

## **Assignment 4 Report**

### **How did you design your opinion extraction module with CoreNLP?**

To use CoreNLP, downloaded the StanfordCoreNLP, installed Java8, started the CoreNLP server on port 9000.

In the `extract_pairs` function, `review_content` & `review_id` is taken as input. Then `nlp` is initialized with `StanfordCoreNLP`. Properties passed for annotating the data are `annotators`, `outputFormat` as `json`, the Open Information Extraction annotator is set to `true` and the `timeout`. Then the `review_content` is passed through the `nlp` to get a json string. Once output is obtained, a for loop is used to iterate through the sentences in the output. In that, the pos of each word is added to a dictionary. Enhanced Dependencies will help in retrieving the `nsubj` and `amod` dependencies. A nominal subject is a noun phrase which is the syntactic subject of a clause. The root of the clause can be an adjective or noun. The opinion is added if the governor is an adjective and the dependent is a noun in case of `nsubj`. `amod` is the Adjectival Modifier that is any adjectival phrase that serves to modify the meaning of the NP. Once the opinion is created, it is added to the extracted opinions dictionary along with the review id.

### **How did you measure the opinion similarity? How do you tune the threshold?**

To measure similarity, `gensim` is used. For each query opinion passed, it is split into attribute and quality. Then extracted opinions is accessed and they are split into attribute and quality as well. If the attribute and quality words are present in the `word2vec` object, then the similarity of the attributes and quality is calculated. If both the values are greater than the cosine value, the respective extracted opinion is added to the similar opinions.

I tuned the threshold by increasing it by 0.01 starting from 0.2

### **Discuss the successful cases that your system can handle.**

As compared to the expected outputs mentioned in the assignment document most of the cases are similar.

As I increase the threshold, the tuples for the 1<sup>st</sup> and 2<sup>nd</sup> case are closer to the required output but the 3<sup>rd</sup> and 4<sup>th</sup> case lose out on opinions

### **Discuss the cases that your system fails.**

On tuning the threshold between 0.2 and 0.3, it gives me `[service, bad]` in the first case and `[service, excellent]`, `[service, great]` in the second case.

**On further tuning, 0.35 works best according to me. It gives 1 incorrect opinion in case 1 and 1 in case 2 but it doesn't eliminate many of the opinions in case 3 and 4. If the threshold is increased, the opinions are eliminated in case 3 and 4.**

[meal, wonderful] appears in review	[ 1 ]
[salad, delicious] appears in review	[ 1 ]
[stick, delicious] appears in review	[ 1 ]
[service, excellent] appears in review	[ 1 , 2 ]
[spots, other] appears in review	[ 1 ]
[restaurant, old] appears in review	[ 2 ]
[tablecloths, white] appears in review	[ 2 ]
[potatoes, mashed] appears in review	[ 2 ]
[meat, flavorful] appears in review	[ 2 ]
[slaw, delicious] appears in review	[ 2 ]
[list, great] appears in review	[ 3 ]
[potions, huge] appears in review	[ 3 ]
[thing, good] appears in review	[ 3 ]
[note, positive] appears in review	[ 4 ]
[sandwich, grilled] appears in review	[ 4 ]
[option, choose] appears in review	[ 4 ]
[selection, great] appears in review	[ 4 ]
[server, rude] appears in review	[ 4 ]
[restaurant, small] appears in review	[ 4 ]
[table, own] appears in review	[ 4 ]

