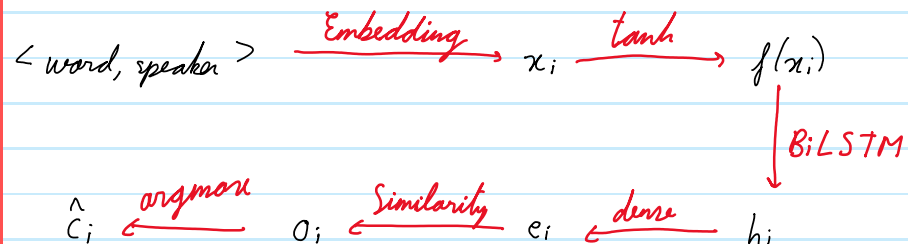


Each word is taken as $\langle \text{Speaker}, \text{word} \rangle$

$$x_i = W_i t_i \parallel \sum_{s \in S_i} W_s s$$



BiLSTM

$$h_i = \text{BiLSTM}(f(x_i), \vec{h}_{i-1}, \overleftarrow{h}_{i+1})$$

Compile:

Loss: NLLLoss
 Optimizer: Adam (lr = 5×10^{-4})

dense

$$h_i \times \begin{matrix} |h| \\ \boxed{W_o} \\ |h| \times |e| \end{matrix} = e_i \quad |e| = k$$

$$e_i = W_o \cdot h_i + b$$

Similarity

$$\text{softmax}(\cosine(E, e_i))$$

$$E = W_s$$

$$|h| = 459$$

$$k = 134$$

train

347 scenes

47 epi

13,280 mentions

372 entities

test

74 scenes

40 epi

2,429 mentions

106 entities

Parameter

→ 5-fold cross validation

→ random search over hyperparameters to maximize F1.

Batch

→ 24 scenes

→ given to model in chunk of 757 tokens

Ws

- Initialized to Word2Vec
 - 300 dim word vec
 - Google News word2vec [skipgram]
- Milkov et al, 2013

Drapant

8×10^{-3} to input of x_i
 13×10^{-4} " " " h_i

Ws

Random initialization

For the final submission of the answers for the test data, we created an ensemble model by averaging the output (Eq. 3) of the five models trained on the different folds.

$$\text{eq 3} \rightarrow o_i = \text{SoftMax}[\text{Cosine}(E, e_i)]$$

$$o_{\text{fin},i} = \sum_{j=1}^5 o_{ij}$$

$i \rightarrow \text{index of word}$
 $j \rightarrow \text{model num}$

$$\hat{c}_{\text{fin},i} = \text{argmax}(o_{\text{fin},i})$$

All entries

67 classes

$66 \rightarrow \{r_i : \text{count}(r_i) \geq 3 ; r_i \in R\}$ \rightarrow set of all references

1 \rightarrow other