IITB-Sentiment-Analysts: Participation in Sentiment Analysis in Twitter SemEval 2013 Task

Karan Chawla, Ankit Ramteke, Pushpak Bhattacharyya

Dept. of Computer Science and Engineering, IIT Bombay

{chawlakaran,ankitr,pb}@cse.iitb.ac.in

Abstract

We propose a method for using discourse relations for polarity detection of tweets. We have focused on unstructured and noisy text like tweets on which linguistic tools like parsers and POS-taggers don't work properly. We have showed how conjunctions, connectives, modals and conditionals affect the sentiments in tweets. We have also handled the commonly used abbreviations, slangs and collocations which are usually used in short text messages like tweets. This work focuses on a Web based application which produces results in real time. This approach is an extension of the previous work (Mukherjee et al. 2012).

1. Introduction

Discourse relation is an important component of natural language processing which connects phrases and clauses together to establish a coherent relation. Linguistic constructs like conjunctions, connectives, modals, conditionals and negation do alter the sentiments of a sentence. For example, *the movie had **quite a few memorable moments** but I still **did not like it***. The overall polarity of the sentence is negative even though it has one positive and one negative clause. This is because of the presence of the conjunction **but** which gives more weightage to the clause following the conjunction.

Traditional works in discourse analysis use a discourse parser (Marcu et al., 2003; Polanyi et al., 2004; Wolf et al., 2005; Welner et al., 2006; Narayanan et al., 2009; Prasad et al., 2010). Many of these works and some other works in discourse (Taboada et al., 2008; Zhou et al., 2011) build on the Rhetorical Structure Theory (RTS) proposed by Mann et al. (1988) which tries to identify the relations between the nucleus

and satellite in the sentence.

Most of the work is based on well-structured text and the methods applied on that text is not suitable for the discourse analysis on micro-blogs because of the following reasons:

1.   Micro-blogs like Twitter restricts a post (tweet) to be of only 140 characters. Thus, users do not use formal language to discuss their views. Thus, there are abundant spelling mistakes, abbreviations, slangs, collocations, discontinuities and grammatical errors.
These differences cause NLP tools like POS-taggers and parsers to fail frequently, as these tools are built for well-structured text. Thus, most of the methods described in the previous works are not well suited for discourse analysis on Micro-blogs like text.
2.   The web-based applications require a fast response time. Using a heavy linguistic resource like parsing increases the processing time and slows down the application.

Most of the previous work on discourse analysis does not take into consideration the conjunctions, connectives, modals, conditionals etc and are based on bag-of-words model with features like part-of-speech information, unigrams, bigrams etc. along with other domain-specific features like emoticons, hashtags etc. Our work harness the importance of discourse connectives like conjunctions, connectives, modals, conditionals etc and show that along with bag-of-words model, it gives better sentiment classification accuracy. This work is the extension of (Mukherjee et al. 2012).

The roadmap for the rest of the paper is as follows: Section 2 studies the effect of discourse relations on sentiment analysis and identifies the

495

critical ones. Section 3 talks about the semantic operators which influence the discourse relations. Section 4 discusses the lexicon based classification approach. Section 5 describes the feature engineering of the important features. Section 6 gives the list of experiments conducted and analysis of the results. Conclusion and Future Work is presented in Section 7.

2. Discourse Relations Critical for Sentiment Analysis

(Mukherjee et al. 2012) showed that that the following discourse relations are critical for SA as all relations are not useful for SA. Table 1 provides examples of various discourse relations.

Violated Expectations and Contrast: In Example 2, a simple bag-of-words feature based classifier will classify it as positive. However, it actually represents a negative sentiment. Such cases need to be handled separately. In Example 5, "memorable" has (+1) score and "not like" has (-1) score and overall polarity is 0 or objective whereas it should be negative as the final verdict following "but" is the deciding factor.

