Sentiment Analysis using Lightweight Discourse Analysis

Shivanshu Gupta, 170020032
Rohan Bansal. 170070058
Ayushraj Mehrotra, 17D070034

December 2020

1 Abstract

We implement a method that will take into account several discourse relations while predicting the sentiment of a sentence. Many sentiment analysis algorithms record the polarity of the word however do not correctly capture the context that leads them to make mistakes. In this paper, we demonstrate how connectives, conditionals and modals can be perceived as extremely valuable context for sentiment prediction. We also highlight the different nuances of dealing with different types of discourse relations. Using a lexicon based and an SVM based approach, we showcase the advantage of using these discourse relations over a simple weighted polarity model.

2 Introduction

Discourse relations are widely prominent in the field of natural language processing. They are of various forms like *connectives, modals, conditionals and negation* and have the power to change the sentiment at both the sentence and the clausal level. Consider the sentence, "*I was happy* **but** *soon became sad*". This sentence portrays an overall negative sentiment although the constituent words comprise of an equal amount of positive and negative. because of this any bag of words model would not be able to classify this sentence properly without considering the discourse relation, "**but**". The presence of the word 'but' puts more emphasis on the following clause (*soon became sad*) as opposed to the preceding clause (*I was happy*). Consider another sentence, "*I was angry* **in spite of** *securing first place in the class*". Here again the sentiment is negative however the discourse relation "in spite of" puts more emphasis on the preceding clause than on the following clause.
We further examine different types of discourse relations using two models, lexicon based classifier and SVM based classifier. We compare our models with

another model which is not capable of handling the discourse relations and draw important conclusions. Our work is primarily based on Mukherjee *et al.* (1)

3 Related Works

Earlier works in discourse analysis include using a discourse parser (2) or a dependency parser (3) which are extensions on the Rhetorical Structure Theory proposed by Mann *et al.* (4) which tries to identify the relations between the nucleus and satellite in the sentence.
The problem with using a parsing is that it is a heavy linguistic resource which increases the processing time whereas web based applications require faster response times. Therefore, there is a need for a lightweight model that can address these issues.

4 Analysis Pipeline

The following discourse relations are handled by our model (1).

Relations	Attributes
Conj_Fol	But, however, nevertheless, otherwise, yet, still, nonetheless
Conj_Prev	Till, until, despite, in spite, though, although
Conj_Infer	Therefore, furthermore, consequently, thus, as a result, subsequently, eventually, hence
Conditionals	If
Strong_Mod	Might, could, can, would, may
Weak_Mod	Should, ought to, need not, shall, will, must
Neg	Not, neither, never, no, nor

Figure 1: List of discourse relations

4.1 Algorithm

We prepare a feature vector that models the discourse relation by selectively assigning weights based on the relation type.
The algorithm takes input a set of m sentences, s_i comprising of n_i words each and returns an n_i sized list per input sentence where each entry is a 4 tuple of the form (w_{ij}, f_{ij}, $flip_{ij}$ and hyp_{ij}) where
w_{ij} : j^{th} word of the i^{th} sentence
f_{ij} : weight of w_{ij}
$flip_{ij}$: A variable indicating whether $w_{ij} \in$ Neg
hyp_{ij} : A boolean variable indicating whether $w_{ij} \in$ Conditionals $\cup$ Strong_Mod
A window size of 5 is used in negation for reversing polarities (Neg_Window)

The algorithm is as follows :

- Initialize all f_{ij}s to 1, $flip_{ij}$ to 1 and hyp_{ij} to 0
- Loop over the words w_{ij}s in a sentence s_i
- If $w_{ij} \in$ Conditionals $\cup$ Strong_Mod, set hyp_{ij} to 1
- If $w_{ij} \in$ Conj_Fol $\cup$ Conj_Infer then for words w_{ik} where k > j and $w_{ik} \notin$ set of discourse relations, increase f_{ik} by 1
- Else if $w_{ij} \in$ Conj_Prev then for words w_{ik} where k < j and $w_{ik} \notin$ set of discourse relations, increase f_{ik} by 1
- If $w_{ij} \in$ Neg then for words w_{ik} where k $\in$ (j, j+Neg_Window) and $w_{ik} \notin$ Conj_Fol $\cup$ Conj_Prev, set $flip_{ik}$ to -1
- return (w_{ij}, f_{ij}, $flip_{ij}$ and hyp_{ij})

4.2 Lexicon Based Classification

The Bing Liu sentiment lexicon(5), which contains 6800 manually tagged words, is used for obtaining the polarity of a word. the output is defined by,

$$\text{output} = \text{sign}(\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij} \times flip_{ij} \times p(w_{ij})) \tag{1}$$

$$\text{where } p(w_{ij}) = \begin{cases} \frac{\text{pol}(\text{w}_{ij})}{1}, & \text{if hyp}_{ij} = 0 \\ \frac{\text{pol}(\text{w}_{ij})}{2}, & \text{if hyp}_{ij} = 1 \end{cases}$$

The $\frac{1}{2}$ scaling factor models the uncertainty introduced by presence of a strong modal or conditional. If output is positive then the sentence represents a positive sentiment and vice versa.

