CS-626
Aspect Based Emotion Analysis

Kunal Verma	203050121
Mallela Niteesh Kumar Reddy	203050065
Nimesh Aggarwal	203050049

December 12, 2020

Contents

1 Introduction

Aspect based emotion analysis aims to extract various aspects of reviews and determine the corresponding emotion for each aspect category. The term ‘aspect’ refers to an attribute or a component of the product .Instead of classifying the emotion of overall review into anger, sadness, happiness, surprise and joy aspect-based analysis allows us to associate specific emotion to different aspects of a product and such a analysis provides greater insight to the emotions expressed in the written reviews.

2 Problem Statement

Extract the emotions corresponding to each aspect present in the sentence. Classify the emotions received into 3 sentiments :

- Positive
- Negative
- Neutral

Then test the result using the sentiments obtained.

3 Dataset

For aspect detection and sentiment detection SemEval-16 dataset is used. SemEval-16 providing benchmark datasets of English reviews and a common evaluation framework. The datasets were annotated with aspect terms (e.g. “hard disk”, “pizza”) and their polarity for laptop and restaurant reviews, as well as coarser aspect categories (e.g., FOOD) and their polarity only for the restaurants domain. We have used the reviews corresponding to only Restaurant category for this project.

4 Proposed Methodology

4.1 Data Pre-processing

We have done preprocessing for all the reviews in the dataset by removing unwanted characters,links,spaces etc.

Example :Not only was the food outstanding, but the little ’perks’ were great.
Output :Not only was the food outstanding but the little perks were great

4.2 Ground truth extraction from XML file

XML file is parsed using Beautiful Soup module and a list of sentences , aspects , polarity is extracted from it for each sentence.


```
{'aspect': ['food', 'portions'],
 'aspect_polarity': {'food': 'negative', 'portions': 'negative'},
 'id': 4,
 'text': 'The food was lousy - too sweet or too salty and the portions tiny.'},
{'aspect': ['NULL'],
 'aspect_polarity': {'null': 'negative'},
 'id': 5,
 'text': 'After all that, they complained to me about the small tip.'},
{'aspect': ['place'],
 'aspect_polarity': {'place': 'negative'},
 'id': 6,
 'text': 'Avoid this place!'},
{'aspect': ['food'],
 'aspect_polarity': {'food': 'positive'},
 'id': 7,
 'text': 'I have eaten at Saul, many times, the food is always consistently, outrageously good.'},
```


4.3 Aspect term Extraction

Aspect term extraction task is viewed as a sequence labelling problem.Each token of review is marked with B,I,O encoding scheme.where B, I and O denote the beginning,inside and outside entities of aspect terms.We have used a CRF(Conditional Random Field) model to classify the aspect terms. The classifier is trained with the following set of features:

- Word information
- Part-of-Speech (PoS) tag information
- Previous chunk label information
- Prefixes and suffixes

We were able to achieve an more than 80 percent accuracy.We have used nltk library for obtaining pos tag information.

Input :Not only was the food outstanding but the little perks were great
Output :[’O’, ’O’, ’O’, ’O’, ’B-A’, ’O’, ’O’, ’O’, ’O’, ’B-A’, ’O’ ’O’]
Aspects : [’food’, ’perks’]

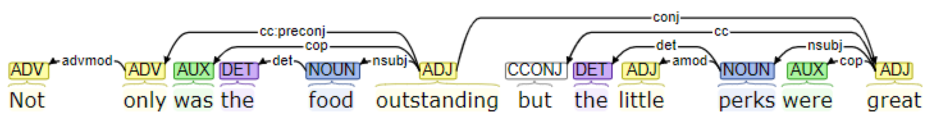
4.3.1 Accuracy and Score

	precision	recall	f1-score	support
O	0.9960	0.9951	0.9956	22497
B-A	0.9419	0.9495	0.9457	1743
I-A	0.9869	0.9947	0.9908	759
accuracy			0.9919	24999
macro avg	0.9750	0.9798	0.9774	24999
weighted avg	0.9920	0.9919	0.9919	24999

4.4 Dependency Relation Extraction

Stanford parser was used to extract the dependency relations of the given input sentence.

