

Structured Representation Learning for Online Debate Stance Prediction

Chang Li Aldo Porco Dan Goldwasser

Department of Computer Science
Purdue University, West Lafayette, IN 47907
{li1873, aporco, dgoldwas}@purdue.edu

Abstract

Online debates can help provide valuable information about various perspectives on a wide range of issues. However, understanding the stances expressed in these debates is a highly challenging task, which requires modeling both textual content and users’ conversational interactions. Current approaches take a collective classification approach, which ignores the relationships between different debate topics.

In this work, we suggest to view this task as a representation learning problem, and embed the text and authors jointly based on their interactions. We evaluate our model over the Internet Argumentation Corpus, and compare different approaches for structural information embedding. Experimental results show that our model can achieve significantly better results compared to previous competitive models.

1 Introduction

In recent years, social media platforms play an increasingly important role in shaping political discourse. Online debate forums allow users to voice their opinions and engage with other users holding different views. Understanding the interactions between the users on these platforms can help provide insight into current political discourse, argumentation strategies and can help gauge public sentiment on policy issues on a large scale. The importance of understanding debate dialog has motivated significant research efforts (Somasundaran and Wiebe, 2010; Anand et al., 2011; Ghosh et al., 2014; Walker et al., 2012b; Hasan and Ng, 2013; Sridhar et al., 2015; Mohammad et al., 2016; Dong et al., 2017).

In this paper, we focus on stance prediction, automatically identifying the stance expressed in debate posts on various issues. For example, Figure 1 describes a short debate dialog about marijuana legalization between three users (denoted a_1 , a_2 , a_3). The content associated with each user is classified as supportive of legalization (PRO), or not (CON).

Author	Discussion Text	Stance
a_1	<i>“There are no deaths related to the actual use for marijuana this past year”</i>	Pro
a_2	<i>“Whether it kills people or not, it still is harmful to your body.”</i>	Con
a_3	<i>“There are many things that are harmful. Alcohol is more harmful than marijuana.”</i>	Pro
a_2	<i>“But that doesn’t mean marijuana should be made legal because the other two are.”</i>	Con
a_3	<i>“If we were to make everything illegal because it was harmful we would be living with nothing.”</i>	Pro

Figure 1: Example of excerpts from a debate between three users about marijuana legalization.

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

Early work took a text classification approach (Somasundaran and Wiebe, 2010; Anand et al., 2011), classifying individual posts using a rich feature set. Since debate posts are not written in isolation, but rather express the conversational interactions between users, modeling these interactions can help alleviate some of the difficulty of this task. More recent work take a collective classification approach (Hasan and Ng, 2013; Sridhar et al., 2015), which models the dependencies between authors and their content and captures the debate structure. For example, the interactions between users can express *agreements* (or *disagreements*), which would entail a similar (or different) stance prediction associated with their content. The stance decision can also be considered as a user level decision, as users tend to maintain the same stance throughout the debate, forcing stance agreement between all of their posts. Unfortunately, despite these efforts, stance classification remains a challenging problem.

In this paper we suggest a new approach for representing the structural dependencies of debate dialogs, by taking a *structured representation learning* approach. Intuitively, our system is designed to exploit the advantages of collective relational classification methods (often discussed in the context of graphical models) and distributed representation learning (often discussed in the context of deep learning and embedding). We suggest a method for combining the two approaches in a single framework that can exploit their complimentary strengths.

Our key intuition is that the embedding function can be trained to respect the relevant structural dependencies. We jointly embed all the debate objects (i.e., authors, stances and textual posts), by considering the relationships between these objects. For example, we model stance classification as a relationship between a post and a given stance label, by measuring the similarity between their embedded representations. We can also model the relationships between input objects; the similarity between the representations of two posts would entail agreement between the labels associated with them, thus allowing us to perform collective classification over all the input instances. Specifically, we define the factor graph corresponding to the dependencies between stance predictions in a debate thread, and use the similarity between the embedded representation of objects as a scoring function for the factors. We explain this process in more detail in Section 3.

