Prior versus Contextual Emotion of a Word in a Sentence

Diman Ghazi
EECS, University of Ottawa
dghaz038@uottawa.ca

Diana Inkpen
EECS, University of Ottawa
diana@eecs.uottawa.ca

Stan Szpakowicz
EECS, University of Ottawa &
ICS, Polish Academy of Sciences
szpak@eecs.uottawa.ca

Abstract

A set of words labelled with their prior emotion is an obvious place to start on the automatic discovery of the emotion of a sentence, but it is clear that context must also be considered. No simple function of the labels on the individual words may capture the overall emotion of the sentence; words are interrelated and they mutually influence their affect-related interpretation. We present a method which enables us to take the contextual emotion of a word and the syntactic structure of the sentence into account to classify sentences by emotion classes. We show that this promising method outperforms both a method based on a *Bag-of-Words* representation and a system based only on the prior emotions of words. The goal of this work is to distinguish automatically between prior and contextual emotion, with a focus on exploring features important for this task.

1 Introduction

Recognition, interpretation and representation of affect have been investigated by researchers in the field of affective computing (Picard 1997). They consider a wide range of modalities such as affect in speech, facial display, posture and physiological activity. It is only recently that there has been a growing interest in automatic identification and extraction of sentiment, opinions and emotions in text.

Sentiment analysis is the task of identifying positive and negative opinions, emotions and evaluations (Wilson, Wiebe, and Hoffmann, 2005). Most of the current work in sentiment analysis has focused on determining the presence of sentiment in the given text, and on determining its polarity – the positive or negative orientation. The applications of sentiment analysis range from classifying positive and negative movie reviews (Pang, Lee, and Vaithyanathan, 2002; Turney, 2002) to opinion question-answering (Yu and Hatzivassiloglou, 2003; Stoyanov, Cardie, and Wiebe, 2005). The analysis of sentiment must, however, go beyond differentiating positive from negative emotions to give a systematic account of the qualitative differences among individual emotion (Ortony, Collins, and Clore, 1988).

In this work, we deal with assigning fine-grained emotion classes to sentences in text. It might seem that these two tasks are strongly tied, but the higher level of classification in emotion recognition task and the presence of certain degrees of similarities between some emotion labels make categorization into distinct emotion classes more challenging and difficult. Particularly notable in this regard are two classes, anger and disgust, which human annotators often find hard to distinguish (Aman and Szpakowicz, 2007). In order to recognize and analyze affect in written text – seldom explicitly marked for emotions – NLP researchers have come up with a variety of techniques, including the use of machine learning, rule-based methods and the lexical approach (Neviarouskaya, Prendinger, and Ishizuka, 2011).

There has been previous work using statistical methods and supervised machine learning applied to corpus-based features, mainly unigrams, combined with lexical features (Alm, Roth, and Sproat, 2005; Aman and Szpakowicz, 2007; Katz, Singleton, and Wicentowski, 2007). The weakness of such methods

Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pages 70–78,
Jeju, Republic of Korea, 12 July 2012.

is that they neglect negation, syntactic relations and semantic dependencies. They also require large (annotated) corpora for meaningful statistics and good performance. Processing may take time, and annotation effort is inevitably high. Rule-based methods (Chaumartin, 2007; Neviarouskaya, Prendinger, and Ishizuka, 2011) require manual creation of rules. That is an expensive process with weak guarantee of consistency and coverage, and likely very task-dependent; the set of rules of rule-based affect analysis task (Neviarouskaya, Prendinger, and Ishizuka, 2011) can differ drastically from what underlies other tasks such as rule-based part-of-speech tagger, discourse parsers, word sense disambiguation and machine translation.

The study of emotions in lexical semantics was the theme of a SemEval 2007 task (Strapparava and Mihalcea, 2007), carried out in an unsupervised setting (Strapparava and Mihalcea, 2008; Chaumartin, 2007; Kozareva et al., 2007; Katz, Singleton, and Wicentowski, 2007). The participants were encouraged to work with *WordNet-Affect* (Strapparava and Valitutti, 2004) and *SentiWordNet* (Esuli and Sebastiani, 2006). Word-level analysis, however, will not suffice when affect is expressed by phrases which require complex phrase- and sentence-level analyses: words are interrelated and they mutually influence their affect-related interpretation. On the other hand, words can have more than one sense, and they can only be disambiguated in context. Consequently, the emotion conveyed by a word in a sentence can differ drastically from the emotion of the word on its own. For example, according to the *WordNet-Affect* lexicon, the word ”afraid” is listed in the ”fear” category, but in the sentence “I am afraid it is going to rain.” the word ”afraid” does not convey fear.