Input: not only was the food outstanding but the little perks were great



Output: [('ROOT', 0, 12), ('neg', 2, 1), ('cc:preconj', 12, 2), ('cop', 12, 3), ('det', 5, 4), ('nsubj', 12, 5), ('amod', 5, 6), ('cc', 5, 7), ('det', 10, 8), ('amod', 10, 9), ('conj', 5, 10), ('cop', 12, 11)]

4.5 Emotion words Extraction

A rule based methodology is used to extract the emotion words related to an aspect. Dependency relations and aspects extracted from Stanford parser and CRF model were used as an input.

```python
txt = "But the staff is so horrible to us"
aspects = ['staff']
print(words(txt,aspects,nlp))
```

```
{'staff': ['horrible']}
```

```python
txt = "The food is best in the restaurant but price is high"
aspects = ['food','price']
print(words(txt,aspects,nlp))
```

```
{'food': ['best'], 'price': ['high']}
```

```python
txt = "The food was good but service was bad"
aspects = ['food','service']
print(words(txt,aspects,nlp))
```

```
{'food': ['good'], 'service': ['bad']}
```

```python
txt = "Not only was the food outstanding, but the little perks were great"
aspects = ['food','perks']
print(words(txt,aspects,nlp))
```

```
{'food': ['outstanding'], 'perks': ['little', 'great']}
```

4.6 Aspect Emotion Detection

We tag all the aspects with a particular emotion based on the dependent words extracted We have used 4 types of emotion tagging methods

4.6.1 NRC Lexicon

The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

4.6.2 text2emotion

text2emotion is the python package which will help us to extract the emotions from the content. It processes any textual message and recognize the emotions embedded in it. Compatible with 5 different emotion categories as Happy, Angry, Sad, Surprise and Fear.

4.6.3 Tf-Idf SVM

The parts of sentences that were extracted from the dependency word extractor are treated as sentence which is converted into a tf-idf vector which passed to a

trained SVM classifier which outputs the emotion.
SVM classifier is trained to detect emotions from a sentence.

- Dataset - dailydialog dataset + ISEAR dataset
- 100k sentences
- 5 labels: joy, sad, anger, fear, and neutral
- SVM (Linear Kernel) F1 score = 0.727

4.6.4 Logistic Regression

The parts of sentences that were extracted from the dependency word extractor are treated as sentence which is converted into a tf-idf vector which passed to a trained logistic regression classifier which outputs the emotion. Logistic regression classifier is trained to detect emotions from a sentence.

- Dataset - dailydialog dataset + ISEAR dataset
- 100k sentences
- 5 labels: joy, sad, anger, fear, and neutral
- F1 score = 0.693

4.7 Mapping Emotion to polarity

Emotions Identified for each aspect were mapped as Positive , Negative and Neutral.

- Trust, surprise, happy, joy - positive
- Fear, anger, disgust, sadness - negative

Input : 'food': 'Happy', 'perks': 'Happy'
Output : 'food': 'positive', 'perks': 'positive'

4.8 Output

```
Sentence         =====>         the food was good but service was bad
Extracted Aspect list    =====>         [['food', 'service']]
Extracted dependency words       =====>         {'food': ['good'], 'service': ['bad']}
##################################################
predicted Polarity tagger NRC-   =====>         {'food': 'positive', 'service': 'negative'}
predicted emotions tagger NRC-   =====>         {'food': 'trust', 'service': 'fear'}
##################################################
predicted Polarity tagger text2emotion-  =====>         {'food': 'positive', 'service': 'negat
predicted emotions tagger text2emotions-  =====>        {'food': 'Happy', 'service': 'Sad'}
##################################################
predicted Polarity tagger SVM-   =====>         {'food': 'positive', 'service': 'negative'}
predicted emotions tagger SVM-   =====>         {'food': 'neutral', 'service': 'sadness'}
##################################################
predicted Polarity tagger LGR-   =====>         {'food': 'positive', 'service': 'negative'}
predicted emotions tagger LGR-   =====>         {'food': 'neutral', 'service': 'sadness'}
##################################################
```

5 Results

We have verified our results on the basis of the sentiment/polarity detection. Our main goal here was to annotate the data based on emotion but since no such data was available for verification, we mapped aspects to emotions and then emotions to the sentiment and reported precision, recall, f1-score for 4 different emotion tagging mechanisms.

