Clos, J., Wiratunga, N., Massie, S., & Cabanac, G. (in press). Shallow techniques for argument mining. *ECA'15: Proceedings of the 1st European Conference on Argumentation: Argumentation and Reasoned Action.* Studies in Logic and Argumentation (Vol. ___, pp. ___-___). London, UK: College Publications.

Shallow techniques for argument mining

JÉRÉMIE CLOS
Robert Gordon University, UK
j.clos@rgu.ac.uk

NIRMALIE WIRATUNGA
Robert Gordon University, UK
n.wiratunga@rgu.ac.uk

STEWART MASSIE
Robert Gordon University, UK
s.massie@rgu.ac.uk

GUILLAUME CABANAC
Université de Toulouse, France
guillaume.cabanac@univ-tlse3.fr

Argument mining has recently emerged as a promising field at the frontiers of the argumentation and text mining communities. However, most techniques developed within that field do not scale to larger amounts of data, depriving us for example of valuable insights in large-scale discussion forums. On two social media datasets, we study different lightweight scalable text mining techniques used within the sentiment analysis community and their applicability to the argument mining problem.

KEYWORDS: argument mining, sentiment analysis, text mining

1. INTRODUCTION

The advent of the Web 2.0 has seen a massive increase in user-generated data in the form of comments and messages such as the ones displayed in Figure 1. Increasingly it is the platform of choice for public debate and conversation, but its traditional "document-centric" focus and associated search and browse methods are less fit for purpose. For example, in Figure 1 it is not sufficient to only search for keywords but instead be influenced by the tree-like thread structure enforced by users responding to other user comments.

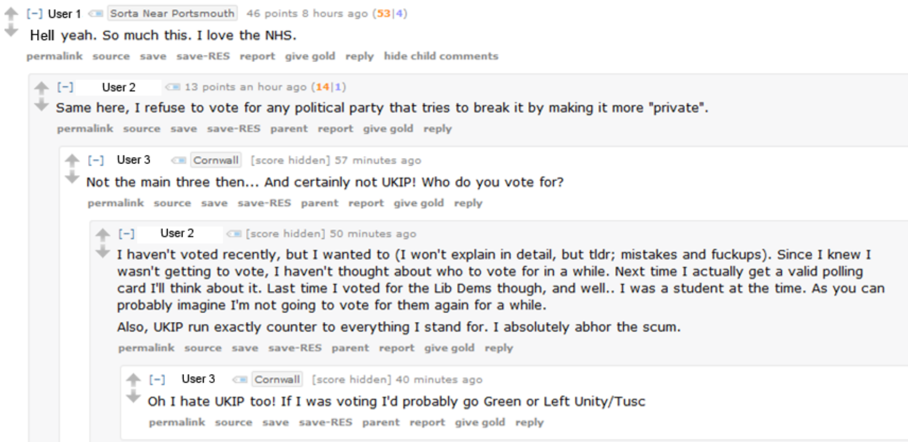

Figure 1: Excerpt from a comment tree where users discuss UK politics.

New research fields such as sentiment analysis and topic modelling have thus emerged in order to fill this need for better and more intuitive ways to help users browse through large amounts of data. While valuable, these techniques inevitably fall short of meeting the representational requirements when dealing with conversational data. For instance, sentiment analysis is only concerned with projecting documents on a negative to positive opinion dimension, and topic modelling is focused on analysing a corpus and identifying central topics Neither are interested in the conversational dynamics.

Argument mining is able to discover knowledge that would allow us to detect justifications for common opinions, generate fine-grained debate graphs for complex political issues or refine common opinion mining algorithms. There are however many challenges in adapting argument mining algorithms to the scale of the Social Web. Current approaches either rely on computationally expensive NLP techniques or on human annotations, neither of which are transferable to a real time analysis setting where large volumes of data, absence of reliable knowledge sources and informal language are the norm.

In this paper we propose to relax the requirements of an argument mining algorithm by restricting the argument mining task to a target (another argument/expression of opinion) detection and stance (whether it supports or attacks the target) classification task, leveraging existing literature in the sentiment analysis and opinion mining. Our contribution is threefold: firstly, we build a novel dataset based on online comments from the Reddit[1] social website and a noisy labelling process. Secondly, we experiment using three standard unsupervised sentiment analysis approaches in order to measure how well they can approximate the stance classification part of the argument mining process. Thirdly and finally, we improve a PMI-based classifier by incorporating contextual clues in a simple but intuitive way into the classification process.

