

Answer 8

Used the following relations

- 1. verb (present vs present continuous)
- 2. singular vs plural
- 3. nation vs national language

Created the following analogies (questions) for words present in vocab

```
links.txt x testq4.py x q7.py
1 : verb-present--continuous
2 eat eating drink drinking
3 laugh laughing cry crying
4 drive driving ride riding
5 : singular-plural
6 man men woman women
7 night nights dream dreams
8 fan fans follower followers
9 : Nation-language
10 england english france french
11 spain spanish germany german
12 france french italy italian
```

Accuracies generated using the dense word embedding vectors using the cosine similarity metric

```
Harshith Guru Prasad@hp MINGW64 ~/Desktop/NLP_HW5
$ python q7.py
Group Best-1 Acc Best-5 Acc Best-10 Acc
verb-present-continuous 0.667 0.667 0.667
singular-plural 0.667 0.667 0.667
Nation-language 0.667 0.667 1.000
```

Observations:

The stem(lemma) or root of a word can be used as a powerful feature in determining the similarity of words. Nations, nationalities, languages are easily derivable from each other. Similarly, for the plural forms of words, the common stem form helps in identifying similarities and associated contexts. Proper nouns are associated with semantically closer words than common nouns. This is because of the limited context in which their meaning is defined and used.

Tried to determine the accuracies using sparse context count vector but the results were disappointing with accuracies close to 0. Sparse context count vectors are less helpful for analogies and suffer from performance issues. The dense vectors predicted 2/3 analogies accurately in each group. The relation nation-language has a best-10 accuracy of 1. This is on account of the similar context in which the national language of a nation is used. It can also be used in the context of nationality or as an adjective.

Alternate word tenses and comparative degrees can be predicted with good accuracies on account of same stem and frequency of similar contexts.

Predicting nations, nationalities, languages, adjectives associated with nations, locations and proper nouns is also relatively easy as a results of same stem and context similarity.