The Role of Knowledge-based Features in Polarity Classification at Sentence Level

Michael Wiegand and **Dietrich Klakow**
Spoken Language Systems
Saarland University
D-66123 Saarbrücken, Germany
{Michael.Wiegand|Dietrich.Klakow}@lsv.uni-saarland.de

Abstract

Though polarity classification has been extensively explored at document level, there has been little work investigating feature design at sentence level. Due to the small number of words within a sentence, polarity classification at sentence level differs substantially from document-level classification in that resulting bag-of-words feature vectors tend to be very sparse resulting in a lower classification accuracy.

In this paper, we show that performance can be improved by adding features specifically designed for sentence-level polarity classification. We consider both explicit polarity information and various linguistic features. A great proportion of the improvement that can be obtained by using polarity information can also be achieved by using a set of simple domain-independent linguistic features.

Introduction

One of the most popular subtasks of opinion mining is polarity classification, i.e. the task of distinguishing between positive and negative utterances. This task has been extensively explored at document level but there has only been comparatively little work at sentence level although the task is an established research problem (Matsumoto, Takamura, and Okumura 2005; Meena and Prabhabkar 2007).

Sentiment information is not evenly distributed across a document. Not only do documents usually comprise both subjective and factual sentences but also the polarity of subjective sentences within a document varies. Thus, sentence-level classification can be used to improve document-level classification (McDonald et al. 2007). Moreover, for tasks demanding fine-grained text analyses, such as text summarization, sentiment classification at sentence level seems more appropriate than document classification.

Due to the small number of words within a sentence, polarity classification at sentence level differs substantially from document-level classification in that resulting feature vectors encoding sentences tend to be much sparser. Therefore, a classifier trained on bag of words performs worse than at document level.

Fortunately, there is a plethora of linguistic features by which a word can be described within a sentence. We consider features, such as *part-of-speech information*, *clause types*, *depth of word constituents*, or *WordNet hypernyms*. At document level, these features have hardly been used. In general, the benefit of these features remains controversial since their extraction is computationally expensive (many of these features require linguistic pre-processing such as part-of-speech tagging or even syntactic parsing) and their contribution in terms of performance is fairly limited since bag-of-words classifiers already pose a robust baseline.

We show that explicit polarity information and a set of simple linguistic features can significantly improve a standard bag-of-words classifier. The additional insight that a standard classifier can be improved by linguistic features in the absence of any polarity information might be useful for situations in which no domain knowledge is available since polarity information is domain-dependent to a great extent.

We consider polarity classification as a binary classification task. That is we assume that each sentence to be classified is subjective. We neglect the distinction between objective and subjective content since this classification is usually solved independently (Pang and Lee 2004; Ng, Dasgupta, and Arifin 2006). Our experiments are carried out on a subset of the MPQA corpus (Wiebe, Wilson, and Cardie 2003).

Related Work

The most closely related work to this are (Wilson, Wiebe, and Hoffmann 2005; Choi and Cardie 2008) which deal with determining contextual polarity of expressions using linguistic information. The crucial difference to these works is that we attempt to determine the overall polarity of a sentence and not the local contextual meaning of individual polar expressions. Sentence-level polarity classification has the benefit that it can harness features derived from sentence structure displaying some form of prominence that cannot be used for expression-level classification (e.g. we consider different clause types, the main predicate of a sentence and depth of word constituents). Unlike (Wilson, Wiebe, and Hoffmann 2005; Choi and Cardie 2008), we also examine in how far linguistic features improve a bag-of-words feature representation in the absence of any polarity information.

(Kudo and Matsumoto 2005) consider polarity and modality classification at sentence level in Japanese. Improvement of a bag-of-words feature set is achieved on both tasks using n-grams based on dependency paths.

(Meena and Prabhabkar 2007) deal with the aspect of conjunctions in polarity classification at sentence level. The results of the heavily domain-dependent rule-based classifier are inconclusive since only sentences with conjuncts are classified more reliably while other sentences are more accurately classified by standard machine learning approaches.

