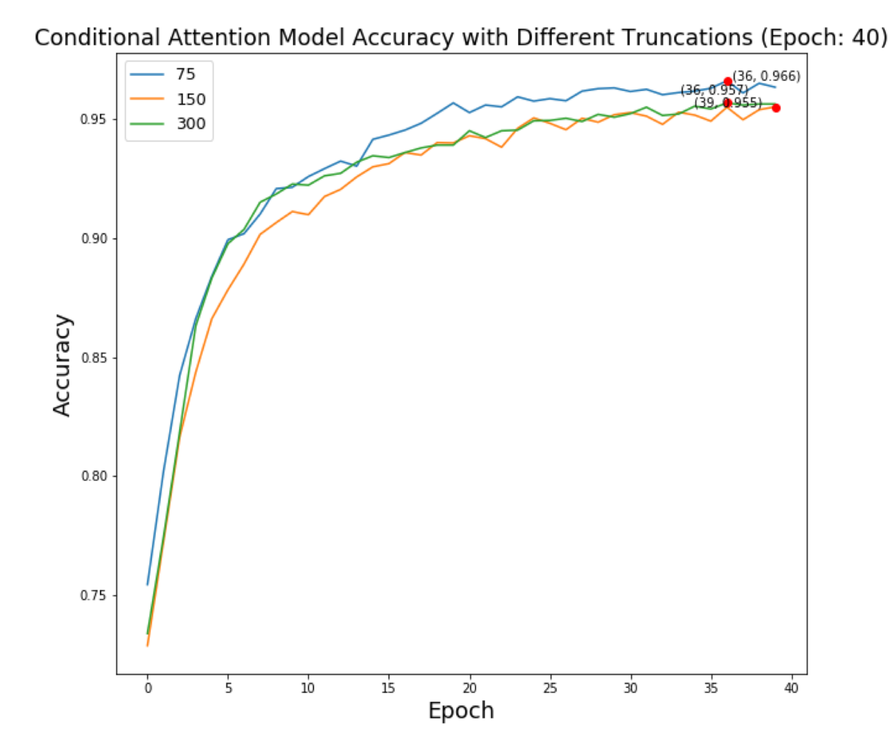

Figure 10. CELA Model Accuracy with Different Truncations

Data

There are two main challenges of FNC-1. First, BodyID Lookup Table contains sentences of various lengths. In training dataset, the longest sentence length is **4876**, and the shortest sentence length is **5**. Another challenge is that the majority of "Stance" of training dataset is unrelated with **72.65**%. The training dataset contains **66,677** rows in total and the details of each stance category is shown in Table 1.

Table 1 Percentage of Each Stance Category in Training Dataset

	Unrelated	Agree	Disagree	Discuss
Training Dataset	72.65%	7.41%	1.94%	18%

Attention Mechanism

- Let the attention window size is L such that first L states produced by the first LSTM.
- Let $Y \in \mathbb{R}^{k \times L}$ be a matrix consisting of output vectors $[h_1,..., h_L]$ such that k is the dimension of the hidden layer.
- Let h_N is the last output vector (final state) such that N is total sequence length.
- Let $e \in \mathbb{R}^L$ be a vector of 1 s.
- Let W^y, W^h, W^p, W^x, w are the trainable matrixes that the attention layer should learn.
- h* is the final state which is computed as follows.

$$M = \tanh(W^y Y + W^h h_N e_L^T)$$
$$\alpha = softmax(w^T M)$$
$$r = Y\alpha^T$$
$$h^* = \tanh(W^p r + W^x h_N)$$

Conclusions

- Conditional encoding LSTM with attention mechanism outperformed than other three models, with highest accuracy and competition weighted score.
- Attention Mechanism and Conditional Encoding can effectively utilize more information from long sequences without heavily influencing model performance.
- Dropout can effectively help with overfitting problem.
- Basic LSTM with GloVe performed better than Basic LSTM with W2V.
- Bidirectional LSTM got high accuracy but got a low weighted score, since it performed bed in predicting agree, disagree, discuss labels.
- In the future, I plan to utilize bidirectional conditional encoding and more complex attention mechanisms such as word-by-word attention mechanism.

Contact

Yuqing Zhao
the University of Waterloo
Email: y479zhao@uwaterloo.ca
Phone: 226-978-6858

References

- Dean Pomerlau and Delip Rao. Post-facto Fake News Challenge.
- Isabelle Augenstein, Tim Rock¨aschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance Detection with Bidirectional Conditional Encoding. Empirical Methods in Natural Language Processing, (2010):876–885.
- Tim Rock¨aschel, Edward Grefenstette, Karl Moritz Hermann, Tom Koˇcisk´y, and Phil Blunsom.Reasoning about Entailment with Neural Attention. 9 2015.