Information Retrieval and Extraction Course Major Project

Aspect Based Sentiment Analysis (ABSA).

Team No:- 13

Introduction

Sentiment analysis is increasingly viewed as a vital task both from an academic and a commercial standpoint. The majority of current approaches, however, attempt to detect the overall polarity of a sentence, paragraph, or text span, regardless of the entities mentioned (e.g., laptops, restaurants) and their aspects (e.g., battery, screen; food, service). By contrast, this task is concerned with aspect based sentiment analysis (ABSA), where the goal is to identify the aspects of given target entities and the sentiment expressed towards each aspect.

Phases of the Project

Phase 1:

Task 1:

Aspect term Extraction

This task deals with extracting the aspect terms from the reviews. Given a set of sentences with pre-identified entities (e.g., restaurants), identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms. An aspect term names a particular aspect of the target entity.

Task 2:

Aspect Polarity Detection

For a given set of aspect terms within a sentence, determine whether the polarity of each aspect term is positive, negative, neutral or conflict (i.e., both positive and negative).

Phase 2:

Task 3:

Given a predefined set of aspect categories (e.g., price, food), identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of Subtask 1, and they do not necessarily occur as terms in the given sentence.

Task 4:

Given a set of pre-identified aspect categories (e.g., {food, price}), determine the polarity (*positive*, *negative*, *neutral* or *conflict*) of each aspect category.

Dataset:

We used the SemEval 2016 dataset available at:http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-restaurant-reviews-trial-data/1790ab94464211e388f5842b2b6a04d79bb0323cac4f49939bf5b99878dc38be/ This dataset consists of restaurant reviews presented in an XML file.

Pre-Processing of dataset:

- The reviews are first sentence segmented.
- Then the segmented sentences are POS-tagged and their dependency tree is generated using the stanford coreNLP parser.
- From the dependency tree the words are lemmatized using NLTK tool.

  Eg: for the word 'sandwiches', it's lemmatized form is 'sandwich'.

- Now the rules are applied to extract aspects.

Task 1:

Extracting Aspect Terms:

- A set of 10 rules were written, assuming the reviews are grammatically correct.
- Assuming the sentences are grammatically correct.
- One major advantage of using Rule Based Approach is based on the fact that English Language follows a standard structure and hence generalized rules can be formulated.

Rule Based Implementation:

Our system is based on 2 general rules.

1) For sentences having a subject verb.
2) For sentences without a subject verb.

Rules: For sentences with subject verb.

1) if t has any adverbial or adjective modifier and the modifier exists in SenticNet, then t is extracted
2) If the sentence does not have auxiliary verb and t modified by adjective or adverb then both h & t are aspects.
3) t is in direct object relation with noun n,
    a) if n not in senticnet -> n is aspect term.
    b) else n is in senticnet -> n is connected to n1(Noun) using any dependency relation then n1 is aspect term
4) if t has any direct object relation with a token n and the POS of the token n is Noun and n exists in SenticNet, then the token n extracted as aspect term. In the dependency parse tree of the sentence, if another token n1 is connected to n using any dependency relation and the POS of n1 is Noun, then n1 is extracted as an aspect.

5) t connected to t1 using open clausal complement relation and t-t1 should exist in opinion lexicon.

    a) Then t-t1 is extracted as aspect.

    b) If n (Noun) connected to t1 then n is aspect term

6) If t is in relation with copular verb then t is an aspect.

7) if h is noun then h is also aspect.

For sentences without subject verb:

1. Adverb h is in clausal complement relation with t,
    a. then h is extracted as adverb
2. if a token h is connected to a noun t using a prepositional relation,
    a. then both h and t are aspects
3. If h is in direct object relation with t ,
    a. then t is aspect term

Using these set of rules we get aspect terms from reviews.

The result from this task is a list of aspect terms which we store in a dictionary with sentence-id as key.

Task 2:

1. Took each aspect term from the previous task, and the dependency relations and POS tags, which we got from CoreNLP parser, we got words which would give polarity to the aspect term.
2. SenticNet and SentiWordNet are the resources which we used to find the polarity of the aspect terms given by the words extracted from above step.