(Moilanen and Pulman 2007) present a symbolic approach using deep linguistic information. The evaluation is done on headlines and noun phrases but not on complete sentences. The method is not compared with a baseline machine learning approach (e.g. using bag of words) either.

At document level, (Gamon 2004) looks at a large set of linguistic features. The performance is increased, but no definite feature subset can be determined to be effective. (Matsumoto, Takamura, and Okumura 2005; Ng, Dasgupta, and Arifin 2006) present syntactically motivated features, most of them based dependency path information. Though some improvement can be achieved with these features, (Ng, Dasgupta, and Arifin 2006) also show that higher-order n-grams are virtually as effective in terms of performance as these linguistic features.

Data

As the data-set for our experiments, we decided to use a subset of the popular MPQA corpus (Wiebe, Wilson, and Cardie 2003) since the corpus is known to have a fairly high inter-annotation agreement. Since the polarity annotation within the MPQA corpus is not at sentence level but expression level, we had to extrapolate the annotation to sentence level. Expressions either labeled as *direct subjective* or *expressive-subjectivity* with attitude-type *positive* or *negative* were identified as polar expressions. The projection to sentence level is straightforward if the annotated polar expressions within one sentence have the same polarity. Sentence (1), for example, illustrates the case where there are two expressions with polarity information, which are both negative. Therefore, the overall polarity of the sentence is also negative.

(1) Their cause was *an unjust one*$^{-}$ and therefore had *little support*$^{-}$.

Of course, there are a lot of sentences in which there are expressions with differing polarity. We manually annotated these sentences (approximately 30% of the final subcorpus we built). Sentence (2) illustrates the case where there are two expressions with different polarity. However, the overall polarity is not mixed. There is a clear preponderance of the second expression which is negative. Therefore, the overall polarity of the sentence is negative.

(2) "The international community can *support*$^{+}$ us so far, but it can *never remove the shackles of repression*$^{-}$", he said.

Moreover, there are also sentences where the overall polarity is mixed as well:

(3) African observers *generally approved*$^{+}$ of his victory while Western governments *denounced*$^{-}$ it.

The number of sentences with mixed polarity is so small that including it for our classification task was not possible. The final corpus we produced was down-sampled to equal class sizes. It contains 2934 sentences in total.

Feature Design

In this work we distinguish between two types of knowledge-based features: *polarity features* and *linguistic features*. The linguistic features have been formulated at two levels: *sentence level* and *word level*. Polarity features have only been formulated at sentence level. Table 1 lists all sentence-level features and Table 2 all word-level features.

Prior Polarity Features

We use the lexicon from (Wilson, Wiebe, and Hoffmann 2005) as it is fairly large compared to other publicly available lexicons. We consider the polarity values *positive*, *negative* and *neutral*[1]. Moreover, the lexicon distinguishes between *strong* and *weak* entries. We exploit this additional information in separate features.

Linguistic Features

A specific linguistic feature at sentence level refers to the overall amount of polar expressions within a sentence whereas linguistic features at word level describe for each word whether or whether not a certain linguistic property holds for it in the context of a particular sentence. For example, if we consider the linguistic property *verb* (one of the *part-of-speech* types explained below), the corresponding features at sentence level are *number of positive verbs*, *number of negative verbs*, and *number of neutral verbs* (within this sentence), whereas the features at word level are for each word *x*: *is x a verb?* (in this sentence). The benefit of using these two levels is that we have both coarse-grained and fine-grained features. Since all features at word level are independent of polarity information[2], we can also evaluate the impact of structural features which do not take polarity information into account. We consider the following linguistic aspects:

Part-of-Speech Information. The predictability towards polarity varies throughout different parts of speech. Many polarity lexicons, for example the one presented in (Nasukawa and Yi 2003), contain mostly adjectives. This means that this part-of-speech tag is more important for polarity classification than others. Apart from that this type of information may also be exploited for some basic word sense disambiguation which can be of help in polarity classification since some important polar expressions are ambiguous. For example, the word *like* can either be a polar verb or just a preposition. In order not to add too much sparse information (in particular with regard to features at word level), we only consider the five part-of-speech tags *noun*, *verb*, *adjective*, *adverb* and *other*.

