Experiments of Opinion Analysis On Two Corpora MPQA and NTCIR-6

Yaoyong Li, Kalina Bontcheva and Hamish Cunningham
Department of Computer Science, The University of Sheffield
211 Portobello Street, Sheffield, S1 4DP, UK
{yaoyong, kalina, hamish}@dcs.shef.ac.uk

Abstract

This paper describes the algorithms and linguistic features used in our participating system for the opinion analysis pilot task at NTCIR-6. It presents and discusses the results of our system on the opinion analysis task. It also presents our experiments of opinion analysis on the two corpora MPQA and NTCIR-6, by using our learning based system. Our system was base on the SVM learning. It achieved state of the art results on the MPQA corpus for the two problems, opinionated sentence recognition and opinion holder extraction. The results using the NTCIR-6 English corpus for both training and testing are certainly among the first ones. Our results on the opinionated sentence recognition sub-task of the NTCIR-6 were encouraging. The results on the English evaluation of the NTCIR-6 opinion analysis task were obtained from the models learned from the MPQA corpus. The lower results on the NTCIR-6 opinion holder extraction subtask, in comparison with those using each corpus for both training and testing, may possibly show that there exist substantial differences between the MPQA corpus and the NTCIR-6 English corpus.

1 Introduction

With the rapid expansion of internet into almost every area of the society, we are able to express opinions about many interesting and/or useful topics (such as commercial products and cultural and political topics) through internet. Actually there are a huge amount of opinions available on-line (such as product reviews from customers, personal comments of on-line news, and on-line community forum and Blogs), and is increasing rapidly. Most of the on-line opinions are the free texts in different languages. Opinion analysis or mining would be very helpful for improving services in business and government and other areas. Manual analysis of the on-line free-text opinions is very time-consuming and slow. Therefore automatic opinion analysis of free-text opinions is desired.

Opinion analysis is an active research topic in data mining and knowledge discovery. So far mainly two strategies have been pursued in opinion analysis. One is based on the linguistic knowledge about subjective language. The first step of this strategy is to obtain a subjective word list by collecting manually or automatically from corpus or/and thesaurus. The word list then is used for performing a variety of opinion analysis tasks on text. Two exemplary works using the strategy were [12] and [4] on English and Chinese, respectively. Another strategy is applying machine learning method, incorporating with linguistic features, to some concrete opinion analysis task. [1] used the conditional random fields for extracting opinion sources, which also showed that both the linguistic feature deduced from a subjective word list and the syntactic features were useful. [9] used the support vector machines (SVM) for recognising opinionated sentences or documents. [8] experimented the SVM and Naive Bayes algorithms for identifying the perspective of an opinionated document. Interestingly [3] explored one different method of combining the machine learning and linguistics for opinion analysis. It first collected the opinion words and the opinion related frames from the FrameNet corpus, then used a maximum entropy based method for labeling the semantic roles of opinion related frames in a sentence, and finally recognised the opinion holder and topic by mapping semantic roles to them.

This paper describes our participating system for the opinion analysis pilot task at NTCIR-6 (please refer to the overview paper of this task in this proceedings for the detailed descriptions of the task). Our system was based on the SVM, a supervised learning algorithm which achieved state of the art results on many natural language processing problems. Section 2 describes the SVM learning for opinion holder extraction and opinionated sentence recognition, two mandatory subtasks of the opinion analysis task at NTCIR-6. Section 3 presents the experimental results on another opinion analysis corpus MPQA (available from http://www.cs.pitt.edu/mpqa/) which we obtained before doing the experiments on the NTCIR-6 corpus. Section 4 discusses our results of the NTCIR-6 opinion analysis task. Section 5 presents some results using the NTCIR-6 English corpus for both training and testing. Section 6 concludes.

2 Linguistic features and learning algorithm

We used the open-source ANNIE system, which is part of GATE [2] (available from http://www.gate.ac.uk/) to process text for obtaining a variety of linguistic features. This produced a number of linguistic features. The features include token form, capitalization information of words, token kind, lemma, part-of-speech (POS) tag, semantic classes from gazetteer lists, and named entity type according to ANNIE's rule-based recognizer. Those features was used as input for the SVM learning. Different types of linguistic features were used for different opinion analysis tasks, which we will discuss later.