In section 2 we detail the relationship between argumentation, argument mining and information retrieval, thus justifying and contextualizing our approach. In section 3 we explain the classification approaches that are being compared, as well as the approach we are proposing as an incremental improvement over a naive technique based on strength of association. In section 4 the experimental methodology is presented, together with details of a new dataset, generated for the purpose of this research. Finally, before concluding in section 6, section 5 will analyse the results from the comparative study.

2. BACKGROUND AND RELATED WORKS

Our approach to argument mining is inspired by text analysis and representation schemes which are commonly used in information retrieval. Rather than linguistic rigor, we aim to use knowledge-light representations that can still provide insight about the discussion. As explained in the previous section, we seek to rebuild the argumentation graph underlying a discussion by making the following simplifying assumptions: all comments have an argumentative value, and the target of a comment is always the comment to which it is replying. For the purpose of building a bipolar argumentation graph, we loosely assign to the "attack" relationship defined by Dung (Dung, 1995) the semantics of *overall disagreement*, and to the "support" relationship defined by Cayrol and Lagasque-Schiex (Cayrol and Lagasquie-Schiex, 2005) the semantics of *overall agreement*.

The need to scale classification to large amounts of data requires a simple conceptual representation of arguments, such as the one proposed in Pragmatic Argumentation Theory (PAT) (Van Eemeren et al., 1996; Hutchby, 2013). Fitting a complex model of argument would be computationally expensive and not fit the colloquial nature of social media content and it is thus deemed preferable to use a more accurate and simpler model.

PAT defines an argument as an opinionated piece of text which can arise in the presence of two elements: (1) a **target**, being some other action by another actor which has been called out ; (2) a **stance**,

[1] http://www.reddit.com

i.e. whether it is supporting or attacking the target. We bypass the target detection step and focus on stance classification, which allows us to use techniques from text mining and sentiment analysis (Pang and Lee, 2008), considering stance of an argument analogous to the sentiment of an opinionated text.

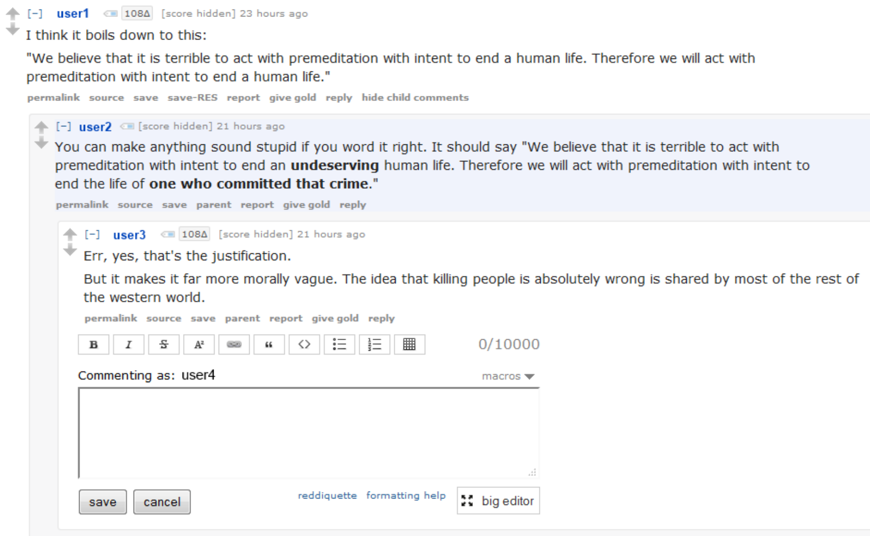

Figure 2: Illustration of how the Reddit commenting system forces the user to place their comment under the contribution

Figure 2 illustrates the way the system incites users to insert their comment under the relevant section of the discussion by presenting it in a threaded structure, allowing us to treat the argument mining problem as a classification task. For example, the comment posted by User 3 in Figure 2 is in agreement with the comment posted by User 2, which makes it a supporting statement.

