Natural Language Inference with Hierarchical BiLSTM Max Pooling Architecture

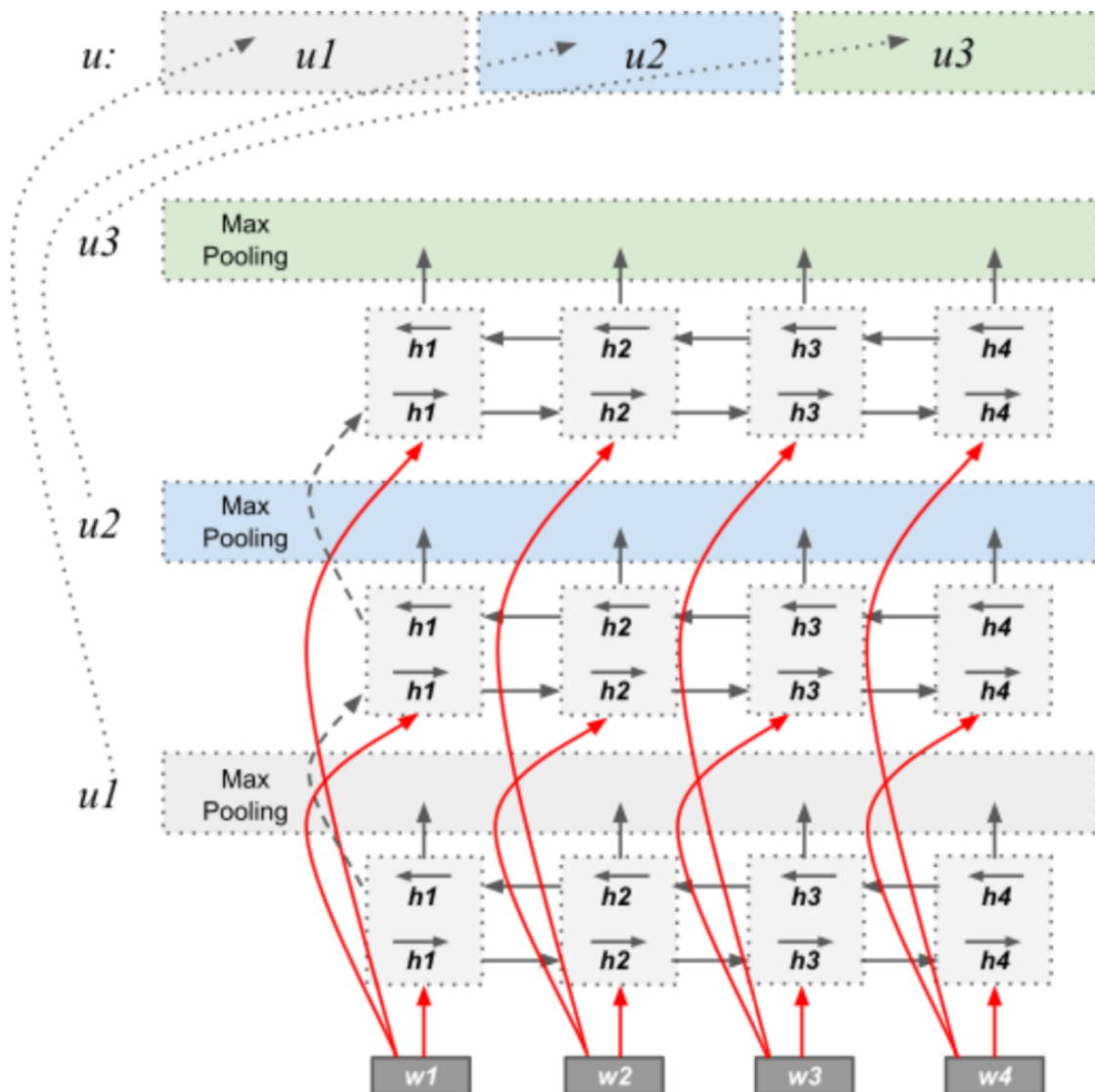| Created	@May 28, 2019 5:50 PM
Property	
Tags	
Updated	@May 31, 2019 8:37 AM