The main strength of distributed representations is in their ability to share information between the represented objects. We exploit this property, and show that by adding additional information to the embedding space, the overall performance of the model improves, *even if this information is not directly relevant to the classification task*. We demonstrate this fact by comparing stance prediction performance, when trained over the multiple topic separately or jointly (thus allowing the model to share information between the representations of multiple debate topics).

We evaluate our approach over the Internet Argument Corpus (Walker et al., 2012a; Abbott et al., 2016), collected from two debate websites, CREATEDEBATE and 4FORUMS. We conduct several experiments, both using in-domain data and out-of-domain data (when we train and test on different debate topics). Our experiments show that formulating the problem as structured representation learning indeed allows debate entities to share information and generalize better, resulting in even larger improvements when multiple stances (corresponding to different output labels) are trained jointly. Furthermore, we show that by using inference over the relationships between the learned representations we can outperform traditional collective classification methods.¹

Our contributions include (1) joint relational embedding for debate entities, allowing the model to share information between related topics and underlying ideologies (2) suggest a collective classification approach, defined over the embedding space, and using it to cast representation learning as a structured prediction problem, and (3) an extensive experimental study in which we evaluate several different modeling choices and information sharing scenarios.

2 Related Work

Stance prediction in online debates is an important subjectivity classification task. Early work viewed the problem as a binary classification task and focused on feature representations (Somasundaran and Wiebe, 2010; Anand et al., 2011), while later work took a collective approach (Walker et al., 2012b;

¹Please refer to <https://github.com/BillMcGrady/StancePrediction> for data and source code.

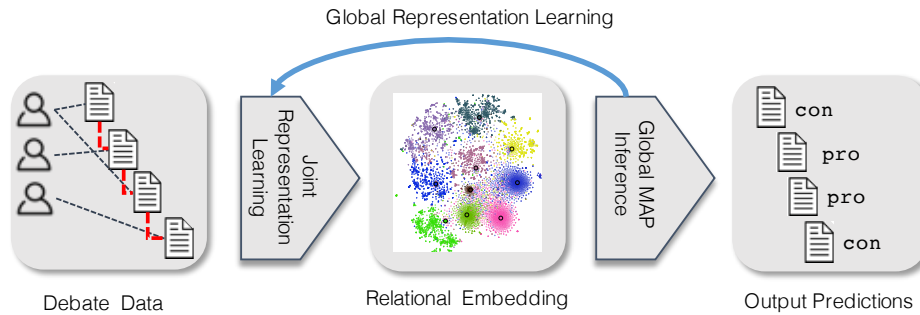


Figure 2: Overall Learning and Inference Processes

Hasan and Ng, 2013; Sridhar et al., 2015). Stance prediction is not limited to online debates, as was also studied in the context of congressional speeches (Bansal et al., 2008; Burfoot et al., 2011) and social media outlets, such as Twitter (Johnson and Goldwasser, 2016; Augenstein et al., 2016; Ebrahimi et al., 2016), including a recent SemEval-16 task (Mohammad et al., 2016). While most work view the task as supervised classification tasks, several work suggest exploiting the interactions between users as a form of distant supervision (Johnson and Goldwasser, 2016; Dong et al., 2017). This task is broadly related to argumentation mining (Ghosh et al., 2014) and stance reason classification (Hasan and Ng, 2014).

Our technical work relies on exploiting distributed representations (i.e., embedding), building on highly influential work on embedding words (Mikolov et al., 2013b; Pennington et al., 2014), sentences (Kiros et al., 2015) and even full documents (Le and Mikolov, 2014). Our work explores the connections between text, users and attributes, attempting to create a common representation for the them. The closest to our work is (Li et al., 2015), that jointly integrates different kinds of cues (text, attribute, graph) into a single latent representation to get user embeddings.

Our work is also broadly related to deep learning methods that capture the structural dependencies between decisions. This can be done either by modeling the dependencies between the hidden representations of connected decisions using RNN/LSTM (Vaswani et al., 2016; Katiyar and Cardie, 2016), or by explicitly modeling the structural dependencies between output predictions (Durrett and Klein, 2015; Lample et al., 2016; Andor et al., 2016). Unlike these work, we formulate our problem as a structured representation learning problem, which to our knowledge is the first work to identify the ties between the two problems.

