

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

Deep Neural Networks with Massive Learned Knowledge: Supplementary Material

A Sentiment Classification

A.1 Transition matrices

$$M_{r=consistent,y=0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_{r=contrastive,y=0} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad M_{r=concessive,y=0} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$
$$M_{r=consistent,y=1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_{r=contrastive,y=1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad M_{r=concessive,y=1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

In practice we add a small number $\epsilon = 0.01$ to the matrices for smoothing.

A.2 Architectures of the clause relation and sentiment networks

The clause sentiment network has the exact same architecture as the base network [Kim, 2014], and the discourse relation network first extracts representations of two input clauses individually using the same convolutional layer architecture, and concatenates the features before feeding into a fully-connected layer. For efficiency we tie the convolutional parameters across all the networks to share the representations, while leaving only the fully-connected layer parameters to be learned separately.

A.3 Conjunction words

- *Consistent*: as well as; and; because; since; similarly; likewise; also; moreover; furthermore; so; therefore; consequently; hence; thus; meanwhile; in particular; in addition; namely; for example; for instance; then; as a result; besides; in other words; above all
- *Contrastive*: but; however; in contrast; whereas; on the contrary; instead; on the other hand
- *Concessive*: though; although; despite; in spite of; even though; even if; even as

A.4 Negation and word polarity lexicons

We use Bing Lius Opinion Lexicon ¹ for word polarity, which contains around 6,800 polarity-carrying words.

We use the LIWC ² lexicon of “negate” category for negation words, which includes around 60 words.

References

[Kim, 2014] Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proc. of EMNLP*.

¹<https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html#lexicon>

²<http://liwc.wpengine.com>