Model Architecture
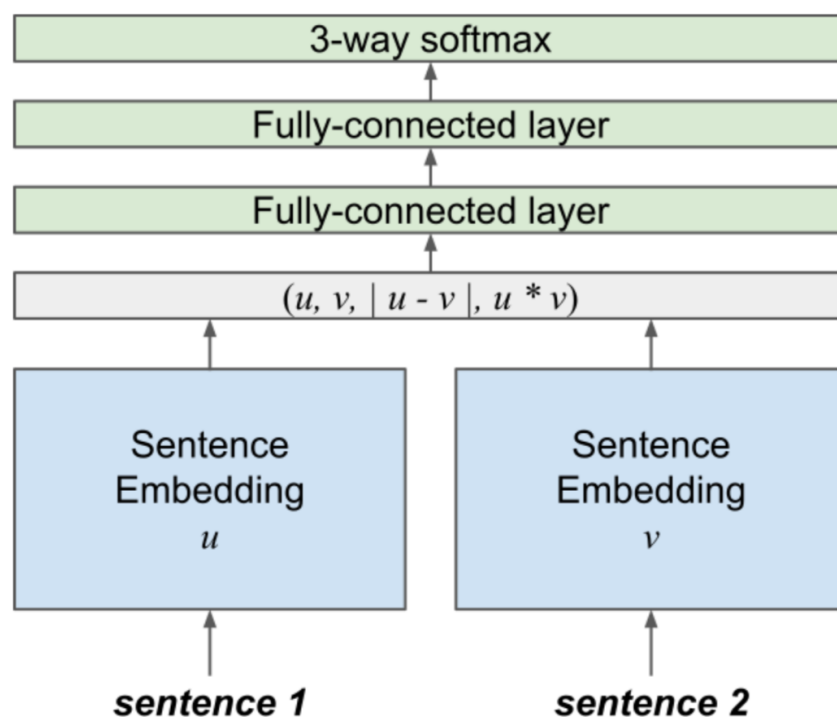
- sentence embeddings are combined

  - concatenation (u, v)

  - absolute element-wise difference $|u - v|$

  - element-wise product $u * v$

- 3-layered multi-layer perceptron (MLP) with a 3-way softmax classifier

- sentence encoder which utilizes BiLSTM with max pooling

$$h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$$

$$\overrightarrow{h}_t = \overrightarrow{LSTM}_t(w_1, \ldots, w_T)$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}_t(w_1, \ldots, w_T).$$

- reads the input sentence as the input

- initialize the initial hidden state and the cell state with the final state of the previous layer

- take the max value over each dimension of the hidden units for each BiLSTM layer

- output of the sentence embedding is the concatenation of each of these max pooling layers

Experimental Results

SNLI

Model	Accuracy
BiLSTM Max Pool (InferSent)[a]	84.5
Distance-based Self-Attention[b]	86.3
ReSA[c]	86.3
600D BiLSTM with generalized pooling[d]	86.6
Our HBMP	**86.6**

MultiNLI

Model	Accuracy (MultiNLI-m)	Accuracy (MultiNLI-mm)
CBOW[a]	66.2	64.6
BiLSTM[a]	67.5	67.1
BiLSTM + enh embed + max pooling[b]	70.7	70.8
BiLSTM + Inner-attention[c]	72.1	72.1
Deep Gated Attn. BiLSTM encoders[d]	73.5	**73.6**
Shortcut-Stacked BiLSTM[e]	**74.5**	73.5
Our HBMP	73.7	73.0

SciTail

Model	Accuracy
DecompAtt[a]	72.3
ESIM[a]	70.6
Ngram[a]	70.6
DGEM w/o edges[a]	70.8
DGEM[a]	77.3
CAFE[b]	83.3
Our LSTM	67.3
Our BiLSTM max pooling	84.9
Our HBMP	**86.0**