

What's with the Attitude? Identifying Sentences with Attitude in Online Discussions

Ahmed Hassan

Vahed Qazvinian

Dragomir Radev

University of Michigan Ann Arbor
Ann Arbor, Michigan, USA

hassanam, vahed, radev@umich.edu

Abstract

Mining sentiment from user generated content is a very important task in Natural Language Processing. An example of such content is threaded discussions which act as a very important tool for communication and collaboration in the Web. Threaded discussions include e-mails, e-mail lists, bulletin boards, newsgroups, and Internet forums. Most of the work on sentiment analysis has been centered around finding the sentiment toward products or topics. In this work, we present a method to identify the attitude of participants in an online discussion toward one another. This would enable us to build a signed network representation of participant interaction where every edge has a sign that indicates whether the interaction is positive or negative. This is different from most of the research on social networks that has focused almost exclusively on positive links. The method is experimentally tested using a manually labeled set of discussion posts. The results show that the proposed method is capable of identifying attitudinal sentences, and their signs, with high accuracy and that it outperforms several other baselines.

1 Introduction

Mining sentiment from text has a wide range of applications from mining product reviews on the Web (Morinaga et al., 2002; Turney and Littman, 2003) to analyzing political speeches (Thomas et al., 2006). Automatic methods for sentiment mining are very important because manual extraction of them is very costly, and inefficient. A new application of sentiment mining is to automatically identify attitudes between participants in an online discussion. An automatic tool to identify attitudes will enable

us to build a signed network representation of participant interaction in which the interaction between two participants is represented using a positive or a negative edge. Even though using signed edges in social network studies is clearly important, most of the social networks research has focused only on positive links between entities. Some work has recently investigated signed networks (Leskovec et al., 2010; Kunegis et al., 2009), however this work was limited to a few number of datasets in which users were allowed to explicitly add negative, as well as positive, relations. This work will pave the way for research efforts to examine signed social networks in more detail. It will also allow us to study the relation between explicit relations and the text underlying those relation.

Although similar, identifying sentences that display an attitude in discussions is different from identifying opinionated sentences. A sentence in a discussion may bear opinions about a definite target (e.g., price of a camera) and yet have no attitude toward the other participants in the discussion. For instance, in the following discussion Alice's sentence has her opinion against something, yet no attitude toward the recipient of the sentence, Bob.

Alice: "You know what, he turned out to be a great disappointment"

Bob: "You are completely unqualified to judge this great person"

However, Bob shows strong attitude toward Alice. In this work, we look at ways to predict whether a sentence displays an attitude toward the text recipient. An attitude is the mental position of one participant with regard to another participant. it could be either positive or negative. We consider features which takes into account the entire structure of sentences at different levels or generalization. Those

features include lexical items, part-of-speech tags, and dependency relations. We use all those patterns to build several pairs of models that represent sentences with and without attitude.

The rest of the paper is organized as follows. In Section 2 we review some of the related prior work on identifying polarized words and subjectivity analysis. We explain the problem definition and discuss our approach in Sections 3 & 4. Finally, in Sections 5 & 6 we introduce our dataset and discuss the experimental setup. Finally, we conclude in Section 7.

2 Related Work

Identifying the polarity of individual words is a well studied problem. In previous work, Hatzivassiloglou and McKeown (1997) propose a method to identify the polarity of adjectives. They use a manually labeled corpus to classify each conjunction of an adjective as “the same orientation” as the adjective or “different orientation”. Their method can label simple in “simple and well-received” as the same orientation and simplistic in “simplistic but well-received” as the opposite orientation of well-received. Although the results look promising, the method would only be applicable to adjectives since noun conjunctions may collocate regardless of their semantic orientations (e.g., “rise and fall”).

In other work, Turney and Littman (2003) use statistical measures to find the association between a given word and a set of positive/negative seed words. In order to get word co-occurrence statistics they use the “near” operator from a commercial search engine on a given word and a seed word.

In more recent work, Takamura et al. (2005) used the spin model to extract word semantic orientation. First, they construct a network of words using definitions, thesaurus, and co-occurrence statistics. In this network, each word is regarded as an electron, which has a spin and each spin has a direction taking one of two values: up or down. Then, they use the energy point of view to propose that neighboring electrons tend to have the same spin direction, and therefore neighboring words tend to have the same polarity orientations. Finally, they use the mean field method to find the optimal solution for electron spin directions.