We used the binary classification SVM algorithm for opinion analysis. For a binary classification problem, the SVM tries to find out a hyperplane in the feature space which separates the positive training examples from the negative ones. The SVM classification hyperplane also has the maximal margin to the training examples. This makes the SVM classification model have good generalisation capability. In the follows we describe our system in detail for the two opinion analysis tasks, opinionated sentence recognition and opinion holder extraction, respectively.

The *opinionated sentence recognition* task is about classifying a sentence as opinionated or not opinionated. Hence it is a sentence classification problem with only two labels, opinionated or not. Hence the binary classification SVM learning can be applied to it directly. For the linguistic features, we used the unigram of tokens with two attributes of token, lemma and POS. In another word, two tokens in a sentence were regarded as the same term if and only if the two tokens had the same lemma and POS. We constructed one feature vector for each sentence with each unique unigram term of the sentence as one component and the commonly used $tf * idf$ value of the term as the value of the component, where tf is the frequency of the term in the sentence and idf is the inverse of the term frequency in all the sentences in training data. The feature vector was normalised and then fed into the SVM algorithm for learning an SVM model as well as applying the model to test instance.

The *opinion holder extraction* is extracting opinion holder(s) from sentence. An opinion holder may be just one word or can span several words. We considered the opinion holder extraction as a special case of the general information extraction (IE) problem, where opinion holder was the only information entity to be extracted. Our SVM based IE system, which was described in [5], was used for the opinion holder extraction. Our system assumed that an opinion holder consisted of a single token or several consecutive tokens. It learned two SVM models respectively for identifying the start token and end token of an opinion holder. If an opinion holder has only one token, the single token was regarded as both the start and end token of the opinion holder.

Thus our system converted the problem of opinion holder extraction into two binary classification problems with respect to the start and end tokens of opinion holder, respectively. A binary SVM model was learned for each of the two problem. In the application of the SVM models, after identifying the start and end tokens in the testing text, some post-processing procedure was applied to obtain opinion holders from the recognised start and end tokens. Basically the post-processing was applied to the situations where two or more entities overlapped with each other – it selected the entity with the highest confidence among the overlapped entities.

Note that our system did not use the standard SVM algorithm. Instead it used the uneven margins SVM, which often achieved better F-measure score than the standard SVM on the imbalanced data where the negative example outnumbered the positive examples. See [7] for detailed description of the uneven margins SVM. Our experiments showed that the uneven margins SVM obtained better results than the standard SVM on document classification and information extraction (see [7] and [5]).

As for the linguistic features used for the opinion holder extraction, we used the features obtained from the ANNIE, such as the token's lemma, POS tag, morphological features of word and entity types as listed above. Moreover, we also experimented with the features obtained from a subjective word list and the dependent tree, which were similar with those used in [1]. In detail, we used a subjective word list included in the opinion analysis system *OpinionFinder* (available from http://www.cs.pitt.edu/mpqa/opinionfinderrelease/). The word list contained 636 subjective words with the POS tags from two separated sources (see [11]). In order to obtain the syntactic features, the text was first processed by the parser Minipar (available at http://www.cs.ualberta.ca/ lindek/minipar.htm) to obtain the syntactic dependent relations between words. Then syntactic chunks were identified and grouped from the dependent tree, each of which consisted of a head word and all the child words of the head word. Finally the following four types of features were deduced for each word w and used in our system.

- the syntactic role of the word t in the chunk where w is a child word;
- whether w's parent chunk has an opinion word;

- whether the chunk in which the word t is the head word is an argument of its parent chunk;
- whether t has at least one child word and all its child words are in one (right or left) side of t in the sentence.

In the following section we will see that those lexical and syntactic features are indeed useful for opinion holder extraction.

