Towards Automated Identification of Fake News: Headline-Article Stance Detection

Sahil Chopra, Saachi Jain, John Sholar

CS 224N: NLP with Deep Learning, Mentor: Danqi Chen

Overview

- Increase in Prevalence of Fake News
- Help Fact Checkers find Potential Fake News
- **Goal:** Label Article-Headline Pair as "Unrelated", "Discuss", "Agree" or "Disagree"

Data + Scoring

# Headlines	# Articles
1648	1683

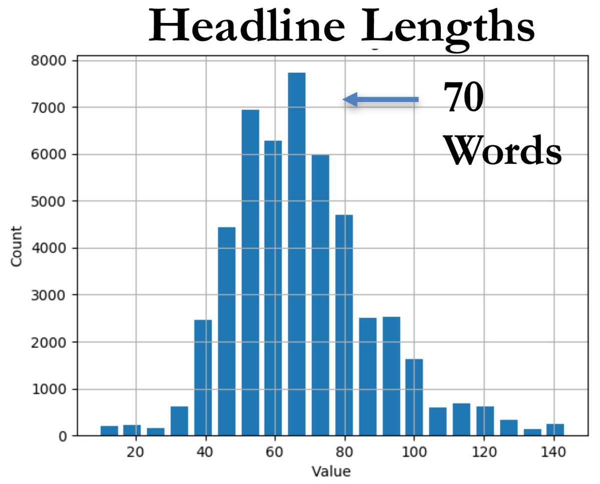
$S_1 = Acc_{Related,Unrelated}$

$S_2 = Acc_{Agree,Disagree,Discuss}$

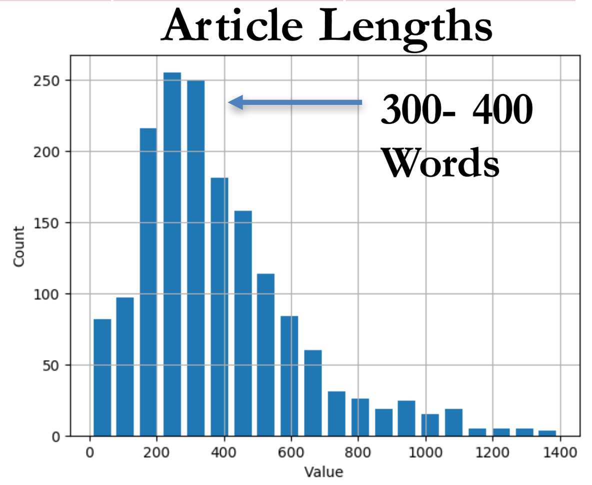
$S_{FNC} = .25\, S_1 + .75\, S_2$

Pairs	Unrelated	Discuss	Agree	Disagree
49972	0.731	0.178	.074	.017

Headline Lengths

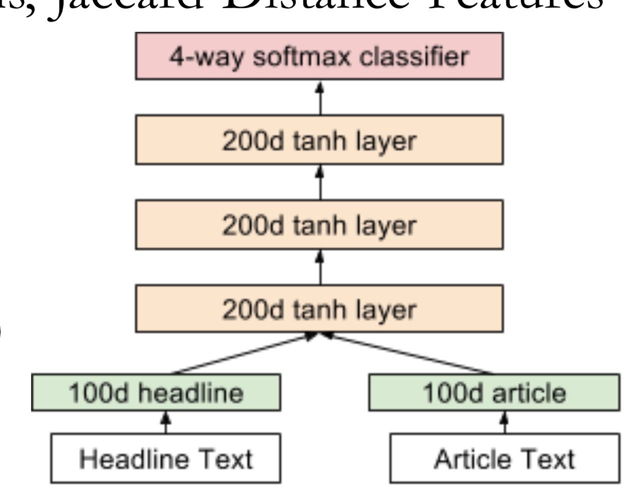


Article Lengths



Baseline Models

- Multinomial Bayes Classifier w. BLEU Score, Unigrams, Crossgrams, Jaccard Distance Features
- Bag of Words MLP w. 300 Dim GloVe Embed. (Bowman et. al. 2015)