3 Model Overview

In this paper we suggest casting stance classification as a structured representation learning task. Our approach revolves around two key ideas.

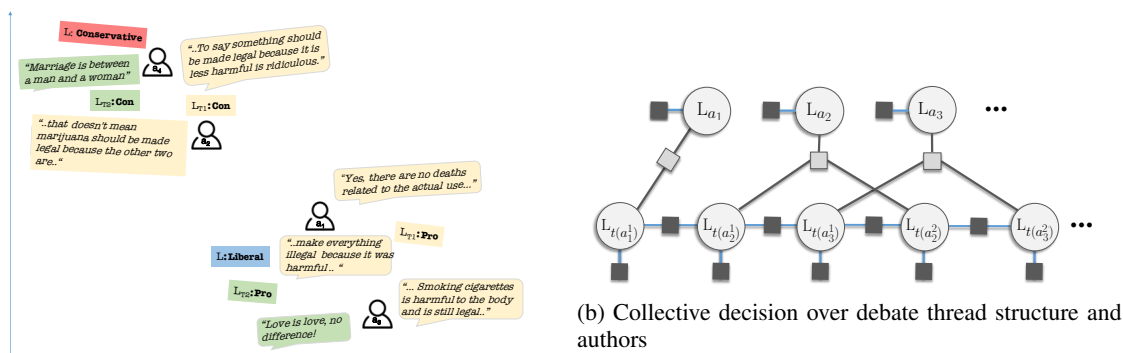
First, stance classification can be done by embedding both the input objects (i.e., posts) and the output labels in the same space. The actual classification is performed by comparing the similarity between the embedded representations of an input object and the competing output labels.

Second, we can augment this representation with additional structural constraints, capturing relevant domain information, such as the connection between posts by the same author, the disagreements between debate participants.

To help clarify these ideas intuitively, consider the debate dialog in Figure 1. Our learning approach uses the structural and textual information in the dialog in three ways, as shown in the process depicted in Figure 2.

1. Joint Representation Learning The embedding learning objective is designed to represent relevant relational information, allowing the representation of different input objects to share information. For example, stances on different topics may share a similar ideology. Figure 3a demonstrates the joint embedding space. The relationship between authors and their posts is preserved by the proximity of their embedded representations. Similar relationships between posts and their corresponding stances and

underlying ideologies are also represented. To accomplish this goal we define a joint objective function over different relations.



(a) Relational embedding representing authors, stances and posts on various topics in the same embedding space

(b) Collective decision over debate thread structure and authors

Figure 3: Model Overview Demonstration

The model is trained to maximize the similarity between corresponding entities (positive examples) compared to irrelevant ones (negative examples). We define the positive examples based on relational information, and negative examples to capture disagreements between authors and posts. We explain this process in detail in Section 5.

2. Global MAP Inference (Collective Classification) Representing the input objects and their labels in the same embedding space allows us to reason about the relationships between them. We view the prediction task as a collective classification, in which all the posts in one or more given debate threads are decided together. We model inference required for the MAP assignment using a factor graph. For example, the graph described in Figure 3b, contains nodes corresponding to author level stance decisions (denoted L_{a_i}), their posts levels stance decision (denoted $L_{t(a_i^j)}$). We score these decisions using the learned embedding. For example, scoring the output assignment PRO to the post corresponding to $L_{t(a_i^j)}$ will be done by observing the similarity (dot product) between their vectorized representations v_{PRO} and $v_{t(a_i^j)}$.

Factor nodes can either have a degree of 1 (scoring the similarity between an author or post and an output label), or 2 (scoring the relationship between consecutive posts in a debate discussion thread). We also allow hard constraints (light gray factors in Figure 3b), which force the model to produce consistent assignments. We explain this process in detail in Section 4.

