

```

1 # Download data
2 !rm -rf sts2017.eval.v1.1.zip* STS2017.eval.v1.1
3 !wget http://alt.qcri.org/semEval2017/task1/data/uploads/sts2017.eval.v1.1.zip
4 !unzip sts2017.eval.v1.1.zip
5 !cp STS2017.eval.v1.1/STS.input.track5.en-en.txt input.txt

```

URL transformed to HTTPS due to an HSTS policy

--2022-03-04 16:15:23-- <https://alt.qcri.org/semEval2017/task1/data/uploads/sts2017.eval.v1.1.zip>

Resolving alt.qcri.org (alt.qcri.org)... 80.76.166.234

Connecting to alt.qcri.org (alt.qcri.org)|80.76.166.234|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 87902 (86K) [application/zip]

Saving to: 'sts2017.eval.v1.1.zip'

sts2017.eval.v1.1.z 100%[=====>] 85.84K 401KB/s in 0.2s

2022-03-04 16:15:24 (401 KB/s) - 'sts2017.eval.v1.1.zip' saved [87902/87902]

Archive: sts2017.eval.v1.1.zip

inflating: STS2017.eval.v1.1/LICENSE.txt

inflating: STS2017.eval.v1.1/README.txt

inflating: STS2017.eval.v1.1/STS.input.track1.ar-ar.txt

inflating: STS2017.eval.v1.1/STS.input.track2.ar-en.txt

inflating: STS2017.eval.v1.1/STS.input.track3.es-es.txt

inflating: STS2017.eval.v1.1/STS.input.track4a.es-en.txt

inflating: STS2017.eval.v1.1/STS.input.track4b.es-en.txt

inflating: STS2017.eval.v1.1/STS.input.track5.en-en.txt

inflating: STS2017.eval.v1.1/STS.input.track6.tr-en.txt

```

1 # Step 1. Preprocessing
2 ## 1.1 Read data
3
4 with open('input.txt') as f:
5     lines = [line.strip() for line in f]
6
7 print(lines)

```

```
['A person is on a baseball team.\tA person is playing basketball on a team.', 'Our current vehicles will be in muse
```

```
1 # 1.2 Preparing words
```

```
2 import nltk
```

```
3 nltk.download('stopwords')
```

```
4 nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

```
[nltk_data] Package stopwords is already up-to-date!
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
```

```
[nltk_data] Unzipping corpora/wordnet.zip.
```

```
True
```

```
1 from nltk.corpus import stopwords
```

```
2 from nltk.stem import WordNetLemmatizer
```

```
3 import string
```

```
4
```

```
5 english_stopwords = stopwords.words('english')
```

```
6 translation_table = str.maketrans('', '', string.punctuation)
```

```
7 lemmatizer = WordNetLemmatizer()
```

```
8
```

```
9 words = dict()
```

```
10 sentences = list()
```

```
11
```

```
12 for line in lines:
```

```
13     s1, s2 = line.split('\t')
```

```
14
```

```
15     v1 = []
```

```
16     v2 = []
```

```
17
```

```
18     for s, v in [(s1, v1), (s2, v2)]:
```

```
19         for w in s.split():
```

```
20             w = w.lower().translate(translation_table)
```

```
21             if w in english_stopwords:
```

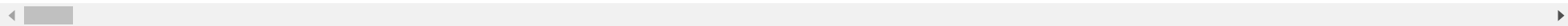
```
22                 continue
```

```
23                 w = lemmatizer.lemmatize(w)
```

```
24         words.setdefault(w, len(words))
25         v.append(words[w])
26
27     sentences.append((v1, v2))
```

```
1 print(words)
2 print(sentences)
```

```
{'person': 0, 'baseball': 1, 'team': 2, 'playing': 3, 'basketball': 4, 'current': 5, 'vehicle': 6, 'museum': 7, 'eve'
 [[0, 1, 2], [0, 3, 4, 2]], ([5, 6, 7, 8, 9], [10, 11, 12]), ([13, 14, 15, 16, 17], [13, 18, 19]), ([20, 21, 22, 13]
```