- LSTM w. Concatenated Headline + Article as Input, 300 Dim GloVe Embed & Softmax Classifier (Bowman et. al. 2015)

Model	FNC-1 Score
Linear Baseline	**.7860**
B.O.W. MLP	.7787
LSTM (Concatenated Input)	.4005

Approach

Subproblem 1: Related vs. Unrelated

- Proposed Solution: Linear Model

Subproblem 2: Agree vs. Disagree vs. Discuss

- Proposed Solution: RNN Variants

Methods

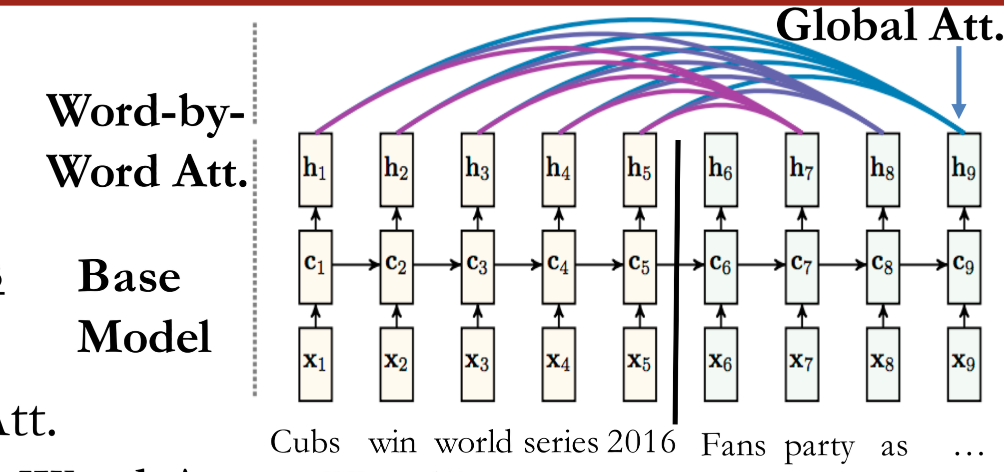
Subproblem 1: Related vs. Unrelated

- SVM Classifier
- Feat: (Article, Headline) TF-IDF Cos. Sim.

