Collective Stance Classification of Posts in Online Debate Forums

Dhanya Sridhar
Computer Science Dept.
UC Santa Cruz
dsridhar@soe.ucsc.edu

Lise Getoor
Computer Science Dept.
UC Santa Cruz
getoor@soe.ucsc.edu

Marilyn Walker
Computer Science Dept.
UC Santa Cruz
maw@soe.ucsc.edu

Abstract

Online debate sites are a large source of informal and opinion-sharing dialogue on current socio-political issues. Inferring users' stance (PRO or CON) towards discussion topics in domains such as politics or news is an important problem, and is of utility to researchers, government organizations, and companies. Predicting users' stance supports identification of social and political groups, building of better recommender systems, and personalization of users' information preferences to their ideological beliefs. In this paper, we develop a novel collective classification approach to stance classification, which makes use of both structural and linguistic features, and which collectively labels the posts' stance across a network of the users' posts. We identify both linguistic features of the posts and features that capture the underlying relationships between posts and users. We use probabilistic soft logic (PSL) (Bach et al., 2013) to model post stance by leveraging both these local linguistic features as well as the observed network structure of the posts to reason over the dataset. We evaluate our approach on 4FORUMS (Walker et al., 2012b), a collection of discussions from an online debate site on issues ranging from gun control to gay marriage. We show that our collective classification model is able to easily incorporate rich, relational information and outperforms a local model which uses only linguistic information.

1 Introduction

Modeling user stance (PRO, CON) in discussion topics in online social media debate is of interest to researchers, corporations and governmental

organizations alike. Predicting a user's stance towards a given issue can support the identification of social or political groups (Gawron et al., 2012; Abu-Jbara et al., 2012; Anand et al., 2011; Qiu et al., 2013; Hasan and Ng, 2013), help develop better recommendation systems, or tailor users' information preferences to their ideologies and beliefs. Stance classification problems consist of a collection of debate-style discussions by authors on different controversial, political topics.

While these may be spoken as in the Congressional Debates corpus (Thomas et al., 2006; Burfoot, 2008), we focus on forum posts on social media debate sites. Users on debate sites share their opinions freely, using informal and social language, providing a rich and much more challenging domain for stance prediction.

Social media debate sites contain online discussions with posts from various authors, where each post is either a response to another post or the root of the discussion (Anand et al., 2011; Walker et al., 2012a). Posts are linked to one another by either rebuttal or agreement links and are labelled for stance, either PRO or CON, depending on the framing of the issue under discussion. Each post reflects the stance and sentiment of its author. Authors may participate in multiple discussions in the same topic, and may discuss multiple topics. For example consider the sample posts from the online discussion forum 4forums.com shown in Fig. 1. Here, we see discussion topics, together with sample quotes and responses, where the response is a direct reply to the quote text. The annotations for stance were gathered using Amazon's Mechanical Turk service with an interface that allowed annotators to see complete discussions. Quotes provide additional context that were used by human annotators in a separate task for annotating agreement and disagreement (Misra and Walker, 2013). Responses can be labeled as either PRO or CON toward the topic. For the example shown in Fig. 1,

109

Quote **Q**, Response **R**	Stance	Topic
Q: I thought I'd start a new thread for those newcomers who don't want to be shocked by sick minded nazi XXXX. Anyway... When are fetuses really alive, and how many fetuses are actually aborted (murdered) before that time? **R**: The heart starts beating 3 weeks after conception, and you can't live without a beating heart, but me personally, I think that as soon as the miracle starts, (egg and sperm combine) that is when life begins. I know it's more of a spiritual thing for me instead of a fact. :)	CON	**Abortion**
Q2: Most americans support a Federal Marriage Amendment. Defining Marriage as a union between a man and a woman. Federal Marriage Amendment. This is the text of the Amend: Marriage in the United States shall consist only of the union of a man and a woman. Neither this constitution or the constitution of any state, nor state or federal law, shall be construed to require that marital status or the legal incidents thereof be conferred upon unmarried couples or groups. **R2**: Debator, why does it bother you so much that some people are gay? Its a sexual prefference. People like certain things when they have sex. Example: A man likes a women with small boobs. Or, a man likes a women with nice legs. People like the way certain things feel (I'm not giving te example for that one;)). So why does it bother people that someone's sexual prefference is just a little kinkier than thiers?	PRO	**Gay Marriage**