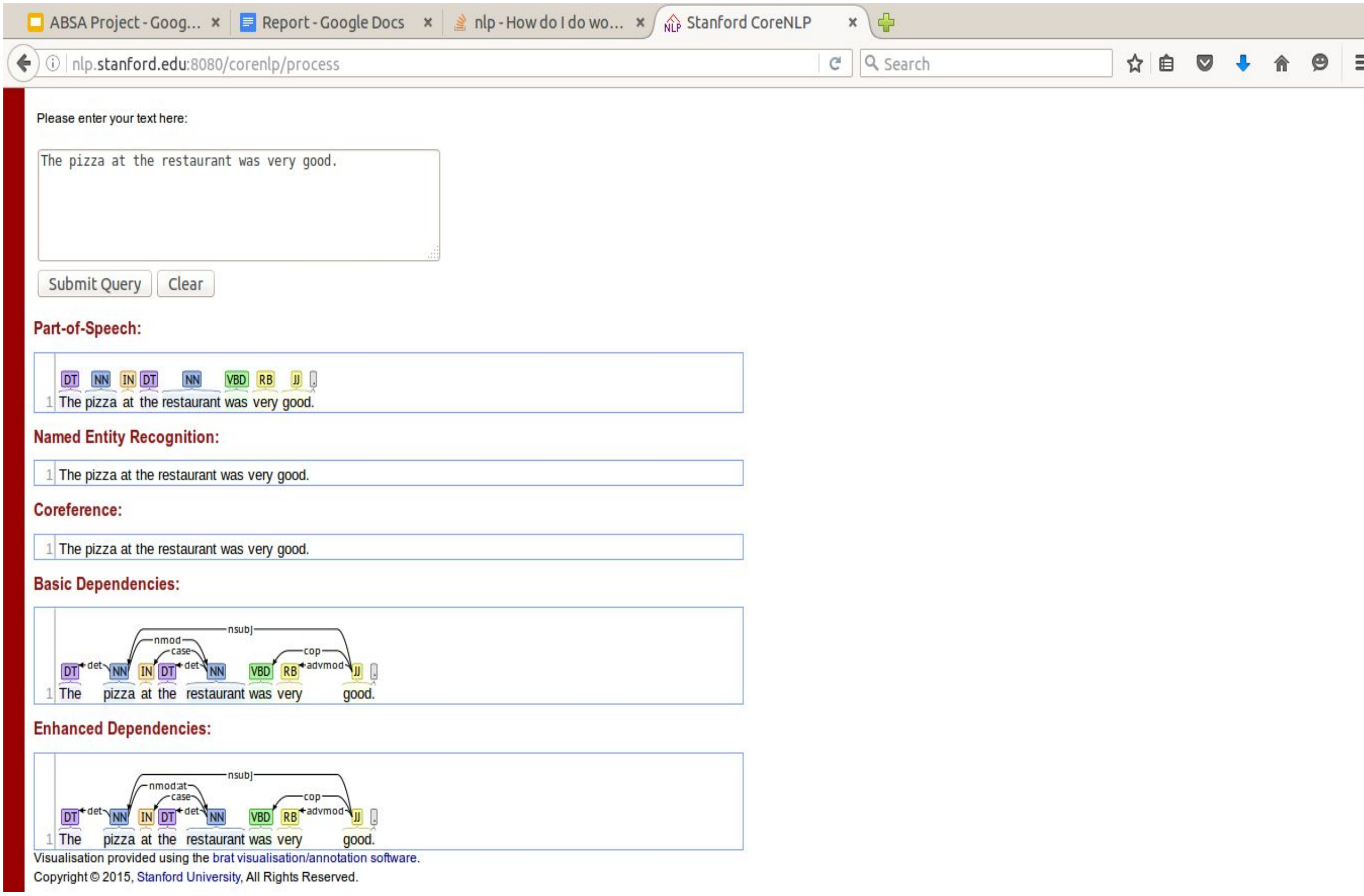
    example :

    For the input string:

‘The pizza at the restaurant was very good.’

We get the following dependencies.

Example



Using this dependency relation we can see that there is a relation between ‘good’ and ‘pizza’ (the aspect term). Sentiment conveyed by ‘good’ is positive and hence we can assign the polarity positive to ‘pizza’.

In this way polarity of each aspect term was assigned.

Task 3:

1. Five categories were given for the aspect terms for the data we used

Namely,   [food, service, ambience, price, miscellaneous]

1. In this task, we made use of Wordnet using which we can find for each aspect term which of the 5 categories are most similar to it.
2. First took each aspect term, which was extracted in Task1 and it's similarity with each one of the 5 categories was calculated using NLTK.Wordnet().
3. Which category gave highest similarity score with the aspect term, that category was assigned to it.

We take each word from the list of aspect terms and calculate the similarity score with each category (namely: price, food, service, ambience, miscellaneous)

The category with which the aspect term gets maximum similarity score is assigned as that aspect category.

Task 4:

1. This task was to find polarity of the category.
2. Using the polarities of the aspect terms from Task2, and categories from Task 3, we calculated the polarities of each category.
3. This was done , by taking the average of the polarity of aspect terms belonging to one category.

