Aspect based Sentiment Analysis

Ankit Singh, 12128[1] and Md. Enayat Ullah, 12407[2]

[1] ankitsin@iitk.ac.in,
[2] enayat@iitk.ac.in
Indian Institute of Technology, Kanpur
Mentor: Amitabha Mukerjee

Abstract. Sentiment Analysis is a widely addressed Natural Language Processing task wherein the semantic orientation of a text unit is adjudged. However, a major challenge in Sentiment Analysis is the identification of entities towards which the opinion is expressed. Aspect based Sentiment Analysis is a two-fold SemEval 2015 task constrained to reviews of two domains: Restaurants and Laptops. The first part involves the extraction of the aspect term from a sentence and secondly the polarity of the opinion corresponding to that aspect is adjudged. We adopted an approach based on Probabilistic Graphical Models(PGMs). A linear-chain CRF is trained with features based on word vectors and text processing techniques(POS, dependency parse) to sequentially label the aspect term in a sentence. A Maximum Entropy classifier then identifies the polarity corresponding to the aspect, with features based on cosine similarity with words from sentiwordnet.

Keywords: Natural Language Processing, Sentiment Analysis, Probabilistic Graphical Models

1 Motivation

Sentiment analysis refers to identification and extraction of subjective impressions from text sources. It aims to determine the attitude of a person with respect to something in particular or the overall contextual polarity of a document. In general, a binary composition of opinions is assumed: for/against, like/dislike, good/bad etc. However, an opinion may also be categorized into a neutral sentiment.

Sentiment analysis finds its application in various disciplines; in Information Extraction, it is used to discard subjective information, in Question-Answering, it identifies opinion-oriented questions; in news sources, detecting if there is bias expressed by the author.

1.1 Related Work

Various approaches have been adopted to identify aspects from sentences. Bing Lui et al. used frequency of noun phrases, followed by a redundancy pruning

to identity the feature corresponding to a review[6].Yejin Choi et al. performed semantic tagging using conditional random fields with features based on Capitalization, syntactic chunking to extract sources of opinions from texts[3].

This task was also part of SemEval 2014[1] and the submissions report showcases the use of a variety of approaches[12]. The best performing one uses a Conditional Random Field with features extracted using named entity recognition, POS tagging and parsing.

We try to augment this approach by using features not only based upon text processing techniques, but also on vector embeddings of words and sentences. The motivation behind this being that the number of candidate aspect words of either the laptop or restaurant domain is rather restrictive, and lie close to each other on the space of all word vectors, owing to similarity in meaning and frequent contextual usage. The task of polarity detection was addressed using various classification techniques like Random Forest classifier, Naive Bayes, SVM etc. We adopted a Maximum Entropy model so that we can perform efficient feature engineering which detects polarity with respect to the aspect term.

2 Problem Statement

2.1 Dataset

The data set provided by SemEval is a subset of Ganu et at(2009)[4]. It is in the XML format, and has separate files for Laptop and Restaurant reviews. The training data contains about 500 reviews for each domains, which is 1386 sentences for Restaurants and 1403 sentences for Laptops. The image attached below shows a part the XML file of the Restaurant data.

```xml
▼<Review rid="1028246">
  ▼<sentences>
    ▼<sentence id="1028246:0">
      ▼<text>
          Went on a 3 day oyster binge, with Fish bringing up the closing, and I am so glad this was the place it O trip ended, because it was so great!
        </text>
      ▼<Opinions>
          <Opinion target="Fish" category="RESTAURANT#GENERAL" polarity="positive" from="35" to="39"/>
        </Opinions>
      </sentence>
    ▼<sentence id="1028246:1">
      ▼<text>
          Service was devine, oysters where a sensual as they come, and the price can't be beat!!!
        </text>
      ▼<Opinions>
          <Opinion target="Service" category="SERVICE#GENERAL" polarity="positive" from="0" to="7"/>
          <Opinion target="oysters" category="FOOD#QUALITY" polarity="positive" from="20" to="27"/>
          <Opinion target="NULL" category="RESTAURANT#PRICES" polarity="positive" from="0" to="0"/>
        </Opinions>
      </sentence>
    ▼<sentence id="1028246:2">
        <text>You can't go wrong here.</text>
      ▼<Opinions>
          <Opinion target="NULL" category="RESTAURANT#GENERAL" polarity="positive" from="0" to="0"/>
        </Opinions>
      </sentence>
    </sentences>
  </Review>
```

Fig. 1: Snipptet of the Restaurant dataset XML file

For reach sentence, we have a target attribute which lists the aspect term, and a corresponding polarity attribute. The distribution of positive, neagtive and neutral sentiments in both the dataset is provided below:

Domain	Positive	Negative	Neutral
Laptop	1104	765	106
Restaurant	1198	408	53

2.2 What is an aspect?

An aspect is an explicit reference of an entity about which an opinion is expressed in a given sentence. The opinion can be positive, negative or even neutral. A few examples of aspects with respect to the laptop and restaurant domains are as follows:

The food of restaurant is amazing.
Aspect : food **Polarity**: Positive

The laptop has an awfully low screen resolution.
Aspect : Screen resolution **Polarity**: Negative.