Figure 1: Sample Quote/Response Pair from `4forums.com` with Mechanical Turk annotations for stance. Both response posts are from the same author.

both response posts are from the same author. We describe the dataset further in Section 4.1.

We believe that models of post stance in on-line debate should capture both the content and the context of author posts. By jointly reasoning over both the content of the post and its relationships with other posts in the discussion, we perform collective classification, as we further define in Section 3 (Sen et al., 2008). Previous work has shown that collective classification models often perform better than content-only approaches. (Burfoot et al., 2011; Hasan and Ng, 2013; Thomas et al., 2006; Bansal et al., 2008; Walker et al., 2012c). Here, we develop a collective classification approach for stance prediction which leverages the sentiment conveyed in a post through its language, and the reply links consisting of agreements or rebuttals between posts in a discussion. We implement our approach using *Probabilistic Soft Logic* (PSL) (Bach et al., 2013), a recently introduced tool for collective inference in relational data. We evaluate our model on data from the 4FORUMS online debate site (Walker et al., 2012b).

Section 2 first presents an overview of our approach and then in Section 3.1 we describe the PSL framework in more detail. Section 4 describes the evaluation data and our results showing that the PSL model improves prediction of post stance in the 4Forums dataset. In Section 5 we describe related work, and compare with our proposed approach. Section 6 summarizes our approach and results.

2 Overview of Approach

Given a set of topics $\{t_1 \ldots t_n\}$, where each topic t_i consists of a set of discussions $\{d_{i1} \ldots d_{ij}\}$, we model each discussion d_k as a collection of posts $\{p_{k0}, \ldots, p_{km}\}$, where each post p_{ki} is mapped to its author a_i.

A discussion $d_i \in \mathcal{D}$ is a tree of posts, starting with the initial post p_{i0}. We distinguish between posts that start a new thread (root) and others (non-root). Each non-root post p_{ij} is the response to some previous post p_{ik}, where $k < j$, and we refer to p_{ik} as the parent of p_{ij}. For a subset of the posts, p_{ij} has been annotated with a real valued number in the interval $[-5, 5]$ that denotes whether the post disagrees or agrees with its parent. Values ≤ 0 are considered disagreement and values ≥ 1, as agreement. We discard the posts where the annotations are in the interval $(0, 1)$ since those indicate high annotator uncertainty about agreement.

Fig. 2 illustrates an example of three discussion trees for two topics where author a_2 participates in multiple discussions of the same topic and a_3 and a_4 participate in multiple topics. An author directly replies with a post to another author's post and either disagrees or agrees.

Each post p_{ij} in discussion d_i is also mapped to $\{x_{ij_1}, \ldots, x_{ij_N}\}$ linguistic features as described in Section 3.2.1 as well as y_{ij}, the stance label (PRO, CON) towards the discussion topic t_i.

We say that a_j participates in topic t_i if there exist any posts $p_j \in d_i$ with author a_j.