2.1 Argument mining

Argument mining has been approached in the literature as the study of methods and techniques to detect argumentative discourse units, their role in the argumentation process and how they relate to other argumentative discourse units (Peldszus and Stede, 2013). Early work focused on representing arguments in a restricted manner (Cohen 1987), but the first steps towards an automated treatment of argument mining (Palau and Moens, 2009; Mochales and Moens, 2011) aimed to mine legal text using supervised learning techniques. More particularly, Palau and Moens (Palau and Moens, 2009) performed a three-step argument analysis by firstly detecting argumentative sentences, secondly identifying whether they were part of a conclusion or a premise, and thirdly classifying the relationships between these sentences, thus trying to mine argumentative structure in these legal texts. Cabrio and Villata (Cabrio and Villata, 2012) on the other hand made use of textual entailment and semantic similarity features in order to train a classifier to recognize attacking arguments, in line with Dung's abstract argumentation framework (Dung, 1995). Similar work along these lines studied the use of context-free grammars (Wyner et al., 2010) to extract arguments but did not handle the detection of their relationships with other arguments, since legal texts mainly deal with cases of monological argumentation. However these approaches do not transfer well to social media because of their reliance on idiosyncratic features and complex learning methods such as support vector machines (SVM) (Bishop, 2006).

Other works from the argument mining community focused on bridging it to the field of opinion mining (Villalba and Saint-Dizier, 2012) and studied the use of reasoning patterns in user-contributed reviews. They did not however direct their study towards the automated detection of arguments themselves within these textual reviews and instead focused on a descriptive analysis, making them not directly relevant to our work.

2.2 Stance classification

Stance classification becomes relevant to argument mining because of its binary classification nature. However most techniques used in the literature base their work on training a complex classifier using a large number of computationally expensive features (Boltuzic and Snajder, 2014; Anand et al., 2011; Abbott et al., 2011; Somasundaran and Wiebe, 2010; Walker et al., 2012a) or are performed on automatically transcribed text (Wang et al., 2011) or non-conversational content (Bousmalis et al., 2013) with limited applicability to Web2.0.

Much closer to our approach, Yin et al. (Yin et al., 2012) used a logistic regression classifier trained on somewhat complex features and related the notion of local and global stances, building the global stance of a post by computing a sequence of local stances between that post and the first appearance of the topic of discussion

Cardie and Wang (Wang and Cardie, 2014) took a different approach to the stance classification problem in that they used an isotonic conditional random field-based technique to detect local stance. However, they still required a significant training phase and need a large collection of idiosyncratic features, which negates the portability of their approach.

2.2 Sentiment analysis and argument mining

The field of sentiment analysis (Pang and Lee, 2008) also treats the binary classification of large text corpora along the axis of positivity/negativity. Two main families of methodologies emerge:

- **Supervised sentiment analysis** involves the use of supervised machine learning techniques in order to perform sentiment classification. Traditional algorithms known for their versatility and performance on text classification tasks are MaxEnt, SVM, and Naive Bayes (Pang et al., 2002). These algorithms are trained on training data in order to produce a model able to classify future test data.
- **Lexicon-based sentiment analysis** involves the learning/building of lexicons, which are look-up tables of terms with strengths of association scores for different classes, as well as combination rules (Taboada et al., 2011). Test data is directly fed to the lexicon-based classifier which uses the combination rules in order to compute the scores of each class (positive and negative) based on the presence of terms from the lexicon.

Our work can be put in context between the argument mining and the stance detection communities, as we aim for the detection of relationships between textual entities with the bipolar semantics of Cayrol and Lagasquie-Schiex (Cayrol and Lagasquie-Schiex, 2005) but do so by attributing different semantics to their relations.

3. SHALLOW TECHNIQUES FOR ARGUMENT MINING

We refer to the relationship between a comment and its direct parent as the conversational context of that comment. Our goal is to take into account, for all comments, a progressively deeper level of their conversational context and study its effect on the overall classification accuracy. As such, we will review the approaches according to their level of context-informedness as well as their degree of supervision. Here context-informedness refers to the extent to which the approach uses information that is external to the comment, and supervision refers to the extent to which the algorithm requires a human-labelled dataset in order to work.