```
1 # Step 2. Graph construction
2 import numpy as np
3
```

```
1 # 2.1 Synonyms
2 from nltk.corpus import wordnet
3
4 synonym_graph = np.zeros((len(words), len(words)))
5
6 def set_of_synonyms(word):
7     synonyms = []
8     for syn in wordnet.synsets(word):
9         for l in syn.lemmas():
10             syn_word = l.name()
11             if syn_word in english_stopwords:
12                 continue
13             syn_word = lemmatizer.lemmatize(syn_word)
14             synonyms.append(syn_word)
15     return set(synonyms)
16
17 for word, idx in words.items():
18     for syn in set_of_synonyms(word):
19         if syn not in words:
20             continue
```

```
21     syn_idx = words[syn]
22     synonym_graph[idx][syn_idx] = 1
23     synonym_graph[syn_idx][idx] = 1
```

```
1 # 2.2 GoW(no)
2
3 window_graph = np.zeros((len(words), len(words)))
4
5 window_size = 3 # Parameter
6
7 for ss in sentences:
8     for s in ss:
9         window = []
10        for i in range(len(s)):
11            window = window[-(window_size - 1):]
12            v1 = s[i]
13            for v2 in window:
14                window_graph[v1][v2] += 1
15                window_graph[v2][v1] += 1
16            window.append(v1)
```

```
1 # 2.3 Merge graphs
2
3 import tqdm
4 import numba
5
6 distances = 1 / (synonym_graph + window_graph)
7 np.fill_diagonal(distances, 0)
8
9 def floyd(dist):
10     N = dist.shape[0]
11
12     @numba.jit
13     def helper(dist, N, k):
14         for i in range(N):
```

```

15         for j in range(N):
16             dist[i][j] = min(dist[i][j], dist[i][k] + dist[k][j])
17
18     for k in tqdm.trange(N):
19         helper(dist, N, k)
20
21 floyd(distances)

```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:6: RuntimeWarning: divide by zero encountered in true_d

100%|██████████| 762/762 [00:00<00:00, 828.96it/s]

```

1 # Step 3. Main part of this project (MDS)
2
3 import numpy.linalg as la
4
5 @numba.jit
6 def mds(matD, k):
7     N = matD.shape[0]
8     matJ = np.eye(N) - np.ones((N, N)) / N
9     matG = -0.5 * matJ @ (matD ** 2) @ matJ
10    vecL, matP = la.eigh(matG)
11    vecL[vecL < 0] = 0
12    return matP[:, -k:], vecL[-k:]
13
14 def embed_graph(distances, k):
15     p, l = mds(distances, k)
16     result = p @ np.diag(np.sqrt(l))
17     result -= np.mean(result, axis=0)
18     return result

