

Assignment 4 report

Shuchen Liu (shl174@pitt.edu)

How to run this project

In order to run this project, please put the "assign4_reviews.txt" file and the "assign4_word2vec_for_python.bin" file into the data folder. Then please download the coreNLP package from this link <http://nlp.stanford.edu/software/stanford-corenlp-full-2018-10-05.zip> and put the unzipped folder in the "src" folder. And run "Assignment4Main.py".

I used the "stanford-corenlp-full-2018-10-05" to run my project, please download the same version. If you want to change it, you can change line 20 in file "ExtractOpinions.py" to run another version of coreNLP.

Note: It takes about 5 minutes to run the project, if you do not see any feedback, please be patient:)

Opinion extraction design

I used two functions of coreNLP to do the opinion extraction, dependency parse and pos tag. For each review, since there are more than one sentences in it, first I split it into several single sentences. Then, I use coreNLP to get the dependency parses and pos tags for each sentence.

In this task, adjectives are pretty important because they represent the opinion of customers. So, I iterate all pos tags, if the current token is tagged as "JJ", I will find all dependency parses of this token and save them as opinions into the dictionary. The disadvantage of this method is that several useless parses would be saved, time and space are wasted. In order to solve this problem, I found that only the dependency parses between none and adjective are useful in this task, so in my final algorithm, I only keep the "[NN,JJ]" parses as key in the dictionary.

Similarity measurement

The function of word2vec can calculate the similarity of two words. However, in this task, the input and keys in the dictionary is not one word but two just like "service, good". So, I split both the keys of dictionary and the input query into two words and calculate the similarity between their former word (written as s1) and the similarity between their latter word (written as s2).

Threshold

At first, I use cosine similarity > 0.8 as the threshold which means only both s1 and s2 are larger than 0.8 can be taken as similar opinion. In this situation, some queries cannot find any similar review. The reason is that 0.8 is a pretty high value, only the same word can reach this bar. For example, the similarity between "good" and "great" is only 0.72915095. And the similarity between "good" and "nice" is only 0.6836092. So, I decided to set the threshold of s1 as 0.5 and the threshold of s2 as 0.3.

successful cases

The search of "service, good", "atmosphere, good", "food, delicious" performed well in my model.

For "service, good" my model can find [Service,excellent, Service,great, service,warm, service,solid, service,great, service,good]

For "atmosphere, good" my model can find [atmosphere,nice, ambiance,nice, atmosphere,great, ambiance,pleasant]

For "food delicious" my model can find [meal,wonderful, meal,delicious, food,great, food,hearty, food,fresh, food,interesting, food,satisfying]

fail cases

the search of "service bad" performed not well in my model. Some positive opinion such as "service bad" will be shown. The reason is that the similarity between "good" and "bad" is 0.7190051 which is too high. If I set the threshold larger than 0.719, some correct search result will be eliminated.

Running result

Threshold: cosine similarity larger than 0.5 for NN, 0.3 for JJ

Running result:

[co-worker,old] appears in review	[1]
[meal,wonderful] appears in review	[1]
[salad,huge] appears in review	[1]
[stick,little] appears in review	[1]
[stick,delicious] appears in review	[1]
[place,great] appears in review	[1]
[experience,upscale] appears in review	[1]
[bar/restaurant,old] appears in review	[2]
[woodwork,ornate] appears in review	[2]
[meat,tender] appears in review	[2]
[slaw,delicious] appears in review	[2]
[slaw,picky] appears in review	[2]
[atmosphere,nice] appears in review	[3]

[world,nice] appears in review [3]
[list,local] appears in review [3]
[list,great] appears in review [3]
[menu,huge] appears in review [3]
[thing,huge] appears in review [3]
[thing,good] appears in review [3]
[note,positive] appears in review [4]
[meal,delicious] appears in review [4]
[sandwich,delicious] appears in review [4]
[sandwich,char-grilled] appears in review [4]
[something,amazing] appears in review [4]
[option,choose-your-own] appears in review [4]
[ambiance,nice] appears in review [4]
[selection,great] appears in review [4]
[restaurant,small] appears in review [4 , 5]
[thing,small] appears in review [4]
[ride,useless] appears in review [5]
[fan,huge] appears in review [5]
[restaurant,nice] appears in review [5]
[atmosphere,great] appears in review [5]
[food,great] appears in review [5]
[rotation,regular] appears in review [6]
[food,Excellent] 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]
[food,Good] appears in review [8]
[food,hearty] appears in review [8]

[service,warm] appears in review [8]
[service,Bad] appears in review [9]
[min,like] appears in review [9]
[beer,Good] 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]
[feeling,warm] appears in review [1 2]
[feeling,great] appears in review [1 2]
[price,generous] appears in review [1 2]
[order,particular] appears in review [1 2]
[course,main] appears in review [1 2]
[soup,French] appears in review [1 2]
[Summer,last] appears in review [1 3]
[time,wonderful] appears in review [1 3]
[food,excellent] appears in review [1 3]
[addition,solid] appears in review [1 3]
[service,solid] appears in review [1 3]
[turkey,roasted] appears in review [1 4]
[service,great] appears in review [1 4]
[food,fresh] 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]
[thing,second] appears in review [1 5]
[group,large] appears in review [1 6]
[food,satisfying] appears in review [1 6]

[food,typical] appears in review [1 6]
[food,American] appears in review [1 6]
[way,bad] appears in review [1 6]
[way,greasy] appears in review [1 6]
[salad,Large] appears in review [1 7]
[Salad,Greek] appears in review [1 7]
[waiter,friendly] appears in review [1 7]
[food,Fast] appears in review [1 8]
[contamination,sick] appears in review [1 8]
[meat,seasoned] appears in review [1 8]
[list,to-eat] appears in review [1 9]
[food,OK] appears in review [1 9]
[service,good] appears in review [2 0]
[ambience,pleasant] appears in review [2 0]

query opinion [service, good] has similar opinions:

[service,warm] appears in review [8]
[service,solid] appears in review [1 3]
[service,great] appears in review [1 4]
[service,good] appears in review [2 0]

query opinion [service, bad] has similar opinions:

[service,solid] appears in review [1 3]
[service,great] appears in review [1 4]
[service,good] appears in review [2 0]

query opinion [atmosphere, good] has similar opinions:

[atmosphere,nice] appears in review [3]
[ambience,nice] appears in review [4]
[atmosphere,great] appears in review [5]
[ambience,pleasant] appears in review [2 0]

query opinion [food, delicious] has similar opinions:

[meal,wonderful] appears in review [1]

[meal,delicious] appears in review [4]

[food,great] appears in review [5]

[food,hearty] appears in review [8]

[food,fresh] appears in review [1 4]

[food,interesting] appears in review [1 4]

[food,satisfying] appears in review [1 6]

Process finished with exit code 0