3 Experiments on MPQA corpus

The first set of experiments on opinion analysis we carried out were on the Multi-Perspective Question Answering (MPQA) corpus (available at http://www.cs.pitt.edu/mpqa/databaserelease/) [13] , partly for the preparation of our system for the opinion analysis task at NTCIR-6. The MPQA corpus has been used as experimental data by many previous researches of opinion analysis. The corpus consists of 535 news articles from 187 different foreign and U.S. news sources. The corpus was manually annotated with a variety of subjective information, such as subjective expression, objective speech event and agent. The annotations were at expression (subsentence) level. The subjective annotations at sentence level can be derived from the low level annotations.

Since we wanted to run experiments at sentence level, we adopted the method described in [10] to obtained subjective sentence from the original MPQA annotations, which basically considered a sentence as subjective if the sentence contained one subjective annotation and the intensity of the annotation was not low. Then we defined the opinion holder(s) as the agent(s) in a subjective sentence. Once we had the opinionated sentences and opinion holders, we can evaluate our sentence level opinion analysis system on the corpus. In the experiments described in this section, 360 documents were selected from the MPQA corpus as training data and the other 175 documents were used for testing. The results presented in this section were obtained by using the evaluation tools of the GATE, which compared the annotations created by the system with the golden standard one and outputted the overall F-measures and the results per document.

Table 1 presents the results of our opinionated sentence recognition experiments on the MPQA corpus. Our results are quite similar with the previous results on the same corpus presented in e.g. [10] and [9]. Note that our experiments used only the lemma and POS tags of tokens in a sentence, which were less and simpler than the features used in the previous works. The table also presents the results of the standard SVM and the uneven margins SVM. We can see that uneven margins SVM resulted in somewhat lower precision but much higher recall and higher F_1.

Table 1. The F-measure results of opinionated sentence recognition on MPQA corpus by using the standard SVM ($\tau = 1.0$) and the uneven margins SVM ($\tau = 0.5$), respectively. τ is the uneven margins parameter.

τ	Precision	Recall	F_1
1.0	0.754	0.772	0.763
0.5	0.678	0.914	0.779

Table 2 presents the results of opinion holder extraction on the MPQA corpus. Our results are comparable to the previous results on the same corpus presented in [1]. Our experiment used the lemma and POS tag of the token and the subjective word feature and syntactic features, as described in Section 2. The previous work in [1] used more linguistic features than ours.

Table 2. The F-measure results of opinion holder extraction on MPQA corpus by using the uneven margins SVM ($\tau = 0.5$).

Precision	Recall	F_1
0.676	0.560	0.613

Table 3 compares the results between different combination of linguistic features. We can see that the syntactic features derived from dependent tree were useful for identifying opinion holder. The capitalisation information and named entity types about token were not as helpful as for the general information extraction (see e.g. [5]). The feature from a subjective word list was somewhat useful on its own.

Table 3. Comparisons of the precision (P), recall (R) and F_1 for different combinations of the token's linguistic features, such as lemma (L), POS (P), capitalisation information (C), named entity information (E), subjective word list (Sub), and syntactic features (Syn).

	L+P	L+P +C	L+P +E	L+P +Sub	L+P +Sub+Syn
P	0.604	0.607	0.607	0.624	0.676
R	0.552	0.550	0.554	0.548	0.560
F_1	0.577	0.577	0.579	0.583	0.613

4 Results on the opinion analysis task at NTCIR-6

We has participated the opinion analysis pilot task at NTCIR-6. We applied our opinion analysis system described above to the English and Chinese corpora of the task. In the follows we explain the training data we used for our system and discuss the results of our submitted runs.

First note that all the results presented in this section were obtained by using the evaluation scripts created by the task organisers. All the results were released by the organisers, except the opinion holder extraction results of the run Gate-2 which we obtained by using the evaluation script released by the task organisers of English evaluation.

For the *Chinese corpus* our system participated in two sub-tasks, opinionated sentence recognition and opinion holder extraction. The only training data our systems used were the annotated Chinese documents of the four topics released by the organisers before releasing the formal evaluation data. We submitted two runs. The run GATE-1 used the so-called lenient annotations on which two of the three annotators agreed. Another run GATE-2 used the strict annotations on which all the three annotators agreed.