We refer to the emotion listed for a word in an emotion lexicon as the word’s *prior* emotion. A word’s *contextual* emotion is the emotion of the sentence in which that word appears, taking the context into account.

Our method combines several way of tackling the problem. First, we find keywords listed in *WordNet-Affect* and select the sentences which include emotional words from that lexicon. Next, we study the syntactic structure and semantic relations in the text surrounding the emotional word. We explore features important in emotion recognition, and we consider their effect on the emotion expressed by the sentence. Finally, we use machine learning to classify the sentences, represented by the chosen features, by their contextual emotion.

happiness	sadness	anger	disgust	surprise	fear	total
398	201	252	53	71	141	1116

Table 1: The distribution of labels in the *WordNet-Affect* Lexicon.

We categorize sentences into six basic emotions defined by Ekman (1992); that has been the choice of most of previous related work. These emotions are happiness, sadness, fear, anger, disgust and surprise. There also may, naturally, be no emotion in a sentence; that is tagged as neutral/non-emotional.

We evaluate our results by comparing our method applied to our set of features with Support Vector Machine (SVM) applied to *Bag-of-Words*, which was found to give the best performance among supervised methods (Yang and Liu, 1999; Pang, Lee, and Vaithyanathan, 2002; Aman and Szpakowicz, 2007; Ghazi, Inkpen, and Szpakowicz, 2010). We show that our method is promising and that it outperforms both a system which works only with prior emotions of words, ignoring context, and a system which applies SVM to *Bag-of-Words*.

Section 2 of this paper describes the dataset and resources used. Section 3 discusses the features which we use for recognizing contextual emotion. Experiments and results are presented in Section 4. In Section 5, we conclude and discuss future work.

2 Dataset and Resources

Supervised statistical methods typically require training data and test data, manually annotated with respect to each language-processing task to be learned. In this section, we explain the dataset and lexicons used in our experiments.

***WordNet-Affect* Lexicon** (Strapparava and Valitutti, 2004). The first resource we require is an emotional lexicon, a set of words which indicate the presence of a particular emotion. In our experiments, we use *WordNet-Affect*, which contains six lists of words corresponding to the six basic emotion categories. It is the result of assigning a variety

Neutral	Negative	Positive	Both
6.9%	59.7%	31.1%	0.3%

Table 2: The distribution of labels in the Prior-Polarity Lexicon.

hp	sd	ag	dg	sr	fr	ne	total
536	173	179	172	115	115	800	2090

Table 3: The distribution of labels in Aman's modified dataset. The labels are *happiness*, *sadness*, *anger*, *disgust*, *surprise*, *fear*, *no emotion*.

of affect labels to each synset in *WordNet*. Table 1 shows the distribution of words in *WordNet-Affect*.

Prior-Polarity Lexicon (Wilson, Wiebe, and Hoffmann, 2009). The prior-polarity subjectivity lexicon contains over 8000 subjectivity clues collected from a number of sources. To create this lexicon, the authors began with the list of subjectivity clues extracted by Riloff (2003). The list was expanded using a dictionary and a thesaurus, and adding positive and negative word lists from the General Inquirer.[1] Words are grouped into strong subjective and weak subjective clues; Table 2 presents the distribution of their polarity.

Intensifier Lexicon (Neviarouskaya, Prendinger, and Ishizuka, 2010). It is a list of 112 modifiers (adverbs). Two annotators gave coefficients for intensity degree – strengthening or weakening, from 0.0 to 2.0 – and the result was averaged.

Emotion Dataset (Aman and Szpakowicz, 2007). The main consideration in the selection of data for emotional classification task is that the data should be rich in emotion expressions. That is why we chose for our experiments a corpus of blog sentences annotated with emotion labels, discussed by Aman and Szpakowicz (2007). Each sentence is tagged by its dominant emotion, or as non-emotional if it does not include any emotion. The annotation is based on Ekman's six emotions at the sentence level. The dataset contains 4090 annotated sentences, 68% of which were marked as non-emotional. The highly unbalanced dataset with non-emotional sentences as by far the largest class, and merely 3% in the fear and surprise classes, prompted us to remove 2000 of the non-emotional sentences. We lowered the number of non-emotional sentences to 38% of all the sentences, and thus reduced the imbalance. Table 3 shows the details of the chosen dataset.

