Cornell System Description for the NTCIR-6 Opinion Task

Eric Breck Yejin Choi Veselin Stoyanov Claire Cardie
Department of Computer Science
Cornell University
Ithaca, NY USA
{ebreck,ychoi,ves,cardie}@cs.cornell.edu

Abstract

We present our opinion analysis system for English that was used in the Opinion Analysis Pilot Task at NTCIR-6. Our goal in developing the system was to use, as much as possible, components and features from our previous work in this area.

Keywords: *Opinion Extraction, Machine Learning.*

1 Introduction

Our goal in the NTCIR-6 Opinion Analysis Pilot Task was, as much as posible, to rely on the natural language learning methods, components, and features developed in our previous work in this area of research. In particular, we have proposed to treat opinion analysis as a standard information extraction task [2, 3, 1]. The traditional goal of information extraction (IE) systems has been to extract information about events, including the participants of the events, from accounts of the events in text (e.g. newspapers). An IE system that extracts information about corporate acquisitions, for example, might identify the company that is doing the acquiring, the company that is being acquired, the date of the acquisition, the terms of the agreement, etc. Similarly, an IE system that extracts information about natural disasters might determine the type of disaster (e.g. a hurricane), the number of victims, the amount of damage to physical property, the date of the event, the locations affected, etc. In previous work, we hypothesized that IE techniques would be well-suited for opinion analysis: namely, statements of opinion can be viewed as a kind of speech event with the source as the agent.

As a result, we have investigated the application of sequence tagging methods (e.g. Lafferty et al. (2001)) and extraction pattern learning (e.g. Riloff (1996)) to the problem of opinion analysis (including opinion holder/source identification) since both natural language learning paradigms have been successful for a variety of IE tasks.

In the sections below, we provide a more detailed description of the methods, features, and training data employed for the NTCIR-6 Opinion Analysis Pilot. Of the four opinion analysis subtasks, we participated only in three:

1. opinionated sentence judgment (Section 2),
2. opinion holder extraction (Section 3), and
3. polarity judgment (Section 4).

We worked only on the English NTCIR-6 opinion data. Section 5 presents and briefly discusses our results.

2 Opinionated Sentence Judgment

Our method for judging whether or not a sentence is "opinionated" is based largely on our work in identifying opinion expressions in context [1]. Unlike that work, however, we make all decisions at the sentence level rather than at the token level. More specifically, we consider two types of opinion expressions as defined in Wiebe et al. 2005 and highlighted in the examples below:

S1: Minister Vedrine **criticized** the White House reaction.

S2: 17 persons were killed by sharpshooters **faithful to** the president.

S3: Tsvangirai **said** the election result was "*illegitimate*" and a clear case of "*highway robbery*".

S4: Criminals have been *preying* on Korean travelers in China.

Direct subjective expressions (DSEs), shown in boldface above, are spans of text that explicitly express an attitude or opinion. "Criticized" and "faithful to" (examples S1 and S2), for example, directly denote negative and positive attitudes towards the

"White House reaction" and "the president", respectively. Speech events like "said" in example S3 can be DSEs if what is being expressed has subjective content. In contrast, *expressive subjective elements* (ESEs), shown in italics in the examples, are spans of text that indicate, merely by the specific choice of words, a degree of subjectivity on the part of the speaker. The phrases "illegitimate" and "highway robbery", for example, indirectly relay Tsvangirai's negative opinion of "the election result" (example S3), and the use of "preying on" (instead of, say, "mugging") indicates the writer's sympathy for the Korean travelers in example S4.

For the NTCIR-6 opinion analysis tasks, we focus on the identification of DSEs and ESEs.

Opinion Expression Classifiers. In recent work [1], we employed a conditional random field approach to the identification of DSEs and ESEs. The NTCIR-6 task, however, requires sentence-level decisions rather than expression-level opinion recognition. As a result, we train three support vector machine (SVM) classifiers[1] to determine whether the sentence contains:

1. a DSE
2. a DSE or an ESE
3. a DSE or an ESE of high or medium intensity

We chose SVMs because they have been very successful in text categorization tasks similar to the sentence-level classification task we tackle here.

The same set of features is used to train each classifier. In addition, we use sentence-level versions of a subset of the features from Breck et al. (2007). For pedagogical reasons, we present the features as categorically valued below, but in our model we encode all features in binary (i.e. a feature (f, v) is 1 for a token t if $f(t) = v$, and 0 otherwise):

words: all words in the sentence. These are encoded into about 18,000 binary features (i.e. the vocabulary size).

semantic class: all *WordNet* synsets that are hypernyms of any word in the sentence that appears in the WordNet hierarchy [4]. This is encoded into about 30,000 binary features, many of which may be 1 for a given token.

Levin verb classes: for all verbs in the sentence, this feature indicates which of Levin's categories of English verbs it falls into [6].

FrameNet: the categories of all nouns and verbs in Framenet[2].

[1] We use SVMlight `http://svmlight.joachims.org/`).

[2] `http://www.icsi.berkeley.edu/~framenet/`

subjectivity indicators: Wilson et al. 2005 identify a set of clues as being either strong or weak cues to the subjectivity of a clause or sentence. We identify any sequence of tokens included on this list, and then define a feature that returns the value '-' if the sentence contains none of recognized clues, or **strong** or **weak** if the sentence contains a recognized clue of that strength. This clue is encoded into two binary features (the '-' case is not encoded).

Training and Model Selection. The classifiers were trained on all 535 documents in the MPQA opinion corpus[3], which contains newswire documents with a variety of subjectivity annotations [10]. In particular, all DSEs and ESEs and their strengths/intensities have been manually identified. These were used to determine the target value for each sentence in the corpus for each of the three classifiers.

For the NTCIR-6 submissions, we employed the third classifier above: a sentence is classified as "opinionated" if it contains a DSE or an ESE of high or medium intensity. We chose this classifier upon examination of its output on the NTCIR sample data: classifier 1 did not identify enough opinionated sentences, and classifier 2 identified too many sentences as opinionated.

3 Polarity Judgment

Polarity judgment was a new task for us. Using the same feature set as above for opinonated sentence classification, we also train two SVM classifiers to determine whether the sentence contains:

1. an expression of negative polarity
2. an expression of positive polarity

The classifiers were trained on all 535 documents from the MPQA corpus using the polarity attributes that are available for all DSEs and ESEs to determine the target class for each sentence.

To assign a polarity value to sentences at prediction time, we use the following heuristic:

1. if the opinion sentence classifier indicates that **no DSE or ESE is present**, return NEUTRAL polarity; else
2. if the negative polarity classifier indicates the **presence of a negative expression**, return NEGATIVE polarity; else
3. if the positive polarity classifier indicates the **presence of a positive expression**, return POSITIVE polarity; else

[3] Available at `http://www.cs.pitt.edu/mpqa/`.

4. if the opinion sentence classifier indicates that a DSE or ESE **is** present, return NEUTRAL polarity.

4 Recognizing Opinion Holders

In previous research, we identified opinion holders, i.e. direct and indirect sources of opinions, emotions, sentiments, and other *private states* that are expressed in text [3]. To illustrate the nature of this problem, consider the examples below:

S1: Taiwan-born voters favoring independence...

S2: According to the report, the human rights record in China is horrendous.

S3: International officers believe that the EU will prevail.

S4: International officers said US officials want the EU to prevail.

In S1, the phrase “Taiwan-born voters” is the direct (i.e., first-hand) source of the “favoring” sentiment. In S2, “the report” is the direct source of the opinion about China’s human rights record. In S3, “International officers” are the direct source of an opinion regarding the EU. The same phrase in S4, however, denotes an indirect (i.e., second-hand, third-hand, etc.) source of an opinion whose direct source is “US officials”.

In Choi et al. (2005), we viewed *source identification* — referred to as *opinion holder identification* in NTCIR-6 — as an information extraction task and tackled the problem using a hybrid approach that combines sequence tagging via graphical models and pattern matching techniques. In particular, we consider Conditional Random Fields [5] and a variation of AutoSlog [8]. While CRFs treat source identification as a token-level sequence tagging task, AutoSlog views the problem as a pattern-matching task, acquiring symbolic patterns that rely on both the syntax and lexical semantics of a sentence. Choi et al. (2005) hypothesized (correctly for the data set under consideration) that a combination of the two techniques would perform better than either one alone.

The CRF approach. We defined the problem of opinion source identification as a sequence tagging task via CRFs as follows. Given a sequence of tokens, $x = x_1x_2...x_n$, we need to generate a sequence of tags, or labels, $y = y_1y_2...y_n$. We define the set of possible label values as `'S'`, `'T'`, `'-'`, where `'S'` is the first token (or Start) of a source, `'T'` is a non-initial token (i.e. a conTinuation) of a source, and `'-'` is a token that is not part of any source.[4]

[4]This is equivalent to the IOB tagging scheme used in syntactic chunkers [7].

We used a large collection of syntactic, semantic, and orthographic lexical features, dependency parse features, and opinion recognition features. The features are described in detail in Choi et al. (2005).

Extraction pattern learning. We also learn patterns to extract opinion sources using a statistical adaptation of the AutoSlog IE learning algorithm. AutoSlog [8] is a supervised extraction pattern learner that takes a training corpus of texts and their associated answer keys as input. A set of heuristics looks at the context surrounding each answer and proposes a lexico-syntactic pattern to extract that answer from the text. The heuristics are not perfect, however, so the resulting set of patterns needs to be manually reviewed by a person. In order to build a fully automatic system that does not depend on manual review, we combined AutoSlog’s heuristics with statistics from the annotated training data to create a fully automatic supervised learner. Again, please see the original paper [3] for details.

The hybrid approach. The extraction patterns provide two kinds of information. First, extraction patterns indicate whether a particular word activates any source extraction pattern. For example, the word *“complained”* activates the pattern *“<subj> complained”* because it anchors the expression. Second, patterns indicate whether a word is extracted by any source pattern. For example, in the sentence “President Jacques Chirac frequently complained about France’s economy”, the words “President”, “Jacques”, and “Chirac” would all be extracted by the *“<subj> complained”* pattern. In the hybrid CRF+AutoSlog IE system, we use both types of information to create additional AutoSlog-based features for the CRF [3].

Our NTCIR-6 system employs the hybrid CRF+Autoslog-based approach to identify spans of text that correspond to opinion holders/sources. For the NTCIR-6 system, we use the same feature set but also incorporated word features for a window of [-4,+4] around each token (as a mechanism for possibly dealing with the new NTCIR-6 data).

Note that our system was trained to identify particular text spans associated with opinion holder entities; we typically rely on a source coreference resolution algorithm to link together all mentions of sources that refer to the same opinion holder [9]. The NTCIR-6 task, on the other hand, is to identify an opinion holder entity for each sentence.

Training. The CRF+AutoSlog IE system was trained on 360 documents from the MPQA opinion corpus, which contains opinion holder annotations for all subjective expressions in the texts.

		Prec	Recall	F1
Opinionated Sentences	strict	0.069	0.662	0.125
	lenient	0.317	0.651	0.427
Polarity	strict	0.010	0.135	0.018
	lenient	0.073	0.197	0.107
Opinion Holder	opinionated	0.163	0.346	0.222
	holder	0.041	0.392	0.074

Table 1. Results

Because our system was not trained to identify when the author of the document was the (implicit) opinion holder, we employ a simple post-processing step for this task: if a sentence does not contain an opinion holder, but does contain an ESE, then return AUTHOR as the opinion holder for this sentence.

5 Discussion of Results

Our results on the English Opinion Task are summarized in Table 1. In general, we suffered from the lack of training data for each of the tasks. This is probably the same for the other sites as well. The MPQA corpus, which we used for training, includes annotations for *all* subjective expressions in its texts, not just opinions. For example, all emotion expressions and argumentative expressions are annotated in addition to opinions. Finding a method to better match our available training data with the NTCIR-6 task data would be one way to increase our system's performance. Another option we might consider in the future is to augment our training data with unlabeled data that is more similar to the NTCIR-6 data.

6 Acknowledgments

This work was supported in part by the Advanced Research and Development Activity (ARDA), by NSF Grants IIS-0535099 and IIS-0208028, and by gifts from Google and the Xerox Foundation.

References

[1] E. Breck, Y. Choi, and C. Cardie. Identifying expressions of opinion in context. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[2] C. Cardie, J. Wiebe, T. Wilson, and D. Litman. Low-level annotations and summary representations of opinions for multi-perspective question answering. In M. Maybury, editor, *New Directions in Question Answering*. 2004.

[3] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 355–362, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

[4] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[5] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*. Morgan Kaufmann, San Francisco, CA, 2001.

[6] B. Levin. *English Verb Classes and Alternations*. University of Chicago Press, 1993.

[7] L. A. Ramshaw and M. P. Marcus. Text Chunking using Transformation-Based Learning. In *Proceedings of the Third Workshop on Very Large Corpora*, pages 82–94. Association for Computational Linguistics, 1995.

[8] E. Riloff. An empirical study of automated dictionary construction for information extraction in three domains. *Artificial Intelligence*, 85, 1996.

[9] V. Stoyanov and C. Cardie. Partially supervised coreference resolution for opinion summarization through structured rule learning. In *Proceedings of EMNLP*, 2006.

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

[11] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*. Vancouver, Canada, 2005.