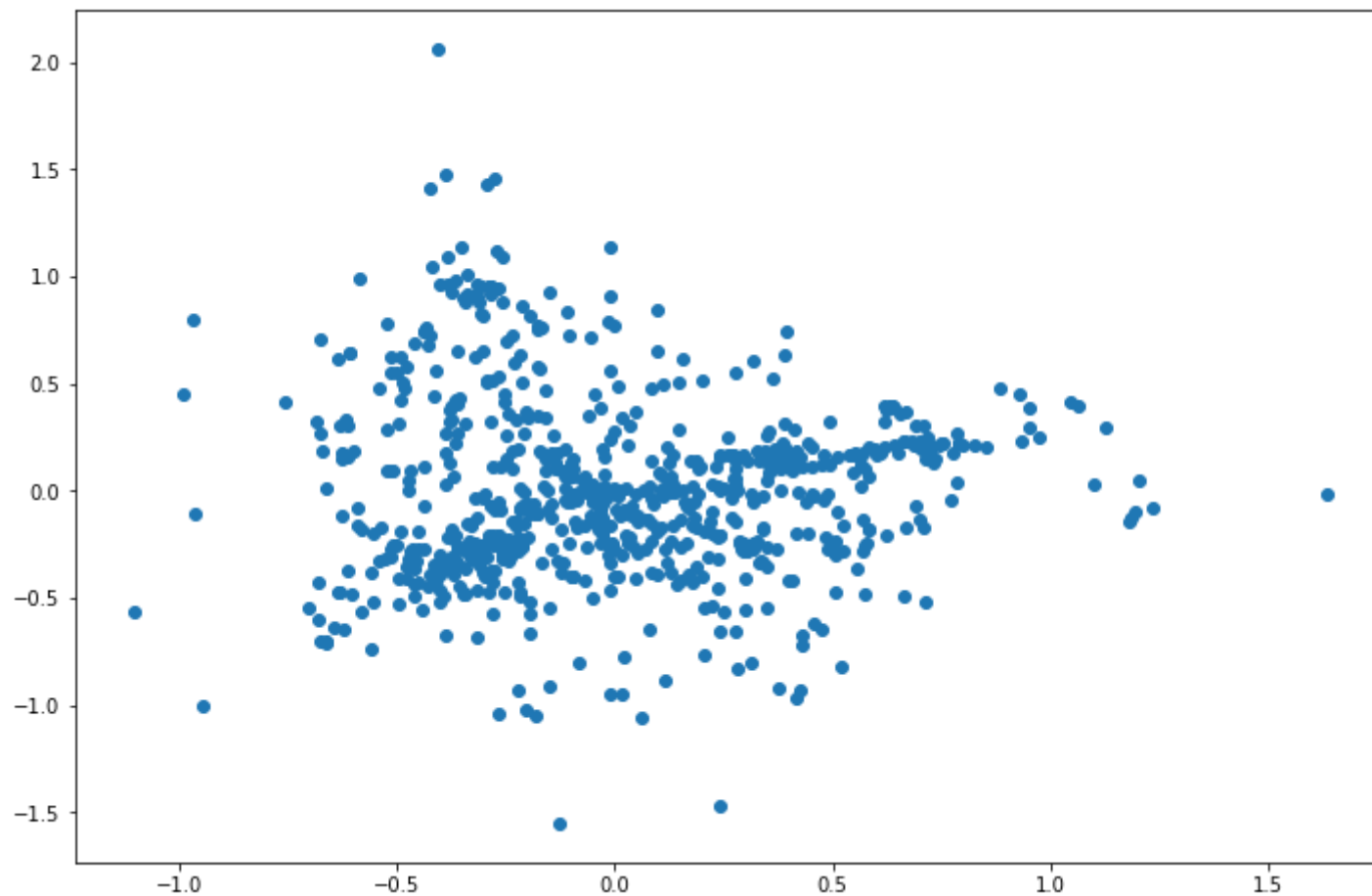
```

```

1 import matplotlib.pyplot as plt
2
3 word_to_vec = embed_graph(distances, 10)
4

```

```
5 plt.figure(figsize=(12,8))
6 plt.scatter(*embed_graph(distances, 2).transpose())
7 plt.show()
```



```
1 import scipy.spatial.distance as dist
2
3 for i, (s1, s2) in enumerate(sentences):
4     v1 = np.mean(word_to_vec[s1,:], axis=0)
5     v2 = np.mean(word_to_vec[s2,:], axis=0)
```