3 Features

The features used in our experiments were motivated both by the literature (Wilson, Wiebe, and Hoffmann, 2009; Choi et al., 2005) and by the exploration of contextual emotion of words in the annotated data. All of the features are counted based on the emotional word from the lexicon which occurs in the sentence. For ease of description, we group the features into four distinct sets: emotion-word features, part-of-speech features, sentence features and dependency-tree features.

Emotion-word features. This set of features are based on the emotion-word itself.

- The emotion of a word according to *WordNet-Affect* (Strapparava and Valitutti, 2004).
- The polarity of a word according to the prior-polarity lexicon (Wilson, Wiebe, and Hoffmann, 2009).
- The presence of a word in a small list of modifiers (Neviarouskaya, Prendinger, and Ishizuka, 2010).

Part-of-speech features. Based on the Stanford tagger's output (Toutanova et al., 2003), every word in a sentence gets one of the Penn Treebank tags.

- The part-of-speech of the emotional word itself, both according to the emotion lexicon and Stanford tagger.
- The POS of neighbouring words in the same sentence. We choose a window of [-2,2], as it is usually suggested by the literature (Choi et al., 2005).

Sentence features. For now we only consider the number of words in the sentence.

Dependency-tree features. For each emotional word, we create features based on the parse tree and its dependencies produced by the Stanford parser (Marneffe, Maccartney, and Manning, 2006). The

[1] www.wjh.harvard.edu/∼inquirer/

dependencies are all binary relations: a grammatical relation holds between a governor (head) and a dependent (modifier).

According to Mohammad and Turney (2010),[2] adverbs and adjectives are some of the most emotion-inspiring terms. This is not surprising considering that they are used to qualify a noun or a verb; therefore to keep the number of features small, among all the 52 different type of dependencies, we only chose the negation, adverb and adjective modifier dependencies.

After parsing the sentence and getting the dependencies, we count the following dependency-tree Boolean features for the emotional word.

- Whether the word is in a "neg" dependency (negation modifier): true when there is a negation word which modifies the emotional word.
- Whether the word is in a "amod" dependency (adjectival modifier): true if the emotional word is (i) a noun modified by an adjective or (ii) an adjective modifying a noun.
- Whether the word is in a "advmod" dependency (adverbial modifier): true if the emotional word (i) is a non-clausal adverb or adverbial phrase which serves to modify the meaning of a word, or (ii) has been modified by an adverb.

We also have several modification features based on the dependency tree. These Boolean features capture different types of relationships involving the cue word.[3] We list the feature name and the condition on the cue word w which makes the feature true.

- Modifies-positive: w modifies a positive word from the prior-polarity lexicon.
- Modifies-negative: w modifies a negative word from the prior-polarity lexicon.
- Modified-by-positive: w is the head of the dependency, which is modified by a positive word from the prior-polarity lexicon.
- Modified-by-negative: w is the head of the dependency, which is modified by a negative word from the prior-polarity lexicon.

[2]In their paper, they also explain how they created an emotion lexicon by crowd-sourcing, but – to the best of our knowledge – it is not publicly available yet.

[3]The terms "emotional word" and "cue word" are used interchangeably.

	hp	sd	ag	dg	sr	fr	ne	total
part 1	196	64	64	63	36	52	150	625
part 2	51	18	22	18	9	14	26	158
part 1+ part 2	247	82	86	81	45	66	176	783

Table 4: The distribution of labels in the portions of Aman's dataset used in our experiments, named part 1, part 2 and part 1+part 2. The labels are *happiness*, *sadness*, *anger*, *disgust*, *surprise*, *fear*, *no emotion*.

- Modifies-intensifier-strengthen: w modifies a strengthening intensifier from the intensifier lexicon.
- Modifies-intensifier-weaken: w modifies a weakening intensifier from the intensifier lexicon.
- Modified-by-intensifier-strengthen: w is the head of the dependency, which is modified by a strengthening intensifier from the intensifier lexicon.
- Modified-by-intensifier-weaken: w is the head of the dependency, which is modified by a weakening intensifier from the intensifiers lexicon.

