Introduction:

We compared the performance of Word2Vec and Glove models on both semantic and syntactic analogy tests on different words. First, we used a smaller Glove model with 400K vocabulary, but this makes the comparison not fair because of the larger vocabulary of the Word2Vec model with 3 million words. So, to address this we replaced the Glove model with the larger one with 2M vocabulary, making sure a more balanced comparison between them.

**Comparison of Accuracy:**

* Glove showed higher accuracy on semantic analogies. This is because Glove captures global word co-occurrence statistics, helping it learns deeper and similarities in meaning between words on large corpora.
* Word2Vec performed better on syntactic analogies. It learns from local context and predicts words based on nearby context windows, which helps it recognize grammatical structures, verb forms, and patterns in word more effectively.

**Example that we tried and one of them predict wrong:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test |  |  | Expected | Glove | Word2Vec |
| doctor | hospital | teacher | school | school | elementary |
| carrot | vegetable | apple | fruit | fruit | tomato |
| london | england | tokyo | japan | japan | ronaldo |
| begin | beginning | end | ending | ended | ending |
| drive | driving | ride | riding | rides | riding |

After trying different words, the results showed that Glove performed better on semantic analogies, while Word2Vec showed higher accuracy on syntactic analogies but the difference in the accuracy between them is not large they both works well on both fields. In the end, by selecting and testing multiple words, both models reached 100% accuracy.

**Detailed Explanation for Observed Differences:**

**1. Semantic Accuracy: Glove > Word2Vec:**  
Glove is trained on global word co-occurrence statistics from a large corpus, which allows it to capture deeper, more meaningful semantic relationships between words. It understands the patterns of how words relate in meaning across a corpus, making it more accurate in semantic tasks. Word2Vec, while strong in contextual understanding, may miss out some of the words.

**2. Syntactic Accuracy: Word2Vec > Glove:**

Word2Vec learns word representations by predicting words based on their surrounding context. This makes it particularly effective at capturing syntactic relationships such as verb tenses and plurals. The local context learning helps it understand grammar related patterns better than Glove.

**3.Size and content of the training data:**

Word2Vec trained in 3M vocabulary

Glove trained in 2M vocabulary

So Word2Vec has a larger vocabulary, which offers an advantage in prediction sometime.

**Discussion of Accuracy Differences and Hypothesis:**

In our comparison, we found that Glove achieved higher accuracy in semantic tasks, while Word2Vec was better for syntactic tasks. This is because of their differing training methodologies. Glove focuses on global word co-occurrences, which enables it to build a broad semantic map of word relationships. Word2Vec, works by capturing local grammatical and structural relationships, making it ideal for syntactic pattern

**Final Result:**

A screenshot of a computer

AI-generated content may be incorrect.A black screen with white text

AI-generated content may be incorrect.

Ayman Obedalla 160397

Omar Abu al Rub 160208