4.3 SVM Based Classification

We used two models , one which could handle the discourse relations and one which could not. Both models were based on a bag of words model with the difference being that the model with discourse had an extra feature vector the size of which was decided by the maximum sized sentence in the dataset which encoded the discourse entities obtained from the algorithm. The dataset used was a restaurant review dataset containing 900 reviews along with manually appended 13 sentences specifically containing discourse relations of various types. Preprocessing included NLTK's tokenization, stemming and removal of stopwords. The sklearn library's SVM class was used as a baseline model.
To tackle the issue of conditionals and strong modals, we introduced a variable α in the binary cross entropy loss function of the SVM class that would model the uncertainty of the above mentioned discourse relations. However, this implementation was not successful.

5 Results and Discussion

Both classification models were tested against custom inputs

5.1 Lexicon Based Classification

- Input sentence = "I am sad"
  Output = -1

The model predicts a negative sentiment which is correct

- Input sentence = "I should feel happy"
  Output = 1

- Input sentence = "I might feel happy"
  Output = 0.5

In the presence of a string modal, the model predicts a weaker verison of the original sentiment. This indicates that the model is able to understand the nuances of modals.

- Input sentence = "I am not sad"
  Output sentence = 1

The model correctly predicts a positive sentiment and hence is able to deal with negation very well.

- Input sentence = "I am happy, relaxed and want to enjoy life"
  Output sentence = 3

Due to presence of multiple positive words, the model predicts a stronger positive sentiment, which indicates that the model performs well on connectives.

5.2 SVM Based Classification

Both models were evaluated on the individual classes, positive and negative

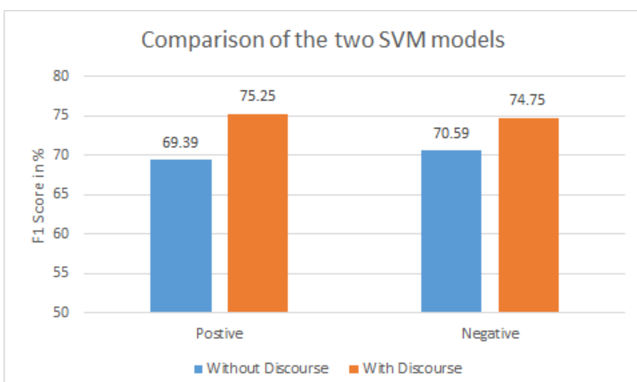

Figure 2: Comparison of f1 scores for both models

Both models were also tested against custom inputs

Without discourse

- Input sentence = “He was relaxed despite the danger”
  Output = **Predicted negative with confidence 62.22%**
- Input sentence = “He is not happy”
  Output = **Predicted neutral**
- Input sentence = “I was happy but soon became very sad”
  Output = **Predicted positive with confidence 79.28%**

With discourse

- Input sentence = “He was relaxed despite the danger”
  Output = **Predicted positive with confidence 72.99%**
- Input sentence = “He is not happy”
  Output = **Predicted positive with confidence 90.95%**
- Input sentence = “I was happy but soon became very sad”
  Output = **Predicted positive with confidence 58.93%**

In both the first and second examples the discourse model outperforms the model without discourse. The third example shows that even when the both the models predict incorrectly, the confidence of discourse model is lower than that of the model without discourse which shows the robustness of the model.

References

[1] Mukherjee, Subhabrata & Bhattacharyya, Pushpak. (2013). Sentiment Analysis in Twitter with Lightweight Discourse Analysis. 24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers.

[2] Marcu, Daniel. 2000. The Theory and Practice of Discourse Parsing and Summarization, MIT Press, Cambridge, MA.

[3] Ng, Vincent and Dasgupta, Sajib and Arifin, S. M. Niaz. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In Proceedings of the COLING/ACL on Main conference poster sessions. 2006

[4] Mann, William C. and Sandra A. Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. Text, 8 (3), 243-281. 1988

[5] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of ACM SIGKDD. 2004