HW1 - Written Report

ap9283@nyu.edu

27 February 2025

1 Problem 4

1.1 4a. Syntactic vs Semantic Relation Types

Semantic relations, involve real-world relationships such as city to country pairs for example *Paris : France :: Berlin : Germany.*

In contrast, syntactic relations pertain to linguistic structures and grammar rules, such as opposites *possible : impossible :: ethical : unethical.*

Embedding Space Accuracy Results (k=1)

Embedding Space	Semantic	Syntactic	Overall
GloVe 50	0.384	0.245	0.295
GloVe 100	0.414	0.247	0.307
GloVe 200	0.313	0.184	0.230

We observe that increasing dimensionality improves accuracy, particularly for semantic relations, with diminishing returns for more complex embeddings spaces.

1.2 4b. Comparing Against k=2

Embedding Space	Semantic	Syntactic	Overall
GloVe 50	0.560	0.504	0.524
GloVe 100	0.634	0.628	0.630
GloVe 200	0.658	0.634	0.642

Increasing leniency (k=2) significantly improves accuracy across all embeddings, reinforcing the trend that higher dimensionality provides better performance, though improvements beyond 100 dimensions are less pronounced. This is also seen in the paper.

1.3 4c. Qualitative Evaluation

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

Overall, the word embeddings correctly capture geographic relationships (syntactic) but struggle with the grammatical analogies, highlighting the limitations within semantics.

1.4 Extra Credit: Problem 1c

The issue arises in the `__getitem__` method implementation. The method expects an iterable (such as a list) but misinterprets a single string input as an iterable of characters. As a result, when querying `embeddings[ẗhe]̈`, the method incorrectly attempts to retrieve embeddings for ẗ¨, ḧ¨, and ë¨, causing a `KeyError` when these characters are not found in the vocabulary.