WordNet Hypernyms *(only used at word level)*. The WordNet ontology (Miller et al. 1990) allows words to be generalized to a certain extent. Our features are inspired

[1] We ignored the value *both* since there are only very few entries with that label (approximately 0.25%).

[2] Note that, on the other hand, all sentence-level features carry polarity information.

by (Scott and Matwin 1998). For each word in a sentence we add all the hypernyms of its synset[3].

Main Predicate & Main Predicate Phrase. We assume that words within a sentence which have a prominent role from a structural perspective are also important words for polarity classification. In this respect, the main predicate of a sentence is of particular importance. We deliberately did not restrict ourselves to verbs since predicative adjectives (*the book is* ***interesting***) seem to be at least equally important. Sentence (4) displays a case where the polarity of the main verb *support*, which is positive, corresponds to the overall polarity of the sentence. The majority of polar expressions, however, is negative. The main predicate feature which is only active on *support* should outweigh the other polar expressions within the sentence with an appropriately learned feature weight.

(4) The Pakistani government *supports*$^+$ President Bush and his *war*$^-$ on *terror*$^-$.

Apart from a feature referring exclusively to the main predicate, we also introduce a more general feature for the entire main predicate phrase, i.e. the entire verbal or adjectival phrase. This should allow polar modifiers within the predicate phrase to be included as well:

(5) The president of the National *Trust*$^+$ [acted *unlawfully*$^-$]$_{predicate\ phrase}$.

We did not consider other grammatical functions for separate features, such as *subject* or *object*, because we assume that these entities are less likely to carry polar information (e.g. these grammatical functions are usually occupied by opinion-holders and opinion-topics).

Depth of Word Constituents. In addition to the previous feature which defines prominence on the basis of grammatical functions (which is fairly restrictive), we also introduce a more general feature which is not bound to any grammatical information. We assume that the depth of a word constituent within a syntax tree (i.e. the length of the path from the leaf node to the root node) can be regarded as another indicator as to how prominent the word is within a sentence. The deeper a constituent is embedded, the less prominent it is and, therefore, the less meaningful it should be for polarity classification. In order to avoid too sparse features we restrict ourself to five depth levels defined in Table 3.

Clause Type. We consider syntactic-based and discourse-based clause types. By syntactic-based type, we distinguish between *main clause* and *other clause* (i.e. adverbial clauses, relative clauses etc.). We assume that words within the main clause of a sentence are more predictive to the overall polarity of a sentence than words in other clause types. By discourse-based types, we also make use of features inspired by (Meena and Prabhbakar 2007) which denote the presence of strengthening discourse connectives (e.g. *but*) and weakening connectives (e.g. *although*).

Both feature types are illustrated by Sentence (6). The polarity of the main clause is also the overall polarity. The strength of the polarity of the subordinate clause is decreased by the presence of the weakening discourse connective *although* and by the fact that this is an *other clause*.

(6) [***Although*** he had *difficulties*$^-$]$_{other}$, [he *successfully*$^+$ managed the job in the end]$_{main}$.

We refrained from defining more specific clause types, e.g. enumerating each subordinate clause since it would have created extremely sparse features.

Intensifiers. Intensifiers are adjectives and adverbs which strengthen the meaning of words. For example, a word, such as *good*, should obtain a higher weight in a sentence if it is modified by an intensifier, such as *extremely*. We took the intensifiers from (Wilson, Wiebe, and Hoffmann 2005). Note that we use this feature also as a word-level feature. A classifier trained on word-level features only (i.e. without the knowledge of polar expressions) might still learn that expressions modified by an intensifier are important since the likelihood of these expressions being polar (in the scope of an intensifier) is quite high.