3.1 Argument mining as lexicon-based classification

In its most basic form, argument mining can be seen as a classification task, where the classes are either support or attack. This basic form entails that any piece of text can be classified by itself without taking into account any notion of conversational context, by simply applying standard text representation techniques and mapping this representation into the class codomain. contains simple lexicon-based classification (referred to as PMILex).

A simple way to reliably classify instances is the use of a lexicon. Because there does not exist a manually built lexicon of local stance, we compute it on a distant-labelled dataset using normalized pointwise mutual information (NPMI) as a measure of strength of association between terms and their class.

$$NPMI(x,y) = \frac{PMI(x,y)}{-\log[p(x,y)]}$$

$$PMI(x,y) = \frac{\log\,(p(x,y))}{p(x)p(y)}$$

This classification rule classifies a user comment x on the basis of a maximized sum of associations between each of its terms t and each class c. Notice that the term class associations can now be exploited as a lexicon.

While this approach is sensible to create a general purpose lexicon, it suffers some flaws in the following cases: (1) if none of the terms used in the child post has an argumentative value or is present within the lexicon, no classification is possible, and (2) some terms might end up with an undeserved score because they accidentally appear more frequently within comments of one class. For example if non-argumentative terms such as "Monday" accidentally co-occur too often within one class, they will be misconstrued as being indicative of that class.

3.2. Context-aware methods for argument classification

The previous method provided a simple mapping from a set of terms to a class label, we now explore different ways to consider context during the classification. This context can originate from either the sentiment contained within the comment, or the conversation that contains the comment. Sentiment-guided methods deal with the sentiment context, i.e. the sentiment that is expressed within the terminology and grammatical structure used in the text. Within the context of a discussion, this sentiment is akin to a global stance taken by the author of that text with respect to the topic at hand. Conversational context-aware methods on the other hand are focused on representing the relationship between a comment and its conversational context, in our situation its parent post. They do so either by detecting attack or support against the author of the parent comment (Wang and Cardie, 2014) or by modifying the importance of some terms based the parent comment (the proposed approaches).

3.2.1 Sentiment-guided methods

Sentiment methods provide a means to use emotive context to infer argument stance, by assuming the stance of a comment as equivalent to its sentiment orientation. We employ a simple sentiment analysis algorithm based on a lexicon (Esuli and Sebastiani, 2006) to which we will refer to as SENTLEX. It operates by looking up positive and negative values of terms present in the comment and summing them separately into a positive and a negative score. The classification rule is based on a simple comparison: a higher positive strength implies a supportive comment, and higher negative strength implies an attacking comment. We use the SMARTSA algorithm (SSA) developed by Muhammad et al. (Muhammad et al., 2013) as an extension of traditional lexicon-based sentiment analysis techniques taking into account additional linguistic factors such as the presence of special terms called modifiers that alter the class values of terms in their vicinity by exaggerating them (amplifiers), reducing them (diminishers) or inversing their polarity (negators).

Sentiment-guided methods assume that global stance of the comment, i.e. how its author feels about the discussion topic, can be used in place of its local stance, i.e. how its author feels about the parent comment. As such they present some flaws whenever (1) those two stances do not align, (2) the stance is expressed in a sentiment-neutral WAY, or (3) the overall sentiment of the sentence ends up being balanced. The following examples illustrate these flaws:

(1) "*I completely disagree with you, this movie was very good and I enjoyed every minute of it.*" Here we can see that the author expresses a positive opinion by disagreeing with the author of the parent comment, thus making an attacking statement.

(2) "*There is nothing in the world that will make me see the situation your way.*" In this comment there is no positive or negative terminology used, while the sentence is written with an attacking stance.

(3) "*I agree that the acting was good, but I am still disappointed that the dialogues were so poorly written*." Here negative and positive

sentiments are equally used, but the stance should be supporting.

3.2.2 Conversational context-aware methods

We explore conversational context-aware methods using the Unsupervised Sentiment Surface algorithm (USS). USS is extended from Cardie and Wang (Wang & Cardie, 2014) and uses a shallow linguistic analysis to compute the average distance between second person pronouns and positive or negative terms. The classification compares those average distances, classifying the comment as attacking if negative terms are on average closer to second person pronouns and supporting otherwise. USS works on the intuition that the stance of the comment is contained within explicit references to the parent posts: such references can be analyzed by detecting second person pronouns (e.g. "you", "your", etc.) and their polarity by searching their grammatical neighborhood for sentiment-bearing terminology.