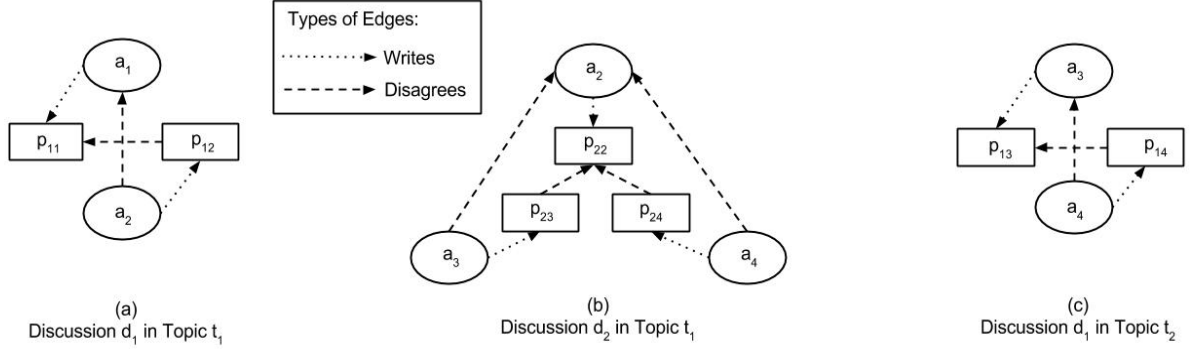Using the tree structure and posts that have annotations for agreement or disagreement, we con-

Figure 2: Example of 3 discussions in (a), (b) and (c). Dotted lines denote the 'writes' relation between authors and posts and dashed lines denote the 'disagrees' relation between posts and between authors. Authors can participate in multiple discussions of the same topic, shown by a_2 in both (a) and (b). Moreover, authors may post in multiple topics, as shown by a_3 and a_4 in both (b) and (c), and may interact with the same authors multiple times, as shown again in (b) and (c).

sider the network graph $\mathcal{G}$ of disagreement and agreement between posts and between authors, where the vertices are posts $\{p_0, \ldots, p_m\}$ and authors $\{a_0, \ldots, a_n\}$. A disagreement edge exists from post p_u to p_v if p_u disagrees with p_v.

A disagreement edge exists from a_w to a_y if any of the posts $\{p_w, \ldots, p_x\}$ mapped to a_w disagree with any posts $\{p_y, \ldots p_z\}$ mapped to a_y. We similarly define agreement edges for both posts and authors.

3 Collective Classification of Stance

Given the discussion structure defined in the previous section, our task is to infer the stance of each post. We make use of both linguistic features and the relational structure in order to *collectively* or *jointly* infer the stance labels. This corresponds to a collective classification setting (Sen et al., 2008), in which we are given a multi-relational network and some partially observed labels, and we wish to infer all of the unobserved labels, conditioned on observed attributes and links. Collective classification refers to the combined classification of a set of interdependent objects (posts, in our domain) using information given by both the local features of the objects and the properties of the objects' neighbors in a network. For the stance classification problem, we infer stance labels for posts using both the correlation between a post and its linguistic attributes $\{x_{ij_1}, \ldots, x_{ij_N}\}$, and the labels and attributes of its neighbors in observed network graph $\mathcal{G}$. We use PSL, described below, to perform collective classification.

3.1 Probabilistic Soft Logic

Probabilistic soft logic (PSL) is a framework for probabilistic modeling and collective reasoning in relational domains (Kimmig et al., 2012; Bach et al., 2013). PSL provides a declarative syntax and uses first-order logic to define a templated undirected graphical model over continuous random variables. Like other statistical relational learning methods, dependencies in the domain are captured by constructing rules with weights that can be learned from data.

But unlike other statistical relational learning methods, PSL relaxes boolean truth values for atoms in the domain to soft truth values in the interval [0,1]. In this setting, finding the *most probable explanation* (MPE), a joint assignment of truth values to all random variable ground atoms, can be done efficiently.