A sentence can have multiple aspects, and and semantic orientation of the opinion expressed with respect to each aspect may or may not be same.

The pizza was delivered cold and the cheese wasn't even fully melted!
Aspect: pizza, **Polarity**: Negative
Aspect: cheese, **Polarity**: Negative

2.3 Aspect term Extraction

The first sub-task requires us to identity the aspect terms in a sentence. There can be multi-word aspects as well as multiple aspects, and every aspect given a sentence needs to extracted. It is also possible that the sentence does not contain any aspect term.

2.4 Polarity Detection

With respect to the aspect term identified above, the polarity of the opinion expressed is now calculated. The polarity of textual unit containing one aspect can either be positive, negative or neutral.In case of no aspect term in the sentence, the overall polarity of the sentence is adjudged.

3 Methodology

3.1 Pre-processing

A word2vec[13] model with skip-gram modeling is trained on 11.7 GB dump of English Wikipedia Corpus to obtain word vector representations each of dimensional 100. The training data provided by SemEval in the XML file is parsed and punctuations are removed from the sentences(excluding " and -). Stanford coreNLP[5] library is then used for Parts of Speech tagging, word tokenising, porter-stemming[9] and constructing dependency parse of sentences[8]. We also shortlisted the 20 most occuring aspect terms for laptops and restaurants individually and computed the word vectors centroids of both of the lists, dubbed as domain centroid. For words missing in the word2vec model, we use the word vector of its lemma as its vector.

3.2 Apsect Term Detection

Conditional Random Fields(or CRFs)[7] are a type of discriminative undirected probabilistic graphical model(PGM) used to encode known relationships between observations and construct consistent interpretations and is majorly used for sequential labellings of entities. A linear chain CRF puts constraints on the relationships between entities such that each entity is only linked to its immediate preceding and succeeding entity thereby forming a linear chain.

The formula below defines the linear-chain CRF: $y = \{y, t\}_{t=1}^{T}$, $x = \{x, t\}_{t=1}^{T}$ are label sequence and observation sequence respectively, and there are K arbitrary feature functions $\{f_k\}_{1 \leq k \leq K}$ and the corresponding weight parameters $\{\theta_k\}_{1 \leq k \leq K}$.

$$P(y \mid x) = \frac{1}{Z(x)} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x, t))$$

Z(x) is a normalizing factor over all classes to ensure that it is a probability function. Below is a schematic for a linear chain CRF which shows the accessible relationships between the tokens as well as the independent features of a token.
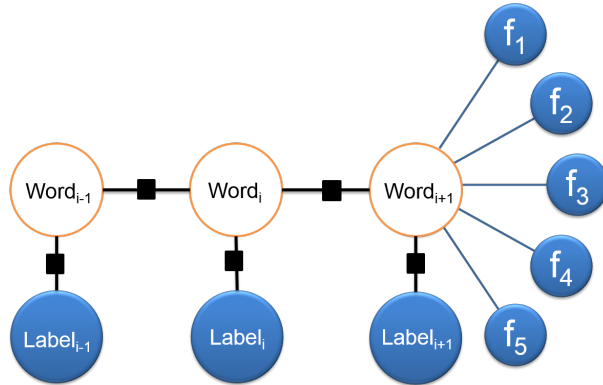
Fig 2: A Linear-chain Conditional Random Field
(Source: Created using MS Word)

3.2.1 Features

In a CRF, one can use as many features he intends, but an increment in the number of features increases the computational cost of the operation. So, we restricted ourselves to 5 features for the CRF. We tested the CRF for a lot of combination of features, some inspired from semEval submission [12] whereas the others based on our own intuition. Below is the list of all the features we used for our CRF

1. Dependent on an opinion word(from Dependency parse tree)
2. Part of Speech tag of the word
3. Cosine Similarity with the Domain Centroid
4. N-gram words(uni-gram was used)
5. Capitalization or Hyphenization in the word
6. Position of the word
7. Is the word a Named Entity?
8. Cosine Similarity with the nearest opinion word

The first five features turned out to give the best result, and the results corresponds to the CRF training on these. Also, in the first feature, we categorize a word as an opinion word if its polarity from SentiWordNet[2](positive or negative) is above a certain user-defined threshold(we used 0.6 as that threshold). .

