CS 6603: AI, Ethics, and Society
AI/ML - II Assignment

Cleo Zhang

yzhang3761@gatech.edu

***Abstract*—this assignment will continue to explore word embeddings and facial recognition applications using AI/ML algorithms. For Task Set #1, I will mainly tackle with Word2Vec, calculate the similarity measures, predict analogies and compute the related correlations, while in Task Set #2, the UTK dataset will be analyzed to evaluate the given algorithm.**

1 THE TASK SET #1

1.1 Q1: Similarity Score for Each Word-target Pair

Using the target words *man* and *woman*, the similarity scores and rankings are listed in *Table 1*.

Table 1 —Similarity Score for Each Word-target Pair.

Target - Man	**Similarity Score**	**Target - Woman**	**Similarity Score**	**Words ranked from the most similar to the least similar**
man	1.0	woman	1.0	1
woman	0.5876938	child	0.5898086	2
child	0.33342198	man	0.5876938	3
doctor	0.28924733	husband	0.44964314	4
wife	0.2834791	birth	0.42030883	5
king	0.26449704	wife	0.30068845	6
husband	0.23411639	nurse	0.25435835	7

Target - Man	Similarity Score	Target - Woman	Similarity Score	Words ranked from the most similar to the least similar
nurse	0.153481	queen	0.22857243	8
birth	0.12343917	teacher	0.2040782	9
scientist	0.11226919	doctor	0.19613354	10
queen	0.110419504	scientist	0.13731061	11
professor	0.107622154	king	0.12252855	12
teacher	0.09874003	professor	0.10519859	13
president	0.09457928	president	0.084626846	14
engineer	0.087363556	engineer	0.044264372	15

1.2 Q2: The Bigger Analogy Test Set (BATS)

1.2.1 *Similarity Scores for E10 [male - female].txt*

This section selects *E10 [male - female].txt* from *BATS_3.0.zip* as an example, chooses the first word as the target word from each row and provides the measure of similarity between the target word and the other words on that row. See *Table 2* for details.

Table 2 — Similarity Measures for E10 [male - female].txt.

Targets	Paired Words	Similarity Score
actor	actress	0.86941457
batman	batwoman	N/A
boar	sow	0.49044377
boy	girl	0.47565967

Targets	Paired Words	Similarity Score
brother	sister	0.73543334
buck	doe	0.15203016
bull	cow	0.4411975
businessman	businesswoman	0.6755803
chairman	chairwoman	N/A
dad	mom/mum	0.74045944
daddy	mommy/mother/mom	0.42494774
duke	duchess	0.6362255
emperor	empress	0.66597676
father	mother	0.832764
fisherman	fisherwoman	N/A
fox	vixen	0.08036125
gentleman	lady/gentlewoman/madam	0.3554653
god	goddess	0.39541078
grandfather	grandmother	0.7366434
grandpa	grandma	0.55757904
grandson	granddaughter	0.6778581
groom	bride	0.3520856
headmaster	headmistress	N/A
heir	heiress	0.5088582
hero	heroine	0.46802154

Targets	Paired Words	Similarity Score
hound	bitch	0.39353663
husband	wife	0.6377156
king	queen	0.5685571
lion	lioness	N/A
man	woman	0.5876938
manager	manageress	N/A
mister	miss/missis/mis-sus/mis'ess/mrs/ms/madam	0.46928492
murderer	murderess	N/A
nephew	niece	0.73131126
poet	poetess	0.52189505
policeman	policewoman	N/A
prince	princess	0.72858447
ram	ewe	0.055268407
rooster	hen	0.29030108
sculptor	sculptress	N/A
sir	madam	0.22270176
son	daughter	0.7831376
stallion	mare	0.35898367
stepfather	stepmother	0.7483355
superman	superwoman	N/A

Targets	Paired Words	Similarity Score
tiger	tigress	0.22913316
uncle	aunt	0.6715928
valet	maid/maidservant/housemaid/chambermaid/handmaid/handmaiden/parlormaid/parlourmaid	0.44579548
waiter	waitress	0.59829265
webmaster	webmistress	N/A

1.2.2 *Similarity Scores for words in E10 [male - female].txt and the protected class of Race (White, Black and Asian)*

Table 3 —Similarity Measures for Targets from *E10 [male - female].txt* and subgroups of Race.

