Nlp mid

14 March 2023 14:14

Each word is taken as $\langle$ Speaker, word $\rangle$

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

(word: $W_i t_i$; speaker: $\sum_{s \in S_i} W_s s$)

$\langle$ word, speaker $\rangle$ → Embedding → x_i → tanh → $f(x_i)$ → BiLSTM → h_i → dense → e_i → Similarity → o_i → argmax → $\hat{c}_i$

BiLSTM

$$h_i = BiLSTM(f(x_i), \overrightarrow{h_{i-1}}, \overleftarrow{h_{i+1}})$$

Compile:

Loss : NLLLoss

Optimizer: Adam ($lr = 5 \times 10^{-4}$)

dense

h_i $_{1 \times |h|}$ × W_o $_{|h| \times |e|}$ = e_i $_{1 \times |e|}$

$|e| = k$

$e_i = W_o h_i + b$

Similarity

$$softmax(cosine(E, e_i))$$

$E = W_s$

$|h| = 459$

$k = 134$

train	test
347 scener	74 scener
47 epi	40 epi
13,280 mentions	2,429 mentions
372 entities	106 entities

Parameter

→ 5-fold cross validation

→ random search over hyperparameters to maximise F-1.

Batch

→ 24 scener

→ given to model in chunks of 757 tokens

W_t

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

W_s

Random initialization

Dropout

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

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.

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

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

i → index of word
j → model num

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

All entries

67 classes
66 → $\{r_i : count(r_i) \geq 3 \; ; \; r_i \in \mathbb{R}\}$ (set of all relations)
1 → other