These kinds of sentences refute the neighboring clause. They can be classified as Conj_Prev in which the clause preceding the conjunction is preferred and Conj_Fol in which the clause following the conjunction is preferred.

Conclusive or Inferential Conjunctions: These are the set of conjunctions, Conj_infer, that tend to draw a conclusion or inference. Hence, the discourse segment following them (subsequently in Example 11) should be given more weight.

Conditionals: In Example 3, "amazing" represent a positive sentiment. But the final polarity should be objective as we are talking of a hypothetical situation.

Other Discourse Relations: Sentences under Cause-Effect, Similarity, Temporal Sequence, Attribution, Example, Generalization and Elaboration, provide no contrasting, conflicting or hy-

pothetical information. They can be handled by taking a simple bag-of-words model.

3. Semantic Operators Influencing Discourse Relations

There are connectives or semantic operators present in the sentences which influence the discourse relation within a sentence. For example, in the sentence *the cannon camera may bad despite good battery life*. The connective *despite* increases the weightage of the previous discourse element *i.e.* bad is weighted up but *may* introduces a certain kind of uncertainty which cannot be ignored.

---

1. (I did not study anything throughout the semester), **so** (I failed in the exams).

2. (Sourav failed to deliver in the penultimate test) **despite** (great expectations).

3. If (I had bought the amazing Nokia phone), I would not be crying).

4. (I love Cannon) **and** (I also love Sony).

5. (The movie had quite a few memorable moments) **but** (I still did not like it).

6. (The theater became interesting) **after a while**.

7. According (to the reviews), (the movie must be bad).

8. (Salman is a bad guy), **for instance** (he is always late).

9. In addition (to the bad battery life), (the camera is also very costly).

10. In general, (cameras from cannon (take great pictures).

11. (They were not in favour of that camera) and **subsequently** (decided not to buy it).

Table 1: Examples of Discourse Coherent Relations

Similarity, in the sentence *He gave his best in the movie, but still it was not good enough to win an Oscar*. The connective *but* increases the weight of the following discourse *i.e. good* and *win* are weighted up but presence of negation operator also cannot be ignored.

1. Modals: Events that are happening or are bound to happen are called *realis events*. And those events that have possibly occurred or have some probability to occur in distant future are known as *irrealis events*. And it is important to distinguish between the two as it also alters the sentiments in a piece of text. Modals depict irrealis events and just cannot be handled by simple majority valence model.

(Mukherjee et al. 2012) divided modals into two categories: Strong_Mod and Weak_Mod.

Strong_Mod is the set of modals that express a higher degree of uncertainty in any situation. Weak_Mod is the set of modals that express lesser degree of uncertainty and more emphasis on certain events or situations.

Like conditionals, sentences with strong modals express higher degree of uncertainty, thus discourse elements near strong modals are weighted down. Thus, in the previous example *the cannon camera may bad despite good battery life* bad is toned down.

Relations	Attributes
Conj_Fol	but, however, nevertheless, otherwise, yet, still, nonetheless
Conj_Prev	till, until, despite, in spite, though, although
Conj_Inf	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

Table 2: Discourse Relations and Semantic Operators Essential for Sentiment Analysis

2. Negation: The negation operator inverts the polarity of the sentence following it. Usually, to handle negation a window (typically 3-5 words) is considered and the polarities of all the words are reversed. We have considered the window size to be 5 and reverse the polarities of all the words within the window, till either a conjunction comes or window size exceeds. For example In the sentence *He gave his best in the movie, but still it was not good enough to win an Oscar* polarities of good and win are reversed.