Modification of Polar Expressions by Other Polar Expressions *(only used at sentence level)*. Polar expressions can modify each other. The consequence of this is that there is a change in the overall meaning. If the polarity of both expressions is the same, there is an intensification (this is similar to the phenomenon described with the previous category type). If the polarity is different, there might be a weakening in strength or even a shift in polarity of the polar expression being modified. The latter phenomenon is illustrated in the following sentence:

(7) Korea has *rejected*$^-$ the framework *agreement*$^+$.

Since the positive expression *agreement* is modified by the negative expression *rejected*, the overall meaning is negative. This sentence also shows that the modifying relation is a long-range relationship that can hardly been captured by higher order n-grams. This feature only operates at sentence level, since it refers to polar expressions which are not considered at word level.

Modal Scope. If an utterance appears within a modal scope[4], semantically, it is not bound to be true. For polar expressions, we assume that words within modal scope are less important than they usually are. Consider, for example, the positive expression *like* in Sentence (8) which is modified by the modal verb *might* and thus semantically weakened.

(8) He ***might*** *like*$^+$ the book, but I'm not sure.

Negation Scope. Usually, if a word appears within the semantic scope of a negation, its meaning is reverted. Apart from using standard negation expressions, such as *no*, *not*, *never*, we also add *polarity shifters* (Wilson, Wiebe, and Hoffmann 2005). Polarity shifters are weaker than negation markers in the sense that they only reverse polarity. They do not fully negate linguistic entities. Most of them usually only change one particular polarity type. For instance the shifter *abate* only turns negative polar expressions into positive polar expressions (as in *abate*$^+$ *the damage*$^-$).

[3] In order to avoid word sense disambiguation, we always map a word onto the first synset in the list of its potential synsets. The first synset usually corresponds to the most frequent sense.

[4] We define the *scope* of constituent x as the set of all constituents which are dominated by x.

Bare Polarity Features
number of positive/negative/neutral expressions
number of strong positive/negative/neutral expressions
number of weak positive/negative/neutral expressions
Linguistic Features
number of positive/negative/neutral nouns
number of positive/negative/neutral verbs
number of positive/negative/neutral adjectives
number of positive/negative/neutral adverbs
number of positive/negative/neutral other (part-of-speech tags)
is main predicate positive/negative/neutral expression?
number of positive/negative/neutral exp. within main predicate phrase
number of positive/negative/neutral exp. with depth level I
number of positive/negative/neutral exp. with depth level II
number of positive/negative/neutral exp. with depth level III
number of positive/negative/neutral exp. with depth level IV
number of positive/negative/neutral exp. with depth level V
number of positive/negative/neutral exp. in main clause
number of positive/negative/neutral exp. in other clause
number of positive/negative/neutral exp. in weak clause
number of positive/negative/neutral exp. in strong clause
number of positive/negative/neutral exp. modified by intensifier
number of positive/negative/neutral exp. modified by positive exp.
number of positive/negative/neutral exp. modified by negative exp.
number of positive/negative/neutral exp. modified by neutral exp.
number of positive/negative/neutral exp. in modal scope
number of negated positive/negative/neutral expressions

Table 1: List of Sentence-Level Features.

Experiments

The results of the following experiments are reported on the basis of a 10-fold crossvalidation. Feature selection was carried out on the training data of each partitioning during the crossvalidation in order to obtain an unbiased set of features. Statistical significance is reported on the basis of a paired t-test with 0.05 as the significance level. We used *SVM-Light* (Joachims 1999) with its standard configuration (linear kernel) for SVMs. All linguistic features were extracted from the output from Charniak's parser (Charniak 2000).

Linguistic Features
is word a noun/verb/adjective/adverb/other?
add hypernym synsets of word
is word the main predicate?
is word within main predicate phrase?
has word depth level I/II/III/IV/V?
is word within main/other clause?
is word within weak/strong clause?
is word preceded by intensifier?
is word within modal scope?
is word negated?