Previous work has also used WordNet, a lexical database of English, to identify word polarity. Specifically, Hu and Liu (2004) use WordNet synonyms and antonyms to predict the polarity of any given word with unknown polarity. They label each word with the polarity of its synonyms and the opposite polarity of its antonyms. They continue in a bootstrapping manner to label all unlabeled instances. This work is very similar to (Kamps et al., 2004) in which a network of WordNet synonyms is used to find the shortest path between any given word, and the words “good” and “bad”. Kim and Hovy (Kim and Hovy, 2004) used WordNet synonyms and antonyms to expand two lists of positive and negative seed words. Similarly, Andreevskaja and Bergler (2006) used WordNet to expand seed lists with fuzzy sentiment categories, in which words could be more central to one category than the other. Finally, Kanayama and Nasukawa (2006) used syntactic features and context coherency, defined as the tendency for same polarities to appear successively, to acquire polar atoms.

All the work mentioned above focus on the task of identifying the polarity of individual words. Our proposed work is identifying attitudes in sentences that appear in online discussions. Perhaps the most similar work to ours is the prior work on subjectivity analysis, which is to identify text that present opinions as opposed to objective text that present factual information (Wiebe, 2000). Prior work on subjectivity analysis mainly consists of two main categories: The first category is concerned with identifying the subjectivity of individual phrases and words regardless of the sentence and context they appear in (Wiebe, 2000; Hatzivassiloglou and Wiebe, 2000; Banea et al., 2008). In the second category, subjectivity of a phrase or word is analyzed within its context (Riloff and Wiebe, 2003; Yu and Hatzivassiloglou, 2003; Nasukawa and Yi, 2003; Popescu and Etzioni,). A good study of the applications of subjectivity analysis from review mining to email classification is given in (Wiebe, 2000). Somasundaran et al. (2007) develop genre-specific lexicons using interesting function word combinations for detecting opinions in meetings. Despite similarities, our work is different from subjectivity analysis because the later only discriminates between opinions and facts. A discussion sentence may display an

opinion about some topic yet no attitude. The language constituents considered in opinion detection may be different from those used to detect attitude. Moreover, extracting attitudes from online discussions is different from targeting subjective expressions (Josef Ruppenhofer and Wiebe, 2008; Kim and Hovy, 2004). The later usually has a limited set of targets that compete for the subjective expressions (for example in movie review, targets could be: director, actors, plot, and so forth). We cannot use similar methods because we are working on an open domain where anything could be a target. A very detailed survey that covers techniques and approaches in sentiment analysis and opinion mining could be found in (Pang and Lee, 2008).

There is also some related work on mining online discussions. Lin et al (2009) proposes a sparse coding-based model simultaneously model semantics and structure of threaded discussions. Shen et al (2006) proposes three clustering methods for exploiting the temporal information in the streams, as well as an algorithm based on linguistic features to analyze the discourse structure information. Huang et al (2007) used an SVM classifier to extract (thread-title, reply) pairs as chat knowledge from online discussion forums to support the construction of a chatbot for a certain domain. Other work has focused on the structure of questions and question-answer pairs in online forums and discussions (Ding et al., 2008; Cong et al., 2008).

3 Problem Definition

Assume we have a set of sentences exchanged between participants in an online discussion. Our objective is to identify sentences that display an attitude from the text writer to the text recipient from those that do not. An attitude is the mental position of one participant with regard to another participant. An attitude may not be directly observable, but rather inferred from what participants say to one another. The attitude could be either positive or negative. Strategies for showing a positive attitude may include agreement, and praise, while strategies for showing a negative attitude may include disagreement, insults, and negative slang. After identifying sentences that display an attitude, we also predict the sign (positive or negative) of that attitude.

4 Approach

In this section, we describe a model which, given a sentence, predicts whether it carries an attitude from the text writer toward the text recipient or not. Any given piece of text exchanged between two participants in a discussion could carry an attitude toward the text recipient, an attitude towards the topic, or no attitude at all. As we are only interested in attitudes between participants, we limit our study to sentences that use second person pronouns. Second person pronouns are usually used in conversational genre to indicate that the text writer is addressing the text recipient. After identifying those sentences, we do some pre-processing to extract the most relevant fragments. We examine these fragments to identify the polarity of every word in the sentence. Every word could be assigned a semantic orientation. The semantic orientation could be either positive, negative, or neutral. The existence of polarized words in any sentence is an important indicator of whether it carries an attitude or not.

