**COMP5423 Natural Language Processing**

Neural Language Models

Due: Tuesday, 4 March, 2025

Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NOTE: Please submit your answers via Blackboard by due time.

Homework:

Both RNN and BERT are effective for sequence labeling tasks like Part-of-Speech Tagging and Name Entity Recognition.

(1) Can these models also be applied to Chinese word segmentation? If yes, how should we prepare the training data, including label annotation, and how can we train the RNN and BERT models for this task?

(2) In addition, are there traditional non-neural-network sequence models that are suitable for Chinese word segmentation? If so, what are they and what features would you like to consider as input?

(3) Furthermore, is it possible to use an encoder-decoder model for Chinese word segmentation? If so, how can this be achieved?

If you answer to the above questions are no, please explain why.

(Maximum 600 words)