Subproblem 2: Agree vs. Disagree vs. Discuss

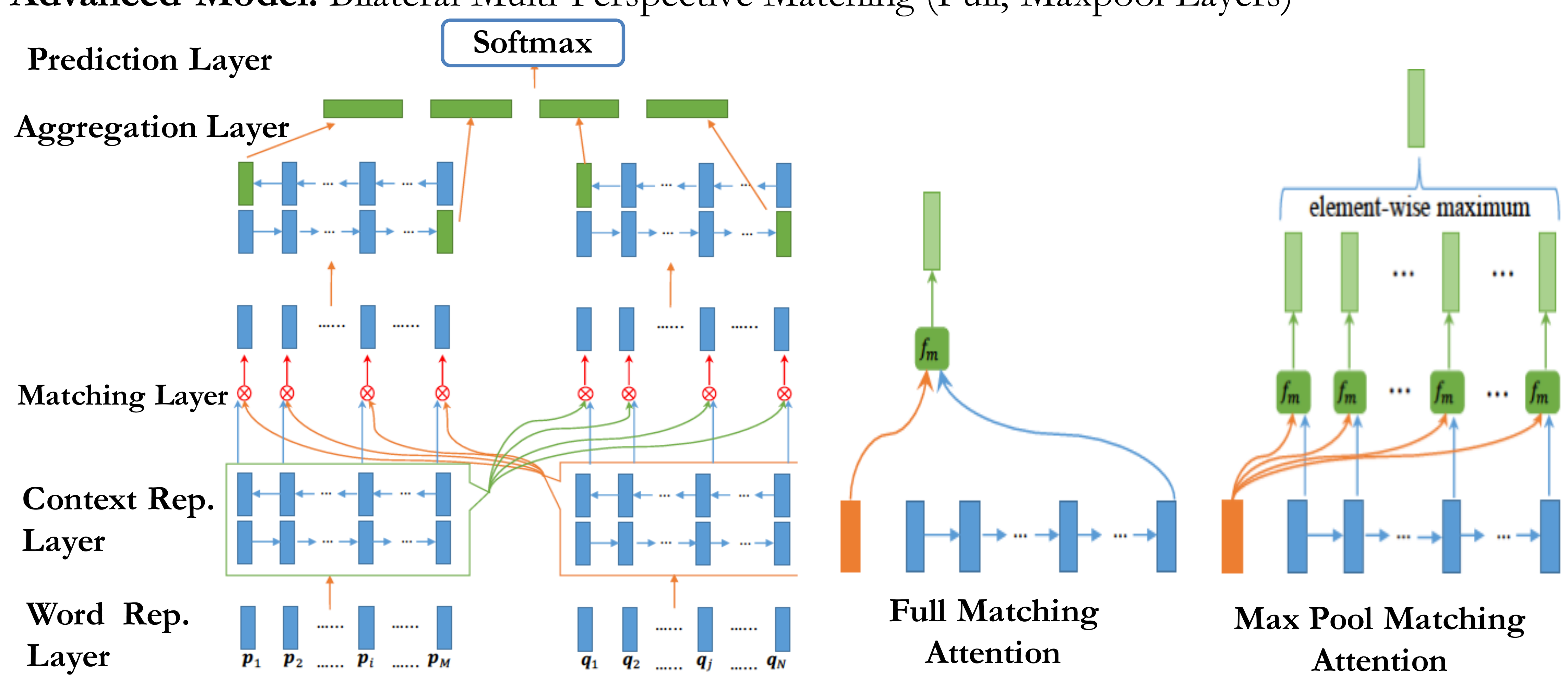
Base Model: Conditionally Encoded (C.E.) LSTM

- Variant 1: C.E. LSTM w. Headline-Article Global Att.
- Variant 2: C.E. LSTM w. Headline-Article Word-by-Word Att.
- Variant 3: C.E. LSTM w. Bidirectional Global Att.
- Variant 4: Bidirectional C.E. LSTMs w. Bidirectional Global Attention
- Variant 5: Multilayered Implementation of Variant 4



(Rocktaschael et. al. 2016)

Advanced Model: Bilateral Multi-Perspective Matching (Full, Maxpool Layers)



(Wang et. al. 2017)

