Supplemental Material to:
Toward Featureless Event Coreference Resolution via Conjoined Convolutional Neural Networks

Anonymous ACL submission

1 Learned Parameters for the CCNN Model, per the Dev Set

- Features: Lemma and Character Embeddings

- # Training Epochs: 20

- Context Window Size: 0 (this agrees with the Choubey's, et. al. findings (2017)).

- #Negative Examples per Positive (training): 5

- Batch Size: 128

- Pool Type: MaxPooling

- Word Embeddings (used for Lemma Embeddings): GloVe, 300 dimension, trained on 6 Billion Token Corpus

- Dropout: 0.0

- Optimizer: Adam

- #Kernels: 64 (at every level)

2 Learned Parameters for the Neural Clustering (NC) Model, per its own Dev Set

- # Hidden Units: 50

- Batch Size: 5

- #Negative Examples per Positive (training): 5

- Initializer: Normal

- Optimizer: Adam

- Learning Rate: 0.001

3 Training / Development / Test Data

Other researchers who have used the ECB+ corpus use Topics 1-22 for training, 23-25 as development, and 26-45 as testing. We adhere to the same; however, since our NC model needs CCNN's predictions as training data, we must remove some of the training data and use it for separate development sets instead. Our full splits are shown in Table 1.

References

Prafulla Kumar Choubey and Ruihong Huang. 2017. Event coreference resolution by iteratively unfolding inter-dependencies among events. In *EMNLP*.

		ECB+ Topics						
		1-18	19	20	21	22	23-25	26-45
CCNN	TRAIN	✓						
	DEV						✓	
	PREDICT		✓	✓	✓	✓		✓
NC (WD)	TRAIN		✓					
	DEV			✓				
	PREDICT				✓	✓		✓
NC (CD)	TRAIN				✓			
	DEV					✓		
	PREDICT							✓

Table 1: The data splits used by each of our models. Note, "predict" means that we use the model to make predictions on the given data.

2