Table 2: List of Word-Level Features.

Feature	Description
Level I	constituents with depth ≤ 5
Level II	constituents with depth ≤ 10
Level III	constituents with depth ≤ 15
Level IV	constituents with depth ≤ 20
Level V	constituents with depth > 20

Table 3: Definition of the Different Depth Features.

Bag-of-Words Feature Set (Baseline)

Following (Pang, Lee, and Vaithyanathan 2002) we encoded all bag-of-words features as binary features indicating the presence or absence of a feature in a sentence. In order to define a strict baseline, we need to find out what subset of bag of words performs best. We tested various amounts using χ^2 feature selection (Yang and Pederson 1997) and found that the best feature set is the one using all words occurring in the training data. This means that a feature selection on this dataset is superfluous.

The average accuracy using the entire set of bag of words with no further normalization than described above is 67.2%. By using the lemmatizer within *WordNet* we increase the performance by approximately 1.4% to 68.6%. (The size of the unlemmatized feature set with approximately 9100 tokens is reduced by approximately 2000 tokens when lemmatization is used.) Comparing this with results of polarity classification at document level, e.g. (Pang, Lee, and Vaithyanathan 2002) report 82.9% on movie reviews using similar features, suggests that polarity at sentence level is much harder and that there is much more room for improvement given this low-performing baseline.

(Linguistic) Word-Level Features

The first extension of the standard feature set we look into are the linguistic word-level features (see Table 2), none of which contains any polarity information. Since polar expressions vary across different domains and common polarity lexicons only capture a unique polarity of polar expressions, the linguistic word-level features should give us a realistic estimate of how good domain-independent features are.

In order to see which features improve the performance of the bag-of-words feature set, we add each feature category (for all words) separately to the standard feature set and measure the increase in performance. We also apply χ^2 feature selection on each separate feature set. Table 4 shows the result of this experiment. The table displays the benefit when the optimal feature size is used. We only display the results of the feature types where we could measure a (notable) increase in performance. Clearly *depth of constituents* is the predominant feature with a contribution of 2.1%. *part of speech*, *clause type*, *WordNet hypernyms* are very similar in their strength. All features with exception of *main predicate (phrase)* are significantly improving the bag-of-words baseline. We were very surprised that *negation* did not notably increase the baseline performance. However, (Pang, Lee, and Vaithyanathan 2002) also report only negligible improvement. We assume this is due to the fact that this feature lacks in recall[5]. We also assume that the same is true for the remaining features referring implicitly to polarity, i.e *intensifier modification* and *modal scope*.

The upper part of Table 5 contrasts the word-level feature set with the other bare bag-of-words feature sets. We applied χ^2 feature selection to the entire linguistic word-level feature set. The classifier using all bag of words and the optimal subset of all linguistic features (i.e. 6000 additional features) outperforms the simplest baseline classifier by 5.9% which is clearly significant and still 4.5% better than the lemmatized bag-of-words feature set. The linguistic word-level features are the only features in our experiments where a feature selection produced a significantly better performance than using the entire feature set. The accuracy of the complete feature set (with approximately $26,000$ active features) is more than 2% worse than the optimal feature set.

Feature Type	Opt. Size of Feat. Set	Benefit (Acc.)
Depth of Constituents	2000	$+2.1\%$
Part of Speech	2000	$+1.3\%$
Clause Type	1000	$+1.2\%$
WordNet hypernyms	1000	$+1.1\%$
Main Predicate (Phrase)	1000	$+0.8\%$

Table 4: Benefit of Individual Word-Level Feature Type Categories *(optimal feature size)* when Added to Bag of Words.

Sentence-Level: Polarity and Linguistic Features

The lower part of Table 5 shows the result of the classifiers using different sentence-level feature sets. A classifier only trained on the prior polarity features (see Table 1) already achieves 70.4% accuracy. If we add all linguistic sentence-level features (see also Table 1), we obtain an increase in performance by 3.4%. This shows that these remaining sentence-level features are encoding other important information than the bare prior polarity features.