4 Experiments

In the experiments, we use the emotion dataset presented in Section 2. Our main consideration is to classify a sentence based on the contextual emotion of the words (known as emotional in the lexicon). That is why in the dataset we only choose sentences which contain at least one emotional word according to *WordNet-Affect*. As a result, the number of sentences chosen from the dataset will decrease to 783 sentences, 625 of which contain only one emotional word and 158 sentences which contain more than one emotional word. Their details are shown in Table 4.

Next, we represent the data with the features presented in Section 3. Those features, however, were defined for each emotional word based on their context, so we will proceed differently for sentences with one emotional word and sentences with more than one emotional word.

- In sentences with one emotional word, we assume the contextual emotion of the emotional word is the same as the emotion assigned to the sentence by the human annotators; therefore all the 625 sentences with one emotional word are represented with the set of features presented in Section 3 and the sentence's emotion will be considered as their contextual emotion.

- For sentences with more than one emotional word, the emotion of the sentence depends on all emotional words and their syntactic and semantic relations. We have 158 sentences where no emotion can be assigned to the contextual emotion of their emotional words, and all we know is the dominant emotion of the sentence.

We will, therefore, have two different sets of experiments. For the first set of sentences, the data are all annotated, so we will take a supervised approach. For the second set of sentences, we combine supervised and unsupervised learning. We train a classifier on the first set of data and we use the model to classify the emotional words into their contextual emotion in the second set of data. Finally, we propose an unsupervised method to combine the contextual emotion of all the emotional words in a sentence and calculate the emotion of the sentence.

For evaluation, we report precision, recall, F-measure and accuracy to compare the results. We also define two baselines for each set of experiments to compare our results with. The experiments are presented in the next two subsections.

4.1 Experiments on sentences with one emotional word

In these experiments, we explain first the baselines and then the results of our experiments on the sentences with only one emotional word.

Baseline

We develop two baseline systems to assess the difficulty of our task. The first baseline labels the sentences the same as the most frequent class's emotion, which is a typical baseline in machine learning tasks (Aman and Szpakowicz, 2007; Alm, Roth, and Sproat, 2005). This baseline will result in 31% accuracy.

The second baseline labels the emotion of the sentence the same as the prior emotion of the only emotional word in the sentence. The accuracy of this experiment is 51%, remarkably higher than the first baseline's accuracy. The second baseline is particularly designed to address the emotion of the sentence only based on the prior emotion of the emotional words; therefore it will allow us to assess the difference between the emotion of the sentence based on the prior emotion of the words in the sentence versus the case when we consider the context and its effect on the emotion of the sentence.

		Precision	Recall	F
SVM + *Bag-of-Words*	Happiness	0.59	0.67	0.63
	Sadness	0.38	0.45	0.41
	Anger	0.40	0.31	0.35
	Surprise	0.41	0.33	0.37
	Disgust	0.51	0.43	0.47
	Fear	0.55	0.50	0.52
	Non-emo	0.49	0.48	0.48
Accuracy	50.72%			
SVM + our features	Happiness	0.68	0.78	0.73
	Sadness	0.49	0.58	0.53
	Anger	0.66	0.48	0.56
	Surprise	0.61	0.31	0.41
	Disgust	0.43	0.38	0.40
	Fear	0.67	0.63	0.65
	Non-emo	0.51	0.53	0.52
Accuracy	58.88%			
Logistic Regression + our features	Happiness	0.78	0.82	0.80
	Sadness	0.53	0.64	0.58
	Anger	0.69	0.62	0.66
	Surprise	0.89	0.47	0.62
	Disgust	0.81	0.41	0.55
	Fear	0.71	0.71	0.71
	Non-emo	0.53	0.64	0.58
Accuracy	66.88%			

Table 5: Classification experiments on the dataset with one emotional word in each sentence. Each experiment is marked by the method and the feature set.

Learning Experiments

In this part, we use two classification algorithms, Support Vector Machines (SVM) and Logistic Regression (LR), and two different set of features, the set of features from Section 3 and *Bag-of-Words* (unigram). Unigram models have been widely used in text classification and shown to provide good results in sentiment classification tasks.