For example, "*I don't agree with you and I think your opinion is wrong*" would be interpreted as an attacking statement because of the overwhelming proximity of negative terms ("*don't*", "*wrong*") near second person pronouns ("*you*", "*your*"). This approach can also give flawed results when there is ambiguity contained in the text in the following: (1) when a positive (respectively negative) term is accidentally closed to a second person pronoun which is semantically linked to a negative (respectively positive) term (2) when a more complex sentence structure is used where the polarity of a term is implicitly negated, or (3) whenever no second person pronouns or sentiment-bearing terms are used. The following examples illustrate the first two cases:

(1) "*I like you, but you are wrong.*" Here we can see that $d(like, you) < d(you, wrong)$ where d is a distance function, which would classify this instance as a supporting statement.

(2) "*I can barely tolerate that you believe yourself to be right.*" Here the sentence structure puts *yourself* very close to *right*, which will classify the sentence as a supporting statement. However, it is clear under a human eye that the sentence has a disapproving tone.

The USS approach is based on the assumption that over a significant number of sentences these errors would cancel each other out and result in a good overall classification accuracy.

3.2.3 Context-informed feature vector enrichment

Term vector representation is a convenient way to work with text data because of its simplicity and versatility (more details can be found in (Manning et al., 2008)). It is compatible with lexicon-based classification as well as more standard supervised classifiers and extremely common in text classification. In this section we use conversational context to alter that feature vector of a comment based on its parent comment. We experiment using two basic variations of this alteration. The classification rule we use is to assign the class label c that maximizes the association score between an instance x and c. The association score is computed differently according to the enrichment scheme.

$$Classification(x) = ArgMax_c[AssociationScore(x, c)]$$

Intersection-based vector enrichment. We used a lexicon computed similarly to the simple PMI lexicon discussed in a previous section, and compute the association score between an instance x and a class c as the sum of NPMI scores between all terms t with c where t is present in both the x and its parent p(x).

$$AssociationScore_1(x, c) = \sum_{t \in x \cap p(x)} NPMI(t, c)$$

Union-based vector enrichment. We modify the classification rule and compute the association score as the sum of NPMI scores between all terms t with a given class c where t is present in either the instance x and its parent p(x).

$$AssociationScore_2(x, c) = \sum_{t \in x \cup p(x)} NPMI(t, c)$$

The algorithm described in Figure 3 represents the general classification algorithm of our approaches, where the changing part is the way the instance I is created from the comments P and C. The classification is done by a simple summation of terms, which is a highly scalable operation with a negligible computational cost.

Data: Child comment C and parent comment P
Result: A class label L
Create instance I from P and C;
$AL \leftarrow 0$;
$DL \leftarrow 0$;

For *Term T in I* **do**
  $AL \leftarrow AL + agreementValue(\mathrm{T})$;
  $DL \leftarrow DL + disagreementValue(\mathrm{T})$;
End
If $DL \leq AL$ **then**
  Return *agreement* ;
Else
  Return *disagreement* ;
End

Figure 3: General classification algorithm

4. EXPERIMENTAL DESIGN

In this section we detail our experimental design, firstly by going over our datasets and the way they were collected, secondly moving on to the pre-processing steps that were performed in order to sanitize the data, thirdly to the metrics used in order to compare approaches, and finally to our experimental methodology.

4.1 Datasets

We performed our experiments on two social media datasets: the Internet Argument Corpus (referred to as IAC) and the Reddit Noisy-Labelled Corpus (referred to as RNLC). Statistics on the corpora can be found in Table 1.

