CS 4803/8803: Big Data and Society - Misuse, Abuse, and Algorithms
Spring 2019

Assignment #4 (110 points)

Due: April 3, 2019

Readings:

- Chapter 6: Weapons of Math Destruction (Sweating Bullets: On the Job)
- "A Few Useful Things to Know about Machine Learning" by Pedro Domingos https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

In this assignment, you'll apply AI/ML algorithms related to two applications – word embedding and facial recognition.

Task Set #1: Here you will use distributional vectors trained using Google's deep learning Word2vec system.

1. Familiarize yourself with the original paper on word2vec - Mikolov et al. (2013) (http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf). To learn more about the system and how to train your own vectors, you can find more information here (https://code.google.com/archive/p/word2vec). To learn about the python wrapper around Word2vec, you can find more information here (https://rare-technologies.com/word2vec-tutorial/)
2. Install Gensim : pip install gensim. | pip install --upgrade gensim
3. Download the reducedvector.bin file which is a a pre-trained Word2vec model based on the Google News dataset (https://code.google.com/archive/p/word2vec/)
```
from gensim.models import Word2Vec
import gensim.models
import nltk
newmodel = gensim.models.KeyedVectors.load_word2vec_format(<path to
reducedvector.bin>, binary=True)
```
4. We can compute similarity measures associated with words within the model. For example, to find different measures of similarity based on the data in the Word2vec model, we can use:
```
# Find the five nearest neighbors to the word man
newmodel.most_similar('man', topn=5)

# Compute a measure of similarity between woman and man
newmodel.similarity('woman', 'man')
```
5. To complete analogies like woman is to king as man is to ??, we can use:
```
newmodel.most_similar(positive=['woman', 'king'], negative=['man'], topn=1)
```

Q1: Take as your target word woman. Use the pre-trained word2vec model to rank the following 10 words from the most similar to the target word to the least similar to the target word. For each word, provide the similarity score.

boy
girl
child

queen
man
marriage
birth
elephant
introspection
pregnant
children

Q2: According to the word embeddings, which word is the most different from all the others? Which two words does word embeddings identify as the most similar to each other?

a. ['tissue', 'papyrus', 'manila', 'newsprint', 'parchment', 'gazette']
b. ['engineer', 'nurse', 'doctor', 'mother', 'father', 'scientist']
c. ['criminal', 'black', 'hispanic', 'man', 'woman']

Q3: Sentences:

man is to woman as king is to ___?
water is to ice as liquid is to ___?
bad is to good as sad is to ___?
nurse is to hospital as teacher is to ___?
usa is to pizza as japan is to ___?
human is to house as dog is to ___?
grass is to green as sky is to ___?

a. Complete the above sentences with your own word analogies. Use the Word2Vec model to find the similarity measure between your pair of words. Provide this information.

Example:

man is to woman as king is to _queen__?
newmodel.similarity('king', 'queen') -> 0.5685571

b. Use the Word2Vec model to find the word analogy and corresponding similarity score. Provide this information.

Example:

man is to woman as king is to ___?
newmodel.most_similar(positive=['man', 'woman'], negative=['king'], topn=1) -> girl, 0.50538

c. Lastly, compute and print the correlation between the vector of similarity scores from your analogies versus the Word2Vec analogy-generated similarity scores. What is the strength of the correlation?
   - .00-.19 “very weak” correlation
   - .20-.39 “weak” correlation
   - .40-.59 “moderate” correlation
   - .60-.79 “strong” correlation
   - .80-1.0 “very strong” correlation

Task Set #2: For the next set of lectures on Fairness/Bias, we'll be using the AI Fairness 360 Open Source Toolkit (https://aif360.mybluemix.net/). For this part of the assignment, go through the Bias in Image based Automatic Gender Classification tutorial (*Only Step 1 and Step 2*) and answer the following questions:

- https://github.com/IBM/AIF360/blob/master/examples/tutorial_gender_classification.ipynb

Q1: Each image in the dataset has a unique value representing age, gender, and race based on the following legend:

- age: indicates the age of the person in the picture and can range from 0 to 116.
- gender: indicates the gender of the person and is either 0 (male) or 1 (female).
- race: indicates the race of the person and can from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).

Compute and document the frequency of images associated with each subgroup for age (subdivide based on the NIST study discussed in lecture - (0, 6]; (6,12]; (12,18]; (18,24]; (24,30]; (30,36]; (36,42]; (42,48]; (48,54], (54,60], (60,66], (66,72], (72,116)), gender (0,1), and race (0 to 4). Which subgroup in each age, gender, and race category has the largest representation? Which subgroup in each age, gender, and race category has the least representation?

Q2: In this tutorial, the researchers restricted the images for training the baseline classifier to White and Others races and set the prediction (i.e. output) based on gender. They then computed (among others) a metric called the Equal Opportunity Difference, which is the difference in the true positive rates between the unprivileged and the privileged groups. For Q2, select a different race combination to train the baseline classifier (i.e. replace the parameters for - unprivileged_groups = [{'race': 4.0}], privileged_groups = [{'race': 0.0}]).

What is the corresponding value for the Equal Opportunity Difference metric? Would you consider this value as showing bias? Why or Why not?