In general, SVM has long been a method of choice for sentiment recognition in text. SVM has been shown to give good performance in text classification experiments as it scales well to the large numbers of features (Yang and Liu, 1999; Pang, Lee, and Vaithyanathan, 2002; Aman and Szpakowicz, 2007). For the classification, we use the SMO algorithm (Platt, 1998) from Weka (Hall et al., 2009), setting *10-fold cross validation* as a testing option. We compare applying SMO to two sets of features, (i) *Bag-of-Words*, which are binary features defining whether a unigram exists in a sentence and (ii) our set of features. In our experiments we use unigrams from the corpus, selected using feature selection methods from Weka.

We also compare those two results with the third experiment: apply SimpleLogistic (Sumner, Frank, and Hall, 2005) from Weka to our set of features, again setting *10-fold cross validation* as a testing option. Logistic regression is a discriminative probabilistic classification model which operates over real-valued vector inputs. It is relatively slow to train compared to the other classifiers. It also requires extensive tuning in the form of feature selection and implementation to achieve state-of-the-art classification performance. Logistic regression models with large numbers of features and limited amounts of training data are highly prone to over-fitting (Aliasi, 2008). Besides, logistic regression is really slow and it is known to only work on data represented by a small set of features. That is why we do not apply SimpleLogistic to *Bag-of-Words* features. On the other hand, the number of our features is relatively low, so we find logistic regression to be a good choice of classifier for our representation method. The classification results are shown in Table 5.

We note consistent improvement. The results of both experiments using our set of features significantly outperform (on the basis of a paired t-test, p=0.005) both the baselines and SVM applied to *Bag-of-Words* features. We get the best result, however, by applying logistic regression to our feature set. The number of our features and the nature of the features we introduce make them an appropriate choice of data representation for logistic regression methods.

4.2 Experiments on sentences with more than one emotional word

In these experiments, we combine supervised and unsupervised learning. We train a classifier on the first set of data, which is annotated, and we use the model to classify the emotional words in the second group of sentences. We propose an unsupervised method to combine the contextual emotion of the emotional words and calculate the emotion of the sentence.

Baseline

We develop two baseline systems. The first baseline labels all the sentences the same: as the emotion of the most frequent class, giving 32% accuracy. The second baseline labels the emotion of the sentence the same as the most frequently occurring prior-emotion of the emotional words in the sentence. In the case of a tie, we randomly pick one of the emotions. The accuracy of this experiment is 45%. Again, as a second baseline we choose a baseline that is based on the prior emotion of the emotional words so that we can compare it with the results based on contextual emotion of the emotional words in the sentence.

Learning Experiments

For sentences with more than one emotional word, we represent each emotional word and its context by the set of features explained in section 3. We do not have the contextual emotion label for each emotional word, so we cannot train the classifier on these data. Consequently, we train the classifier on the part of the dataset which only includes sentences with one emotional word. In these sentences, each emotional word is labeled with their contextual emotion – the same as the sentence's emotion.

Once we have the classifier model, we get the probability distribution of emotional classes for each emotional word (calculated by the logistic regression function learned from the annotated data). We add up the probabilities of each class for all emotional words. Finally, we select the class with the maximum probability. The result, shown in Table 6, is compared using supervised learning, SVM, with *Bag-of-Words* features, explained in previous section, with setting *10-fold cross validation* as a testing

		Precision	Recall	F
SVM + *Bag-of-Words*	Happiness	0.52	0.60	0.54
	Sadness	0.35	0.33	0.34
	Anger	0.30	0.27	0.29
	Surprise	0.14	0.11	0.12
	Disgust	0.30	0.17	0.21
	Fear	0.44	0.29	0.35
	Non-emo	0.23	0.35	0.28
Accuracy	36.71%			
Logistic Regression + unsupervised + our features	Happiness	0.63	0.71	0.67
	Sadness	0.67	0.44	0.53
	Anger	0.50	0.41	0.45
	Surprise	1.00	0.22	0.36
	Disgust	0.80	0.22	0.34
	Fear	0.60	0.64	0.62
	Non-emo	0.37	0.69	0.48
Accuracy	54.43%			

Table 6: Classification experiments on the dataset with more than one emotional word in each sentence. Each experiment is marked by the method and the feature set.

option.[4]

By comparing the results in Table 6, we can see that the result of learning applied to our set of features significantly outperforms (on the basis of a paired t-test, p=0.005) both baselines and the result of SVM algorithm applied to *Bag-of-Words* features.