Dataset	*IAC*	*RNLC*
Number of comments	1856	3086
Average terms/sentence	40.3	35
Average sentences/comment	2.9	7.8
Instances of agreement	928	1543
Instances of disagreement	928	1543
Common vocabulary size	6036	
Total vocabulary size	25004	

Table 1: Descriptive statistics on the dataset

The Internet Argument Corpus (IAC). The IAC (Walker et al., 2012b) is a corpus of forum comments manually labelled by 5 annotators on a degree of agreement/disagreement with their parent comment on a scale of -5 to 5. A subset of this dataset was used for our experiment, by selecting the comments that ensured disjoint class membership (meaning filtering out comments with an average score close to 0).

The Reddit Noisy-Labelled Corpus (RNLC). The RNLC is a new corpus of comments extracted from the Reddit and automatically labelled with a binary class using evidence contained within the comments. Explicit expressions such as "*I [positive adverb] agree*" and "*I [positive adverb] disagree*" variations were used to detect evidence of a comment belonging to a class. In the case of the presence of conflicting evidence, i.e. expressions acting as strong evidence towards both classes, the comments were not considered. Remaining comments were automatically assigned to their respective class and the corresponding sentences were deleted from the comments in order to avoid a class bias advantage. That labelling process is inspired from distant supervision learning (Mintz et al., 2009) whereby highly discriminative expressions are used as class label proxies.

Both datasets were pre-processed by removing comments that were deemed as non-constructive because of their limited length. A threshold was empirically chosen based on a human observation of the data and all comments composed of less than 20 words and/or with a vocabulary of less than 10 words were considered as noise and removed from the data.

No stemming was applied due to the unreliable vocabulary used in social media, meaning that a rule-based procedure for would reunify terms that are semantically distant and thus remove information from the datasets. For the same reason no lemmatization was applied, since dictionary-based lemmatizers would at best be ineffective and at worst detrimental to our approach.

Finally, in the absence of information about the real class distribution, we artificially enforced a uniform class distribution by subsampling the majority class.

4.2 Evaluation metrics and experimental protocol

We chose classification accuracy as our evaluation metric because balanced data renders other threshold metrics (such as F_1-Score) less meaningful and used the standard 10-Fold cross-validation protocol for machine learning experiments (Bishop, 2006) for both our distantly learned approaches and our unsupervised approaches in order to preserve as fair a comparison as possible.

5. RESULT AND DISCUSSION

The results shown in Table 2 show that while a standard lexicon built on a background corpus with a simple bag-of-words representation does not significantly outperform standard sentiment analysis or stance classification techniques, changing the representation of the instances by adding some form of context improves our classification accuracy again by a significant margin (+5.7% compared to a similar approach without context and +7.8% compared to the best baseline).

Dataset/Method	*SentLex*	*SentLex*	*SmartSA*	*USS*	PMILex	PMILex+inter	PMILex+union
IAC	0.5042	0.5260	0.5147	0.5061	0.5362	**0.5899**	0.5696
RNLC	0.4670	0.4664	0.4522	0.4718	0.4843	**0.5304**	0.5043

Table 2: Accuracy of the compared approaches.

We note that between the two approaches to introduce context in the bag of words representation, the best accuracy was given by the approach using the intersection of the bag of words representations of child and parent comments rather than the union (which preserves more information). A possible reason for this is that the intersection of bags of words preserve topically relevant terms which can then be used in the classification process. In the context of classifying argumentative stance, such result implies that evidence of this stance is contained within common information present in both child and parent comment.

However, this does not account for the fact that the union of bags of words outperforms the standard methods while potentially adding noise into the representation. This leads us to think that further refinement of the model could be done by using a term weighting scheme as a middle ground between adding information using the union operator and filtering information using the intersection operator.

The accuracy scores however while being an improvement over the baselines are too low for an operational context. More work is required in improving the naive way in which we added context at classification time.

6. CONCLUSION

Argument mining in the context of classification has much to gain from shallow techniques borrowed from information retrieval, text mining and sentiment analysis research. A comparative analysis of representation techniques for classifying the parent-child relationships in threaded posts show that sentiment analysis approaches can be successfully adopted for argument mining provided that the conversational-context is captured in representation schemes. We show that the simple bag-of-words representation contextualised by parent-child vocabulary intersection leads to significant improvements over comparable baseline approaches. Following on from these results we aim to explore conversational-contextual enrichments that can further improve representation for argument classification. For this in addition to the parent-child single level relationship, we intend exploring further levels of context such as from siblings to ancestors.