3. Global Representation Learning A natural extension is to combine the previous two steps, and adopt a global training approach that uses joint prediction during training. In this case the loss function used when learning the embedding is defined with respect to the structural dependencies imposed by the factor graph. This approach is similar in spirit to deep structured learning approaches (Andor et al., 2016), however, in this case the structured learning process is defined directly over the embedding space. This process is explained in Section 5.5.

4 Collective Classification

Our joint embedding model maps authors, attributes, and text into the same space. Thus it allows us to compute the similarity between any pair of authors, texts, attributes, or their combination. This is a very useful property, as information from all aspects can now be used for predicting the target of interest. For example, more information are available for identifying the stance of a post by using its author and neighboring posts comparing to the post's embedding alone. We exploit this property by defining the classification as a global inference process, enforcing the constraints and preferences on all of the predictions.

ILP Formulation We exploit the dependencies described above, using joint prediction over the different aspects. We formulate the decision as an Integer Linear Programming (ILP) which allows us to enforce the consistency or preferences between decisions. The ILP objective function is defined over the similarity scores between objects’ vector representation in the joint embedding space. Since integer linear constraints over 0-1 variables can represent logical constraints, we define the ILP constraints using both representations to help improve readability.

In the stance prediction task, all the posts from multiple debate threads that potentially share authors form a single ILP instance. The ILP global optimization objective is defined over authors a_i , the textual content (posts) $\{t_i^0, \dots, t_i^k\}$ associated with a_i , and other textual posts $\{t_m^p, \dots, t_l^q\}$, responding to or responded by a_i ’s posts.

We create different types of boolean decision variables corresponding to the decision tasks above. We assign a boolean variable $\text{AuthorLabel}(a_i, r_j)$ to represent author a_i has attribute r_j (i.e., its stance), and associate a score $\text{sim}(e_{a_i}, e_{r_j})$ with that variable. Similarly, we assign a boolean variable $\text{TextLabel}(t_i^k, r_j)$ to represent that the text t_i^k is labeled with an attribute r_j , and associate a score $\text{sim}(e_{t_i^k}, e_{r_j})$ with that variable.

To ensure the consistency of the predicted variables, we define two types of constraints.

1. *Single output value on a debate topic:*

$$\forall i \sum_j \text{AuthorLabel}(a_i, r_j) = 1$$

$$\forall i, k \sum_j \text{TextLabel}(t_i^k, r_j) = 1$$

2. *Output consistency:*

$$\forall i, j, k \text{ AuthorLabel}(a_i, r_j) = \text{TextLabel}(t_i^k, r_j)$$

Note that in the debate domain, this constraint forces agreement between all the posts by the same author.

We also add variables capturing the dependencies between connected posts. For debate threads, a boolean variable $\text{Disagree}(t_i^p, t_l^q)$ is created for any two posts t_i^p, t_l^q when t_l^q is a response to t_i^p , and associate a score $\text{disagree_parameter}$ with that variable. This score is a hyper-parameter for local models, capturing the preference towards disagreement between consecutive posts in a debate. It is set according to the training set. When using global learning, it is also included in training, such that similarity scores of consecutive posts will be adjusted appropriately (similar intuition as a margin constraint).

$$\forall t_i^p, t_l^q \text{ Disagree}(t_i^p, t_l^q) \wedge \text{TextLabel}(t_i^p, r_j) \rightarrow \neg \text{TextLabel}(t_l^q, r_j)$$

The set of all possible decisions for the three set of variables are denoted as A for AuthorLabel , B for TextLabel , Γ for Disagree .

Given these variables, our prediction function can be define as follows -

$$\arg \max_{\alpha, \beta, \gamma} \sum_{\alpha \in A} \alpha \cdot \text{score}(\alpha) + \sum_{\beta \in B} \beta \cdot \text{score}(\beta) + \sum_{\gamma \in \Gamma} \gamma \cdot \text{score}(\gamma)$$

Subject To **C**

Where **C** is a set of constraints defined above.

