CS 8803: AI, Ethics, and Society: Assignment 4

Sagarika Srishti (ssrishti3@gatech.edu)

April 2020

1 Task Set 1

1.1 Answers to Q1

- Similarity with target word 'man', ranked from most similar to least similar:

Word	Similarity Score
man	1.0
woman	0.588
girl	0.523
boy	0.339
child	0.333
doctor	0.289
king	0.265
marriage	0.204
nurse	0.153
birth	0.123
scientist	0.112
queen	0.110
teacher	0.099
president	0.085
engineer	0.087

- Similarity with target word 'woman', ranked from most similar to least similar.

Word	Similarity Score
woman	1.0
child	0.589
man	0.588
girl	0.568
Continued on next page	

Table 1 – continued from previous page

Word	Similarity Score
marriage	0.429
birth	0.420
boy	0.312
nurse	0.254
queen	0.229
teacher	0.204
doctor	0.196
scientist	0.137
king	0.123
president	0.085
engineer	0.044

1.2 Answers to Q2

For this part, I chose the file 'E10 [male - female].txt' from the folder 3_Encyclopedic_semantics.

- Given below are the measures of similarities between the words on each row. Words which were not present in the vocabulary were omitted.

Word1	Word2	Similarity Score
actor	actress	0.86941457
boar	sow	0.49044377
boy	0.girl	0.47565967
brother	sister	0.73543334
buck	doe	0.15203016
bull	cow	0.44119757
businessman	businesswoman	0.6755804
dad	mom	0.74045944
dad	mum	0.57334465
daddy	mommy	0.42494777
daddy	mother	0.25613183
daddy	mom	0.61681795
duke	duchess	0.6362255
emperor	empress	0.66597676
father	mother	0.832764
fox	vixen	0.08036125
gentleman	lady	0.3554653
gentleman	madam	0.4857473

Continued on next page

Table 2 – continued from previous page

Word 1	Word 2	Similarity Score
god	goddess	0.39541078
grandfather	grandmother	0.7366434
grandpa	grandma	0.55757904
grandson	granddaughter	0.6778581
groom	bride	0.3520856
heir	heiress	0.5088582
hero	heroine	0.46802157
hound	bitch	0.39353663
husband	wife	0.6377156
king	queen	0.5685571
man	woman	0.5876938
mister	miss	0.4692849
mister	mrs	0.49664253
mister	ms	0.15184526
mister	madam	0.6186389
nephew	niece	0.73131126
poet	poetess	0.52189505
prince	princess	0.7285844
ram	ewe	0.055268407
rooster	hen	0.29030108
sir	madam	0.22270176
son	daughter	0.7831377
stallion	mare	0.3589837
stepfather	stepmother	0.7483355
tiger	tigress	0.22913313
uncle	aunt	0.6715928
valet	maid	0.44579548
valet	handmaid	0.5327758
waiter	waitress	0.5982927

- For the second part of this exercise, I have chosen three words related to the protected class Race, namely: black, white, and asian. Given below are the similarity measures of the first word in row and the three target words.

Word	Similarity with 'black'	Similarity with 'white'	Similarity with 'asian'
actor	0.11718847	0.112154886	0.13501917
batman	0.12345499	0.07052541	0.05379804
boar	0.2584654	0.25323373	-0.009064779
boy	0.20971368	0.17648886	0.101264946
brother	-0.0031859055	0.009758614	-0.004879918
buck	0.20390256	0.18934101	0.034704495
bull	0.214998	0.3062246	0.06472506
businessman	0.0144363865	0.0057638353	0.032001115
chairman	0.056033432	0.042048335	0.07964939
dad	0.016293392	0.055559866	-0.073422946
daddy	0.2812515	0.19687036	-0.028832044
duke	-0.016956173	0.022518605	-0.03930489
emperor	0.0071171746	-0.017861646	0.04829904
father	0.06559749	0.09366952	-0.019721184
fisherman	0.19385749	0.11725126	0.03571578
fox	0.21027085	0.21393023	0.1580783
gentleman	0.0607187	0.107854046	-0.07444217
god	-0.03965548	-0.025656909	-0.06705533
grandfather	0.006285418	0.029033124	0.010667495
grandpa	0.093665116	0.04892784	-0.012741977
grandson	-0.040286995	-0.029455937	-0.12832639
groom	0.056712423	0.07927805	-0.07606879
headmaster	0.020722339	0.017382141	-0.14455618
heir	-0.005706014	-0.040829297	-0.040426217
hero	0.09830142	0.07290647	0.02172384
hound	0.2606691	0.29115742	0.16440864
husband	0.028756931	0.037834853	0.01817239
king	0.055144865	0.067879334	-0.043314017
lion	0.39581382	0.41088715	0.08285851
man	0.19195054	0.2250261	0.02412987
manager	-0.015603829	-0.024080247	-0.0032726089
mister	0.17819813	0.19546364	-0.12864235
murderer	0.11469218	0.088783264	-0.112529926
nephew	-0.09416974	-0.061811283	-0.102951765
poet	0.027736388	0.01990934	0.05144231
policeman	0.047845848	0.04271179	-0.086236015
prince	0.098829485	0.086296335	0.008170079
ram	-0.07121579	-0.026666922	-0.16180189
rooster	0.104399785	0.062109508	-0.07390449
sculptor	0.03023398	0.019829229	0.061279356
sir	0.046330474	0.07660729	-0.084699735
son	-0.029756725	-0.0070160106	-0.02929385
stallion	0.27896073	0.28766167	0.024622679
stepfather	0.01157003	0.01850352	-0.16458842
superman	0.11782715	0.12685847	0.043981194
tiger	0.29617834	0.30093005	0.08696288
uncle	0.025488578	0.06250765	-0.079233244
valet	-0.013618473	-0.043459453	-0.16665864
waiter	0.08911914	0.09162006	-0.07081735
webmaster	-0.112115294	-0.07853847	-0.09830445