```

6 d = dist.cosine(v1, v2)
7 print(i, "{:1.8f}".format(d), lines[i], sep='\t')
0.00074201 man on steps A man sits on steps.
0.10076416 A dog under the stairs A dog is resting on the stairs.
0.05087777 There is a young girl. There is a young boy with the woman.
0.06979246 Two women give a demonstration at a Renaissance Fair. Two people are doing a demonstration at
0.01583403 A woman paints a picture of a large building which can be seen in the background. A person
0.00109936 Some guy sitting on a couch watching television. A guy is sitting on the couch watching
1.32430769 A woman is posing for a Christmas card. A girl is taking a photo.
0.00263167 Two kids are playing a game of Foosball. The kids are playing a game with each other.
0.09505302 The man is in a deserted field. The man is outside in the field.
0.00148784 A little boy is drinking from a juice packet. A boy is drinking from a juice packet.
0.50733697 A man has his skis ready for the field to get iced out A man on a boat getting ready to pull in
0.00150115 The biker is doing dangerous stunts. A biker is doing stunts.
0.55808072 A musician is smearing jam on his white guitar at a concert. Trombonist playing the her inst
0.14609345 A man is working on his farm. An old guy is working on his farm and a cow kicks him.
0.72870073 A pair of young boys in t-shirts are hiding in the woods with one looking aghast. Two smi
0.27674725 A group of people playing soccer on a soccer field in front of a crowd. A group of men playing
0.04698902 A man is enjoying watching three asian children dance. A man is watching three asian kids dance
0.01407355 A girl is talking to her dad on a cellphone. a girl is talking on her phone.
0.00589431 One person wields an umbrella. One person is holding an umbrella.
1.06522902 A woman walking down the side of a highway. A man is standing on one feet on a bridge.
0.03777908 A group of teenagers jumping and having a good time. A group of kids having a good time.
0.07436863 A man calling for a taxi at the airport on a sunny day. A man is outside on a July day.
0.05580595 A man is playing his guitar at Joe's Cafe. The man is playing the drums for his mom.
0.14891026 A young boy has a red fishing pole. A young boy stands over a bridge with a fishing pole.
0.01570235 A player bounces a ball. A player catching a ball.
0.07574282 Someone is on a blanket. The person is making a blanket.
0.02391269 Men competing in a contest. Two women compete in a contest.
0.63606242 The puppy is outdoor. A man in printed board shorts is doing a yoga pose on the beach.
0.07428404 The man has a orange shirt. The woman has a orange shirt.
0.83520845 The two kids look at the products sold inside the subway shop. The kids are stealing from a sto
0.07650532 The baby boy wants his mother. A baby boy is happy to see his mother.
0.07440555 A young girl in a pink jacket is playing a board game. A boy in a jacket playing.
0.27557761 A smiling Caucasian male wearing a blue and black striped shirt it about to take a photograph.
0.24641036 A person is scaling a rock wall. A person and a horse are above a fence.
0.03285989 A boy is very close to a girl. A girl is close to a boy whose face is not shown.
0.12358054 Two men are playing a game of Scrabble together. The two women are playing a game.
0.19566073 Six children are cleaning a room. Two male children cleaning up leaves in a parking lot.
0.38517735 The girl is very skilled and practices a lot. The man is very skilled.
0 83970461 Many guards are standing in front of the starting line of a race Two men in business dre

```

0.000000001 many guards are standing in front of the burning line of a tree. Two men in business suits.
 0.03390828 There is a cook preparing food. A cook is making food.
 0.18471989 Two men are helping a boy. Two men are with a young child.
 0.00000000 A man at a farmers market. A man is at a farmers market.
 0.38908337 four people walk toward a tower Two men walk toward a dome-shaped building.
 1.43434488 A little girl and boy are reading books. An older child is playing with a doll while gaze.
 0.08616144 The child would like to swim. The children want to swim.
 0.00483424 Two skiing people are doing cross country skiing. People skiing cross country.
 0.87382301 A dog taking a poop in the street. A man making balloon animals for two children on a street.
 0.02235202 Men are fighting after a basketball game. The men are playing a game of basketball.
 0.06516118 A kid is talking in class. A girl is going to class.
 0.67470939 The women are trying to sell something to the individual. The men are trying to make some
 0.40616945 People are waiting for the fireworks to start. Three people are waiting for the rain to stop.
 0.43357966 A dog prepares to herd three sheep with horns. A dog and sheep run together.
 0.08163325 People are sitting around dimly lit tables while smiling and laughing. People are sitting on benches.
 0.01351550 A dog is chasing cows. A white dog is chasing cows.
 0.00606166 One football player tries to tackle a player on the opposing team. A football player attempts.
 0.24091769 Three men are trying to pull a Subaru out of a ditch along a mountain road. Two men are waiting.
 1.09974223 A woman doesn't have her arms up because she's paralyzed from the neck down. A woman is sitting.
 0.02717833 People gathered in a room. People gathered together in a room.

✓ 0 сек. выполнено в 18:43