For example, a typical PSL rule looks like the following:

$$P(A, B) \wedge Q(B, C) \rightarrow R(A, C)$$

where P, Q and R are *predicates* that represent observed or unobserved attributes in the domain, and A, B, and C are *variables*. For example, in our 4FORUMS domain, we consider an observed attribute such as *writesPost*(A, P) and infer an unobserved attribute (or label) such as *isProPost*(P, T). Instantiation of predicates with data is called grounding (e.g. *writesPost(A2, P7)*), and each ground predicate, often called ground atom, has a soft truth value in the interval [0,1]. To build a PSL model for stance classification, we represent posts

111

$isProPost(P, T) \wedge writesPost(A, P)$	$\rightarrow$	$isProAuth(A, T)$
$\neg\, isProPost(P, T) \wedge writesPost(A, P)$	$\rightarrow$	$\neg\, isProAuth(A, T)$
$agreesPost(P, P2) \wedge isProPost(P, T)$	$\rightarrow$	$isProPost(P2, T)$
$agreesPost(P, P2) \wedge \neg\, isProPost(P, T)$	$\rightarrow$	$\neg\, isProPost(P2, T)$
$disagreesPost(P, P2) \wedge isProPost(P, T)$	$\rightarrow$	$\neg\, isProPost(P2, T)$
$disagreesPost(P, P2) \wedge \neg\, isProPost(P, T)$	$\rightarrow$	$isProPost(P2, T)$
$agreesAuth(A, A2) \wedge isProAuth(A, T)$	$\rightarrow$	$isProAuth(A, T)$
$agreesAuth(A, A2) \wedge \neg\, isProAuth(A, T)$	$\rightarrow$	$\neg\, isProAuth(A2, T)$
$disagreesAuth(A, A2) \wedge isProAuth(A, T)$	$\rightarrow$	$\neg\, isProAuth(A2, T)$
$disagreesAuth(A, A2) \wedge \neg\, isProAuth(A, T)$	$\rightarrow$	$isProAuth(A2, T)$
$hasLabelPro(P, T)$	$\rightarrow$	$isProPost(P, T)$
$\neg\, hasLabelPro(P, T)$	$\rightarrow$	$\neg\, isProPost(P, T)$

Table 1: Rules for PSL model, where P = post, T = Topic, and A = Author.

and authors as variables and specify predicates to encode different interactions, such as *writes*, between them. Domain knowledge is captured by writing rules with weights that govern the relative importance of the dependencies between predicates. The groundings of all the rules result in an undirected graphical model that represents the joint probability distribution of assignments for all unobserved atoms, conditioned on the observed atoms.

Triangular norms, which are continuous relaxations of logical AND and OR, are used to combine the atoms in the first-order clauses. As a result of the soft truth values and the triangular norms, the underlying probabilistic model is a *hinge-loss Markov Random Field* (HL-MRF). Inference in HL-MRFs is a convex optimization, which leads to a significant improvement in efficiency over discrete probabilistic graphical models. Thus, PSL offers a very natural interface to compactly represent stance classification as a collective classification problem, along with methods to reason about our domain.

3.2 Features

We extract both linguistic features that capture the content of a post and features that capture multiple relations from our dataset.

3.2.1 Linguistic Features

To capture the content of a post, on top of a *bag-of-words* representation with unigrams and bigrams, we also consider basic lengths, discourse cues, repeated punctuation counts and counts of lexical categories based on the Linguistic Inquiry and Word Count tool (LIWC) (Pennebaker et al.,

2001). Basic length features capture the number of sentences, words, and characters, along with the average word and sentence lengths for each post. The discourse cues feature captures frequency counts for the first few words of the post, which often contain discourse cues. To capture the information in repeated punctuation like "!!", "??" or "?!" we include the frequency count of the given punctuation patterns as a feature of each post (Anand et al., 2011). LIWC counts capture sentiment by giving the degree to which the post uses certain categories of subjective language.

3.2.2 Relational Information

As our problem domain contains relations between both authors and posts, for our PSL model, we consider the relations between authors, between posts and between authors and posts. As described above, in PSL, we model these relations as first-order predicates. In Section 3.3, we describe how we populate the predicates with observations from our data.

Author Information We observe that authors participate in discussions by writing posts. For a subset of authors, we have annotations for their interactions with other authors as either disagreement or agreement, as given by network graph $\mathcal{G}$. We encode this with the following predicates: *writesPost*(A, P), *disagreesAuth*(A1, A2), *agreesAuth*(A1, A2).