We first segmented every Chinese sentence into a sequence of Chinese words using our Chinese word segmenting system [6]. We then regarded Chinese words as being equivalent to the tokens in English and implemented the learning algorithm described in Section 2 on the Chinese data for the two sub-tasks. However, it is worth noting that, unlike the English data for which we used other linguistic features besides the token form, we used the word form as the only feature for the Chinese evaluation. For the opinionated sentence recognition we formed a feature vector for one sentence of which each component corresponded to one unique Chinese word and the value of the component was $tf * idf$. We converted the opinion holder extraction into two binary classification problems – recognising the first and last Chinese words of an opinion holder.

Table 4 presents our results on the Chinese corpus. The Gate-1'1 results on opinionated sentence recognition were among the best ones. But the results on the opinion holder extraction were quite low in comparison to the best one. However, it is worth noting that the results from different participants may not be comparable to each other, because they may use different material for developing or training their systems. Comparing our two runs, the Gate-1 achieved higher F_1 than the Gate-2 for the lenient measure, while the Gate-2 obtained better F_1 than the Gate-1 for the strict measure. That was consistent with the training procedures of the two runs — as said above, the Gate-1 was trained by lenient annotations and the Gate-2 was trained by using the strict annotations.

For the *English corpus*, the whole MPQA corpus was used for training our system. We also submitted two runs. The Gate-1 run was the direct application of the learned models from the MPQA corpus to the NTCIR-6 English corpus. In the Gate-2 run we post-processed the opinion holder results of the Gate-1 run. The post-processing procedure had two steps. First we found out all the noun chunks (NPs) of the sentences using an NP recogniser. Then we matched the opinion holders identified by our system to the noun chunks. If the text of an opinion holder was a part of an NP, we replaced the opinion holder with the NP as opinion holder, as we believed that an opinion holder should be an NP. However if an opinion holder contained an NP already, we did not made any change on it.

Table 5 presents the results of our two submitted runs on the English corpus. As said earlier, the opinion holder extraction results of the Gate-2 were obtained by us using the evaluation script released by the task organisers[1]. The Gate-2 obtained slightly better results than the Gate-1 for the opinion holder extraction, showing that our post-processing using NP chunks was helpful.

In comparison with the results on the MPQA corpus discussed in Section 3, the results on the NTCIR-6 corpus were quite low, especially for the opinion holder extraction. Since we used the same learning algorithm and the same linguistic features and applied the models learned from the MPQA corpus, we suspected that the lower F-measures on the NTCIR-6 English corpus may be due to the differences between the two corpora, including the different annotation schemes and different sources of documents. We have done the experiments by using the NTCIR-6 English corpus for both training and testing. We will present those experimental results in next section.

We also participated the relevance judgment subtask of the English corpus, which made a decision for each sentence of a document whether or not the sentence was relevant to the topic that the document belonged to. Our method for the relevance judgment was quite simple — we first constructed a normalised $tf * idf$ feature vector for each topic description, then assessed the correlation between the sentence and the topic by computing the inner product between the topic's feature vector and the sentence's feature vector, and finally made a judgment by comparing the inner product against a pre-defined threshold (which was simply set as 0.02 in our submitted runs) — the sen-

[1] While running the evaluation script on our submission Gate-2, the script asked us 69 questions for our judgments on the matches of the extracted opinion holders and the sentences. We did not make any judgment and answer all the questions with "N", meaning that we may missed some correct matches of those the evaluation script asked us for manual judgment. Hence, the results for the opinion holder extraction of the Gate-2 listed in this paper may be a bit lower than the true ones.

Table 4. Results on the NTCIR-6 Chinese corpus of our two submitted runs for the opinionated sentence recognition (OS) and opinion holder extraction (OP) sub-tasks.

Sub-tasks	Runs	Lenient			Strict		
		P	R	F_1	P	R	F_1
OS	Gate-1	0.643	0.933	0.762	0.253	0.979	0.402
	Gate-2	0.746	0.591	0.659	0.330	0.696	0.448
OH	Gate-1	0.427	0.154	0.227	0.419	0.156	0.227
	Gate-2	0.373	0.046	0.082	0.368	0.052	0.091

Table 5. Results on the NTCIR-6 English corpus of our two submitted runs for the opinionated sentence recognition (OS) and opinion holder extraction (OP) sub-tasks.