If a category has multiple aspect terms assigned to it, we take the average polarity of each aspect term under it and this average polarity score is assigned to the category.

Eg: I really *liked* the pizza at the restaurant but their service was *disappointing.*

The aspect terms extracted were pizza and service.

Category{pizza}:: food -> positive

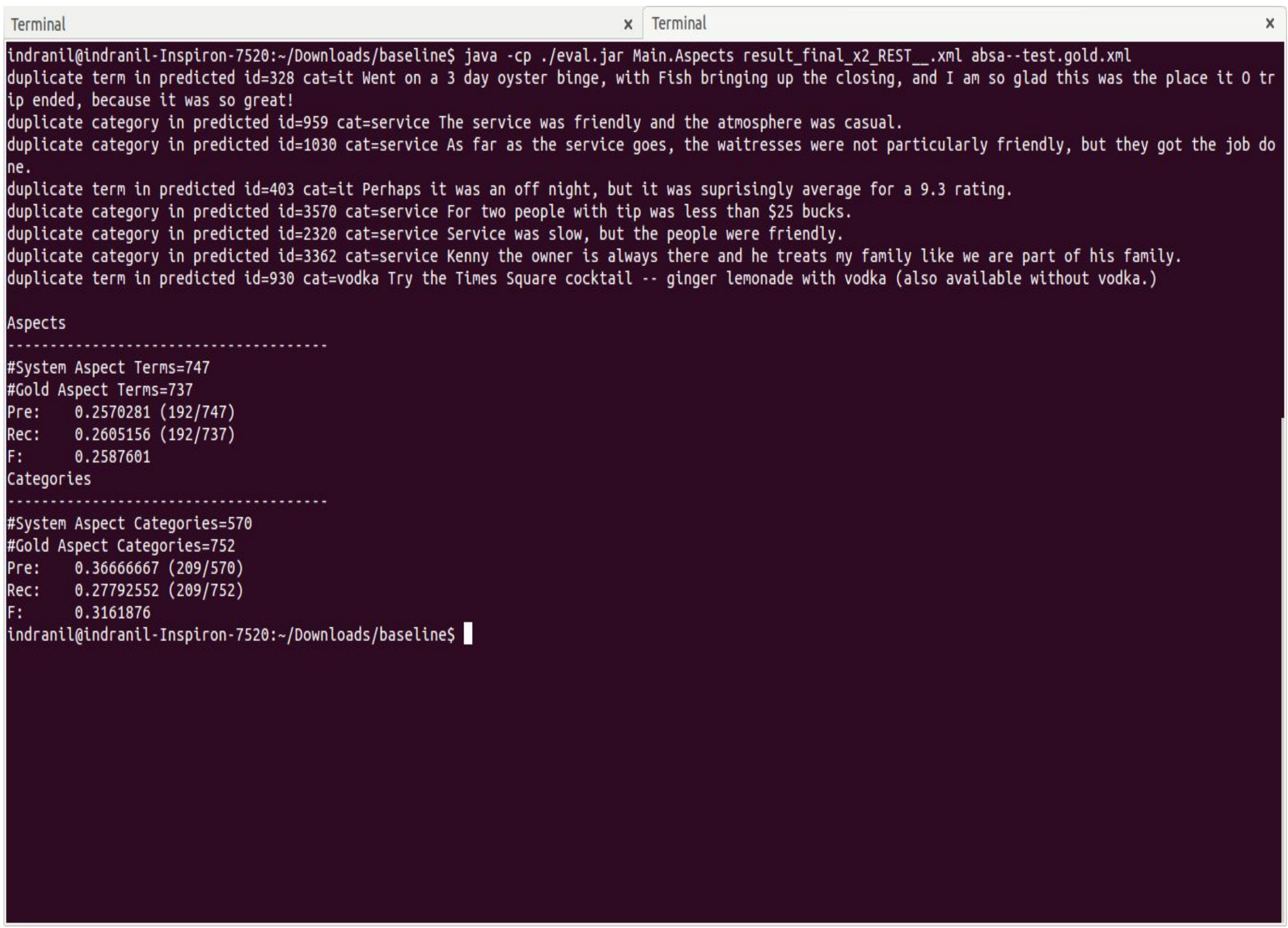
Category{service}:: service -> negative

Results:

The system was run on SemEval 2016 dataset for restaurant reviews.

The evaluation outputs are obtained as follows.