REFERENCES

Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., & King, J. (2011, June). How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proc of the Workshop on Languages in Social Media* (pp. 2-11). Association for Computational Linguistics.

Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R., & Minor, M. (2011, June). Cats rule and dogs drool!: Classifying stance in online debate. In *Proc 2nd workshop on computational approaches to subjectivity and sentiment analysis* (pp. 1-9). Association for Computational Linguistics.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Boltuzic, F., & Šnajder, J. (2014, June). Back up your stance: Recognizing arguments in online discussions. In *Proceedings of the First Workshop on Argumentation Mining* (pp. 49-58).

Bousmalis, K., Mehu, M., & Pantic, M. (2013). Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing*, *31*(2), 203-221.

Cabrio, E., & Villata, S. (2012, July). Combining textual entailment and argumentation theory for supporting online debates interactions. In *Proc of the 50th Annual Meeting of the ACL: Short Papers-Volume 2* (pp. 208-212). Association for Computational Linguistics.

Cayrol, C., & Lagasquie-Schiex, M. C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. In *Symbolic and quantitative approaches to reasoning with uncertainty* (pp. 378-389). Springer Berlin Heidelberg.

Cohen, R. (1987). Analyzing the structure of argumentative discourse. *Computational linguistics*, *13*(1-2), 11-24.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, *77*(2), 321-357.

Esuli, A., & Sebastiani, F. (2006, May). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proc of LREC* (Vol. 6, pp. 417-422).

Hutchby, I. (2013). *Confrontation talk: Arguments, asymmetries, and power on talk radio*. Routledge.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1, p. 496). Cambridge: Cambridge university press.

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009, August). Distant supervision for relation extraction without labelled data. In *Proc of the Joint Conf of*

the 47th Annual Meeting of the ACL and the 4th Intl Joint Conf on NLP of the AFNLP: Volume 2-Volume 2 (pp. 1003-1011). ACL.
Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, *19*(1), 1-22.
Muhammad, A., Wiratunga, N., Lothian, R., & Glassey, R. (2013). Domain-Based Lexicon Enhancement for Sentiment Analysis. In *SMA@ BCS-SGAI* (pp. 7-18).
Palau, R. M., & Moens, M. F. (2009, June). Argumentation mining: the detection, classification and structure of arguments in text. In *Proc of the 12th intl conf on AI and law* (pp. 98-107). ACM.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.
Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proc of the ACL-02 conference on Empirical methods in NLP Volume 10*(pp. 79-86). ACL.
Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *Int Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *7*(1), 1-31.
Somasundaran, S., & Wiebe, J. (2010, June). Recognizing stances in ideological on-line debates. In *Proc of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (pp. 116-124). ACL.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.
Van Eemeren, F. H., Grootendorst, R., Jackson, S., & Jacobs, S. (1993). Reconstructing argumentative discourse.
Villalba, M. P. G., & Saint-Dizier, P. (2012, September). Some Facets of Argument Mining for Opinion Analysis. In *COMMA* (pp. 23-34).
Walker, M. A., Anand, P., Abbott, R., & Grant, R. (2012, June). Stance classification using dialogic properties of persuasion. In *Proc of the North American Chapter of the ACL: Human Language Technologies* (pp. 592-596). Association for Computational Linguistics.
Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., & King, J. (2012). A Corpus for Research on Deliberation and Debate. In *LREC* (pp. 812-817).
Wang, L., & Cardie, C. (2014). Improving agreement and disagreement identification in online discussions with a socially-tuned sentiment lexicon. *ACL 2014*, 97.
Wang, W., Yaman, S., Precoda, K., Richey, C., & Raymond, G. (2011, June). Detection of agreement and disagreement in broadcast conversations. In *Proc of the 49th Annual Meeting of the ACL: Human Language Technologies: short papers-Volume 2* (pp. 374-378). ACL.
Wyner, A., Mochales-Palau, R., Moens, M. F., & Milward, D. (2010).*Approaches to text mining arguments from legal cases* (pp. 60-79). Springer Berlin Heidelberg.
Yin, J., Thomas, P., Narang, N., & Paris, C. (2012, July). Unifying local and global agreement and disagreement classification in online debates. In *Proc of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 61-69). ACL.