4. Lexicon Based Classification

We have used Senti-WordNet (Esuli et al. 2006), Inquirer (Stone et. al 1996) and the Bing Liu sentiment lexicon (Hu et al. 2004) to find out the word polarities. To compensate the bias effects introduced by the individual lexicons, we have used three different lexicons. The polarities of the reviews are given by (Mukherjee et al. 2012)

$$sign \left(\sum_{i=1}^{m} \sum_{i=1}^{n_i} f_{ij} * flip_{ij} * p(w_{ij}) \right)$$

$$where \ p(w_{ij}) = pol(w_{ij}) \ if \ hyp_{ij} = 0$$

$$= \frac{pol(w_{ij})}{2} \ if \ hyp_{ij} = 1$$

Above equation finds the weighted, signed polarity of a review. The polarity of each word, $pol(w_{ij})$ being *+1 or -1*, is multiplied with its discourse weight f_{ij} and all the weighted polarities are added. *Flip_{ij}* indicates if the polarity of w_{ij} is to be negated.

In case there is any conditional or strong modal in the sentence (indicated by $hyp_{ij} = 1$), then the polarity of every word in the sentence is toned down, by considering half of its assigned polarity ($\frac{+1}{2}$, $\frac{-1}{2}$)

Thus, if *good* occurs in the user post twice, it will contribute a polarity of $+1 \times 2 = +2$ to the overall review polarity, if $hyp_{ij} = 0$. In the presence of a *strong modal* or *conditional*, it will contribute a polarity of $\frac{+1}{2} * 2 = +1$.

All the stop words, discourse connectives and modals are ignored during the classification phase, as they have a zero polarity in the lexicon. We have handled commonly used slangs, abbreviations and collocations by manually tagging them as positive, negative or neutral.

5. Feature Engineering

The features specific for lexicon based classification for the task sentiment Analysis, identified in Section 2.4, are handled as follows:

a) The words following the Conj_Fol (Table 2) are given more weightage. Hence their frequency count is incremented by 1.
We follow a naive weighting scheme whereby we give a (+1) weightage to every word we consider important. In Example 5, "memorable" gets (+1) score, while "did not like" gets a (-2) score, making the overall score (-1) i.e. the example suggests a negative sentiment.

b) The weightage of the words occurring before the Conj_Prev (Table 2) is increased by 1. In Example 2, "failed" will have polarity (-2) instead of (-1) and "great expectations" will have polarity (+1), making the overall polarity (-1), which conforms to the overall sentiment.

c) The weightage of the words in the sentences containing conditionals (if) and strong modals (might, could, can, would, may) are toned down.

e) The polarity of all words appearing within a window of 5 from the occurrence of a negation operator (not, neither, nor, no, never) and before the occurrence of a violating expectation conjunction is reversed.

f) Exploiting sentence position information, the words appearing in the first k and last k sentences, are given more weightage. The value of k is set empirically.

g) The Negation Bias factor is treated as a parameter which is learnt from a small set of negative polarity tagged documents. The frequency count of all the negative words (in a rule based system) is multiplied with this factor to give negative words more weightage than positive words.

6. Experiments and Evaluation

For the lexicon-based approach, we performed two types of experiments- sentiment pertaining to a particular instance in a tweet (SemEval-2013 Task A) and generic sentiment analysis of a tweet (SemEval-2013 Task B). We treat both the tasks similarly.

6.1 Dataset

We performed experiments on two Datasets:

1) SemEval-2013-task 2 Twitter Dataset A containing 4435 tweets without any external data.
2) SemEval-2013-task 2 Twitter Dataset B containing 3813 tweets without any external data.

6.2 Results on the Twitter Dataset A and B

The system performs best for the positive class tweets as shown in Table 3 and Table 4 and performs badly for the negative class which is due to the fact that negative tweets can contain sarcasm which is a difficult phenomenon to capture. Also the results of the neutral category are very less which suggests that our system is biased towards subjective tweets and we wish to give the majority sentiment in the tweets.