The next step is to extract several patterns at different levels of generalization representing any given sentence. We use those patterns to build two Markov models for every kind of patterns. The first model characterizes the relation between different tokens for all patterns that correspond to sentences that have an attitude. The second model is similar to the first one, but it uses all patterns that correspond to sentences that do not have an attitude. Given a new sentence, we extract the corresponding patterns and estimate the likelihood of every pattern being generated from the two corresponding models. We then compare the likelihood of the sentence under the two models and use this as a feature to predict the existence of an attitude. A pair of models will be built for every kind of patterns. If we have n different patterns, we will have n different likelihood ratios that come from n pairs of models.

4.1 Word Polarity Identification

Identifying the polarity of words is an important step for our method. Our word identification module is similar to the work in (Annon, 2010). We construct a graph where each node represent a word/part-of-speech pair. Two nodes are linked if the words are related. We use WordNet (Miller, 1995) to link re-

lated words based on synonyms, hypernyms, and similar to relations. For words that do not appear in Wordnet, we used Wiktionary, a collaboratively constructed dictionary. We also add some links based on co-occurrence statistics between words as from a large corpus. The resulting graph is a graph $G(W, E)$ where W is a set of word/part-of-speech pairs, and E is the set of edges connecting related words.

We define a random walk model on the graph, where the set of nodes correspond to the state space of the random walk. Transition probabilities are calculated by normalizing the weights of the edges out of every node. Let $S+$ and $S-$ be two sets of vertices representing seed words that are already labeled as either positive or negative respectively. We used the list of labeled seeds from (Hatzivassiloglou and McKeown, 1997) and (Stone et al., 1966). For any given word w , we calculate the mean hitting time between w , and the two seed sets $h(w|S+)$, and $h(w|S-)$. The mean hitting time $h(i|k)$ is defined as the average number of steps a random walker, starting in state $i \neq k$, will take to enter state k for the first time (Norris, 1997). If $h(w|S+)$ is greater than $h(w|S-)$, the word is classified as negative, otherwise it is classified as positive. We also use the method described in (Wilson et al., 2005) to determine the contextual polarity of the identified words. The set of features used to predict contextual polarity include word, sentence, polarity, structure, and other features.

4.2 Identifying Relevant Parts of Sentences

The writing style in online discussion forums is very informal. Some of the sentence are very long, and punctuation marks are not always properly used. To solve this problem, we decided to use the grammatical structure of sentences to identify the most relevant part of sentences that would be the subject of further analysis. Figure 1 shows a parse tree representing the grammatical structure of a particular sentence. If we closely examine the sentence, we will notice that we are only interested in a part of the sentence that includes the second person pronoun "you". We extract this part, by starting at the word of interest, in this case "you", and go up in the hierarchy till we hit the first sentence clause. Once, we reach a sentence clause, we extract the corre-

sponding text if it is grammatical, otherwise we go up one more level to the closest sentence clause. We used the Stanford parser to generate the grammatical structure of sentences (Klein and Manning, 2003).

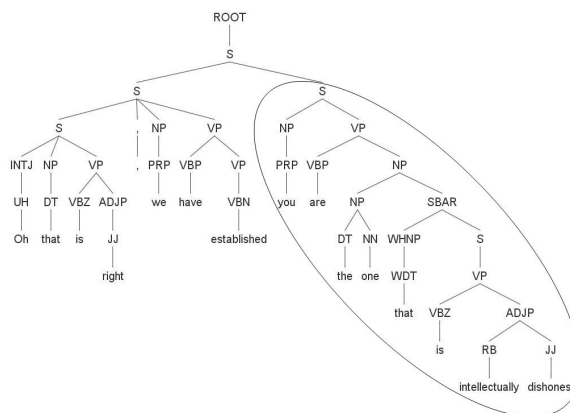


Figure 1: An example showing how to identify the relevant part of a sentence.