For Aspect Terms:



```
indranil@indranil-Inspiron-7520:~/Downloads/baseline$ java -cp ./eval.jar Main.Aspects result_final_x2_REST__.xml absa--test.gold.xml
duplicate term in predicted id=328 cat=it Went on a 3 day oyster binge, with Fish bringing up the closing, and I am so glad this was the place it O tr
ip ended, because it was so great!
duplicate category in predicted id=959 cat=service The service was friendly and the atmosphere was casual.
duplicate category in predicted id=1030 cat=service As far as the service goes, the waitresses were not particularly friendly, but they got the job do
ne.
duplicate term in predicted id=403 cat=it Perhaps it was an off night, but it was suprisingly average for a 9.3 rating.
duplicate category in predicted id=3570 cat=service For two people with tip was less than $25 bucks.
duplicate category in predicted id=2320 cat=service Service was slow, but the people were friendly.
duplicate category in predicted id=3362 cat=service Kenny the owner is always there and he treats my family like we are part of his family.
duplicate term in predicted id=930 cat=vodka Try the Times Square cocktail -- ginger lemonade with vodka (also available without vodka.)

Aspects
--------------------------------------
#System Aspect Terms=747
#Gold Aspect Terms=737
Pre:    0.2570281 (192/747)
Rec:    0.2605156 (192/737)
F:      0.2587601
Categories
--------------------------------------
#System Aspect Categories=570
#Gold Aspect Categories=752
Pre:    0.36666667 (209/570)
Rec:    0.27792552 (209/752)
F:      0.3161876
indranil@indranil-Inspiron-7520:~/Downloads/baseline$
```

Aspects

#System Aspect Terms=747

#Gold Aspect Terms=737

Pre: 0.2570281 (192/747)

Rec: 0.2605156 (192/737)

F: 0.2587601

Categories

#System Aspect Categories=570

#Gold Aspect Categories=752

Pre: 0.36666667 (209/570)

Rec: 0.27792552 (209/752)

F: 0.3161876

For Aspect Polarities:

```
Terminal                                    Terminal
777 food
777 food
875 anecdotes/miscellaneous
671 table
671 pot of boiling water
671 meats
671 vegetables
671 rice
671 glass noodles
671 food
617 anecdotes/miscellaneous
Aspects
-------------------------------------------------------------------------
Accuracy:       0.44791666 (86/192)

-------------------------------------------------------------------------
label\measure    |Precision        |Recall           |F-measure        |
-------------------------------------------------------------------------
conflict         |NaN(0/0)         |0(0/6)           |NaN              |
negative         |0.3333(11/33)    |0.3235(11/34)    |0.3284           |
neutral          |0.1389(10/72)    |0.5(10/20)       |0.2174           |
positive         |0.7471(65/87)    |0.4924(65/132)   |0.5936           |
-------------------------------------------------------------------------



Categories
-------------------------------------------------------------------------
Accuracy:       0.3110048 (65/209)

-------------------------------------------------------------------------
label\measure    |Precision        |Recall           |F-measure        |
-------------------------------------------------------------------------
conflict         |NaN(0/0)         |0(0/9)           |NaN              |
negative         |0.2143(6/28)     |0.1667(6/36)     |0.1875           |
neutral          |0.087(10/115)    |0.5(10/20)       |0.1481           |
positive         |0.7424(49/66)    |0.3403(49/144)   |0.4667           |
-------------------------------------------------------------------------
indranil@indranil-Inspiron-7520:~/Downloads/baseline$
```

Aspects

Accuracy: 0.44791666 (86/192)

label\measure	Precision	Recall	F-measure
conflict	NaN(0/0)	0(0/6)	NaN
negative	0.3333(11/33)	0.3235(11/34)	0.3284
neutral	0.1389(10/72)	0.5(10/20)	0.2174
positive	0.7471(65/87)	0.4924(65/132)	0.5936

Categories

Accuracy: 0.3110048 (65/209)

label\measure	Precision	Recall	F-measure
conflict	NaN(0/0)	0(0/9)	NaN
negative	0.2143(6/28)	0.1667(6/36)	0.1875

neutral	0.087(10/115)	0.5(10/20)	0.1481	
positive	0.7424(49/66)	0.3403(49/144)	0.4667	

Challenges:

- Aspect term extraction,
    - We used rule based method, and some of the aspect terms weren't extracted correctly by the rules.
    - We assumed all sentences are grammatically correct which is not always the real life scenario.
- Aspect term polarity,
    - In sentences with many aspect terms, it was difficult to assign polarity.

Contributors:-

Team No: 13

Indranil Mukherjee

Saujanya Reddy

Satyendra Venkat

Github:-

-https://github.com/SaujanyaReddy/Aspect-Based-Sentiment-Analysis-IRE-Major-Project slideshare:-http://www.slideshare.net/IndranilMukherjee20/absa-project-60961283

dropBox:--https://www.dropbox.com/sh/krpv30cwdakgr90/AAC-cQ-Vgkm1OpWaokZIEZlba?dl=0