Results

Subproblem 1: Related vs. Unrelated

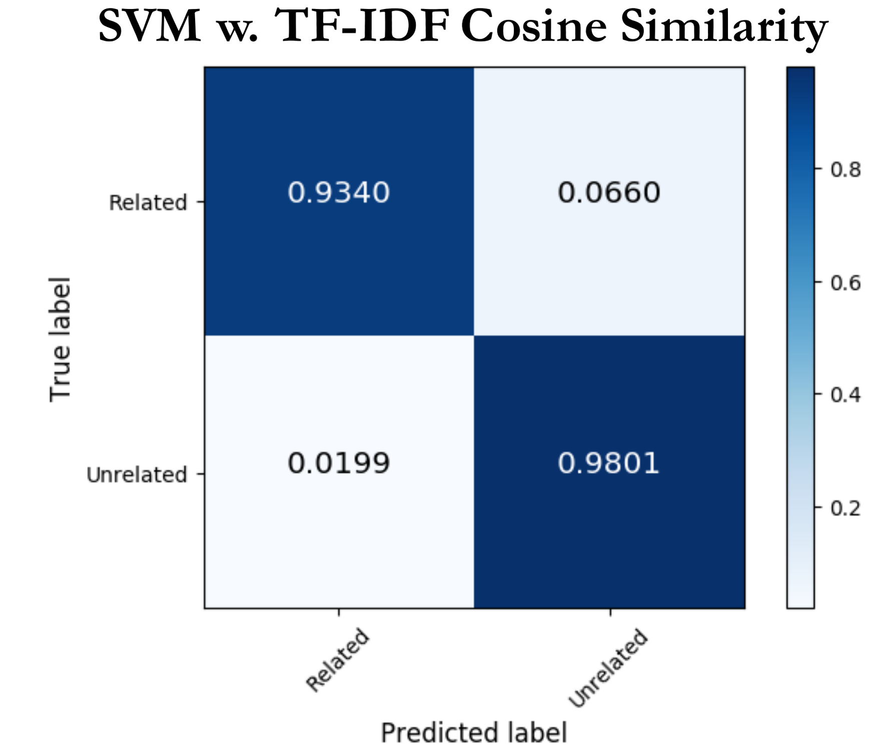
Model	F1-Score
Headline-Article TF-IDF Cosine Similarity	.9712

Subproblem 2: Agree vs. Disagree vs. Discuss

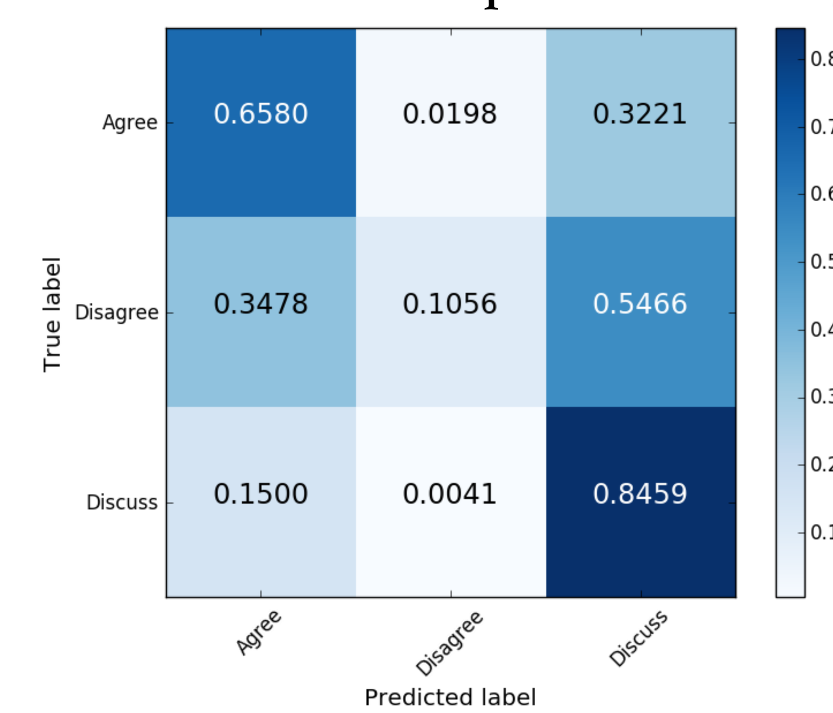
Model	F1-Score	S_2	S_{FNC}
Conditionally Encoded LSTM (Base Model)	.730	.7859	------
C.E. LSTM w. Headline-Article Global Attention (Var. 1)	.753	.8144	------
C.E. LSTM w. Headline-Article Word-by-Word Attention (Var. 2)	.768	.8263	------
C.E. LSTM w. Bidirectional Global Attention (Var. 3)	**.777**	.8324	------
Bidirectional C.E. LSTMs w. Bidirectional Global Attention (Var. 4)	.761	**.8507**	**.8658**
Multilayered Implementation of Variant 4 (Var. 5)	.761	.8209	------
Bilateral Multi-Perspective Matching (Full, Maxpool Layers)	.760	.819	.8501