Target words	Race		
	White	**Black**	**Asian**
actor	0.112154886	0.117188476	0.13501917
batman	0.070525415	0.12345498	0.05379803
boar	0.25323373	0.25846544	-0.009064786
boy	0.17648886	0.20971368	0.10126495
brother	0.009758618	-0.0031859092	-0.004879916
buck	0.189341	0.20390254	0.03470449
bull	0.30622458	0.21499804	0.064725064
businessman	0.00576384	0.014436396	0.032001123
chairman	0.042048324	0.056033432	0.07964939

Target words	Race		
	White	Black	Asian
dad	0.055559866	0.016293382	-0.07342293
daddy	0.19687034	0.2812515	-0.028832044
duke	0.022518612	-0.016956175	-0.039304893
emperor	-0.01786165	0.0071171783	0.048299044
father	0.09366953	0.0655975	-0.019721195
fisherman	0.117251284	0.19385749	0.035715796
fox	0.21393022	0.21027084	0.15807828
gentleman	0.10785405	0.0607187	-0.07444218
god	-0.025656916	-0.03965547	-0.06705533
grandfather	0.029033132	0.006285426	0.010667495
grandpa	0.04892783	0.09366512	-0.0127419755
grandson	-0.029455945	-0.040287	-0.12832639
groom	0.079278044	0.05671243	-0.0760688
headmaster	0.017382141	0.020722328	-0.1445562
heir	-0.0408293	-0.005706018	-0.040426202
hero	0.07290649	0.09830141	0.021723844
hound	0.29115742	0.2606691	0.16440864
husband	0.037834864	0.02875693	0.018172387
king	0.06787935	0.05514486	-0.043314025
lion	0.41088712	0.39581382	0.0828585

Target words	Race		
	White	Black	Asian
man	0.22502609	0.19195053	0.024129866
manager	-0.02408023	-0.015603832	-0.0032726112
mister	0.19546364	0.17819811	-0.12864235
murderer	0.08878327	0.11469219	-0.11252994
nephew	-0.061811276	-0.09416974	-0.102951765
poet	0.019909333	0.027736388	0.051442318
policeman	0.042711798	0.047845833	-0.08623602
prince	0.086296335	0.09882948	0.008170075
ram	-0.026666924	-0.0712158	-0.1618019
rooster	0.06210951	0.10439977	-0.07390449
sculptor	0.019829215	0.030233975	0.061279375
sir	0.07660729	0.046330467	-0.084699735
son	-0.007016003	-0.029756727	-0.029293839
stallion	0.28766167	0.27896076	0.024622686
stepfather	0.01850353	0.011570036	-0.16458844
superman	0.12685849	0.117827155	0.043981183
tiger	0.30093005	0.29617834	0.08696288
uncle	0.062507644	0.02548857	-0.079233244
valet	-0.043459445	-0.013618459	-0.16665865
waiter	0.09162004	0.08911913	-0.07081734

Target words	Race		
	White	Black	Asian
webmaster	-0.07853848	-0.112115294	-0.098304436

1.3 Q3: Sentences

1.3.1 *My word analogies and similarity measures*

Table 4 shows the words I filled in to complete the sentence and the corresponding similarity score computed from it. "N/A" means the keyword is not present in the given model.

Table 4 —Complete the given sentences and compute the similarity between the pair of words.

