School of Computing and Information Systems The University of Melbourne COMP90042

NATURAL LANGUAGE PROCESSING (Semester 1, 2020)

Workshop exercises: Week 7

Discussion

- 1. What are **contextual representations**?
- 2. How does a **transformer** captures dependencies between words? What advantages does it have compared to RNN?
- 3. What is **discourse segmentation**? What do the segments consist of, and what are some methods we can use to find them?
- 4. What is an **anaphor**?
 - (a) What is anaphora resolution and why is it difficult?
 - (b) What are some useful heuristics (or features) to help resolve anaphora?

Programming

- 1. In the iPython notebook 10-bert, we provide an example on how we can use a pre-trained BERT model and fine-tune it for a sentiment analysis task. As we'll need a GPU to train BERT, we'll be running the notebook on colab, which provides one free GPU. So the first step is to go to: https://colab.research.google.com/ and sign up or login to a Google account. Next go to "File > Upload Notebook" and upload the notebook (10-bert.ipynb) to colab.
 - Fine-tune the model with more epochs (e.g. 4), and take the best model (based on development performance) and measure its performance on the test set.
 - Modify the code so that you can freeze the BERT parameters from updating during fine-tuning. What performance do you now get?

Get ahead

- Extend the notebook 10-bert for other tasks:
 - Sentence similarity (STS 2017): http://alt.qcri.org/semeval2017/ task1/index.php?id=data-and-tools
 - Question answering (SQuAD v1.1): https://rajpurkar.github.io/SQuAD-explorer/