3.3 Polarity Detection

A Maximum Entropy model[10] defines the conditional distribution of the class (y) given an observation vector x where θ_k is a weight parameter to be estimated for the corresponding feature function f_k

$$P(y \mid x) = \frac{1}{Z(x)} \exp(\sum_{k=1}^{K} \theta_k f_k(y, x))$$

Z(x) is a normalizing factor over all classes to ensure that it is a probability function.

3.3.1 Features

The features selected to train the MaxEnt model were purely based on intuition, and are all dependent on the aspect term extracted in the first part about which the sentiment polarity is being calculated. Features we trained our MaxEnt model on:

1. Nearest Adjective and its polarity(from sentiwordnet)
2. Aspect term dependent on an opinion word(from Dependency parse tree)?
3. Minimum Cosine similarity between the adjective and words from sentiwordnet.

The third feature makes up for the fact that if the adjective is not listed in the sentiwordnet list, we obtain the polarity from the polarity of the word nearest to it.

4 Results

The trained CRF and maxEnt models are tested on test data provided by SemEval. The test data XML file contains 787 sentences for restaurants and 865 sentences for laptops , each tagged with one or more aspect terms and the corresponding polarity. Following tables sums up the results obtained in both the sub-tasks in both the domains.

Domain	Precision	Recall	Accuracy	F1 Score
Laptop	0.5254	0.6823	0.9132	0.5936
Restaurant	0.5769	0.7443	0.9429	0.6522

Table 1: Sub-Task 1: Aspect Term Detection

Polarity	Domain	Precision	Recall	Accuracy	F1 Score
Positive	Laptop	0.8363	0.8255	0.7832	0.7941
	Restaurant	0.8796	0.7542	0.7656	0.8120
Negative	Laptop	0.2891	0.2738	0.7888	0.0.2812
	Restaurant	0.2029	0.2031	0.7686	0.2029

Table 2: Sub-Task 2: Polarity Detection

Restaurant	Original	
Predicted	Positive	Negative
	465	160
	341	8214

Table 3: Confusion Matrix for Aspect Term Detection(Restaurant)

Laptop	Original	
Predicted	Positive	Negative
	319	147
	289	6813

Table 4: Confusion Matrix for Aspect Term Detection(Laptop)

Restaurant	Original	
Predicted	Positive	Negative
	402	131
	55	216

Table 5: Confusion Matrix for Polarity Detection(Restaurant)

Laptop	Original	
Predicted	Positive	Negative
	455	96
	89	236

Table 6: Confusion Matrix for Polarity Detection(Laptop)

SemEval provided us a baseline algorithm which was an SVM with a Linear Kernel [11] in the guidelines. Evaluation scores of the baseline algorithm is tabulated below:

Domain	Task	Score
Laptop	Aspect Term Detection	F1: 0.3858
	Polarity Detection	Accuracy:0.7647
Restaurant	Aspect Term Detection	F1: 0.4868
	Polarity Detection	Accuracy:0.7174

Table 7: Result with baseline algorithm(provided by Semeval)

Following is a graph depicting the variation of F-1 scores with the different features taken for CRF for Restaurant domain, thus establishing that Cosine Similarity with the Domain Centroid and is perhaps the most important features. Also, the capitalization and hyphenization feature does not seem to bring about a marked increase in the F1 score.
The schema for the x-axis label is:

Wi: With only feature i.
Wi': Without feature i.
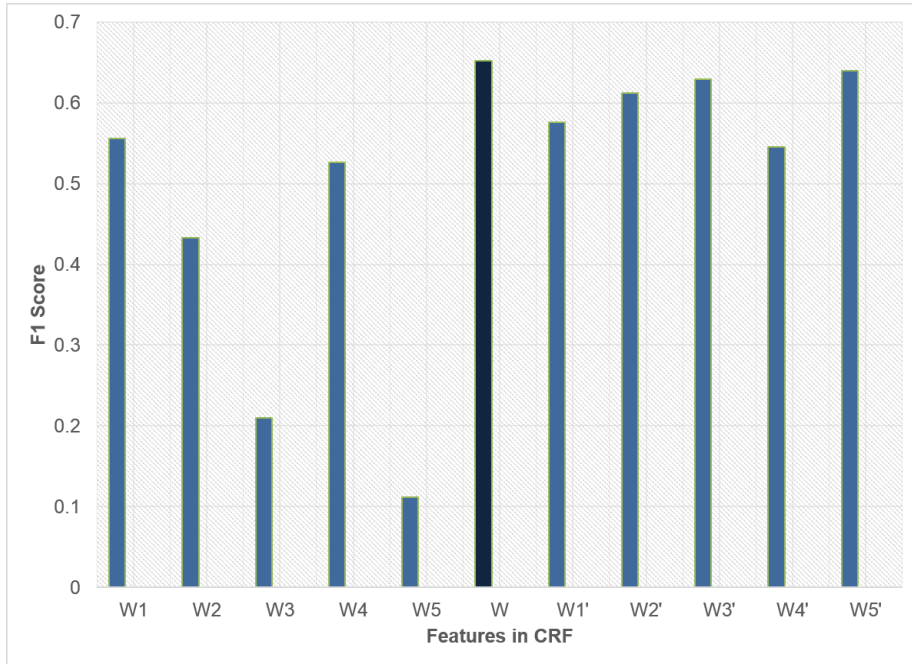**W**: Using all the 5 features.