Results

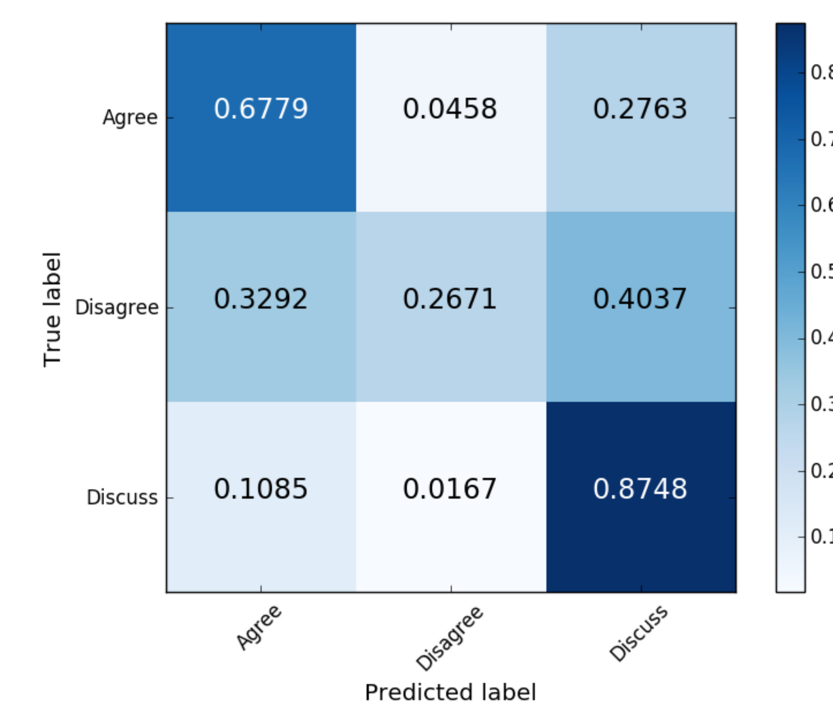
SVM w. TF-IDF Cosine Similarity



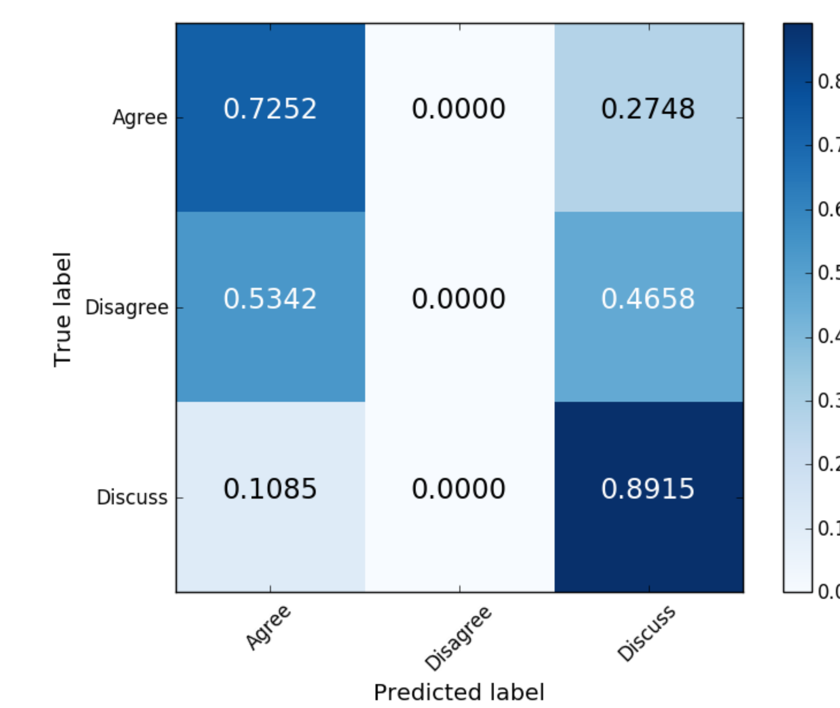
Bilateral Multi-Perspective Matching



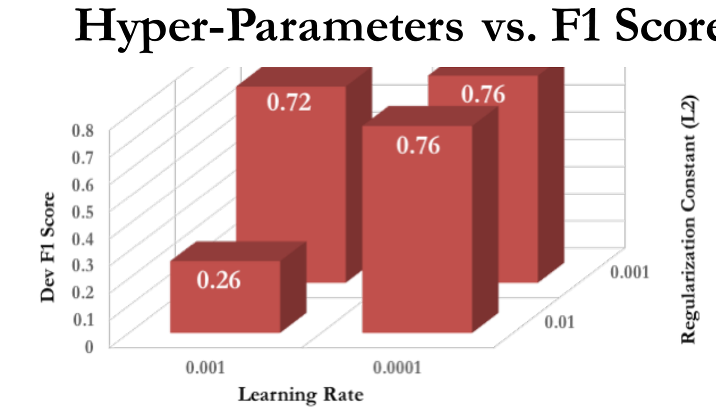
Suboptimal Bidirectional C.E. LSTM w. Bidirectional Global Attention



Optimal Bidirectional C.E. LSTM w. Bidirectional Global Attention

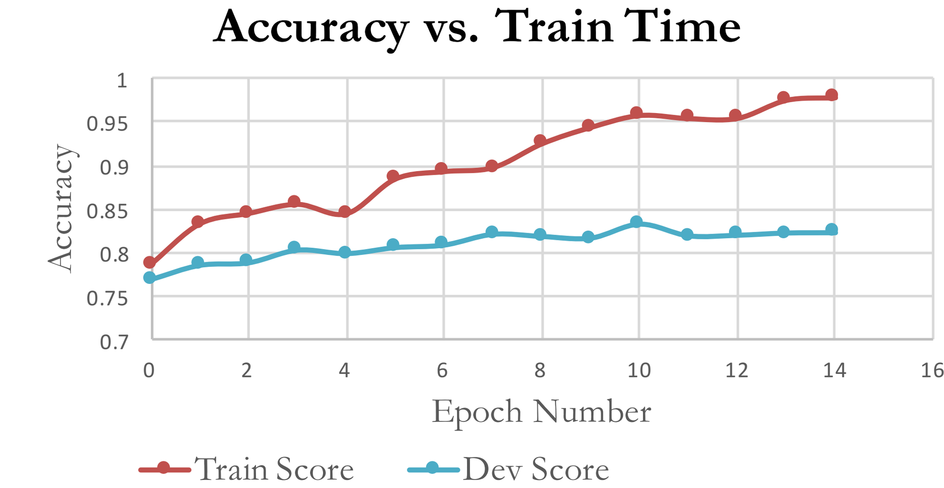


Hyper-Parameters vs. F1 Score



Learning Rate and L2 Regularization Hyper-Parameters vs. F1 Score for Bidirectional C.E. LSTM w. Bidirectional Global Attention. We Also Tuned Drop Out Rates.

Accuracy vs. Train Time



Bidirectional C.E. LSTM w. Bidirectional Global Attention

Discussion + Future Work

- Can Successfully Tune Hyper-Parameters to Classify "Disagree"
- Implemented Attentive-Matching and Max-Attentive-Matching Layers of the Bilateral Multi-Perspective Matching Model but Ran into GPU OOM
- No Leaderboard for FNC-1 Challenge; People Report S_{FNC} Scores of ~.7 - .8
- Future Work: Run Complete Bilateral Multi-Perspective Matching Model with Hyper-Parameter Tuning, Explore Dynamic Co-Attention, & Further Analysis into Improving "Disagree" Classification