Abdulellah Abualshour

Introduction to Data Science

Vector Representations Assignment

April 20th, 2018

Assignment Report

Task 1

I have implemented a program in which the user can choose a word, and the program will look through all of the data provided by GloVe to find the 5 nearest neighbors of that word using cosine similarity. This implies that these words are similar in meaning and/or general context overall. I chose the word "Beyonce" since my name is not present in the data. The results were as follows:

```
5 Nearest Neighbors:
beyoncÃ©,    Cosine similarity: 0.6839737560684697
Knowles,     Cosine similarity: 0.6339364248584699
blige,       Cosine similarity: 0.6282776868369527
mariah,      Cosine similarity: 0.5845779723596153
kanye,       Cosine similarity: 0.5634624493270185
```

Task 2

I have implemented a program that takes a sentence S_0 which is "I love watching movies at night" and computes its vector representation by calculating the component-wise sum of all features of the sentence, and then taking the average. After that, the program takes S_0's vector representation and finds the 5 nearest neighbors from GloVe's data using cosine similarity. The results were as follows:

```
5 Nearest Neighbors:
i,            Cosine similarity: 0.7506543042896137
you,          Cosine similarity: 0.7446401160038179
watching,     Cosine similarity: 0.7317324082291837
just,         Cosine similarity: 0.7200654932100319
n't,          Cosine similarity: 0.7152451209454037
```

Task 3

I have implemented a program that computes the similarities between S_0 which is "I love watching movies at night" and S_1 which is "I love watching movies during the day", and between S_0 and S_2 which is "Donald Trump is the president of the United States", where S_0 is similar in meaning to S_1, and S_0 is dissimilar in meaning to S_2. Looking at the results, we can conclude that they are reasonable:

Sentences	Cosine Similarity
S_0 and S_1	0.933744545977081
S_0 and S_2	0.522049967488389

Since S_0 and S_1 are similar in meaning, they have a higher cosine similarity than S_0 and S_2.