IDS 422: Homework 3

Due Oct 11, 2023

Problem.

In this problem, apply two inference-based (CBOW and Skip-Gram) methods to the Penn Treebank dataset (PTB) dataset. The example of using code/dataset/ptb.py is as follows:

```
import sys
sys.path.append('..')
from dataset import ptb


corpus, word_to_id, id_to_word = ptb.load_data('train')
```

The above code shows how to load the training data. Similarly, you can load the test data and validation data using ptb.load_data('test') and ptb.load_data('valid'), respectively.

- Train a CBOW model and a Skip-Gram Model using the training data
- Use the distributed representations obtained from these two models to find the 5 words that are most similar to 'stock', 'revenue', 'savings' and 'debt', and compare the results
- Use the parameters learned from the above two models to solve the following analogy problems:
  1. take:took = go:?
  2. car:cars = child:?
  3. good:better = bad:?