4.3 Discussion

We cannot directly compare our results with the previous results achieved by Aman and Szpakowicz (2007), because the datasets differ. F-measure, precision and recall for each class are reported on the whole dataset, but we only used part of that dataset. To show how hard this task is, and to see where we stand, the best result from (Aman and Szpakowicz, 2007) is shown in Table 7.

In our experiments, we showed that our approach and our features significantly outperform the baselines and the SVM result applied to *Bag-of-Words*. For the final conclusion, we add one more comparison. As we can see from Table 6, the accuracy result of applying SVM to *Bag-of-Words* is really low. Because supervised methods scale well on large datasets, one reason could be the size of the data we use in this experiment; therefore we try to compare

	Precision	Recall	F
Happiness	0.813	0.698	0.751
Sadness	0.605	0.416	0.493
Anger	0.650	0.436	0.522
Surprise	0.723	0.409	0.522
Disgust	0.672	0.488	0.566
Fear	0.868	0.513	0.645
Non-emo	0.587	0.625	0.605

Table 7: Aman's best result on the dataset explained in Section 2.

the results of the two experiments on all 758 sentences with at least one emotional word.

For this comparison, we apply SVM with *Bag-of-Words* features to all of 758 sentences and we get an accuracy of 55.17%. Considering our features and methodology, we cannot apply logistic regression with our features to the whole dataset; therefore we calculate its accuracy by counting the percentage of correctly classified instances in both parts of the dataset, used in the two experiments, and we get an accuracy of 64.36%. We also compare the results with the baselines. The first baseline, which is the percentage of most frequent class (happiness in this case), results in 31.5% accuracy. The second baseline based on the prior emotion of the emotional words results in 50.13% accuracy. It is notable that the result of applying LR to our set of features is still significantly better than the result of applying SVM to *Bag-of-Words* and both baselines; this supports our earlier conclusion. It is hard to compare the results mentioned thus far, so we have combined all the results in Figure 1, which displays the accuracy obtained by each experiment.

We also looked into our results and assessed the cases where the contextual emotion is different from the prior emotion of the emotional word. Consider the sentence "Joe said it does not happen that often so it does not bother him." Based on the emotion lexicon, the word "bother" is classified as angry; so is the emotion of the sentence if we only consider the prior emotion of words. In our set of features, however, we consider the negation in the sentence, so the sentence is classified as non-emotional rather than angry. Another interesting sentence is the rather simple "You look like her I guess." Based on the lexicon, the word "like" is in the happy category, while

[4] Since SVM does not return a distribution probability, we cannot apply SVM to our features in this set of experiments.

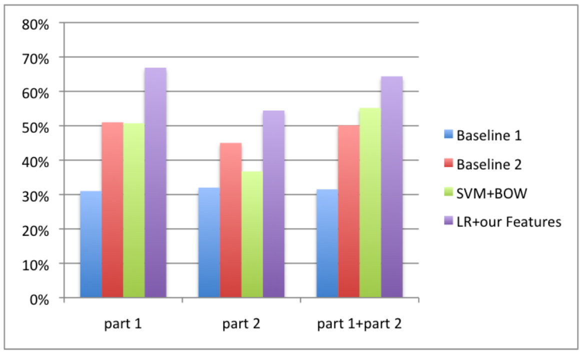

Figure 1: The comparison of accuracy results of all experiments for sentences with one emotional word (part 1), sentences with more than one emotional words (part 2), and sentences with at least one emotional word (part 1+part 2).

the sentence is non-emotional. In this case, the part-of-speech features play an important role and they catch the fact that “like" is not a verb here; it does not convey a happy emotion and the sentence is classified as non-emotional.

We also analyzed the errors, and we found some common errors due to:

- complex sentences or unstructured sentences which will cause the parser to fail or return incorrect data, resulting in incorrect dependency-tree information;
- limited coverage of the emotion lexicon.

These are some of the issues which we would like to address in our future work.

5 Conclusion and Future Directions

The focus of this study was a comparison of prior emotion of a word with its contextual emotion, and their effect on the emotion expressed by the sentence. We also studied features important in recognizing contextual emotion. We experimented with a wide variety of linguistically-motivated features, and we evaluated the performance of these features using logistic regression. We showed that our approach and features significantly outperform the baseline and the SVM result applied to *Bag-of-Words*.