4.3 Sentences as Patterns

The fragments we extracted earlier are more relevant to our task and are more suitable for further analysis. However, these fragments are completely lexicalized and consequently the performance of any analysis based on them will be limited by data sparsity. We can alleviate this by using more general representations of words. Those general representations can be used a long with words to generate a set of patterns that represent each fragment. Each pattern consists of a sequence of tokens. Examples of such patterns could use lexical items, part-of-speech (POS) tags, word polarity tags, and dependency relations.

We use three different patterns to represent each fragments:

- Lexical patterns: All polarized words are replaces with the corresponding polarity tag, and all other words are left as is.
- Part-of-speech patterns: All words are replaced with their POS tags. Second person pronouns are left as is. Polarized words are replaced with their polarity tags and their POS tags.
- Dependency grammar patterns: the shortest path connecting every second person pronoun

to the closed polarized word is extracted. The second person pronoun, the polarized word tag, and the types of the dependency relations along the path connecting them are used as a pattern. It has been shown in previous work on relation extraction that the shortest path between any two entities captures the information required to assert a relationship between them (Bunescu and Mooney, 2005). Every polarized word is assigned to the closest second person pronoun in the dependency tree. This is only useful for sentences that have polarized words.

Table 1 shows the different kinds of representations for a particular sentence. We use text, part-of-speech tags, polarity tags, and dependency relations. The corresponding patterns for this sentence are shown in Table 2.

4.4 Building the Models

Given a set of patterns representing a set of sentences, we can build a graph $G = V, E, w$ where V is the set of all possible token that may appear in the patterns. $E = V \times V$ is the set of possible transitions between any two tokens. $w : E \rightarrow [0..1]$ is a weighting function that assigns to every pair of states (i, j) a weight $w(i, j)$ representing the probability that we have a transition from state i to state j .

This graph corresponds to a Markovian model. The set of states are the vocabulary, and the transition probabilities between states are estimated using Maximum Likelihood estimation as follows:

$$P_{ij} = \frac{N_{ij}}{N_i}$$

where N_{ij} is the number of times we saw a transition from i to state j , and N_i is the total number of times we saw state i in the training data. This is similar to building a language model over the language of the patterns.

We build two such models for every kind of patterns. The first model is built using all sentences that appeared in the training dataset and was labeled as having an attitude, and the second model is built using all sentences in the training dataset that do not have an attitude. If we have n kinds of patterns, we

will build one such pair for every kind of patterns. Hence, we will end up with $2n$ models.

4.5 Identifying Sentences with Attitude

We split our training data into two splits; the first containing all sentences that have an attitude and the second containing all sentences that do not have an attitude. Given the methodology described in the previous section, we build n pairs of Markov models. Given any sentence, we extract the corresponding patterns and estimate the log likelihood that this sequence of tokens was generated from every model.

Given a model M , and sequence of tokens $T = (T_1, T_2, \dots, T_{S_n})$, the probability of this token sequence being generated from M is:

$$P_M(T) = \prod_{i=2}^n P(T_i | T_1, \dots, T_{i-1}) = \prod_{i=2}^n W(T_{i-1}, T_i)$$

where n is the number of tokens in the pattern, and W is the probability transition function.

The log likelihood is then defined as:

$$LL_M(T) = \sum_{i=2}^n \log W(T_{i-1}, T_i)$$

For every pair of models, we may use the ratio between the two likelihoods as a feature:

$$f = \frac{LL_{M_{att}}(T)}{LL_{M_{noatt}}(T)}$$

where T is the token sequence, $LL_{M_{att}}(T)$ is the log likelihood of the sequence given the attitude model, and $LL_{M_{noatt}}(T)$ is the log likelihood of the pattern given the no-attitude model.

Given the n kinds of patterns, we can calculate three different features. A standard machine learning classifier is then trained using those features to predict whether a given sentence has an attitude or not.