Sentences	Words filled in the blank	Similarity Score
king is to throne as judge is to ___?	court	0.6077864
giant is to dwarf as genius is to ___?	idiot	0.34426275
college is to dean as jail is to ___?	jailor	N/A
arc is to circle as line is to ___?	triangle	0.25559562
French is to France as Dutch is to ___?	Netherlands	0.41922885
man is to woman as king is to ___?	queen	0.5685571
water is to ice as liquid is to ___?	solid	0.6546474
bad is to good as sad is to ___?	happy	0.44885093
nurse is to hospital as teacher is to ___?	school	0.53265685
usa is to pizza as japan is to ___?	sushi	0.011866331

Sentences	Words filled in the blank	Similarity Score
human is to house as dog is to ___?	kennel	0.28415978
grass is to green as sky is to ___?	blue	0.44396985
video is to cassette as computer is to ____?	floppy disk	N/A
universe is to planet as house is to ____?	room	0.25021723
poverty is to wealth as sickness is to ___?	health	0.19527602

1.3.2 *Analogies generated from Word2Vec*

Table 5 —Complete the given sentences using Word2Vec and compute the similarity between the pair of words.

Sentences	Highest Similarity Score	Words filled in the blank
king is to throne as judge is to ___?	0.5186458230018616	prosecution
giant is to dwarf as genius is to ___?	0.428088903427124	theorist
college is to dean as jail is to ___?	0.5444425344467163	peress
arc is to circle as line is to ___?	0.4287526309490204	lines
French is to France as Dutch is to ___?	0.6044681072235107	netherlands
man is to woman as king is to ___?	0.5532454252243042	queen
water is to ice as liquid is to ___?	0.4500039219856262	solid
bad is to good as sad is to ___?	0.4403817653656006	glory
nurse is to hospital as teacher is to ___?	0.48289814591407776	institution

Sentences	Highest Similarity Score	Words filled in the blank
usa is to pizza as japan is to ___?	0.576350748538971	dishes
human is to house as dog is to ___?	0.4231664538383484	hound
grass is to green as sky is to ___?	0.5478643178939819	blue
video is to cassette as computer is to ____?	0.6654506921768188	peripherals
universe is to planet as house is to ____?	0.42647024989128113	houses
poverty is to wealth as sickness is to ___?	0.49606096744537354	impious

1.3.3 *Correlation between the vector of similarity scores from your analogies versus the Word2Vec analogy-generated similarity scores*

Table 6 — The correlation of the analogies' vector of similarity scores versus the Word2Vec analogy-generated similarity scores.

Correlation Coefficient	Strength of the correlation
0.01904545	"very weak" correlation

2 TASK SET #2

2.1 Frequency Tables

Frequency tables of images associated with age, gender and race are summarized in *Figure 1*, and we can conclude that:

- For age, the subgroup (0-20) has the most significant representation while (81-116) has the least.
- For gender, the female subgroup (1) has the largest representation, while the male (0) has the least representation.

- For race, the subgroup of white (0) has the largest representation, while Black (1) has the least.

age	0-20	21-40	41-60	61-80	81-116
	4267	2533	1665	969	346

gender	0	1
	4372	5407

race	0	1	2	3	4
	5265	405	1553	1452	1103

Figure 1 — Frequency of images associated with each subgroup for age, gender and race.

Figure 2 shows the entire distribution of the UTK dataset subgroups.

age	0-20	21-40	41-60	61-80	81-116	Total
Female	2326	1632	751	466	232	5407
Male	1941	901	914	502	114	4372
White	1931	1034	1252	793	255	5265
Black	160	100	75	55	15	405
Asian	1017	349	88	47	52	1553
Indian	607	598	162	63	22	1452
Others	552	452	88	9	2	1103
Total	4267	2533	1665	969	232	9666

Figure 2 - Full distribution of the UTK dataset.

2.2 Mostly Heavily Impacted Group

According to Figure 2, the total sample size is 9666, and the "others" race aged 81-116 takes the least representation (about 0.0207% of the sample). Therefore, if an algorithm is trained based on this dataset, the "others" race aged 81-116 will be impacted the most as the data is lacking in this group. The trained algorithm is likely to reinforce the existing stereotypes or discrimination against this minority of the population.