Fig. 2: Features vs the corresponding F1 scores for Sub-task 1(restaurant)

5 Conclusion

Using the proposed model, we get an F1 score of 0.5936 and 0.6522 for the Laptop and Restaurant domains respectively for the first sub-task. This is a significant improvement over the baseline scores provided, but is still far behind the best submission to SemEval 2014 where they achieved an F1 score of 0.7451 and 0.8401 for laptops and restaurants respectively. In the second task, the accuracy we obtained is 0.7832 and 0.7656 for the Laptop and Restaurant domain respectively. This is comparable to the baseline result, and it shows that the features we handpicked for the task are not good enough, and there is a considerable room for improvement in this terrain. A likely reason for this is that the proportion of positive and negative sentiment is the training data is highly skewed, and thus the prior distribution learnt by the CRF is skewed as well. The best result for the second sub-task in SemEval 2014 was achieved using an SVM with features based on parse trees, named entity recognition and POS tags.

5.1 Future Work

There is a lot of room for improvements to the approach. The features need to be fine-tuned so as to account for the missing links. Features based on punctuations(like exclamation marks(!)) can be useful in this task, also our model fails

to adjudge the polarity in cases like " The food is not delicious"(no feature on negation words). These cases need to handled well by improving the features. Moreover, we used just one feature based on word2vec(which significantly enhanced the result),so properties of word vector representations can be exploited to give better features.

5.2 Acknowledgement

6 Source Code

The source code of the work is available at the following link:
`http://goo.gl/qdTRhf`

References

[1] Semeval 2014. "SemEval-2014 Task 4". In: (2014). DOI: `http://alt.qcri.org/semeval2014/task4/`.

[2] F Sebastian A Esuli. "Sentiwordnet: A publicly available lexical resource for opinion mining". In: (2006). DOI: `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.7217`.

[3] Yejin Choi, Ellen Riloff Claire Cardi, and Siddharth Patwardhan. "Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns". In: (2005). DOI: `http://www.cs.utah.edu/~riloff/pdfs/emnlp05.pdf`.

[4] Amelie Marian Gayatree Ganu Noemie Elhadad. "Beyond the Stars: Improving Rating Predictions using Review Text Content". In: (2009). DOI: `http://eden.rutgers.edu/~gganu/resources/WebDB.pdf`.

[5] The Stanford Natural Language Processing Group. "Stanford CoreNLP". In: (). DOI: `http://nlp.stanford.edu/software/corenlp.shtml`.

[6] M. Hu and B. Liu. "Mining and Summarizing Customer Reviews". In: (2004). DOI: `http://dx.doi.org/10.1002/andp.19053221004`.

[7] Fernando C.N. Pereira John Lafferty Andrew McCallum. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data". In: (2001). DOI: `http://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers`.

[8] and Christopher D. Manning Marie-Catherine de Marneffe Bill MacCartney. "Generating Typed Dependency Parses from Phrase Structure Parses". In: (2011). DOI: `http://nlp.stanford.edu/~wcmac/papers/td-lrec06.pdf`.

[9]   M.F. Porter. "Snowball: A language for stemming algorithms". In: (2001). DOI: http://snowball.tartarus.org/texts/introduction.html.

[10] Adwait Ratnaparkhi. "A Maximum Entropy Model for Part-Of-Speech Tagging". In: (2011). DOI: http://www.ling.helsinki.fi/kit/2011s/clt350/docs/Ratnaparkhi-tagging96.pdf.

[11] SemEval. "SemEval Guide". In: (2015). DOI: http://alt.qcri.org/semeval2015/task12/data/uploads/baseevalvalid1.zip.

[12] Semeval 2014 submission. "SemEval-2014 Task 4". In: (2014). DOI: http://anthology.aclweb.org/S/S14/S14-2.pdf#page=47.

[13] Greg Corrado Jeffrey Dean Tomas Mikolov Kai Chen. "Efficient Estimation of Word Representations in Vector Space". In: (2013). DOI: http://arxiv.org/pdf/1301.3781.pdf.