Post Information Posts are linked to the topic of their given discussion, and to other posts in their discussion through disagreement or agreement. Additionally, we include a predicate for post stance towards its topic as predicted by a classifier

that only uses linguistic features, as described in Section 3.3, as prior information. We capture these relations with the following predicates: *hasLabelPro*(P, T), *hasTopic*(P, T), *disagreesPost*(P1, P2), *agreesPost*(P1, P2).

3.2.3 Target attributes

Our goal is to 1) predict the stance relation between a post and its topic, namely, PRO or CON and 2) predict the stance relation between an author and a topic. In our PSL model, our target predicates are *isProPost*(P, T) and *isProAuth*(A, T).

3.3 PSL Model

We construct our collective stance classification model in PSL using the predicates listed above. For disagreement/agreement annotations in the interval [-5, 5], we consider values [-5,0] as evidence for the *disagreesAuth* relation and values [1, 5] as evidence for the *agreesAuth* relation. We discard observations with annotations in the interval [0,1] because it indicates a very weak signal of agreement, which is already rare on debate sites. We populate *disagreesPost* and *agreesPost* in the same way as described above.

For each relation, we populate the corresponding predicate with all the instances that we observe in data and we fix the truth value of each observation as 1. For all such predicates where we observe instances in the data, we say that the predicate is closed. For the relations *isPostPro* and *isAuthPro* that we predict through inference, a truth value of 1 denotes a PRO stance and a truth value of 0 denotes a CON stance. We say that those predicates are open, and the goal of inference is to jointly assign truth values to groundings of those predicates.

We use our domain knowledge to describe rules that relate these predicates to one another. We follow our intuition that agreement between nodes implies that they have the same stance, and disagreement between nodes implies that they have opposite stances. We relate post and author nodes to each other by supposing that if a post is PRO towards its topic, then its author will also be PRO towards that topic.

We construct a classifier that takes as input the linguistic features of the posts and outputs predictions for stance label of each post. We then consider the labels predicted by the local classifier as a prior for the inference of the target attributes in our PSL model. Table 1 shows the rules in our PSL model.

Topic	Authors	Posts
Abortion	385	8114
Evolution	325	6186
Gun Control	319	3899
Gay Marriage	316	7025
Death Penalty	170	572

Table 2: Overview of topics in 4FORUMSdataset.

4 Experimental Evaluation

We first describe the dataset we use for evaluation and then describe our evaluation method and results.

4.1 Dataset

We evaluate our proposed approach on discussions from `https://www.4forums.com`, an online debate site on social and political issues. The dataset is publicly available as part of the Internet Argument Corpus, an annotated collection of 109,533 forum posts (Walker et al., 2012b; Walker et al., 2012c). On 4FORUMS, a user initiates a discussion by posting a new question or comment under a topic, or participate in an ongoing discussion by replying to any of the posts in the thread. The discussions were given to English speaking Mechanical Turk annotators for a number of annotation tasks to get labels for the stances of discussion participants towards the topic, and scores for each post in a discussion indicating whether it is in agreement or disagreement with the preceding post.

The scores for agreement and disagreement were on a 11 point scale [-5, 5] implemented using a slider, and annotators were given quote/response pairs to determine if the response text agreed or disagreed with the quote text. We use the mean score across the 5-7 annotators used in the task. A more negative value indicates higher inter-annotator confidence of disagreement, and a more positive value indicates higher confidence of agreement. The gold-standard annotation used for the stance label of each post is given by the majority annotation among 3-8 Mechanical Turk annotators performed as a separate task, using entire discussions to determine the stance of the authors in the discussion towards the topic. We use the stance of each post's author to determine the post's stance. For our experiments, we use all posts with annotations for stance, and about 90% of these posts also have annotations for agree-

ment/disagreement.