In order to find out which features are most discriminative and additive at sentence level, we do a best-first forward selection. Unlike χ^2 feature selection, forward selection has the advantage of selecting features encoding disjunct information[6]. The feature selection on the sentence-level features did not significantly improve performance. After all, there are far fewer features in this feature set (less than 100 features) than in the previous word-level features set ($26,000$ active features) and, therefore, less noise is expected to be in that feature set. Table 6 displays the result of this feature selection. As far as linguistic features are concerned, the results are similar to the feature analysis of the word-level features. The fact that adjectives belong to the most important part-of-speech tag was to be expected (see discussion above). It is no surprise either that only depth levels I and II occur in the optimal feature set since these two levels usually denote a high level of prominence. With the occurrence of *main predicate*, *main predicate phrase* and *main clause*, our analysis proves that virtually all syntactically prominent constituents within a sentence can be effective features for polarity classification.

Adding lemmatized bag of words instead of the other sentence-level features results in an even higher improvement by 5% to 75.4% showing that bag of words and the prior polarity features are complementary and extremely additive. This number, however, may be optimistic since the polarity lexicon we are using does not have to have such a high coverage on other domains.

Finally, we test in how far we can increase the performance of a feature set comprising prior polarity information and bag of words. Performance is increased by adding either the remaining sentence-level features or word-level features. Adding either set of features results in a statistically significant improvement by 1.3% and 1.4%, respectively. When both levels are added, the gain in performance by 2.1% is even better. Comparing this number with the simplest feature set we used (i.e. bag of words - *not lemmatized* in Table 5) we have an increase by 10.3%.

Feature Sets Using No Polarity Information					
Features	*Class*	*Rec.*	*Prec.*	*F.*	*Acc.*
bag-of-words (*not lemmatized*)	+	72.9	65.5	69.0	67.2
	−	61.5	69.5	65.2	
bag-of-words	+	63.2	71.0	66.8	68.6
	−	74.1	66.8	70.3	
bag-of-words + linguistic word-level features	+	68.2	75.8	71.7	**73.1**
	−	78.8	71.0	74.4	
Feature Sets Using Polarity Information					
Features	*Class*	*Rec.*	*Prec.*	*F.*	*Acc.*
prior-polarity	+	68.0	71.5	69.7	70.4
	−	72.9	69.6	71.1	
prior-polarity + linguistic sentence-level features	+	70.9	75.2	72.9	73.8
	−	76.6	72.6	74.5	
prior-polarity + bag of words	+	74.0	76.1	75.0	75.4
	−	76.8	74.8	75.7	
prior-polarity + bag of words + linguistic word-level features	+	74.6	78.0	76.2	76.7
	−	78.9	75.7	77.2	
prior-polarity + bag of words + linguistic sentence-level features	+	74.9	77.9	76.3	76.8
	−	78.7	75.9	77.2	
prior-polarity + bag of words + all linguistic features	+	75.2	78.8	76.9	**77.5**
	−	79.7	76.3	78.0	

Table 5: Performance of Different Feature Sets.

Other Levels of Representation

We tested two alternative types of feature representations: *bigrams* and *tree-kernels*. However, all these features did not improve the performance of our baseline. Bigrams can be a means of capturing more local structure and are known to improve the quality of polarity classification at document level (Ng, Dasgupta, and Arifin 2006). We presume that this representation does not work at sentence level due to

[5]The features from those categories which positively contributed to the overall performance fire in every sentence. However, we only found a negation in 19% of the sentences.

[6]Please note that we could not use this feature selection method for the word-level features since it would have been computationally prohibitive.

