Mini Project 3

WHEN AI MET YOU

Model Executor class

```python
 8  class ModelExecutor:
 9
10      def __init__(self):
11          pass
12
13      def runModel(self, model_name, binary=True):
14          # create output directory if does not already exist
15          if not os.path.exists(os.path.dirname('output')):
16              try:
17                  os.makedirs('output')
18              except OSError as exc:
19                  if exc.errno != errno.EEXIST:
20                      raise
21
22          df = pd.read_csv("synonyms.csv")
23          destination = "output/" + model_name + '-details.csv'
24          f = open(destination, 'w', newline='')
25          writer = csv.writer(f)
26
27          model = KeyedVectors.load_word2vec_format('models/' + model_name + '.gz', binary=binary)
28          answer = ''
29
30          for row in df.values:
31              highestScore = 0.0
32              for i in range(2, 6):
33                  word = row[i]
34                  # verify that both the word and answer are in the model's vocabulary
35                  if word in model.index_to_key and row[0] in model.index_to_key:
36                      print(f"Similarity for {row[0]},{word} = {model.similarity(row[0], word)}")
37                      if model.similarity(row[0], word) > highestScore:
38                          highestScore = model.similarity(row[0], word)
39                          answer = word
40
41              if row[0] not in model.index_to_key or len(list(filter(lambda x: x in model.index_to_key, row))) < 2:
42                  result = "guess"
43              elif row[1] == answer:
44                  result = "correct"
45              else:
46                  result = "wrong"
47              writer.writerow([row[0], row[1], answer, result])
48          f.close()
49          return model
```

Reads through the synonym file

Line 30: iterate through rows of synonyms

Line 32: iterate through the list of synonyms (first two indexes are word and correct synonym.

Line 35: Check if word is in the word2vec model for null safety

Line 37: Check if the similarity is larger than the largest one, this value gets reset at the start of every row

Lines 40-46: Determine if it is a guess, correct or wrong

Line47: Write to the file the results for that row

Analysis Builder class

```python
 7  class AnalysisBuilder:
 8
 9      def __init__(self):
10          pass
11
12      destination = 'output/analysis.csv'
13
14      def writeAnalysis(self, model, model_name):
15          """
16          Given the model and model name,
17          calculates/determines the following information and outputs it to an analysis.csv file:
18
19          (a) the model name (clearly indicating the source of the corpus and the vector size)
20          (b) the size of the vocabulary (the number of unique words in the corpus1)
21          (c) the number of correct labels
22          (d) the number of questions that your model answered without guessing
23          (e) the accuracy of the model (i.e. (c)/(d))
24
25          :return:
26          """
27
28          # create output directory if does not already exist
29          if not os.path.exists(os.path.dirname('output')):
30              try:
31                  os.makedirs('output')
32              except OSError as exc:
33                  if exc.errno != errno.EEXIST:
34                      raise
35
36          # read from the model's details csv (created by the Model Executor)
37          model_filename = 'output/' + model_name + '-details.csv'
38          df = pd.read_csv(model_filename)
39          f = open(self.destination, 'a+', newline='')
40          writer = csv.writer(f)
41
42          vocab_size = len(model.index_to_key)
43          correct = 0
44          questions_answered = 0
45
46          for row in df.values:
47              if row[3] == 'correct':
48                  correct += 1
49  |
50              if row[3] != 'guess':
51                  questions_answered += 1
52
53          if questions_answered == 0:
54              accuracy = 0
55          else:
56              accuracy = correct/questions_answered
57          writer.writerow([model_name, vocab_size, correct, questions_answered, accuracy])
58          f.close()
59
```

Use the model created by the ModelExecutor class to figure out the attributes on line 57.

A question is considered answered if it is correct or wrong, hence the negation for the guess on line 50

Avoid divide by zero exception with if statement on line 53

Task1

```python
from AnalysisBuilder import AnalysisBuilder
from ModelExecutor import ModelExecutor

model_executor = ModelExecutor()
google_model_300 = model_executor.runModel('word2vec-google-news-300')

analyzer = AnalysisBuilder()
analyzer.writeAnalysis(google_model_300, 'word2vec-google-news-300')
```

Execute the google news 300 word2vec model which is downloaded from online.