The discussions span many topics, and Table 2 gives a summary of the topics we consider in our experiments and the distribution of posts across these topics. Each post in a discussion comes as a quote-response pair, where the quote is the text that the post is in response to, and the response is the post text. We refer to (Walker et al., 2012b) for a full description of the corpus and the annotation process.

4.2 Evaluation

In order to evaluate our methods, we split the dataset into training and testing sets by randomly selecting half the authors from each topic and their posts for the training set and using the remaining authors and their posts for the test set. This way, we ensure that no two authors appear in both training and test sets for the same topic, since stance is topic-dependent. We create 10 randomly sampled train/test splits for evaluation. Each split contains about 18,000 posts. For each train/test split, we train a linear SVM for each topic, with the L2-regularized-L1-loss SVM implemented in the LibLINEAR package (Fan et al., 2008). We use only the linguistic features from the posts, for each topic in the training set. We refer to the baseline model which only uses the the output of the SVM as the LOCAL model. We output the predictions from LOCAL model and get stance labels for posts in both the training and test sets. We use the predictions as prior information for the true stance label in our PSL model, with the *hasLabel* predicate.

We use the gold standard stance annotation (PRO, CON) for each post as ground truth for weight learning and inference. A truth value of 1 for *isPostPro* and *isAuthPro* denotes a PRO stance and a truth value of 0 denotes a CON stance. We learn the weights of our PSL model (initially set to 1) for each of our training sets and perform inference on each of the test sets.

Table 3 shows averages for F1 score for the positive class (PRO), area under the precision-recall curve (AUC-PR) for the negative class (CON) and area under the ROC curve (AUROC) over the 10 train/test splits. For the PSL model, the measures are computed for joint inference over all topics in the test sets. For the per-topic linear SVMs (LOCAL model), we compute the measures individually for the predictions of each topic in the test sets and take a weighted average over the

topics. Our PSL model outperforms the LOCAL model, with statistically significant improvements in the F1 score and AUC-PR for the negative class. Moreover, our model completes weight learning and inference on the order of seconds, boasting an advantage in computational efficiency, while also maintaining model interpretability.

Table 4 shows the weights learned by the PSL model for the rules in one of the train/test splits of the experiment. The first two rules relating post stance and author stance are weighted more heavily, in part because the *writesPost* predicate has a grounding for each author-post pair and contributes to lots of groundings of the rule. The rules that capture the alternating disagreement stance also have significant weight, while the rules denoting agreement both between posts and between authors are weighted least heavily since there are far fewer instances of agreement than disagreement. This matches our intuition of political debates.

We also explored variations of the PSL model by removing the first two rules relating post stance and author stance and found that the weight learning algorithm drove the weights of the other rules close to 0, worsening the performance. We also removed rules 3-10 that capture agreement/disagreement from the model, and found that the model performs poorly when disregarding the links between nodes entirely. The PSL model learns to weight the first and second rule very highly, and does worse than when considering the prior alone. Thus, the combination of the rules gives the model its advantage, allowing the PSL model to make use of a richer structure that has multiple types of relations and more information.

5 Related Work

Over the last ten years, there has been significant progress on modeling stance. Previous work covers three different debate settings: (1) congressional debates (Thomas et al., 2006; Bansal et al., 2008; Yessenalina et al., 2010; Balahur et al., 2009); (2) company-internal discussion sites (Murakami and Raymond, 2010; Agrawal et al., 2003); and (3) online social and political public forums (Somasundaran and Wiebe, 2009; Somasundaran and Wiebe, 2010; Wang and Rosé, 2010; Biran and Rambow, 2011; Walker et al., 2012c; Anand et al., 2011). Debates in online public forums (e.g. Fig. 1) differ from debates in congress and on company discussion sites because the posts are

Classifier	F1 Score	AUC-PR negative class	AUROC
LOCAL	0.66 ± 0.015	0.44 ± 0.04	0.54 ± 0.02
PSL	0.74 ± 0.04	0.511 ± 0.04	0.59 ± 0.05

