DS-GA 1012: Natural Language Understanding and Computational Semantics (Spring 2025)

Allen George Ajith (aa12938)

February 25, 2025

Problem 1c: (Extra Credit)

```
      4 # This line of code doesn’t work:
----> 5 print(embeddings["the"])

File ~/Desktop/NLU/nlu_s25/hw1/embeddings.py:48, in Embeddings.__getitem__(self, words)
     46 embed_array = []
     47 for word in words:
---> 48     index = self.indices[word]
     49     embed_array.append(index)
     51 return self.vectors[embed_array]


     KeyError: ’h’
```

The above line of code does not work because the `__getitem__` function of the embedding class expects a list of words (which is a type of Iterable[str]). The logic implemented in that method is to initialize an embeddings object and index each element in the iterable. Since it gets only **’the’**, it checks for ’t’, ’h’ and ’e’. The error happens at ’h’ because ’t’ is present in the embedding text file so the next character ’h’ which is not present in the embedding file throws the error.

Problem 4a:

Embedding Space	Semantic	Syntactic	Overall
GloVe 50	40	27.6	33.2
GloVe 100	44.5	27.8	35.4
GloVe 200	31.7	21.7	26.2

Table 1: Total analogy question accuracies with k = 1

The below image of Table 4 from the paper shows the semantic, syntactic and overall accuracies respectively of CBOW and Skip-gram approaches. They are mostly higher than our GloVe results except for the Semantic accuracy of CBOW (15.5) which is significantly lower than our GloVe embeddings across all dimensions.

CBOW	300	783M	15.5	53.1	36.1
Skip-gram	300	783M	**50.0**	55.9	**53.3**

The Skip-Gram approach performs better than CBOW and GloVe in understanding word meanings for analogy tasks. This is because Skip-Gram predicts nearby words for a given word, helping it learn more details, especially for rare words. In contrast, CBOW averages multiple words, which can lose important information, and GloVe focuses on overall word co-occurrence, which may not capture fine details.

Dimensionality of the Embedding space

Yes, we can see in Table 1 that increasing the dimensionality from 50 to 100 leads to a slight improvement in overall accuracy (33.5% → 34.6%), but jumping to 200 dimensions actually results in a decrease (down to 26.2%). This suggests that simply using more dimensions does not guarantee better analogy performance. Other training details, hyperparameters, and corpus differences can all influence whether higher-dimensional embeddings improve or hurt accuracy.

Problem 4b:

Embedding Space	Semantic	Syntactic	Overall
GloVe 50	56.6	53.6	55
GloVe 100	66.5	65.9	66.2
GloVe 200	70.5	67.2	68.7

Table 2: Total analogy question accuracies with k = 2

We can clearly see better results with more lenience. It makes sense because there is a greater chance for the target word to be in the set of 2 closest words that we get when set k = 2.

When allowing some flexibility in evaluation (k = 2), GloVe consistently performs better than both Word2Vec models. At 200 dimensions, it reaches 68.7% overall accuracy, compared to Skip-Gram's 53.3%. This suggests that while Skip-Gram is better under strict conditions, GloVe captures broader word relationships, making it more effective when small ranking mistakes are allowed.

Dimensionality of the Embedding space

Yes, we can see in Table 2 that increasing the dimensionality from 50 to 200 leads to a decent improvement in overall accuracy (55% → 68.7%) as well the semantic and syntactic accuracies.

Problem 4c:

Analogy Question	Gold Answer	GloVe 50	GloVe 100	GloVe 200
france : paris :: italy : x	rome	rome	rome	rome
france : paris :: japan : x	tokyo	tokyo	tokyo	tokyo
france : paris :: florida : x	tallahassee	miami	florida	florida
big : bigger :: small : x	smaller	larger	larger	smaller
big : bigger :: cold : x	colder	cold	cold	cold
big : bigger :: quick : x	quicker	quick	quick	quick

Table 3: Analogy qualitative evaluation results.

'rome', 'tokyo' and 'smaller' were predicted correctly and are consistent with the gold answers.