

3 Preposition Prediction

In this section, you will use your language model to answer what preposition to use in a sentence. We provide you with three files at <https://github.com/qtli/COMP7607-Fall2023/tree/master/assignments/A1/data/prop>:

- [dev.in](#): The whitespace-tokenized sentences for developing your language model, where 5 prepositions (at, in, of, for on) are replaced with a special token <PREP>.
- [dev.out](#): The correct answers for the dev set.
- [test.in](#): Input sentences for testing in the same format as dev.in, which will be released a week before the deadline.

Your task is to use your language model to predict for each special token <PREP> in the validation (and soon the test) set, which preposition is the proper one there. You should optimize the accuracy of your model on the dev set and predict the prepositions of sentences in the test set. Please strictly follow the format of dev.out for your output test.out as well.

Please explain your method, report the accuracy of your model on dev sets, and discuss the results. Based on your hyperparameters and results, explain the relationship between the test perplexity model and performance in applications.

Go for it, Zoie. The journey of language modeling is about to start!