Table 3: Averages and standard deviations for F1 score for the positive class, area under PR curve for the negative class, and area under ROC curve for post stance over 10 train/test splits.

isProPost(P, T) $\wedge$ writesPost(A, P)	$\rightarrow$	*isProAuth(A, T)* : 10.2
$\neg$ *isProPost(P, T) $\wedge$ writesPost(A, P)*	$\rightarrow$	$\neg$ *isProAuth(A, T)* : 8.5
agreesPost(P, P2) $\wedge$ isProPost(P, T)	$\rightarrow$	*isProPost(P2, T)* : 0.003
agreesPost(P, P2) $\wedge \neg$ isProPost(P, T)	$\rightarrow$	$\neg$ *isProPost(P2, T)* : 0.003
disagreesPost(P, P2) $\wedge$ isProPost(P, T)	$\rightarrow$	$\neg$ *isProPost(P2, T)* : 0.06
disagreesPost(P, P2) $\wedge \neg$ isProPost(P, T)	$\rightarrow$	*isProPost(P2, T)* : 0.11
agreesAuth(A, A2) $\wedge$ isProAuth(A, T)	$\rightarrow$	*isProAuth(A, T)* : 0.001
agreesAuth(A, A2) $\wedge \neg$ isProAuth(A, T)	$\rightarrow$	$\neg$ *isProAuth(A2, T)* : 0.0
disagreesAuth(A, A2) $\wedge$ isProAuth(A, T)	$\rightarrow$	$\neg$ *isProAuth(A2, T)* : 0.23
disagreesAuth(A, A2) $\wedge \neg$ isProAuth(A, T)	$\rightarrow$	*isProAuth(A2, T)* : 0.6
hasLabelPro(P, T)	$\rightarrow$	*isProPost(P, T)* : 2.2
$\neg$ *hasLabelPro(P, T)*	$\rightarrow$	$\neg$ *isProPost(P, T)* : 4.8

Table 4: Weights learned by the model for the PSL rules in train/test split 2 of experiments

shorter and the language is more informal and social. We predict that this difference makes it more difficult to achieve accuracies as high for 4FORUMS discussions as can be achieved for the congressional debates corpus.

Work by (Somasundaran and Wiebe, 2009) on ideological debates very similar to our own show that identifying argumentation structure improves performance; their best performance is approximately 64% accuracy over all topics. Research by (Thomas et al., 2006; Bansal et al., 2008; Yessenalina et al., 2010; Balahur et al., 2009) classifies the speaker's stance in a corpus of congressional floor debates. This work combines graph-based and text-classification approaches to achieve 75% accuracy on Congressional debate siding over all topics. Other work applies MaxCut to the reply structure of company discussion forums (Malouf and Mullen, 2008; Murakami and Raymond, 2010; Agrawal et al., 2003). Murakami & Raymond (2010) show that rules for identifying agreement, defined on the textual content of the post improve performance.

More recent work has explicitly focused on the benefits of collective classification in these settings (Burfoot et al., 2011; Hasan and Ng, 2013; Walker et al., 2012c), and has shown in each case that collective classification improves performance. The results reported here are the first to

apply the PSL collective classification framework to the forums conversations from the IAC corpus (Anand et al., 2011; Walker et al., 2012c).

6 Discussion and Future Work

Here, we introduce a novel approach to classify stance of posts from online debate forums with a collective classification framework. We formally construct a model, using PSL, to capture relational information in the network of authors and posts and our intuition that agreement or disagreement between users correlates to their stance towards a topic. Our initial results are promising, showing that by incorporating more complex interactions between authors and posts, we gain improvements over a content-only approach. Our approach is ideally suited to collective inference in social media. It easily extendable to use additional relational information, and richer behavioral and linguistic information.

Acknowledgments

References

Amjad Abu-Jbara, Mona Diab, Pradeep Dasigi, and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Association for Computational Linguistics (ACL)*, pages 399–409.