[distance, walking] appears in review	[ 5 ]
[meals, best] appears in review	[ 5 ]
[fan, huge] appears in review	[ 5 ]
[service, great] appears in review	[ 5 ]
[sandwiches, french] appears in review	[ 5 ]
[food, great] appears in review	[ 5 ]
[rotation, regular] appears in review	[ 6 ]
[food, excellent] appears in review	[ 6 , 1 3 ]
[deserts, mini] appears in review	[ 6 ]
[deal, great] appears in review	[ 6 ]
[selection, good] appears in review	[ 6 ]
[quality, bad] appears in review	[ 7 ]
[day, bad] appears in review	[ 7 ]
[prices, decent] appears in review	[ 8 ]
[service, warm] appears in review	[ 8 ]
[fish, battered] appears in review	[ 8 ]
[good, bad] appears in review	[ 8 ]
[service, bad] appears in review	[ 9 ]
[fingers, homemade] appears in review	[ 1 0 ]
[beer, priced] appears in review	[ 1 0 ]
[value, great] appears in review	[ 1 0 ]

[menu, large] appears in review	[ 1 1 ]
[quality, excellent] appears in review	[ 1 1 ]
[menu, entire] appears in review	[ 1 1 ]
[specialities, other] appears in review	[ 1 1 ]
[favorites, local] appears in review	[ 1 2 ]
[feeling, warm] appears in review	[ 1 2 ]
[deal, cheapest] appears in review	[ 1 2 ]
[order, particular] appears in review	[ 1 2 ]
[soup, french] appears in review	[ 1 2 ]
[time, wonderful] appears in review	[ 1 3 ]
[group, larger] appears in review	[ 1 3 ]
[turkey, roasted] appears in review	[ 1 4 ]
[food, interesting] appears in review	[ 1 4 ]
[taco, good] appears in review	[ 1 5 ]
[waiter, slow] appears in review	[ 1 5 ]
[time, long] appears in review	[ 1 5 ]
[food, cold] appears in review	[ 1 5 ]
[group, large] appears in review	[ 1 6 ]
[way, bad] appears in review	[ 1 6 ]
[options, vegan] appears in review	[ 1 6 ]
[salad, greek] appears in review	[ 1 7 ]

---

[fries, great] appears in review	[ 1 7 ]
[waiter, attentive] appears in review	[ 1 7 ]
[wines, decent] appears in review	[ 1 8 ]
[food, fresh] appears in review	[ 1 8 ]
[hopes, high] appears in review	[ 1 9 ]
[food, ok] appears in review	[ 1 9 ]
[ambience, pleasant] appears in review	[ 2 0 ]
[pittsburghers, older] appears in review	[ 2 0 ]

-----  
query opinion [service, good] has similar opinions:

[service, excellent] appears in review	[ 1 , 2 ]
[service, great] appears in review	[ 5 ]
[service, bad] appears in review	[ 9 ]

query opinion [service, bad] has similar opinions:

[service, great] appears in review	[ 5 ]
[service, bad] appears in review	[ 9 ]

query opinion [atmosphere, good] has similar opinions:

[ambience, pleasant] appears in review	[ 2 0 ]
--	---------

query opinion [food, delicious] has similar opinions:

[meal, wonderful] appears in review	[ 1 ]
[salad, delicious] appears in review	[ 1 ]

---

[food, fresh] appears in review [ 1 8 ]  
[hopes, high] appears in review [ 1 9 ]  
[food, ok] appears in review [ 1 9 ]  
[ambience, pleasant] appears in review [ 2 0 ]  
[pittsburghers, older] appears in review [ 2 0 ]

-----  
query opinion [service, good] has similar opinions:

[service, excellent] appears in review [ 1 , 2 ]  
[service, great] appears in review [ 5 ]  
[service, bad] appears in review [ 9 ]

query opinion [service, bad] has similar opinions:

[service, great] appears in review [ 5 ]  
[service, bad] appears in review [ 9 ]

query opinion [atmosphere, good] has similar opinions:

[ambience, pleasant] appears in review [ 2 0 ]

query opinion [food, delicious] has similar opinions:

[meal, wonderful] appears in review [ 1 ]  
[salad, delicious] appears in review [ 1 ]  
[meat, flavorful] appears in review [ 2 ]  
[turkey, roasted] appears in review [ 1 4 ]  
-----