This will write to the file showing the correct guesses for the google news model

We then pass this model to the AnalysisBuilder which note down the specific analysis for this model including its: Model Name, Vocabulary Size, Correctly guessed words, Total amount of questions/rows, Accuracy

Task2

```
1  from AnalysisBuilder import AnalysisBuilder
2  from ModelExecutor import ModelExecutor
3
4
5  model_executor = ModelExecutor()
6  twitter_model_100 = model_executor.runModel('glove-twitter-100', binary=False)
7  wiki_model_100 = model_executor.runModel('glove-wiki-gigaword-100', binary=False)
8  wiki_model_50 = model_executor.runModel('glove-wiki-gigaword-50', binary=False)
9  wiki_model_200 = model_executor.runModel('glove-wiki-gigaword-200', binary=False)
10
11 analyzer = AnalysisBuilder()
12 analyzer.writeAnalysis(twitter_model_100, 'glove-twitter-100')
13 analyzer.writeAnalysis(wiki_model_100, 'glove-wiki-gigaword-100')
14 analyzer.writeAnalysis(wiki_model_50, 'glove-wiki-gigaword-50')
15 analyzer.writeAnalysis(wiki_model_200, 'glove-wiki-gigaword-200')
16
17
```

Same as previous slide, except this time using different models

Analysis.csv

	A	B	C	D	E
1	word2vec-google-news-300	3000000	69	78	0.88461538
2	glove-twitter-100	1193514	38	77	0.49350649
3	glove-wiki-gigaword-100	400000	64	79	0.81012658
4	glove-wiki-gigaword-50	400000	56	79	0.70886076
5	glove-wiki-gigaword-200	400000	67	79	0.84810127
6					

A = Model Name

B = Vocabulary Size

C = Correctly guessed words

D = Total amount of questions/rows

E = Accuracy

Word2vec has the most, with glove coming in close second.

More embeddings results in higher accuracy because there is more preciseness

Twitter's score is low because tweets usually contains slang

Synonyms and correct answers

	A	B	C	D	E	F
1	question	answer	0	1	2	3
2	enormously	tremendously	appropriately	uniquely	tremendously	decidedly
3	provisions	stipulations	stipulations	interrelations	jurisdictions	interpretations
4	haphazardly	randomly	dangerously	densely	randomly	linearly
5	prominent	conspicuous	battered	ancient	mysterious	conspicuous
6	zenith	pinnacle	completion	pinnacle	outset	decline
7	flawed	imperfect	tiny	imperfect	lustrous	crude
8	urgently	desperately	typically	conceivably	tentatively	desperately
9	consumed	eaten	bred	caught	eaten	supplied
10	advent	coming	coming	arrest	financing	stability
11	concisely	succinctly	powerfully	positively	freely	succinctly
12	salutes	greetings	information	ceremonies	greetings	privileges
13	solitary	alone	alert	restless	alone	fearless
14	hasten	accelerate	permit	determine	accelerate	accompany
15	perseverance	endurance	endurance	skill	generosity	disturbance
16	fanciful	imaginative	familiar	imaginative	apparent	logical
17	showed	demonstrated	demonstrated	published	repeated	postponed
18	constantly	continually	instantly	continually	rapidly	accidentally
19	issues	subjects	training	salaries	subjects	benefits
20	furnish	supply	supply	impress	protect	advise
21	costly	expensive	expensive	beautiful	popular	complicated
22	recognized	acknowledged	successful	depicted	acknowledged	welcomed
23	spot	location	climate	latitude	sea	location
24	make	earn	earn	print	trade	borrow
25	often	frequently	definitely	frequently	chemically	hardly
26	easygoing	relaxed	frontier	boring	farming	relaxed
27	debate	argument	war	argument	election	competition
28	narrow	thin	clear	freezing	thin	poisonous
29	arranged	planned	planned	explained	studied	discarded
30	infinite	limitless	limitless	relative	unusual	structural
31	showy	striking	striking	prickly	entertaining	incidental
32	levied	imposed	imposed	believed	requested	correlated
33	deftly	skillfully	skillfully	prudently	occasionally	humorously
34	distribute	circulate	commercialize	circulate	research	acknowledge
35	discrepancies	differences	weights	deposits	wavelengths	differences
36	prolific	productive	productive	serious	capable	promising
37	unmatched	unequaled	unrecognized	unequaled	alienated	emulated
38	peculiarly	uniquely	partly	uniquely	patriotically	suspiciously
39	hue	color	glare	contrast	color	scent
40	hind	rear	curved	muscular	hairy	rear