Sub-tasks	Runs	Lenient			Strict		
		P	R	F_1	P	R	F_1
OS	Gate-1	0.324	0.905	0.477	0.070	0.940	0.130
	Gate-2	0.324	0.905	0.477	0.070	0.940	0.130
OH	Gate-1	0.121	0.349	0.180	0.029	0.398	0.055
	Gate-2	0.125	0.359	0.185	0.031	0.414	0.057

tence was regarded as being relevant to the topic if the inner product between their feature vectors was bigger than the threshold. Table 6 presents the results of the run Gate-1 (another run Gate-2 had the same results). The results were not particularly good. As we applied a simple method, the results can certainly be improved by using more complicated techniques and features.

Table 6. The results of the sentence relevance sub-task on the NTCIR-6 English corpus for the run Gate-1.

	Precision	Recall	F_1
Lenient	0.286	0.632	0.393
Strict	0.112	0.579	0.188

5 Using the NTCIR-6 English corpus for both training and testing

As discussed earlier, we applied the models learned from the MPQA corpus to the NTCIR-6 English corpus. The results obtained so far ware quite low results in comparison with those on the MPQA corpus. It looked that the models learned from one corpus were not applicable very well to another. We guessed that it may be due to the differences between the two corpora, such as different annotation details and different sources of documents. That motivated us to carry out some more experiments by using the NTCIR-6 corpus for both training and testing.

In order to make the results comparable with those on the MPQA corpus, we computed the results of the experiments by using the evaluation tool of the GATE, which we used for obtaining the results of the MPQA corpus.

To use the GATE evaluation tool, we had to put the opinion holder annotations, which were in a separated file from the text in the original NTCIR-6 English corpus, into the documents using the so-call inline markups. In other words, the original English corpus listed all the opinion holder(s) of a sentence in a separated file and we had to tag each sentence for its opinion holder(s) in the annotation file. We have done it only for one of the three annotation files (respectively from three annotators), the annotation file “OAT2006_formalrun_english_a1.csv”. What we did was to scan a sentence from left to right to look for the match between one opinion holder and the text of the sentence. Once we found out a match, we added one XML markup tag around the text to tag it as one opinion holder.

When we were doing this kind of tagging, we found that there were 212 opinion holders (among the 2355 opinion holders) in the file “OAT2006_formalrun_english_a1.csv” which had no match within the corresponding sentences. We had checked those non-matches and found that some cases were caused by the typos in the annotation file or in the original text, and other cases were due to the fact that the opinion holder in the annotation file was a short or summarisation of the opinion holder text in the sentence, which was allowed according to the annotation scheme. For each of the non-match cases we made some necessary changes on the annotation file so that the opinion holder in the annotation file was a piece of the text from the document. We denoted the annotation file with those changes as "OAT2006_formalrun_english_a1_changes.csv", which was used as golden standard in all the experiments presented in this section.

First it was interesting to see how the results of our submitted run became when using the GATE evaluation tool and the revised annotation file "OAT2006_formalrun_english_a1_changes.csv". Table 7 presents the results of one of our submitted run Gate-1 for the English corpus, using the GATE evaluation tool and the revised annotation file. In comparison with the results computed by the NTCIR-6 evaluation scripts, the results of the opinionated sentence recognition became lower, but the results of the opinion holder extraction was a slightly higher, in particular for the precision.

Table 7. The F-measure results of the run Gate-1 computed by using the GATE evaluation tool and the revised annotation file from the original annotation file "OAT2006_formalrun_english_a1.csv". OS stands for opinionated sentence recognition sub-task and OH is for opinion holder extraction.

Sub-tasks	Precision	Recall	F_1
OS	0.293	0.496	0.323
OH	0.175	0.314	0.183

In our main experiments we selected 300 documents from the NTCIR-6 English corpus as training data and other 139 documents as testing data. All the documents had the annotations from the revised annotation file.

