AI, Ethics, and Society

Spring 2020

Homework Project #4

Readings:
- Chapter 7: 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 embeddings 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 (Example: pip install gensim. | pip install --upgrade genism)
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: We will use the target words - *man* and *woman*. Use the pre-trained word2vec model to rank the following 15 words from the most similar to the least similar to each target word. For each word-target word pair, provide the similarity score. Provide your results in table format.

boy
girl
child
queen

king
man
woman
marriage
birth
doctor
nurse
teacher
engineer
scientist
president

Q2: The Bigger Analogy Test Set (BATS) Word analogy task has been one of the standard benchmarks for word embeddings since 2013 (https://vecto.space/projects/BATS/). Select any file from the downloaded dataset (BATS_3.0.zip) and provide the measure of similarity between words on each row (Remember to document the file used). Select three target words that identify membership associated with one of the protected classes: race, color, religion, or national origin. Compute the similarity between each of the three target words and one word selected from each row. Indicate when there are noticeable differences in the similarity scores based on membership in the protected class. Provide your results in table format.

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 ___?
king is to throne as judge is to ___?
giant is to dwarf as genius is to ___?
college is to dean as jail is to ___?
arc is to circle as line is to ___?
French is to France as Dutch is to ___?
video is to cassette as computer is to ____?
universe is to planet as house is to ____?
poverty is to wealth as sickness 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 your results.

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 your results.

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 this part of the assignment, we will work with the UTK dataset (UTKface_cropped.tar.gz) downloaded from https://susanqq.github.io/UTKFace/

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 - (0-20), (21,40), (41,60), (61,80), (81, 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? Recreate a table of the age group, gender, and race distributions of subjects based on the UTK dataset subgroups (inspired by the one discussed in the lecture and reposted below). Based on what you've learned so far, if an algorithm is trained based on this dataset, which group(s) will be impacted the most? Explain why.

Age group	0–20	21–40	41–60	61+	Total
Female	1, 248	1, 685	1, 011	165	4, 109
Male	1, 427	2, 501	5, 021	2, 641	11, 590
Black	40	532	354	219	1, 145
White	1, 497	3, 368	5, 140	2, 368	12, 373
Asian	1, 126	284	537	219	2, 166
Unknown	12	2	1	0	15
Total	2, 675	4, 186	6, 032	2, 806	15, 699

http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf