

# Recognizing Arguing Subjectivity and Argument Tags

Alexander Conrad, Janyce Wiebe, and Rebecca Hwa

Department of Computer Science

University of Pittsburgh

Pittsburgh PA, 15260, USA

{conrada,wiebe,hwa}@cs.pitt.edu

## Abstract

In this paper we investigate two distinct tasks. The first task involves detecting arguing subjectivity, a type of linguistic subjectivity on which relatively little work has yet to be done. The second task involves labeling instances of arguing subjectivity with argument tags reflecting the conceptual argument being made. We refer to these two tasks collectively as “recognizing arguments”. We develop a new annotation scheme and assemble a new annotated corpus to support our learning efforts. Through our machine learning experiments, we investigate the utility of a sentiment lexicon, discourse parser, and semantic similarity measures with respect to recognizing arguments. By incorporating information gained from these resources, we outperform a unigram baseline by a significant margin. In addition, we explore a two-phase approach to recognizing arguments, with promising results.

## 1 Introduction

Subjectivity analysis is a thriving field within natural language processing. However, most research into subjectivity has focused on sentiment with respect to concrete things such as product debates (e.g., (Somasundaran and Wiebe, 2009), (Yu et al., 2011)) and movie reviews (e.g., (He et al., 2011), (Maas et al., 2011), (Pang and Lee, 2004)). Analysis often follows the opinion-target paradigm, in which expressions of sentiment are assessed with respect to the aspects of the object(s) under consideration towards which they are targeted. For example,

in the domain of smartphone reviews, aspects could include product features such as the keyboard, screen quality, and battery life.

Although sentiment analysis is interesting and important in its own right, this paradigm does not seem to be the best match for fine-grained analysis of ideological domains. While sentiment is also present in documents from this domain, previous work (Somasundaran and Wiebe, 2010) has found that arguing subjectivity, a less-studied form of subjectivity, is more frequently employed and more relevant for a robust assessment of ideological positions. Whereas sentiment conveys the polarity of a writer’s affect towards a topic, arguing subjectivity is a type of linguistic subjectivity in which a person expresses a controversial belief about what is true or what action ought to be taken regarding a central contentious issue (Somasundaran, 2010). For example, consider this sentence about health care reform:

- (1) Almost everyone knows that we must start holding insurance companies accountable and give Americans a greater sense of stability and security when it comes to their health care.

In a traditional opinion-target or sentiment-topic paradigm, perhaps this sentence could be labeled as containing a negative sentiment towards a topic representing “insurance companies”, or a positive sentiment towards a topic representing “stability” or “security”. However, a reader of a political editorial or blog may be more interested in *why* the author is negative to-

wards insurers, and *how* the author proposes to improve stability of the healthcare system. By focusing on the arguments conveyed through arguing subjectivity, we aim to capture these kind of conceptual reasons an author provides when arguing for his or her position.

However, identifying when someone is arguing is only part of the challenge. Since arguing subjectivity is used to express arguments, the next natural step is to identify the argument being expressed through each instance of arguing subjectivity. To illustrate this distinction, consider the following three example spans:

- (2) the bill is a job destroyer
- (3) President Obamas signature domestic policy will throw 100,000 people out of work come January
- (4) he can't expand his business because he can't afford the burden of Obamacare

Each of these examples contains arguing subjectivity, but more importantly, each expresses roughly the same idea, namely, that the recently-passed health care reform bill will cause economic harm. This latent, shared idea giving rise to each of the three spans is what we mean by “argument tag”.

However, although all three are related, example spans (2) and (3) are more similar than (4) in terms of the notions they convey: while the first two explicitly are concerned with the loss of jobs, the last focuses on business expansion and the economy as a whole. If we were to tag these three spans with respect to the argument that each is making, should they all receive the same tag, or should (4)'s tag be different?

To address these challenges, we propose in this work a new annotation scheme for identifying arguing subjectivity and a hierarchical model for organizing “argument tags”. In our hierarchical model, (4) would receive a different tag from (2) and (3), but because of the tags' relatedness all would share the same parent tag.

In addition to presenting this new scheme for labeling arguing subjectivity, we also explore sentiment, discourse, and distributional similar-

ity as tools to enhance identification and classification of arguing subjectivity. Finally, we also investigate splitting the arguing subjectivity detection task up into two distinct phases: identifying expressions of arguing subjectivity, and labelling each such expression with an appropriate argument tag.

Since no corpora annotated for arguing subjectivity yet exist, we gather and annotate a corpus of blog posts and op-eds about a controversial topic, namely, the recently-passed “ObamaCare” health care reform bill.

## 2 Annotation Scheme

We designed our annotation scheme with two goals in mind: identifying all spans of text which express arguing subjectivity, and labelling each such span with an argument tag. To address the first goal, our annotators manually identified and annotated spans of text containing arguing subjectivity using the GATE environment<sup>1</sup>. Annotators were instructed to identify spans of 1 sentence or less in which a writer “conveys a controversial private state concerning what she believes to be true or what action she believes should be taken” concerning the health care reform debate. To train our annotators to recognize arguing subjectivity, we performed several rounds of practice on a separate dataset. Between each round, our annotators met to discuss their annotations and resolve disagreements.

As a heuristic to help distinguish between borderline sentences, we advised our annotators to imagine disputants from each side writing the sentence in isolation. If a disputant from either side could conceivably write the sentence, then the sentence is likely objective. For example, statements of accepted facts and statistics generally fall into this category. However, if only one side could conceivably be the author of the sentence, it is highly likely that the sentence expresses a controversial belief relevant to the debate and thus should be labeled as subjective.

Next, the annotators labeled each arguing span with an argument tag. As illustrated in earlier examples, an argument tag represents a

---

<sup>1</sup><http://gate.ac.uk/>

controversial abstract belief expressed through arguing subjectivity. Since the meanings of many tags may be related, we organize these tags in a hierarchical “stance structure”. A stance structure is a tree-based data structure containing all of the argument tags associated with a particular debate, organizing those tags using “is-a” relationships. Our stance structure contains two levels of argument tags: upper-level “primary” argument tags and lower-level “secondary” tags. Each primary tag has one of the stances (either “pro” or “anti” in our case) as its parent, while each secondary tag has a primary tag as its parent<sup>2</sup>.

Political science “arguing dimension” approaches to debate framing analysis served, in part, as an inspiration for our stance structure (Baumgartner et al., 2008). Also, as illustrated in Section 1, this approach permits us additional flexibility, supporting classification at different levels of specificity depending on the task at hand and the amount of data available. We envision a future scenario in which a community of users collaboratively builds a stance structure to represent a new topic or debate, or in which analysts build a stance structure to categorize the issues expressed towards a proposed law, such as in the context of e-rulemaking (Cardie et al., 2008).

Because each stance contains a large number of argument tags, we back-off from each secondary argument tag to its primary argument parent for the classification experiments. We chose to do this in order to ensure that we have a sufficient amount of data with which to train the classifier.

### 3 Dataset

For this study, we chose to focus on online editorials and blog posts concerning the ongoing debate over health insurance reform legislation in the United States. Our intuition is that blogs and editorials represent a genre rich in both

<sup>2</sup>Our stance structure contains an additional “aspect” level consisting of a-priori categories adopted from political science research. However, we do not utilize this level of the stance structure in this work.

“pro” documents	37
“pro” sentences	1,222
“anti” documents	47
“anti” sentences	1,456
total documents	84
total sentences	2,678

Table 1: Dataset summary statistics.

arguing subjectivity	
objective	683
subjective	588

argument labels	
no label	683
improves_healthcare_access	130
improves_healthcare_affordability	104
people_dont_know_truth_about_bill	75
controls_healthcare_costs	54
improves_quality_of_healthcare	52
helps_economy	51
bill_should_be_passed	43
other argument	79

Table 2: Arguing and argument label statistics for the “pro” stance.

subjectivity and arguments. We collected documents written both before and after the passage of the final “Patient Protection and Affordable Care Act” bill using the “Google Blog Search”<sup>3</sup> and “Daily Op Ed”<sup>4</sup> search portals. By choosing a relatively broad time window, from early 2009 to late 2011, we aimed to capture a wide range of arguments expressed throughout the debate.

The focus of this paper is on sentence-level argument detection rather than document-level stance classification (e.g., (Anand et al., 2011), (Park et al., 2011), (Somasundaran and Wiebe, 2010), (Burfoot et al., 2011)). We treat stance classification as a separate step preceding arguing subjectivity detection, and thus provide oracle stance labels for our data.

We treat documents written from the “pro”

<sup>3</sup><http://www.google.com/blogsearch>

<sup>4</sup><http://www.dailyoped.com/>

arguing subjectivity	
objective	913
subjective	575

argument labels	
no label	913
diminishes_quality_of_care	122
too_expensive	67
unpopular	60
hurts_economy	55
expands_govt	52
bill_is_politically_motivated	44
other_reforms_more_appropriate	35
other argument	140

Table 3: Arguing and argument label statistics for the “anti” stance.

stance and documents written from the “anti” stance as separate datasets. Being written from different positions, the two stances will have different argument labels and may employ different styles of arguing subjectivity. Table 1 provides an overview of the size of this dataset. Summary statistics concerning the density of arguing and argument labels in the two sides of the dataset is presented in Tables 2 and 3. However, since it can be difficult to summarize a complex argument in a short phrase, many of these labels by themselves do not clearly convey the meaning they are meant to represent. To better illustrate the meanings of some of the more ambiguous labels, Table 4 presents several annotated example spans for some of the more unclear ambiguous argument labels.

#### 4 Agreement Study

One of our authors performed annotation of our corpus, the broad outlines of which are sketched in the previous section. However, to assess inter-annotator agreement for this annotation scheme, we recruited a non-author to independently annotate a subset of our corpus consisting of 384 sentences across 10 documents. This non-author both identified spans of arguing subjectivity and assigned argument tags. She was given a stance structure from which to select argument tags.

improves_healthcare_access	“Our reform will prohibit insurance companies from denying coverage because of your medical history.”
	“Let’s also not overlook the news from last week about the millions of younger Americans who are getting coverage thanks to consumer protections that are now in place.”
improves_healthcare_affordability	“ new health insurance exchanges will offer competitive, consumer-centered health insurance marketplaces...”
	“Millions of seniors can now afford medication they would otherwise struggle to pay for.”
people_dont_know_truth_about_bill	“...the cynics and the naysayers will continue to exploit fear and concerns for political gain.”
	“Republican leaders, who see opportunities to gain seats in the elections, have made clear that they will continue to peddle fictions about a government takeover of the health care system and about costs too high to bear.”
unpopular	“The 1,000-page monstrosity that emerged in various editions from Congress was done in by widespread national revulsion...”
	“Support for ObamaCare’s repeal is broad, and includes one group too often overlooked during the health care debate: America’s doctors.”
expands_govt	“...the real goal of the health care overhaul was to enact the largest entitlement program in history...”
	“the new bureaucracy the health care legislation creates is so complex and indiscriminate that its size and cost is ‘currently unknowable.’ ”
bill_is_politically_motivated	“...tawdry backroom politics were used to sell off favors in exchange for votes.”
	“From the wildly improper gifts to senators like Nebraska’s Ben Nelson to this week’s backroom deals for unions...”

Table 4: Example annotated spans for several argument labels.

metric	recall	precision	f-measure
<i>agr</i>	0.677	0.690	0.683
kappa for overlapping annotations			0.689

Table 5: Inter-annotator span *agr* (top) and argument label kappa on overlapping spans (bottom).

In assessing inter-annotator agreement on this subset of the corpus, we must address two levels of agreement, arguing spans and argument tags.

At first glance, how to assess agreement of annotated arguing spans is not obvious. Because our annotation scheme did not enforce strict boundaries, we hypothesized that both annotators would both frequently see an instance of arguing subjectivity within a local region of text, but would disagree with respect to where the arguing begins and ends. Thus, we adopt from (Wilson and Wiebe, 2003) the *agr* directional agreement metric to measure the degree of annotation overlap. Given two sets of spans  $A$  and  $B$  annotated by two different annotators, this metric measures the fraction of spans in  $A$  which at least partially overlap with any spans in  $B$ . Specifically, agreement is computed as:

$$agr(A || B) = \frac{|A \text{ matching } B|}{|A|}$$

When  $A$  is the gold standard set of annotations, *agr* is equivalent to recall. Similarly, when  $B$  is the gold standard, *agr* is equivalent to precision. For this evaluation, we treat the dataset annotated by our primary annotator as the gold standard. Table 5 presents these *agr* scores and f-measures for the arguing spans.

Second, we measure agreement with respect to the argument tags assigned by the two annotators. Continuing to follow the methodology of (Wilson and Wiebe, 2003), we look at each pair of annotations, one from each annotator, which share at least a partial overlap. For each such pair, we assess whether the two spans share the same primary argument tag. Scores for primary argument label agreement in terms of Cohen’s kappa are also presented in Table 5. Since this kappa score falls within the

range of  $0.67 \leq K \leq 0.8$ , according to Krippendorff’s scale (Krippendorff, 2004) this allows us to draw tentative conclusions concerning a significant level of tag agreement.

## 5 Methods

As discussed earlier, recognizing arguments can be thought of in terms of two related but different tasks: recognizing a type of subjectivity, and labeling instances of that subjectivity with tags. We refer to the binary arguing subjectivity detection task as “arg”, and to the multi-class argument labeling task as “tag”. For the “tag” task, we create eight classes: one for each of the seven most-frequent labels, and an eighth into which we agglomerate the remaining less-frequent labels. We only consider the sentences known to be subjective (via oracle information) for the “tag” task.

We also perform a “combined” task. This third task is conceptually similar to the “tag” task, except that all sentences are considered rather than only the subjective sentences. In addition to the eight classes used by “tag”, “combined” adds an additional class for non-arguing sentences. Finally, we also perform a two-stage “arg+tag” task. In this two-stage task, the instances labeled as subjective by the “arg” classifier are passed as input to the “tag” classifier. The intuition behind this two-phase approach is that the features most useful for identifying arguing subjectivity may not be the most useful for discriminating between argument tags, and vice versa. For all of our classification tasks, we treat both the “pro” and “anti” stances separately, building separate classifiers for each stance for each of the above tasks.

In general, we perform single-label classification at the sentence level. However, sentences containing multiple labels pose a challenge. Since this was an early exploratory work on a very difficult task, we decided to handle this situation by splitting sentences containing multiple labels into separate instances for the purpose of learning, assigning a single label to each instance. However, only about 3% of the sentences in our corpus contained multiple la-

bels. Thus, replacing this splitting step in the future with another method that does not require oracle information, such as choosing the label which covers the most words in the sentence, is a reasonable simplification of the task.

Since discourse actions, such as contrasting, restating, and identifying causation, play a substantial role in arguing, we hypothesize that information about the discourse roles played by a span of text will help improve classification. Although discourse parsers historically haven't been found to be effective for subjectivity analysis, a new parser (Lin et al., 2010) trained on the Penn Discourse TreeBank (PDTB) tagset (Prasad et al., 2008) has recently been released. Previous work has demonstrated that this parser can reliably detect discourse relationships between adjacent sentences (Lin et al., 2011), and the PDTB tagset, being relatively flat, is conducive to feature engineering for our task.

To give a feeling for the kind of discourse relations identified by this parser, the following example illustrates a concession relation identified in the corpus by the parser. The italicized text represents the concession, while the bolded text indicates the overall point that the author is making. The underlined word was identified by the parser as an explicit concessionary clue.

(7) the health care reform legislation that President Obama now seems likely to sign into law , *while an unlovely mess* , **will be remembered as a landmark accomplishment** .

Using this automatic information, we define features indicating the discourse relationships by which the instance is connected to surrounding text. Specifically, the class of discourse relationship connecting the target instance to the previous instance, the relationship connecting it to the following instance, and any internal discourse relationships by which the parts of the instance are connected to each other are each added as features. Since PDTB contains many fine-grained discourse relations, we replace each discourse relationship type inferred by the discourse parser with the parent top-level PDTB

discourse relationship class. We arrive at a total of 15 binary discourse relationship features: (4 top-level classes + "other") x (connects to previous + connects to following + internal connection) = 15. We refer to these features as "rels".

As illustrated in our earlier examples, while arguing subjectivity is different from sentiment, the two types of subjectivity are often related. Thus, we investigate incorporating sentiment information based on the presence of unigram clues from a publically-available sentiment lexicon<sup>5</sup> (Wilson, 2005). Each clue in the lexicon is marked as being either "strong" or "weak".

We found that this lexicon was producing many false hits for positive sentiment. Thus, a span containing a minimum of two positive clues of which at least one is marked as "strong", or three positive "weak" clues, is augmented with a feature indicating positive sentiment. For negative sentiment the threshold is slightly lower, at one "strong" clue or two "weak" clues. These features are referred to as "senti".

A challenge to argument tag assignment is the broad diversity of language through which individual entities or specific actions may be referenced, as illustrated in Examples (2-4) from Section 1. To address this problem, we investigate expanding each instance with terms that are most similar, according to a distributional model generated from Wikipedia articles, to the nouns and verbs present within the instance (Pantel et al., 2009). We refer to these features as "expr $n$ ", where  $n$  is the number of most-similar terms with which to expand the instance for each noun or verb. We experiment with values of  $n = 5$  and  $n = 10$ .

Subjectivity classification of small units of text, such as individual microblog posts (Jiang et al., 2011) and sentences (Riloff et al., 2003), has been shown to benefit from additional context. Thus, we augment the feature representation of each target sentence with features from the two preceding and two following sentences. These additional features are modified so that they do not fall within the same feature space

<sup>5</sup>downloaded from [http://www.cs.pitt.edu/mpqa/subj\\_lexicon.html](http://www.cs.pitt.edu/mpqa/subj_lexicon.html)

feat. abbrev.	elaboration
unigram	
senti	2 binary features indicating positive or negative sentiment based on presence of lexicon clues
rels	15 binary features indicating kinds of discourse relationships and how they connect instance to surrounding text
exp5	for each noun and verb in instance, expand instance with top 5 most distributionally similar words
exp10	for each noun and verb in instance, expand instance with top 10 most distributionally similar words

Table 6: Overview of features used in the arguing and argument experiments.

as the features representing the target sentence.

Using the Naive Bayes classifier within the WEKA machine learning toolkit (Hall et al., 2009), we explore the impact of the features described above on our four experiment configurations. We perform our experiments using  $k$ -fold cross-validation, where  $k$  equals the number of documents within the stance. The test set for each fold consists of a single document’s instances. For the “pro” dataset  $k = 37$ , while for the “anti” dataset  $k = 47$ .

## 6 Results

Table 7 presents the accuracy scores from each of our stand-alone classifiers across combinations of feature sets. Each feature set consists of unigrams augmented with the designated additional features, as described in Section 5. To evaluate the “tag” classifier in isolation, we use oracle information to provide this classifier with only the subjective instances. To assess significance of the performance differences between feature sets, we used the Pearson Chi-squared test with Yates continuity correction.

Expansion of nouns and verbs with distributionally-similar terms (“exp5”, “exp10”) plays the largest role in improving classifier

features	arg	tag	comb.
unigram baseline	0.610	0.425	0.458
senti	0.614	0.426	0.459
rels	0.614	0.422	0.462
senti, rels	0.618	0.424	0.465
exp5	<i>0.635</i>	<b>0.522</b>	<i>0.482</i>
exp5, senti	<b>0.638</b>	<b>0.515</b>	<b>0.486</b>
exp5, rels	<b>0.640</b>	<b>0.522</b>	<i>0.484</i>
exp5, senti, rels	<b>0.643</b>	<b>0.516</b>	<i>0.484</i>
exp10	<b>0.645</b>	<b>0.517</b>	<b>0.488</b>
exp10, senti	<b>0.647</b>	<b>0.515</b>	<b>0.489</b>
exp10, rels	<b>0.642</b>	<b>0.512</b>	<b>0.490</b>
exp10, senti, rels	<b>0.644</b>	<b>0.513</b>	<b>0.490</b>

Table 7: Classifier accuracy for differing feature sets. Significant improvement ( $p < 0.05$ ) over baseline is boldfaced ( $0.05 < p < 0.1$  italicized). Underline indicates best performance per column.

performance. While differences between configurations using “exp5” versus “exp10” were generally not significant, all of the configurations incorporating some version of term expansion outperformed the unigram baseline by either a statistically significant margin ( $p < 0.05$ ) or by a margin that approached significance ( $0.05 < p < 0.1$ ).

Sentiment features consistently produce improvements in accuracy for the “arg” and “combined” tasks. While these improvements are promising, the lack of a significant margin of improvement when incorporating sentiment is surprising. Since sentiment lexicons are known to be highly domain-dependent (Pan et al., 2010), it may be the case that, having been learned from a general news corpus, the sentiment lexicon employed in this work is not the best match for the domain of “ObamaCare” blogs and editorials. Similarly, the discourse features also fail to produce significant improvements in accuracy.

Finally, we aim to test our hypothesis that separating the “arg” and “tag” phases results in improvement beyond treating the two in a single “combined” phase. The first step of our hierarchy involves normal classification of all sentences using the “arg” classifier. Next, all sentences judged to contain arguing subjectivity by “arg”

arg features	tag features	acc.
exp5, senti, rels	exp5	0.506
	exp5, rels	0.506
	exp10	0.501
exp10	exp5	<u>0.514</u>
	exp5, rels	0.513
	exp10	0.512
exp10, senti	exp5	<u>0.514</u>
	exp5, rels	0.513
	exp10	0.512

Table 8: Accuracies of two-stage classifiers across different combinations of feature sets for the “arg” and “tag” phases. Italics indicate improvement over the top “combined” configuration which approaches significance ( $0.05 < p < 0.1$ ). Underline indicates best overall performance.

are passed to the “tag” classifier to have an argument tag assigned. We choose three promising feature sets for the “arg” and “tag” phases, based on best performance in isolation.

Results of this hierarchical experiment are presented in Table 8. We evaluate the hierarchical system against the best-performing “combined” single-phase systems from Table 7. While all of the hierarchical configurations beat the best “combined” classifier, none beats the top combined classifier by a significant margin, although the best configurations approach significance ( $0.05 < p < 0.1$ ).

## 7 Related Work

Much recent work in ideological subjectivity detection has focused on detecting a writer’s stance in domains of varying formality, such as online forums, debating websites, and op-eds. (Anand et al., 2011) demonstrates the usefulness of dependency relations, LIWC counts (Pennebaker et al., 2001), and information about related posts for this task. (Lin et al., 2006) explores relationships between sentence-level and document-level classification for a stance-like prediction task.

Among the literature on ideological subjectivity, perhaps most similar to our work is (Somasundaran and Wiebe, 2010). This paper investigates the impact of incorporating arguing-based

and sentiment-based features into binary stance prediction for debate posts. Also closely related to our work is (Somasundaran et al., 2007). To support answering of opinion-based questions, this work investigates the use of high-precision sentiment and arguing clues for sentence-level sentiment and arguing prediction.

Another active area of related research focuses on identifying important aspects towards which sentiment is expressed within a domain. (He et al., 2011) approaches this problem through topic modeling, extending the joint sentiment-topic (JST) model which aims to simultaneously learn sentiment and aspect probabilities for a unit of text. (Yu et al., 2011) takes a different approach, investigating thesaurus methods for learning aspects based on groups of synonymous nouns within product reviews.

## 8 Conclusion

In this paper, we explored recognizing arguments in terms of arguing subjectivity and argument tags. We presented and evaluated a new annotation scheme to capture arguing subjectivity and argument tags, and annotated a new dataset. Utilizing existing sentiment, discourse, and distributional similarity resources, we explored ways in which these three forms of knowledge could be used to enhance argument recognition. In particular, our empirical results highlight the important role played by distributional similarity in all phases of detecting arguing subjectivity and argument tags. We have also provided tentative evidence suggesting that addressing the problem of recognizing arguments in two separate phases may be beneficial to overall classification accuracy.

## 9 Acknowledgments

This material is based in part upon work supported by National Science Foundation award #0916046. We would like to thank Patrick Pantel for sharing his thesaurus of distributionally similar words from Wikipedia with us, Amber Boydston for insightful conversations about debate frame categorization, and the anonymous reviewers for their useful feedback.



## References

- Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *WASSA*, pages 1–9, Portland, Oregon, June.
- F.R. Baumgartner, S.D. Boef, and A.E. Boydston. 2008. *The decline of the death penalty and the discovery of innocence*. Cambridge University Press.
- Clinton Burfoot, Steven Bird, and Timothy Baldwin. 2011. Collective classification of congressional floor-debate transcripts. In *ACL*, pages 1506–1515, Portland, Oregon, USA, June.
- Claire Cardie, Cynthia Farina, Adil Aijaz, Matt Rawding, and Stephen Purpura. 2008. A study in rule-specific issue categorization for e-rulemaking. In *DG.O*, pages 244–253.
- Mark Hall, 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.
- Yulan He, Chenghua Lin, and Harith Alani. 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *ACL*, pages 123–131, Portland, Oregon, USA, June.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *ACL*, pages 151–160, Portland, Oregon, USA, June.
- K. Krippendorff. 2004. *Content analysis: an introduction to its methodology*. Sage.
- Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on?: identifying perspectives at the document and sentence levels. In *CoNLL*, pages 109–116.
- Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2010. A pdtb-styled end-to-end discourse parser. *CoRR*.
- Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2011. Automatically evaluating text coherence using discourse relations. In *ACL*, pages 997–1006, Portland, Oregon, USA, June.
- Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *ACL*, pages 142–150, Portland, Oregon, USA, June.
- Sinno Jialin Pan, Xiaochuan Ni, Jian tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *WWW*.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL*, pages 271–278, Barcelona, Spain, July.
- Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu, and Vishnu Vyas. 2009. Web-scale distributional similarity and entity set expansion. In *EMNLP*, pages 938–947, Morristown, NJ, USA.
- Souneil Park, Kyung Soon Lee, and Junehwa Song. 2011. Contrasting opposing views of news articles on contentious issues. In *ACL*, pages 340–349, Portland, Oregon, USA, June.
- James W Pennebaker, Roger J Booth, and Martha E Francis. 2001. Linguistic inquiry and word count (liwc): Liwc2001. *Linguistic Inquiry*, (Mahwah, NJ):0.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *LREC*, May.
- Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *CoNLL*, pages 25–32.
- Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *ACL-AFNLP*, pages 226–234.
- Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *CAAGET*, pages 116–124.
- Swapna Somasundaran, Theresa Wilson, Janyce Wiebe, and Veselin Stoyanov. 2007. Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In *ICWSM*.
- Swampa Somasundaran. 2010. *Discourse-Level Relations for Opinion Analysis*. Ph.D. thesis, University of Pittsburgh, USA.
- Theresa Wilson and Janyce Wiebe. 2003. Annotating opinions in the world press. In *SIGdial*, pages 13–22.
- Theresa Wilson. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP*, pages 347–354.
- Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua. 2011. Aspect ranking: Identifying important product aspects from online consumer reviews. In *ACL*, pages 1496–1505, Portland, Oregon, USA, June.