5 Representation Learning

5.1 Embedding Perspectives

Let A and T denote the set of all authors and text respectively, let R denote the set of all attributes for those authors and text. Stance on various topics are the major attributes considered in this paper. For each topic, we have an embedding vector for the Pro stance and another vector for the Con stance, such as $\text{Pro}_{\text{abortion}}$ and $\text{Con}_{\text{abortion}}$. We train our embedding over multiple views of the data, each view connecting users and their content.

Author vs. Text: This objective is to predict text t_j linked with author a_i given the author representation. Each post is a text unit in our experiments.

$$L_{AT} = \sum_{i=1}^n \sum_{j=1}^{text_{a_i}} \log P(t_j|a_i) \quad (1)$$

Author vs. Attribute: This objective is to predict attribute r_j linked with author a_i given the author representation. Stance on different topics and user profile information form the attributes set in debate datasets. Each user attribute value (e.g. male or female in gender attribute) is represented by a vector.

$$L_{AR} = \sum_{i=1}^n \sum_{j=1}^{attri_{a_i}} \log P(r_j|a_i) \quad (2)$$

Text vs. Attribute: This objective is to predict attribute r_j of the text given text t_i . In our experiments, we only used the stance label as attributes of text. However, it may also be possible to inherit attributes from the author of the text.

$$L_{TR} = \sum_{i=1}^m \sum_{j=1}^{attri_{t_i}} \log P(r_j|t_i) \quad (3)$$

Text vs. Text: This objective is to predict text t_j given the text t_i that share the same attribute. It is used to promote similarity between posts sharing the same stance on a certain topic.

$$L_{TT} = \sum_{i=1}^m \sum_{j=1}^{text_{t_i}} \log P(t_j|t_i) \quad (4)$$

All the conditional probabilities can be computed using a softmax function. Taking $P(t_j|a_i)$ as an example:

$$P(t_j|a_i) = \frac{\exp(e_{a_i}^T e_{t_j})}{\sum_{k \in T} \exp(e_{a_i}^T e_{t_k})}$$

5.2 Embedding Initialization

In our model, the embedding for each author and attribute can be randomly initialized. The text is a special case since there are complex structures involved. One way to capture this is to use an pre-trained text embedding model to get an initial representation, and then learn a neural network to map it to one in the shared space. Note that this also allows our model to generate embedding for unseen text in the new space.

Specifically, for a text input x , we can compute its embedding e using M hidden layers $l_i, i = [0, M-1]$. The first hidden layer l_0 is computed from the input x :

$$l_0 = f(W_0x + b_0)$$

Subsequent layers are computed recursively:

$$l_i = f(W_i l_{i-1} + b_i), i = 1, \dots, M-1$$

Then the output from the final layer produces the embedding:

$$e = l_{M-1}$$

f is the non-linear activation function. We used hyperbolic tangent (tanh) in our experiments.

Note that our model offers the flexibility to use more complex neural network structures, including CNN and RNN, to learn a mapping from the initial word embedding sequences of the text to an embedding in the joint space.

5.3 Joint Embedding Learning

Our objective is to learn a semantic embedding for authors, text and attributes associated with them so that they are close in the embedding space if they are semantically close to each other.

Joint Embedding Loss Function: We can combine these embedding losses from Equations 1-4 into a joint training objective:

$$L_{Joint}(A, T, R) = \sum_{i \in (AT, AR, TR, TT)} \lambda_i L_i \quad (5)$$

where λ_i is the coefficient for each view, indicating the relative importance in the loss function. We set all λ_i to the default value 1 in all our experiments.

This is the general framework. Additional views may be added or removed for a certain dataset. For example, we can add a term representing the author vs. author view in the loss function if links between them are available.

5.4 Model Optimization

We train our model using mini-batch Adam optimizer to minimize the loss in Eq.5. However, computing gradient for Eq.1, and Eq.4 is expensive due to the size of the authors or text. To address this problem, we refer to the popular negative sampling approach (Mikolov et al., 2013a), which reduce the time complexity to be proportional to the number of positive example pairs.