5.1 Aspect extraction

- Precision = 0.941
- Recall = 0.949
- f1-score = 0.945

5.2 NRC lexicon emotion tagging

- Precision = 0.697
- Recall = 0.497
- f1-score = 0.580

5.3 text2emotion emotion tagging

- Precision = 0.643
- Recall = 0.458
- f1-score = 0.535

5.4 SVM emotion tagging

- Precision = 0.620
- Recall = 0.442
- f1-score = 0.516

5.5 Logistic Regression emotion tagging

- Precision = 0.674
- Recall = 0.480
- f1-score = 0.561

6 Analysis

We can see that the accuracy measures for all the parts are not high for polarity/sentiment detection part. Although the aspect extraction module is highly efficient in extracting the aspects and parts of speech which the review is actually talking about. On looking at the output generated by the emotion taggers we saw that the frequency of the emotion 'neutral' was the most frequent one. However this wasn't the case in the original dataset. This was attributed because the dependency words extracting module was not not able to take care of all the relations that were required by the taggers to classify the emotions. So the emotions were tagged as neutral and hence decreasing the recall and accuracy measure. Figure 1, 2, 3 and 4 show the frequency distribution of the tagged emotions and we can see the trend high neutral tagged emotions which were the main reason for low performance in terms of tagging the emotions. This can be vastly improved by taking in consideration the whole context of the sentence to predict the emotion.

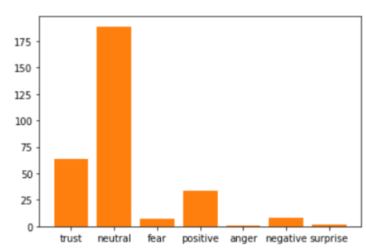


Figure 1: Frequency distribution of tagged emotions - NRC-lexicon

Figure 2: Frequency distribution of tagged emotions - text2emotion

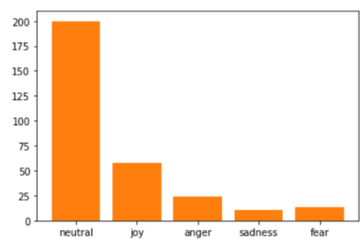


Figure 3: Frequency distribution of tagged emotions - SVM

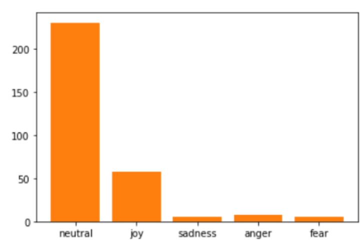


Figure 4: Frequency distribution of tagged emotions - Logistic Regression

7 Future Scope

- The dependency word extractor can be modified using Deep learning based methods such as attention framework to take into consideration all the context present around the aspects
- Combining with the current sentiment analysis techniques to have a better classification of emotion
- Explainability: explaining why certain aspects are tagged as positive and negative to have a better understanding of how underlying system works and what features play important role in classification of sentiment and

emotions

- Aspect based Emotion tagged dataset can be useful for various paradigms of emotion and sentiment analysis

8 References

1. https://github.com/lukasgarbas/nlp-text-emotion - referenced code for ML based emotion tagging
2. https://medium.com/@Intellica.AI/aspect-based-sentiment-analysis-everything-you-wanted-to-know-1be41572e238
3. https://medium.com/analytics-vidhya/aspect-based-sentiment-analysis-a-practical-approach-8f51029bbc4a
4. https://github.com/soujanyaporia/aspect-extraction
5. https://swatimeena989.medium.com/beginners-guide-for-preprocessing-text-data-f3156bec85ca
6. https://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools
7. https://github.com/PoorvaRane/Emotion-Detector/blob/master/ISEAR.csv