glove-twitter-100-details.csv

	A	B	C	D
1	enormously	tremendously	tremendously	correct
2	provisions	stipulations	jurisdictions	wrong
3	haphazardly	randomly	densely	wrong
4	prominent	conspicuous	mysterious	wrong
5	zenith	pinnacle	pinnacle	correct
6	flawed	imperfect	imperfect	correct
7	urgently	desperately	desperately	correct
8	consumed	eaten	eaten	correct
9	advent	coming	coming	correct
10	concisely	succinctly	succinctly	correct
11	salutes	greetings	greetings	correct
12	solitary	alone	restless	wrong
13	hasten	accelerate	accelerate	correct
14	perseverance	endurance	endurance	correct
15	fanciful	imaginative	imaginative	correct
16	showed	demonstrated	repeated	wrong
17	constantly	continually	instantly	wrong
18	issues	subjects	benefits	wrong
19	furnish	supply	advise	wrong
20	costly	expensive	expensive	correct
21	recognized	acknowledged	acknowledged	correct
22	spot	location	location	correct
23	make	earn	trade	wrong
24	often	frequently	frequently	correct
25	easygoing	relaxed	relaxed	correct
26	debate	argument	election	wrong
27	narrow	thin	clear	wrong
28	arranged	planned	planned	correct
29	infinite	limitless	relative	wrong
30	showy	striking	prickly	wrong
31	levied	imposed	imposed	correct
32	deftly	skillfully	humorously	wrong
33	distribute	circulate	commercialize	wrong
34	discrepancies	differences	differences	correct
35	prolific	productive	promising	wrong
36	unmatched	unequaled	unrecognized	wrong
37	peculiarly	uniquely	unrecognized	guess
38	hue	color	color	correct
39	hind	rear	hairy	wrong

A: actual word

B: Correct synonym

C: Guessed synonym

D: Correct, wrong or guess

1: B and C are the same -> correct

2: B and C are not the same -> incorrect

No guesses in first 40

glove-wiki-gigaword-50-details.csv

	A	B	C	D
1	enormously	tremendously	tremendously	correct
2	provisions	stipulations	stipulations	correct
3	haphazardly	randomly	randomly	correct
4	prominent	conspicuous	ancient	wrong
5	zenith	pinnacle	pinnacle	correct
6	flawed	imperfect	imperfect	correct
7	urgently	desperately	desperately	correct
8	consumed	eaten	eaten	correct
9	advent	coming	coming	correct
10	concisely	succinctly	succinctly	correct
11	salutes	greetings	greetings	correct
12	solitary	alone	restless	wrong
13	hasten	accelerate	accelerate	correct
14	perseverance	endurance	generosity	wrong
15	fanciful	imaginative	imaginative	correct
16	showed	demonstrated	demonstrated	correct
17	constantly	continually	continually	correct
18	issues	subjects	subjects	correct
19	furnish	supply	advise	wrong
20	costly	expensive	expensive	correct
21	recognized	acknowledged	successful	wrong
22	spot	location	location	correct
23	make	earn	earn	correct
24	often	frequently	frequently	correct
25	easygoing	relaxed	relaxed	correct
26	debate	argument	argument	correct
27	narrow	thin	clear	wrong
28	arranged	planned	planned	correct
29	infinite	limitless	limitless	correct
30	showy	striking	prickly	wrong
31	levied	imposed	imposed	correct
32	deftly	skillfully	skillfully	correct
33	distribute	circulate	circulate	correct
34	discrepancies	differences	differences	correct
35	prolific	productive	promising	wrong
36	unmatched	unequaled	unequaled	correct
37	peculiarly	uniquely	uniquely	correct
38	hue	color	color	correct
39	hind	rear	hairy	wrong
40	highlight	accentuate	alter	wrong

A: actual word

B: Correct synonym

C: Guessed synonym

D: Correct, wrong or guess

1: B and C are the same -> correct

4: B and C are not the same -> incorrect

No guesses in first 40

glove-wiki-gigaword-100-details.csv

	A	B	C	D
1	enormously	tremendously	tremendously	correct
2	provisions	stipulations	stipulations	correct
3	haphazardly	randomly	randomly	correct
4	prominent	conspicuous	ancient	wrong
5	zenith	pinnacle	pinnacle	correct
6	flawed	imperfect	imperfect	correct
7	urgently	desperately	desperately	correct
8	consumed	eaten	eaten	correct
9	advent	coming	coming	correct
10	concisely	succinctly	succinctly	correct
11	salutes	greetings	greetings	correct
12	solitary	alone	restless	wrong
13	hasten	accelerate	accelerate	correct
14	perseverance	endurance	generosity	wrong
15	fanciful	imaginative	imaginative	correct
16	showed	demonstrated	demonstrated	correct
17	constantly	continually	continually	correct
18	issues	subjects	subjects	correct
19	furnish	supply	advise	wrong
20	costly	expensive	expensive	correct
21	recognized	acknowledged	successful	wrong
22	spot	location	location	correct
23	make	earn	earn	correct
24	often	frequently	frequently	correct
25	easygoing	relaxed	relaxed	correct
26	debate	argument	argument	correct
27	narrow	thin	thin	correct
28	arranged	planned	planned	correct
29	infinite	limitless	limitless	correct
30	showy	striking	prickly	wrong
31	levied	imposed	imposed	correct
32	deftly	skillfully	skillfully	correct
33	distribute	circulate	commercialize	wrong
34	discrepancies	differences	differences	correct
35	prolific	productive	productive	correct
36	unmatched	unequaled	unequaled	correct
37	peculiarly	uniquely	uniquely	correct
38	hue	color	color	correct
39	hind	rear	hairy	wrong
40	highlight	accentuate	accentuate	correct