5.5 Global Embedding Learning

Although different views of the data are captured in our joint loss function, it does not ensure that the information they provide at inference time will be “cooperative”, i.e., it will result in consistent global prediction over all the debate outputs. One potential problem is that examples associated with one view will dominate the training and skew the prediction, when constraints are applied. To handle this issue, we included the inference procedure during training. Instead of making sure the loss for each local view is minimized, the global objective promotes the rank of all the gold predictions *jointly*. For instance, at training, posts, together with their author and neighboring posts (if available) are used to infer their stance based on the inference procedure described in section 4. Then structured hinge loss can be used to define the prediction loss as in Eq.6.

$$L_{pred} = \sum_{i \in instances} \max(0, \max_{y \in Y} (\Delta(y, t_i) + score(x_i, y)) - score(x_i, t_i)) \quad (6)$$

where x_i and t_i are the problem instances and corresponding gold predictions, Y denotes all possible predictions, and $score(\cdot)$ is the inference score function. $\Delta(\cdot, \cdot)$ is the hamming loss. It measures the difference between two predictions and is used to create a margin between gold and other predictions.

The loss function used for updating the parameters in the global model is defined as follows.

$$L_{Global} = \lambda_{pred} L_{pred} + \lambda_{AT} L_{AT} + \lambda_{TT} L_{TT} \quad (7)$$

The coefficients for different terms in the global loss function can adjust the contribution of prediction objective (L_{pred}) versus information sharing objective (L_{AT} and L_{TT}). Again, we set all λ to the default value 1 in experiments of this paper. We leave the exploration of different coefficient settings to future work. Since the inference is used during global training, the scores for text stance and author stance are part of the inference scores. So they will get updated according to L_{pred} . Therefore we do not include L_{AR} and L_{TR} explicitly in the global loss function.

In our experiments, a mini-batch of debate threads is regarded as an instance during training. To reduce the computational cost, we used the parameters learned with the joint embedding loss function as the starting point for the global training.

Dataset	Topic	Posts	Users
CREATEDEBATE	Abortion	1741	340
	Gay Rights	1376	370
	Marijuana	626	258
	Obama	985	278
4FORUMS	Abortion	7937	342
	Evolution	6069	311
	Gay Marriage	6897	296
	Gun Control	3755	281

Table 1: Data Statistics for 4FORUMS and CREATEDEBATE

6 Experiments

To evaluate the different properties of our model and demonstrate their advantages, we evaluate the quality of our structured embedding model on two datasets 4FORUMS and CREATEDEBATE for stance classification tasks at post level, consisting of eight topics in total. The datasets are taken from the Internet Argumentation Corpus (Abbott et al., 2016). Table 1 shows statistics about these datasets.

Experimental Design Our experiments are designed to evaluate the different properties of our model. To accomplish that, we compare different variations of the model, corresponding to (1) only the joint embedding (denoted *Joint*), (2) using inference at test time, over the joint embedding (denoted *Inference*), and (3) using global training (denoted *Global*), which also uses inference during training. We report the results of these experiments in the two datasets in Tables 2 and 3.

Our second set of experiments are designed to evaluate the joint embedding model’s ability to share information between the representations of different objects. In this case we compare the performance of the *Joint* and *Inference* models when additional information is available. We compare three settings (1) In-Domain, when the available training and testing data are from the same domain (2) In+Out Domain, where we augment the In-Domain training data with additional debate threads from other topics. In this case the model can represent the relationships between stances on different topics and potentially generalize better. Finally, (3) User-Attribute, where we augment the author attributes with profile information extracted from the debate website. We conduct all of these experiments over the CREATEDEBATE dataset, and report the results in Table 4.

6.1 Experimental Settings:

We used PyTorch (Paszke et al., 2017) to implement the embedding model and Gurobi (Gurobi Optimization, 2016) as our ILP solver. Each debate post is initially represented using the Skip-Thought Vectors (Kiros et al., 2015), and then mapped to an embedding in the shared space through one hidden layer. We do not add more layers as both datasets are relatively small. Hyperbolic tangent (\tanh) is used as non-linear activation function. All other embeddings are randomly initialized following a normal distribution with variance $1/\sqrt{dim}$. The embedding size dim for all experiments is 300. For the training of the neural network, we used mini-batch Adam optimizer to update parameters. Dropout with probability 0.7 is used as regularization. The termination criteria is convergence on training loss. Five epoch of non-improvement on loss is considered as convergence for joint models, and one epoch for global models. Other parameters in our model includes negative sample size $k=5$, mini-batch size $b=10$.