There are a few instances of noticeable differences based on these target words. For example, the word 'murderer' is significantly more similar to 'black' than it is to 'white' or 'asian'. Similarly, the word 'headmaster' is negatively correlated to 'asian'.

1.3 Answers to Q3

- Given below are the sentences completed with words given by me, and the similarity scores. The word 'Dutch' was not present in the vocabulary so I omitted that sentence.

Sentence	Similarity Score
man is to woman as king is to queen	0.5685571
water is to ice as liquid is to solid	0.65464735
bad is to good as sad is to happy	0.44885093
nurse is to hospital as teacher is to school	0.5326568
usa is to pizza as japan is to sushi	0.011866331
human is to house as dog is to kennel	0.28415978
grass is to green as sky is to blue	0.44396985
king is to throne as judge is to bench	0.30267337
giant is to dwarf as genius is to stupid	0.35719013
college is to dean as jail is to jailer	0.22684216
arc is to circle as line is to polygon	0.22809681
video is to cassette as computer is to cd	0.23679918
universe is to planet as house is to stuff	-0.0067000575
poverty is to wealth as sickness is to health	0.19527602

- Given below are the analogies as completed by the Word2Vec model, and the corresponding similarity scores.

Sentence	Similarity Score
man is to woman as king is to queen	0.553
water is to ice as liquid is to solid	0.45
bad is to good as sad is to glory	0.441
nurse is to hospital as teacher is to institution	0.483
usa is to pizza as japan is to dishes	0.576
human is to house as dog is to hound	0.423
grass is to green as sky is to blue	0.548
king is to throne as judge is to prosecution	0.519
giant is to dwarf as genius is to theorist	0.429
college is to dean as jail is to peress	0.544
arc is to circle as line is to lines	0.429
Continued on next page	

Table 3 – continued from previous page

Sentence	Similarity Score
video is to cassette as computer is to peripherals	0.665
universe is to planet as house is to houses	0.426
poverty is to wealth as sickness is to impious	0.496

- The correlation between the vector of similarity scores from my analogies and the vector of similarity scores from the Word2Vec generated analogies is -0.106. The correlation strength is very weak.

2 Task Set 2

Given below in tabular form is the age, gender and race distribution of subjects based on the UTK dataset subgroups.

Age Group	Male	Female	White	Black	Asian	Indian	Others	Total
0-20	1941	2326	1931	160	1017	607	552	4267
20-40	901	1632	1034	100	349	598	452	2533
40-60	914	751	1252	75	88	162	88	1665
60-80	502	466	793	55	47	63	9	968
80-100	114	229	255	14	51	21	2	343
100+	0	3	0	1	1	1	0	3
Total	4372	5407	5265	405	1553	1452	1103	9779

- Given below are the highest gender and race representations in age group.
  - Highest representation in 0-20 age group: Female and White
  - Highest representation in 20-40 age group: Female and White
  - Highest representation in 40-60 age group: Male and White
  - Highest representation in 60-80 age group: Male and White
  - Highest representation in 80-100 age group: Female and White
  - Highest representation in 100+ age group: Female and Black/Asian/Indian
- Given below are the lowest gender and race representations in age group.
  - Lowest representation in 0-20 age group: Male and Black
  - Lowest representation in 20-40 age group: Male and Black
  - Lowest representation in 40-60 age group: Female and Black
  - Lowest representation in 60-80 age group: Female and Others
  - Lowest representation in 80-100 age group: Male and Others

– Lowest representation in 100+ age group: Male and White/Others

If an algorithm is made based on this data, I think that Black and the races included in Others such as Hispanic, Middle Eastern, Latino, etc. are the people that are most likely to be impacted, as they have the minimum representation in almost every age group. For example, if these images are being used to train a facial recognition model, the model is quite likely to incorrectly identify or miss images of people belonging to the Black or Others category, as the model would not have had enough examples from these categories to learn properly.