Table 8 presents the results of our experiments using the NTCIR-6 corpus for both training and testing. Comparing with the results using the MPQA corpus as training data and the NTCIR-6 corpus as testing data, we obtained much improved results by using the NTCIR-6 corpus for training and testing, showing that there really exist differences between the two corpora, both of which were designed and created for opinion analysis. The differences are in the annotation details as well the sources of the documents. The results listed in Table 8 were still worse than those using MPQA corpus for training and testing (presented in Section 3), probably showing that the NTCIR-6 English corpus is harder than the MPQA corpus for a machine learning based opinion analysis system.

Table 8. The results using the NTCIR-6 English corpus for both training and testing. The 300 documents were selected as training data and other 139 documents were as testing data. OS stands for opinionated sentence recognition sub-task and OH is for opinion holder extraction.

Sub-tasks	Precision	Recall	F_1
OS	0.648	0.610	0.628
OH	0.489	0.346	0.405

6 Conclusions

This paper describes the algorithms and linguistic features used in our participating system for the opinion analysis pilot task at NTCIR-6. It presents and discusses the results of our system on the English evaluation as well the Chinese evaluation of the task. We used the models learned from the MPQA corpus for the English evaluation on the NTCIR-6 English corpus. The results we obtained were quite low in comparison with those using one corpus for both training and testing, possibly meaning that there exist substantial differences between the two corpora. Our results of the opinionated sentence recognition on the Chinese evaluation were encouraging. But the results of opinion holder extraction were also low, possibly because of relatively small number of training data used.

The paper also presents the experiments on the two opinion analysis corpora, the MPQA and the NTCIR-6 English corpora, by using our SVM based learning system for opinion analysis. The results on the MPQA corpus were promising. But the results on the NTCIR-6 English corpus was not as good as those on the MPQA corpus, showing that the former was harder than the latter for a machine learning based opinion analysis system.

In all, the results of our system on the opinionated sentence recognition were encouraging, but those on the opinion holder extraction were not satisfied. Our system treated the opinion holder extraction as an information extraction problem and the results we obtained were not as good as those on information extraction. In comparison with the general information extraction, opinion holder extraction has some unique features, which we should take into account in order to achieve good performance. Hence one of our future works would investigate the unique features of the opinion holder extraction.

7 Acknowledgements

This work is supported by the EU-funded SEKT (IST-2004-506826) and KnowledgeWeb (IST-2004-507482) projects.

References

[1] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In *Proceedings of the 2005 Human Language Technology Conference / Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-05)*, 2005.

[2] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.

[3] S. Kim and E. Hovy. Extracting Opinions Expressed in Online News Media Text with Opinion Holders and Topics. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text at the joint COLING-ACL conference*, 2006.

[4] L.-W. Ku, Y.-T. Liang, and H.-H. Chen. Opinion extraction, summarization and tracking in news and blog Corpora. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, 2006.

[5] Y. Li, K. Bontcheva, and H. Cunningham. SVM Based Learning System For Information Extraction. In *Proceedings of Sheffield Machine Learning Workshop*, Lecture Notes in Computer Science. Springer Verlag, 2005.

[6] Y. Li, C. Miao, K. Bontcheva, and H. Cunningham. Perceptron Learning for Chinese Word Segmentation. In *Proceedings of Fourth SIGHAN Workshop on Chinese Language processing (Sighan-05)*, pages 154–157, Korea, 2005.

[7] Y. Li and J. Shawe-Taylor. The SVM with Uneven Margins and Chinese Document Categorization. In *Proceedings of The 17th Pacific Asia Conference on Language, Information and Computation (PACLIC17)*, Singapore, Oct. 2003.

[8] W. Lin, T. Wilson, J. Wiebe, and A. Hauptmann. Which Side are You on? Identifying Perspectives at the Document and Sentence Levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116, 2006.

[9] E. Riloff, S. Patwardhan, and J. Wiebe. Feature Subsumption for Opinion Analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-06)*, 2006.

[10] E. Riloff, J. Wiebe, and T. Wilson. Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In *Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-03)*, 2003.

[11] J. Wiebe, E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff, and T. Wilson. NRRC Summer Workshop on Multiple-Perspective Question Answering Final Report. Technical report, 2002.

[12] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin. Learning Subjective Language. *Computational Linguistics*, 30(3):277–308, 2004.

[13] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation (formerly Computers and the Humanities)*, 39(2-3):165–210, 2005.