COMP 472 - Mini Project 3

●●●
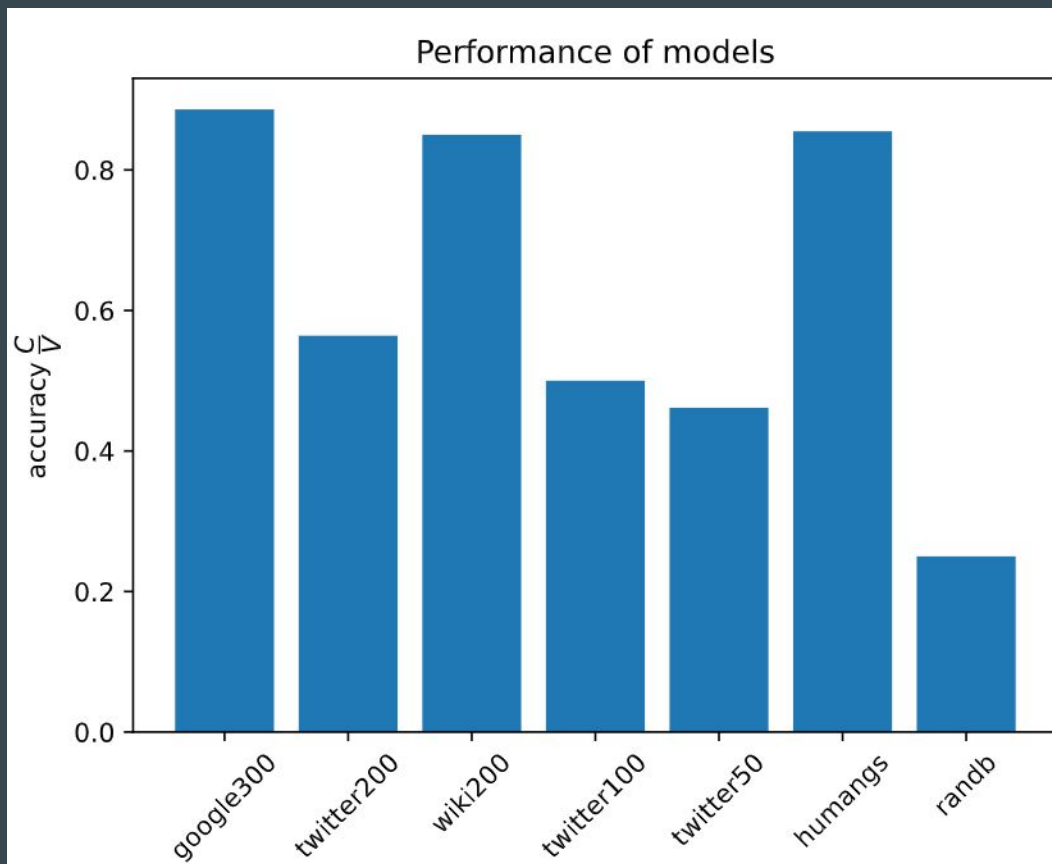
Carl Randyl Tuquero - Lars Martin

Guessing

- Use try-except to handle words not known in the model
- Set flag (-1) for respective cosine value
- If maximum cosine is negative → question word or all answers unknown → guess word

```python
        try:
            cosines[i, j] = round(self.model.similarity(
                dataset[i, 0], dataset[i, 2+j]), 4)
        except Exception as e:
            cosines[i, j] = -1
            pass
```

```python
if all([round(x, 4) < 0 for x in cosines[i, :]]):
    idx_guess = random.randint(0, 3)
    label = 'guess'
```

Results

Performance of models

Analysis

name	voc-size	# correct	# answered w/o guessing	accuracy
word2vec-google-news-300	3000000	70	79	0.8861
glove-twitter-200	1193514	44	78	0.5641
glove-wiki-gigaword-200	400000	68	80	0.85
glove-twitter-100	1193514	39	78	0.5
glove-twitter-50	1193514	36	78	0.4615

Speculations

- google: higher vocabulary size and embedding size → highest accuracy
- twitter: higher embedding size → higher accuracy
- Complex words are not used on Twitter
- Many words on twitter are slang/abbreviations

Contributions

Lars: Task 1

Carl: Task 2