4.6 Identifying the Sign of an Attitude

To determine the orientation of an attitude sentence, we tried two different methods. The first method assumes that the orientation of an attitude sentence is directly related to the polarity of the words it contains. If the sentence has only positive and neutral

Table 1: Tags used for building patterns

Text	That makes your claims so ignorant
POS	That/DT makes/VBZ your/PRP\$ claims/NNS so/RB ignorant/JJ
Polarity	That/O makes/O your/O claims/O so/O ignorant/NEG
Dependency	your ^{poss} → claims ^{nsubj} → ignorant

Table 2: Sample patterns

Lexical pattern	That makes your claims so NEG
POS pattern	DT VBZ your_PRP\$ NNS RB NEG_JJ
Dependency pattern	your poss nsubj NEG

words, it is classified as positive. If the sentence has only negative and neutral words, it is classified as negative. If the sentence has both positive and negative words, we calculate the summation of the polarity scores of all positive words and that of all negative words. The polarity score of a word is an indicator of how strong of a polarized word it is. If the former is greater, we classify the sentence as positive, otherwise we classify the sentence as negative.

The problem with this method is that it assumes that all polarized words in a sentence with an attitude target the text recipient. Unfortunately, that is not always correct. For example, the sentence "You are completely unqualified to judge this great person" has a positive word "great" and a negative word "unqualified". The first method will not be able to predict whether the sentence is positive or negative. To solve this problem, we use another method that is based on the paths that connect polarized words to second person pronouns in a dependency parse tree. For every positive word w , we identify the shortest path connecting it to every second person pronoun in the sentence then we compute the average length of the shortest path connecting every positive word to the closest second person pronoun. We repeat for negative words and compare the two values. The sentence is classified as positive if the average length of the shortest path connecting positive words to the closest second person pronoun is smaller than the corresponding value for negative words. Otherwise, we classify the sentence as negative.

5 Data

Our data was randomly collected from a set of discussion groups. We collected a large number of

threads from the first quarter of 2009 from a set of Usenet discussion groups. All threads were in English, and had 5 posts or more. We parsed the downloaded threads to identify the posts and senders. We kept posts that have quoted text and discarded all other posts. The reason behind that is that participants usually quote other participants text when they reply to them. This restriction allows us to identify the target of every post, and raises the probability that the post will display an attitude from its writer to its target. We plan to use more sophisticated methods for reconstructing the reply structure like the one in (Lin et al., 2009). From those posts, we randomly selected approximately 10,000 sentences that use second person pronouns. We explained earlier how second person pronouns are used in discussions genres to indicate the writer is targeting the text recipient. Given a random sentence selected from some random discussion thread, the probability that the sentence does not have an attitude is significantly larger than the probability that it will have an attitude. Hence, restricting our dataset to posts with quoted text and sentences with second person pronouns is very important to make sure that we will have a considerable amount of attitudinal sentences. The data was tokenized, sentence-split, part-of-speech tagged with the OpenNLP toolkit. It was parsed with the Stanford dependency parser (Klein and Manning, 2003).

5.1 Annotation Scheme

The goals of the annotation scheme are to distinguish sentences that display an attitude from those that do not. Sentences could display either a negative or a positive attitude. Disagreement, insults, and negative slang are indicators of negative attitude.

	A	B	C	D
A	-	82.7	80.6	82.1
B	81.0	-	81.9	82.9
C	77.8	78.2	-	83.8
D	78.3	77.7	78.6	-

Table 3: Inter-annotator agreement

Agreement, and praise are indicators of positive attitude. Our annotators were instructed to read every sentence and assign two labels to it. The first specifies whether the sentence displays an attitude or not. The existence of an attitude was judged on a three point scale: attitude, unsure, and no-attitude. The second is the sign of the attitude. If an attitude exists, annotators were asked to specify whether the attitude is positive or negative. To evaluate inter-annotator agreement, we use the *agr* operator presented in (Wiebe et al., 2005). This metric measures the precision and recall of one annotator using the annotations of another annotator as a gold standard. The process is repeated for all pairs of annotators, and then the harmonic mean of all values is reported. Formally:

$$agr(A|B) = \frac{|A \cap B|}{|A|} \quad (1)$$

where A , and B are the annotation sets produced by the two reviewers. Table 3 shows the value of the *agr* operator for all pairs of annotators. The harmonic mean of the *agr* operator is 80%. The *agr* operator was used over the Kappa Statistic because the distribution of the data was fairly skewed.

6 Experiments

6.1 Experimental Setup

We performed experiments on the data described in the previous section. The number of sentences with an attitude was around 20% of the entire dataset. The class imbalance caused by the small number of attitude sentences may hurt the performance of the learning algorithm (Provost, 2000). A common way of addressing this problem is to artificially rebalance the training data. To do this we down-sample the majority class by randomly selecting, without replacement, a number of sentences without an attitude that equals the number of sentences with an

attitude. That resulted in a balanced subset, approximately 4000 sentences, that we used in our experiments.