A: actual word

B: Correct synonym

C: Guessed synonym

D: Correct, wrong or guess

1: B and C are the same -> correct

4: B and C are not the same -> incorrect

37: peculiarly word or all four guessed words not found in model

glove-wiki-gigaword-200-details.csv

	A	B	C	D
1	enormously	tremendously	tremendously	correct
2	provisions	stipulations	stipulations	correct
3	haphazardly	randomly	densely	wrong
4	prominent	conspicuous	ancient	wrong
5	zenith	pinnacle	pinnacle	correct
6	flawed	imperfect	imperfect	correct
7	urgently	desperately	desperately	correct
8	consumed	eaten	eaten	correct
9	advent	coming	coming	correct
10	concisely	succinctly	succinctly	correct
11	salutes	greetings	ceremonies	wrong
12	solitary	alone	alone	correct
13	hasten	accelerate	accelerate	correct
14	perseverance	endurance	generosity	wrong
15	fanciful	imaginative	imaginative	correct
16	showed	demonstrated	demonstrated	correct
17	constantly	continually	continually	correct
18	issues	subjects	subjects	correct
19	furnish	supply	supply	correct
20	costly	expensive	expensive	correct
21	recognized	acknowledged	acknowledged	correct
22	spot	location	location	correct
23	make	earn	earn	correct
24	often	frequently	frequently	correct
25	easygoing	relaxed	relaxed	correct
26	debate	argument	argument	correct
27	narrow	thin	thin	correct
28	arranged	planned	planned	correct
29	infinite	limitless	limitless	correct
30	showy	striking	prickly	wrong
31	levied	imposed	imposed	correct
32	deftly	skillfully	skillfully	correct
33	distribute	circulate	circulate	correct
34	discrepancies	differences	differences	correct
35	prolific	productive	productive	correct
36	unmatched	unequaled	unequaled	correct
37	peculiarly	uniquely	uniquely	correct
38	hue	color	color	correct
39	hind	rear	curved	wrong
40	highlight	accentuate	accentuate	correct

A: actual word

B: Correct synonym

C: Guessed synonym

D: Correct, wrong or guess

1: B and C are the same -> correct

3: B and C are not the same -> incorrect

No guesses in first 40

word2vecgoogle-news-300-details.csv

	A	B	C	D
1	enormously	tremendously	tremendously	correct
2	provisions	stipulations	stipulations	correct
3	haphazardly	randomly	randomly	correct
4	prominent	conspicuous	conspicuous	correct
5	zenith	pinnacle	pinnacle	correct
6	flawed	imperfect	imperfect	correct
7	urgently	desperately	desperately	correct
8	consumed	eaten	eaten	correct
9	advent	coming	coming	correct
10	concisely	succinctly	succinctly	correct
11	salutes	greetings	ceremonies	wrong
12	solitary	alone	restless	wrong
13	hasten	accelerate	accelerate	correct
14	perseverance	endurance	generosity	wrong
15	fanciful	imaginative	imaginative	correct
16	showed	demonstrated	demonstrated	correct
17	constantly	continually	continually	correct
18	issues	subjects	subjects	correct
19	furnish	supply	impress	wrong
20	costly	expensive	expensive	correct
21	recognized	acknowledged	acknowledged	correct
22	spot	location	location	correct
23	make	earn	earn	correct
24	often	frequently	frequently	correct
25	easygoing	relaxed	relaxed	correct
26	debate	argument	argument	correct
27	narrow	thin	thin	correct
28	arranged	planned	planned	correct
29	infinite	limitless	limitless	correct
30	showy	striking	prickly	wrong
31	levied	imposed	imposed	correct
32	deftly	skillfully	skillfully	correct
33	distribute	circulate	commercialize	wrong
34	discrepancies	differences	differences	correct
35	prolific	productive	productive	correct
36	unmatched	unequaled	unequaled	correct
37	peculiarly	uniquely	uniquely	correct
38	hue	color	color	correct
39	hind	rear	rear	correct
40	highlight	accentuate	accentuate	correct

A: actual word

B: Correct synonym

C: Guessed synonym

D: Correct, wrong or guess

1: B and C are the same -> correct

11: B and C are not the same -> incorrect

No guesses in first 40