6.2 Results

Our results on stance classification are described in Table 2 and Table 3. The results are computed using 5-fold cross-validation. For the CREATEDEBATE dataset, We used the same five data folds as in (Hasan and Ng, 2013) to ensure our results are directly comparable with theirs. For the 4FORUM dataset, we randomly divided debate threads into five folds since the data split is not available in (Sridhar et al., 2015). We regarded the same user in the training and test folds as different ones to avoid leaking label

Model	Abortion	Gay Rights	Marijuana	Obama	Average
Majority	56.2	64.5	72.0	56.1	62.2
NB (Hasan and Ng, 2013)	73.3	67.0	72.4	67.0	70.0
CRF (Hasan and Ng, 2013)	74.7	69.9	75.4	71.1	72.8
PSL (Sridhar et al., 2015)	66.8	72.7	69.1	63.7	68.1
Joint	62.1	63.1	69.2	57.4	63.0
Inference(AC)	70.4	62.7	66.3	62.2	65.4
Inference(Consecutive)	67.2	65.0	66.8	61.0	65.0
Inference(Both)	81.1	75.6	75.0	64.7	74.1
Global	81.0	77.2	77.6	64.8	75.2

Table 2: Average Accuracy on CREATEDEBATE dataset

Model	Abortion	Evolution	Gay Marriage	Gun Control	Average
Majority	56.8	65.8	66.0	67.8	64.1
PSL (Sridhar et al., 2015)	77.0	80.3	80.5	69.1	76.7
Joint	64.1	67.2	68.5	66.5	66.6
Inference(AC)	72.9	66.8	68.6	68.4	69.2
Inference(Consecutive)	67.5	67.7	72.3	69.6	69.3
Inference(Both)	85.9	80.9	88.1	81.6	84.1
Global	86.5	82.2	87.6	83.1	84.9

Table 3: Average Accuracy on 4FORUMS dataset

information from training to test. *NB* and *CRF* stands for the best local and collective models in (Hasan and Ng, 2013). Note that their system also uses the author constraints, as well as a highly engineered feature set and additional weakly-supervised data that we did not use. Despite that fact, our global model significantly outperforms their model with the exception of Obama domain. In the 4FORUM experiments we compare our models to result with *PSL* based model (Sridhar et al., 2015) which performs similar collective classification defined over a feature-rich representation. In this case as well, our Global model achieves the best overall performance.

We evaluate the contribution of two constraints sets, Author constraints (*AC*) enforce author and their posts share the same stance. *Consecutive* will encourage disagreement in stance between neighboring posts as introduced in section 3. The addition of these two constraints leads to significant increase in performance when used together. This is because *AC* and *Consecutive* add agreement and disagreement constraints between test posts, grouping them into clusters and making it easier to be predicted correctly. For instance, given multiple posts from the same author, the model can make correct decisions on all of them even if the prediction based on each individual post may be wrong. Finally, we observe that structured representation learning (i.e., *Global*) leads to a performance improvement compared to inference over the joint embedding objective (*Inference*). This shows the effectiveness of the global learning.

Table 4 shows the result on CREATEDEBATE when additional information is available. We extracted user profile information (User-Attribute) from the website², consisting of five attributes (Gender, Marital Status, Political Party, Religion, and Education). Clearly richer user information results in a better representation, both for users and as a result, also for the text they author, leading to improved performance. A similar trend occurs when out-of-domain data is available (*In+Out Domain*). Stances over different debate topics impact the text and author representations. Interestingly, when this data is available, our model is able to outperform the collective approach of (Hasan and Ng, 2013) in all debate topics, showing that our model can indeed exploit the information shared by the underlying ideologies.