We used Support Vector Machines (SVM) as a classifier. We optimized SVM separately for every experiment. We used 10-fold cross validation for all tests. We evaluate our results in terms of precision, recall, accuracy, and F1. Statistical significance was tested using a 2-tailed paired t-test. All reported results are statistically significant at the 0.05 level. We compare the proposed method to several other baselines that will be described in the next subsection. We also perform experiments to measure the performance if we mix features from the baselines and the proposed method.

6.2 Baselines

The first baseline is based on the hypothesis that the existence of polarized words is a strong indicator that the sentence has an attitude. As a result, we use the number of polarized word in the sentence, the percentage of polarized words to all other words, and whether the sentences has polarized words with mixed or same sign as features to train an SVM classifier to detect attitude.

The second baseline is based on the proximity between the polarized words and the second person pronouns. We assume that every polarized word is associated with the closest second person pronoun. Let w be a polarized word and $p(w)$ be the closes second person pronoun, and $surf_dist(w, p(w))$ be the surface distance between w and $p(w)$. This baseline uses the minimum, maximum, and average of $surf_dist(w, p(w))$ for all polarized words as features to train an SVM classifier to identify sentences with attitude.

The next baseline uses the dependency tree distance instead of the surface distance. We assume that every polarized word is associated to the second person pronoun that is connected to it using the smallest shortest path. The $dep_dist(w, p(w))$ is calculated similar to the previous baselines but using the dependency tree distance. The minimum, maximum, and average of this distance for all polarized words are used as features to train an SVM classifier.

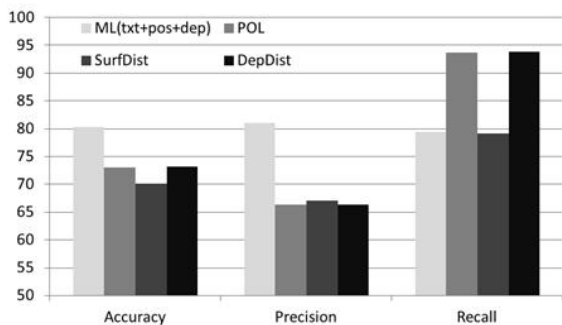


Figure 2: Accuracy, Precision, and Recall for the Proposed Approach and the Baselines.

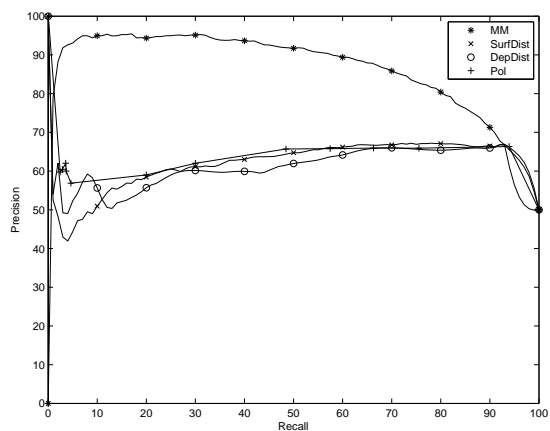


Figure 3: Precision Recall Graph.

6.3 Results and Discussion

Figure 2 compares the accuracy, precision, and recall of the proposed method (ML), the polarity based classifier (POL), the surface distance based classifier (Surf_Dist), and the dependency distance based classifier (Dep_Dist). The values are selected to optimize F1. The figure shows that the surface distance based classifier behaves poorly with low accuracy, precision, and recall. The two other baselines behave poorly as well in terms of precision and accuracy, but they do very well in terms of recall. We looked at some of the examples to understand why those two baselines achieve very high recall. It turns out that they tend to predict most sentences that have polarized words as sentences with attitude. This results in many false positives and low true negative rate. Achieving high recall at the expense of losing precision is trivial. On the other hand, we notice that