Even though the features we presented did quite well on the chosen dataset, in the future we would like to show the robustness of these features by applying them to different datasets.

Another direction for future work will be to expand our emotion lexicon using existing techniques for automatically acquiring the prior emotion of words. Based on the number of instances in each emotion class, we noticed there is a tight relation between the number of words in each emotion list in the emotion lexicon and the number of sentences that are derived for each emotion class. It follows that a larger lexicon will have a greater coverage of emotional expressions.

Last but not least, one of the weaknesses of our approach was the fact that we could not use all the instances in the dataset. Again, the main reason was the low coverage of the emotion lexicon that was used. The other reason was the limitation of our method: we had to only choose the sentences that have one or more emotional words. As future work, we would like to relax the restriction by using the root of the sentence (based on the dependency tree result) as a cue word rather than the emotional word from the lexicon. So, for sentences with no emotional word, we can calculate all the features regarding the root word rather than the emotional word.

References

Alias-i. 2008. Lingpipe 4.1.0., October.

Alm, Cecilia Ovesdotter, Dan Roth, and Richard Sproat. 2005. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In *HLT/EMNLP*.

Aman, Saima and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In *Proc. 10th International Conf. Text, Speech and Dialogue*, pages 196–205. Springer-Verlag.

Chaumartin, François-Regis. 2007. UPAR7: a knowledge-based system for headline sentiment tagging. In *Proc. 4th International Workshop on Semantic Evaluations*, SemEval ’07, pages 422–425.

Choi, Yejin, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proc. Human Language Technology and Empirical Methods in Natural Language Processing*, HLT ’05, pages 355–362.

Ekman, Paul. 1992. An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200.

Esuli, Andrea and Fabrizio Sebastiani. 2006. SENTI-WORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proc. 5th Conf. on Language Resources and Evaluation LREC 2006*, pages 417–422.

Ghazi, Diman, Diana Inkpen, and Stan Szpakowicz. 2010. Hierarchical approach to emotion recognition and classification in texts. In *Canadian Conference on AI*, pages 40–50.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November.

Katz, Phil, Matthew Singleton, and Richard Wicentowski. 2007. SWAT-MP: the SemEval-2007 systems for task 5 and task 14. In *Proc. 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 308–313.

Kozareva, Zornitsa, Borja Navarro, Sonia Vázquez, and Andrés Montoyo. 2007. UA-ZBSA: a headline emotion classification through web information. In *Proc. 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 334–337.

Marneffe, Marie-Catherine De, Bill Maccartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proc. LREC 2006*.

Mohammad, Saif M. and Peter D. Turney. 2010. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In *Proc. NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 26–34.

Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. 2010. AM: textual attitude analysis model. In *Proc. NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 80–88.

Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. 2011. Affect Analysis Model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17(1):95–135.

Ortony, Andrew, Allan Collins, and Gerald L. Clore. 1988. *The cognitive structure of emotions*. Cambridge University Press.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86.

Platt, John C. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.

Riloff, Ellen. 2003. Learning extraction patterns for subjective expressions. In *Proc. 2003 Conf. on Empirical Methods in Natural Language Processing*, pages 105–112.

Stoyanov, Veselin, Claire Cardie, and Janyce Wiebe. 2005. Multi-perspective question answering using the opqa corpus. In *Proc. Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 923–930.

Strapparava, Carlo and Rada Mihalcea. 2007. SemEval-2007 Task 14: Affective Text. In *Proc. Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74, Prague, Czech Republic, June.

Strapparava, Carlo and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proc. 2008 ACM symposium on Applied computing*, SAC '08, pages 1556–1560.

Strapparava, Carlo and Alessandro Valitutti. 2004. WordNet-Affect: an Affective Extension of WordNet. In *Proc. 4th International Conf. on Language Resources and Evaluation*, pages 1083–1086.

Sumner, Marc, Eibe Frank, and Mark A. Hall. 2005. Speeding Up Logistic Model Tree Induction. In *Proc. 9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 675–683.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proc. HLT-NAACL*, pages 252–259.

Turney, Peter D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proc. 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. HLT-EMNLP*, pages 347–354.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, 35(3):399–433.

Yang, Yiming and Xin Liu. 1999. A re-examination of text categorization methods. In *Proc. 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 42–49.

Yu, Hong and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proc. 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 129–136.