Collective Classification of Stance and Disagreement in Online Debate Forums

Dhanya Sridhar **James Foulds**
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064
{dsridhar, jfoulds}@ucsc.edu

Bert Huang
Department of Computer Science
University of Maryland
College Park, MD 20742
bert@cs.umd.edu

Marilyn Walker **Lise Getoor**
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064
{maw, getoor}@soe.ucsc.edu

1 Introduction

In online debate websites such as DEBATE.ORG, CREATEDEBATE, CONVINCEME.NET, and 4FORUMS, users debate and share their opinions on a variety of social and political issues. The debates typically consist of online discussion threads, each focused on a specific question or issue. Modeling the stances of the users towards these issues is of interest to researchers, internet companies and governmental organizations alike. Prediction of user stance can support the identification of social or political groups, and can provide valuable user modeling information for recommender systems. A growing body of work on classification for stance has found it to be a challenging problem [1, 3, 4]. As the interactions on these social media debate websites are inherently dialogic in nature, they have also proved useful for the computational modeling of dialogue. In particular, debates provide a fertile ground for the study of *disagreement* between authors [5].

Stance and disagreement are closely related. Disagreement between authors in the text of their posts is strong evidence that the authors may have opposing stances with regards to the topic of the discussion thread, and vice-versa. The relational nature of disagreement (and dialogue in general) suggests a statistical relational approach for modeling stance and disagreement together, leveraging the relational information present in the data to improve prediction. In this work, we propose a collective classification approach for jointly modeling stance and disagreement.

2 Collective Classification Model

Our approach begins by constructing local predictors of author stance and author-author disagreement. We use linguistic features to construct logistic regression classifiers for the stance of each user (FOR or AGAINST), and for the presence of a disagreement in the text of each post (TEXTAGREE or TEXTDISAGREE), trained on Amazon Mechanical Turk annotations. Note that a TEXTDISAGREE label for a post indicates a disagreement in the discussion but does not necessarily indicate that its author has the opposite stance to the author of the parent post. For this reason, we then train a logistic regression classifier for *stance disagreement*, i.e. opposite stance values, between pairs of authors who interact via replies to each others posts (STANCEAGREE or STANCEDISAGREE). The stance disagreement classifier uses the output of the textual disagreement classifier, as well as linguistic features, and extra-linguistic features such as the number of replies between the authors.

We then define a Markov random field model over the stance and disagreement predictions which allows these predictors to inform each other. More specifically, the model uses a hinge-loss MRF [2] formulation to perform the collective classification. Hinge-loss MRFs admit efficient, scalable inference, as finding the most probable explanation (MPE) is a convex optimization algorithm which can

Topic	**Local Classifier**			**Collective Classifier**		
4Forums	**Accuracy**	**AUC**	**LL**	**Accuracy**	**AUC**	**LL**
Abortion	60.2 ± 2.8	0.57 ± 0.02	-157.4	60.7 ± 3.2	**0.61 ± 0.02**	**-99.4**
Evolution	74.8 ± 3.9	0.59 ± 0.03	-120.6	**75.9 ± 3.1**	**0.64 ± 0.05**	**-66.1**
Gun Control	64.3 ± 2.4	0.54 ± 0.03	-120.5	64.7 ± 2.6	**0.59 ± 0.06**	**-70.2**
Gay Marriage	69.4 ± 4.7	0.55 ± 0.04	-142.4	**70.6 ± 4.6**	**0.63 ± 0.06**	**-68.5**
CreateDebate	**Accuracy**	**AUC**	**LL**	**Accuracy**	**AUC**	**LL**
Abortion	59.4 ± 3.7	0.59 ± 0.04	-86.1	60.3 ± 3.5	**0.65 ± 0.07**	**-43.5**
Gay Rights	64.1 ± 8.2	0.59 ± 0.06	-123.1	**65.4 ± 8.1**	**0.7 ± 0.1**	**-52.9**
Obama	58.2 ± 5.6	0.58 ± 0.06	-69.4	59.5 ± 6.3	0.62 ± 0.08	**-29.3**
Marijuana	61.1 ± 7.6	0.49 ± 0.05	-57.7	60.8 ± 8.8	**0.56 ± 0.08**	**-32.9**

Table 1: Averages and standard deviations for classification accuracy, area under the ROC curve, and log-likelihood for author stance, computed over 10 train/test splits. Results in bold indicate statistically significant differences at $\alpha = 0.05$.

be rapidly solved using ADMM. Furthermore, they are defined over continuous random variables, making them useful for modeling classification probabilities, and they are easy to specify using an intuitive logical language called *Probabilistic Soft Logic* (PSL). We specify the MRF by using PSL rules to define hinge-loss potential functions which encourage the global predictions to be similar to the local predictions, while simultaneously encouraging stance and disagreement to be consistent with each other.

3 Evaluation / Discussion

We evaluated our model on four forum topics from the 4Forums online debate website [6], and four topics from the CreateDebate website [4]. Our proposed collective classification approach improved AUC and log-likelihood for stance prediction over the content only approach in all the datasets, and improved accuracy in all but one, with the majority of these improvements being statistically significant. Due to the efficient nature of the HL-MRF formulation, the MRF inference and weight learning were completed in just over a minute on average.

In relation to the approach of [4], our method uses a model-based framework for collective classification, reasons over the full network structure instead of treating the network as sequence data, and incorporates prediction of disagreement. While [3] also use an MRF for collective classification of stance, our approach studies disagreement in a more nuanced way, and our MRF formulation admits exact inference. An empirical comparison to other state of the art approaches is ongoing work.

References

[1] Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. Cats Rule and Dogs Drool: Classifying Stance in Online Debate. In *ACL Workshop on Sentiment and Subjectivity*, 2011.

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

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

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

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

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