Class	Precision	Recall	F-score
Positive	0.6706	0.5958	**0.6310**
Negative	0.4124	0.5328	0.4649
Neutral	0.0667	0.0063	0.0114

Table 3: Results on Twitter Dataset A

Class	Precision	Recall	F-score
Positive	0.4809	0.5941	**0.5316**
Negative	0.1753	0.5374	0.2643
Neutral	0.6071	0.0104	0.0204

Table 4: Results on Twitter Dataset B

6.3 Discussion

The lexicon based classifier suffers from the problem of lexeme space where it is not able handle all the word senses. Also, short-noisy text like tweets often contain various spelling mistakes like *great* can be grt, g8t *etc.* or *tomorrow* can be tom, tomm, tommrrw *etc.* which will not be detected and handled properly.

We suggest that a supervised approach comprising of the discourse features along with the bag-of-words model and the sense based features will improve the results.

7. Conclusion and Future Work

We have showed that discourse connectives, conjunctions, negations and conditionals do alter the sentiments of a piece of text. Most of the work on Micro-blogs like twitter is build on bag-of-words model and does not incorporate discourse relations. We discussed an approach where we can incorporate discourse relations along-with bag-of-words model for a web-application where parsers and taggers cannot be used as the results are required in real time.

We need to take into consideration word senses and a supervised approach to use all the features collectively. Also, a spell checker would really help in the noisy text like in tweets.

References

A Agarwal and Pushpak Bhattacharyya. 2005. Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be classified. *International Conference on Natural Language Processing (ICON 05)*, IIT Kanpur, India, December

AR Balamurali, Aditya Joshi and Pushpak Bhattacharyya. 2011. Harnessing WordNet Senses for Supervised Sentiment Classification. *In Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*

A Esuli and F Sebastiani, 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings from International Conference on Language Resources and Evaluation (LREC),* Genoa.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. *In Proc. of ACM SIGKDD.*

Aditya Joshi, AR Balamurali, Pushpak Bhattacharyya and R Mohanty. 2010. C-Feel-It: A Sentiment Analyzer for Micro-blogs', *Annual Meeting of the Association of Computational Linguistics (ACL 2011)*, Oregon, USA.

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

R Narayanan, Bing Liu and A Choudhary. 2009. Sentiment Analysis of Conditional Sentences. *In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-09).*

L Polanyi and A Zaenen. 2004. Contextual Valence Shifters. *In James G. Shanahan, Yan Qu, Janyce Wiebe (eds.), Computing Attitude and Affect in Text: Theory and Applications*, pp. 1-10.

BP Ramesh, R Prasad and H Yu. 2010. Identifying explicit discourse connective in biomedical text. *In Annual Symposium proceedings, AMIA Symposium*, Vol. 2010, pp. 657-661.

R Soricut and D Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. *In Proc. of HLT-NAACL*

PJ Stone, DC Dunphy, MS Smith, DM Ogilvie and Associates. 1996. The General Inquirer: A Computer Approach to Content Analysis. *The MIT Press*

Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Sentiment Analysis in Twitter with Lightweight Discourse Analysis. *In Proceedings of COLING 2012*
Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Sentiment Analysis in Twitter with Lightweight Discourse Analysis. *In Proceedings of the 21st ACM Conference on Information and Knowledge Management (CIKM)*, short paper.

Subhabrata Mukherjee, AR Balamurali, Akshat Malu and Pushpak Bhattacharyya. 2012. TwiSent: A Ro-

bust Multistage System for Analyzing Sentiment on Twitter. *In Proceedings of the 21st ACM Conference on Information and Knowledge Management (CIKM),* poster paper.

Maite Taboada, Julian Brooke and Kimberly Voll. 2008. *Extracting Sentiment as a Function of Discourse Structure and Topicality*. Simon Fraser Univeristy School of Computing Science Technical Report.

B Wellner, J Pustejovski, A Havasi, A Rumshiskym and R Suair. 2006. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. *In Proc. of SIGDIAL*

F Wolf and E Gibson. 2005. Representing Discourse Coherence: A Corpus-based Study. *Computational Linguistics*, 31(2), pp. 249-287.

Lanjun Zhou, Binyang Li, Wei Gao, Zhongyu Wei and Kam-Fai Wong. 2011. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. *In Proceedings of EMNLP.*