R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu. 2003. Mining newsgroups using networks arising from social behavior. In *International Conference on World Wide Web (WWW)*, pages 529–535. ACM.

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 *ACL Workshop on Sentiment and Subjectivity*.

Stephen H. Bach, Bert Huang, Ben London, and Lise Getoor. 2013. Hinge-loss markov random fields: Convex inference for structured prediction. In *Uncertainty in Artificial Intelligence (UAI)*.

A. Balahur, Z. Kozareva, and A. Montoyo. 2009. Determining the polarity and source of opinions expressed in political debates. *Computational Linguistics and Intelligent Text Processing*, pages 468–480.

M. Bansal, C. Cardie, and L. Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *COLING*, pages 13–16.

O. Biran and O. Rambow. 2011. Identifying justifications in written dialogs. In *IEEE International Conference on Semantic Computing (ICSC)*, pages 162–168.

Clinton Burfoot, Steven Bird, and Timothy Baldwin. 2011. Collective classification of congressional floor-debate transcripts. In *Association for Computational Linguistics (ACL)*, pages 1506–1515.

C. Burfoot. 2008. Using multiple sources of agreement information for sentiment classification of political transcripts. In *Australasian Language Technology Association Workshop*, volume 6, pages 11–18.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

J.M. Gawron, D. Gupta, K. Stephens, M.H. Tsou, B. Spitzberg, and L. An. 2012. Using group membership markers for group identification in web logs. In *AAAI Conference on Weblogs and Social Media (ICWSM)*.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. *International Joint Conference on Natural Language Processing*.

Angelika Kimmig, Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2012. A short introduction to probabilistic soft logic. In *NIPS Workshop on Probabilistic Programming: Foundations and Applications*.

R. Malouf and T. Mullen. 2008. Taking sides: User classification for informal online political discourse. *Internet Research*, 18(2):177–190.

Amita Misra and Marilyn A Walker. 2013. Topic independent identification of agreement and disagreement in social media dialogue. In *Conference of the Special Interest Group on Discourse and Dialogue*, page 920.

A. Murakami and R. Raymond. 2010. Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions. In *International Conference on Computational Linguistics (ACL)*, pages 869–875.

J. W. Pennebaker, L. E. Francis, and R. J. Booth, 2001. *LIWC: Linguistic Inquiry and Word Count*.

Minghui Qiu, Liu Yang, and Jing Jiang. 2013. Modeling interaction features for debate side clustering. In *ACM International Conference on Information & Knowledge Management (CIKM)*, pages 873–878.

Prithviraj Sen, Galileo Mark Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI Magazine*, 29(3):93–106.

S. Somasundaran and J. Wiebe. 2009. Recognizing stances in online debates. In *ACL and AFNLP*, pages 226–234.

S. Somasundaran and J. Wiebe. 2010. Recognizing stances in ideological on-line debates. In *NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124.

M. Thomas, B. Pang, and L. Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 327–335.

Marilyn Walker, Pranav Anand, Rob Abbott, Jean E. Fox Tree, Craig Martell, and Joseph King. 2012a. That's your evidence?: Classifying stance in online political debate. *Decision Support Sciences*.

Marilyn Walker, Pranav Anand, Robert Abbott, and Jean E. Fox Tree. 2012b. A corpus for research on deliberation and debate. In *Language Resources and Evaluation Conference, LREC2012*.

Marilyn Walker, Pranav Anand, Robert Abbott, and Richard Grant. 2012c. Stance classification using dialogic properties of persuasion. In *Meeting of the North American Association for Computational Linguistics. NAACL-HLT12*.

Y.C. Wang and C.P. Rosé. 2010. Making conversational structure explicit: identification of initiation-response pairs within online discussions. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 673–676.

A. Yessenalina, Y. Yue, and C. Cardie. 2010. Multi-level structured models for document-level sentiment classification. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1046–1056.