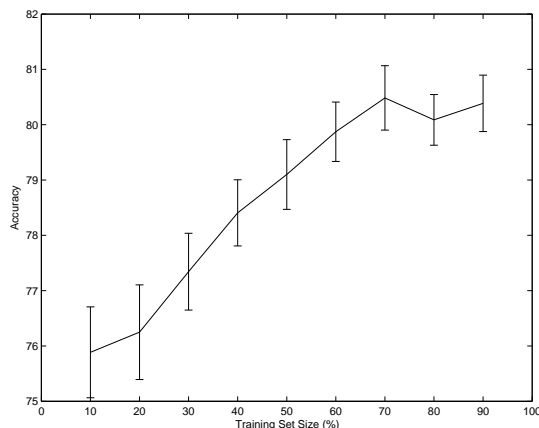


Figure 4: Accuracy Learning Curve for the Proposed Method.

the proposed method results in very close values of precision and recall at the optimum F1 point.

To better compare the performance of the proposed method and the baseline, we study the precision-recall curves for all methods in Figure 3. We notice that the proposed method outperforms all baselines at all operating points. We also notice that the proposed method provides a nice trade-off between precision and recall. This allows us some flexibility in choosing the operating point. For example, in some applications we might be interested in very high precision even if we lose recall, while in other applications we might sacrifice precision in order to get high recall. On the other hand, we notice that the baselines always have low precision regardless of recall.

Table 4 shows the accuracy, precision, recall, and F1 for the proposed method and all baselines. It also shows the performance when we add features from the baselines to the proposed method, or merge some of the baselines. We see that we did not get any improvement when we added the baseline features to the proposed method. We believe that the proposed method captures all the information captured by the baselines and more.

Our proposed method uses three different features that correspond to the three types of patterns we use to represent every sentence. To understand the contributions of every feature, we measure the performance of every feature by itself and also all possible combinations of pairs of features. We compare that

to the performance we get when using all features in Table 5. We see that the part-of-speech patterns performs better than the text patterns. This makes sense because the former suffers from data sparsity. Dependency patterns performs best in terms of recall, while part-of-speech patterns outperform all others in terms of precision, and accuracy. All pairs of features outperform any single feature that belong to the corresponding pair in terms of F1. We also notice that using the three features results in better performance when compared to all other combinations. This shows that every kind of pattern captures slightly different information when compared to the others. It also shows that merging the three features improves performance.

One important question is how much data is required to the proposed model. We constructed a learning curve, shown in Figure 4, by fixing the test set size at one tenth of the data, and varying the training set size. We carried out ten-fold cross validation as with our previous experiments. We see that adding more data continues to increase the accuracy, and that accuracy is quite sensitive to the training data. This suggests that adding more data to this model could lead to even better results.

We also measured the accuracy of the two methods we proposed for predicting the sign of attitudes. The accuracy of the first model that only uses the count and scores of polarized words was 95%. The accuracy of the second method that used dependency distance was 97%.

6.4 Error Analysis

We had a closer look at the results to find out what are the reasons behind incorrect predictions. We found two main reasons. First, errors in predicting word polarity usually propagates and results in errors in attitude prediction. The reasons behind incorrect word polarity predictions is ambiguity in word senses and infrequent words that have very few connection in thesaurus. A possible solution to this type of errors is to improve the word polarity identification module by including word sense disambiguation and adding more links to the words graph using glosses or co-occurrence statistics. The second reason is that some sentences are sarcastic in nature. It is so difficult to identify such sentences. Identifying sarcasm should be addressed as a separate prob-

Method	Accuracy	Precision	Recall	F1
ML	80.3	81.0	79.4	80.2
POL	73.1	66.4	93.9	77.7
ML+POL	79.9	77.9	83.4	80.5
SurfDist	70.2	67.1	79.2	72.7
DepDist	73.1	66.4	93.8	77.8
SurfDist+	73.1	66.4	93.8	77.7
DepDist				
ML+SurfDist	73.9	67.2	93.6	78.2
ML+DepDist	72.8	66.1	93.8	77.6
ML+SurfDist+	74.0	67.2	93.4	78.2
DepDist				
SurfDist+	73.1	66.3	93.8	77.7
DepDist+POL				
ML+SurfDist+	73.0	66.2	93.8	77.6
DepDist+POL				

Table 4: Precision, Recall, F1, and Accuracy for the proposed method, the baselines, and different combinations of proposed method and the baselines features

Method	Accuracy	Precision	Recall	F1
txt	75.5	74.1	78.6	76.2
pos	77.7	78.2	76.9	77.5
dep	74.7	70.4	85.1	77.0
txt+pos	77.8	77.0	79.4	78.1
txt+dep	79.4	79.6	79.2	79.4
pos+dep	80.4	79.1	82.5	80.7
txt+pos+dep	80.3	81.0	79.4	80.2

Table 5: Precision, Recall, F1, and Accuracy for different combinations of the proposed method’s features.

lem. A method that utilizes holistic approaches that takes context and previous interactions between discussion participants into consideration could be used to address it.