Bare Polarity Features
number of positive/negative expressions
number of strong positive/negative expressions
Linguistic Features
number of positive/negative adjectives
number of negative verbs
number of positive/negative expressions with depth level I
number of positive/negative expressions with depth level II
is main predicate a positive expression?
number of negative expressions in predicate phrase
number of positive/negative expressions in main clause
number of positive expressions modified by positive/neutral expressions

Table 6: Best Sentence-Level Features According to Best-First Forward Selection.

the greater data sparseness. The potential of tree-kernels is that structural features are automatically (implicitly) computed and do not have to be explicitly defined. We used SVMLight-TK (Moschitti 2006)[7] for our experiments. The reason for the lacking improvement might be due to too much irrelevant information encoded in syntax trees.

The results of these two experiments may be opposed to the findings in (Kudo and Matsumoto 2005), but we assume that this is due to the different settings of the experiments[8].

Conclusion

In this paper, we have shown that the baseline performance of polarity classifiers of news text at sentence level using bag of words can be significantly improved by applying both linguistic features and polarity information. Unlike polarity classification at document level, just using bag of words produces a fairly low performance.

Though adding prior polarity information to bag of words already gives a significant boost to the baseline performance at sentence level, adding linguistic features can increase this performance even further significantly. In total, our baseline is improved by up to 10.3%. We also showed that in the absence of any polar information, simple and domain-independent structural features can already improve the performance of bag-of-word feature sets by approximately 6%.

Acknowledgements

We would like to thank Josef Ruppenhofer for interesting discussions. Michael Wiegand was funded by the German research council DFG through the International Research Training Group between Saarland University and University of Edinburgh.

References

Charniak, E. 2000. A Maximum-Entropy-Inspired Parser. In *Proc. of NAACL*.

[7]We always tested within the hybrid mode which combines the tree-kernel with the standard bag-of-words features.

[8](Kudo and Matsumoto 2005) report results on Japanese text, they use twice as much data and consider a closed domain (reviews for Personal Handyphone System) presumably comprising more repetitive language than the multi-topic MPQA news corpus.

Choi, Y., and Cardie, C. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proc. of EMNLP*.

Gamon, M. 2004. Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis. In *Proc. of COLING*.

Joachims, T. 1999. Making Large-Scale SVM Learning Practical. In Scholkopf, B.; Burges, C.; and Smola, A., eds., *Advances in Kernel Methods - Support Vector Learning*. MIT Press.

Kudo, T., and Matsumoto, Y. 2005. A Boosting Algorithm for Classification of Semi-Structured Text. In *Proc. of EMNLP*.

Matsumoto, S.; Takamura, H.; and Okumura, M. 2005. Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees. In *Proc. of PAKDD*.

McDonald, R.; Hannan, K.; Neylon, T.; Wells, M.; and Reynar, J. 2007. Structured Models for Fine-to-Coarse Sentiment Analysis. In *Proc. of ACL*.

Meena, A., and Prabhabkar, T. 2007. Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis. In *Proc. of ECIR*.

Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; and Miller, K. 1990. Introduction to Wordnet: An On-line Lexical Database. *International Journal of Lexicography*.

Moilanen, K., and Pulman, S. 2007. Sentiment Construction. In *Proc. of RANLP*.

Moschitti, A. 2006. Making Tree Kernels Practical for Natural Language Learning. In *Proc. of EACL*.

Nasukawa, T., and Yi, J. 2003. Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In *Proc. of K-CAP*.

Ng, V.; Dasgupta, S.; and Arifin, S. M. N. 2006. Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. In *Proc. of ACL*.

Pang, B., and Lee, L. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Miminum Cuts. In *Proc. of ACL*.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proc. of EMNLP*.

Scott, S., and Matwin, S. 1998. Text Classification Using WordNet Hypernyms. In *Proc. of the Workshop on Usage of WordNet in NLP Systems (COLING-ACL)*.

Wiebe, J.; Wilson, T.; and Cardie, C. 2003. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*.

Wilson, T.; Wiebe, J.; and Hoffmann, P. 2005. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proc. of HLT/EMNLP*.

Yang, Y., and Pederson, J. 1997. A Comparative Study on Feature Selection in Text Categorization. In *Proc. of ICML*.