7 Conclusions

We have shown that training a supervised Markov model of text, part-of-speech, and dependency patterns allows us to identify sentences with attitudes from sentences without attitude. This model is more accurate than several other baselines that use features based on the existence of polarized word, and proximity between polarized words and second person pronouns both in text and dependency trees. This method allows to extract signed social networks from multi-party online discussions. This opens the door to research efforts that go beyond standard social network analysis that is based on positive links

only. It also allows us to study dynamics behind interactions in online discussions, the relation between text and social interactions, and how groups form and break in online discussions.

Acknowledgments

This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the U.S. Army Research Lab. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the U.S. Government.

References

- Alina Andreevskaia and Sabine Bergler. 2006. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In *EACL'06*.
- Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *LREC'08*.
- Razvan C. Bunescu and Raymond J. Mooney. 2005. A shortest path dependency kernel for relation extraction. In *HLT '05*, pages 724–731, Morristown, NJ, USA. Association for Computational Linguistics.
- Gao Cong, Long Wang, Chin-Yew Lin, Young-In Song, and Yueheng Sun. 2008. Finding question-answer pairs from online forums. In *SIGIR '08*, pages 467–474.
- Shilin Ding, Gao Cong, Chin-Yew Lin, and Xiaoyan Zhu. 2008. Using conditional random fields to extract contexts and answers of questions from online forums. In *ACL'08*, pages 710–718.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *EACL'97*, pages 174–181.
- Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *COLING*, pages 299–305.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD'04*, pages 168–177.
- Jizhou Huang, Ming Zhou, and Dan Yang. 2007. Extracting chatbot knowledge from online discussion forums. In *IJCAI'07*, pages 423–428.
- Swapna Somasundaran Josef Ruppenhofer and Janyce Wiebe. 2008. Finding the sources and targets of subjective expressions. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*.
- Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using wordnet to measure semantic orientations of adjectives. In *National Institute for*, pages 1115–1118.
- Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *EMNLP'06*, pages 355–363.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *COLING*, pages 1367–1373.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *ACL'03*, pages 423–430.
- Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. 2009. The slashdot zoo: mining a social network with negative edges. In *WWW'09*, pages 741–750, New York, NY, USA.
- Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *WWW '10*, pages 641–650, New York, NY, USA. ACM.
- Chen Lin, Jiang-Ming Yang, Rui Cai, Xin-Jing Wang, and Wei Wang. 2009. Simultaneously modeling semantics and structure of threaded discussions: a sparse coding approach and its applications. In *SIGIR '09*, pages 131–138.
- George A. Miller. 1995. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41.
- Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. 2002. Mining product reputations on the web. In *KDD'02*, pages 341–349.
- Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: capturing favorability using natural language processing. In *K-CAP '03: Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77.
- J. Norris. 1997. Markov chains. Cambridge University Press.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Ana-Maria Popescu and Oren Etzioni. Extracting product features and opinions from reviews. In *HLT-EMNLP'05*.
- Foster Provost. 2000. Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI Workshop on Imbalanced Data Sets*.
- Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *EMNLP'03*, pages 105–112.
- Dou Shen, Qiang Yang, Jian-Tao Sun, and Zheng Chen. 2006. Thread detection in dynamic text message streams. In *SIGIR '06*, pages 35–42.
- Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2007. Detecting arguing and sentiment in

- meetings. In *Proceedings of the SIGdial Workshop on Discourse and Dialogue*.
- Philip Stone, Dexter Dunphy, Marchall Smith, and Daniel Ogilvie. 1966. The general inquirer: A computer approach to content analysis. *The MIT Press*.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *ACL'05*, pages 133–140.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *EMNLP 2006*, pages 327–335.
- Peter Turney and Michael Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21:315–346.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2):0.
- Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP'05*, Vancouver, Canada.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP'03*, pages 129–136.