²www.createdebate.com

	Model	Abortion	Gay Rights	Marijuana	Obama	Average
	CRF	74.7	69.9	75.4	71.1	72.8
In-Domain	Joint	62.1	63.1	69.2	57.4	63.0
	Inference	81.1	75.6	75.0	64.7	74.1
In-Domain + User-Attribute	Joint	63.9	63.6	69.5	59.9	64.2
	Inference	81.0	80.0	75.6	65.7	75.6
In+Out Domain	Joint	63.0	62.9	70.3	61.8	64.5
	Inference	79.0	74.5	77.1	76.2	76.7
In+Out Domain + User-Attribute	Joint	63.3	65.3	71.0	57.7	64.3
	Inference	79.9	80.2	75.1	77.4	78.2

Table 4: Average Accuracy on CREATEDEBATE dataset with additional information

7 Conclusions and Future Work

We study the problem of stance prediction, a challenging text classification problem, which requires connecting textual content analysis, conversational interactions and author information. Traditionally, this is done using a graphical model, which learns a scoring function for each aspect, over a fixed feature representation. We follow the observation that all of these problems are connected, and allow the model to capture these dependencies by allowing it to *learn* a representation for all these aspects jointly, rather than using a fixed representation. We show that by formulating the decision problem over the representation directly, and requiring the representation to respect the global dependencies between these aspects, our model can generalize better and exploit additional information even when it is not directly relevant.

To the best of our knowledge, this work is the first to cast representation learning as a structured prediction problem, we believe that this approach is applicable to many other domains where the input has complex inter-connected structure. Such domains can include other conversation analysis tasks, shared representations of text and images and information networks such as citations graphs and social network analysis. In the future we intend to study additional domains and evaluate whether including additional aspects, can help provide better generalization. Providing sufficient supervision is one of the main bottlenecks of NLP, we intend to apply our approach in weakly and distantly supervised settings, to help alleviate this difficulty.

References

- Rob Abbott, Brian Ecker, Pranav Anand, and Marilyn A Walker. 2016. Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*.
- 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 *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 1–9. Association for Computational Linguistics.
- Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, pages 2442–2452.
- Isabelle Augenstein, Tim Rocktschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Conference on Empirical Methods in Natural Language Processing*, pages 876–885.
- Mohit Bansal, Claire Cardie, and Lillian Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *COLING 2008: Companion Volume: Posters*, pages 15–18.
- Clinton Burfoot, Steven Bird, and Timothy Baldwin. 2011. Collective classification of congressional floor-debate transcripts. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, pages 1506–1515. Association for Computational Linguistics.

- Rui Dong, Yizhou Sun, Lu Wang, Yupeng Gu, and Yuan Zhong. 2017. Weakly-guided user stance prediction via joint modeling of content and social interaction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1249–1258. ACM.
- Greg Durrett and Dan Klein. 2015. Neural crf parsing. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, pages 302–312.
- Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2016. Weakly supervised tweet stance classification by relational bootstrapping. In *Proc. of the Conference on Empirical Methods for Natural Language Processing (EMNLP)*.
- Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui. 2014. Analyzing argumentative discourse units in online interactions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 39–48.
- Inc. Gurobi Optimization. 2016. Gurobi optimizer reference manual.
- Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356.
- Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proc. of the Conference on Empirical Methods for Natural Language Processing (EMNLP)*, pages 751–762.
- Kristen Johnson and Dan Goldwasser. 2016. ” all i know about politics is what i read in twitter”: Weakly supervised models for extracting politicians stances from twitter. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2966–2977.
- Arzoo Katiyar and Claire Cardie. 2016. Investigating lstms for joint extraction of opinion entities and relations. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, pages 919–929, Berlin, Germany, August.
- Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. Skip-thought vectors. In *The Conference on Advances in Neural Information Processing Systems (NIPS)*, pages 3294–3302.
- Guillaume Lample, Miguel Ballesteros, Kazuya Kawakami, Sandeep Subramanian, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*, pages 1–10.
- Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196.
- Jiwei Li, Alan Ritter, and Dan Jurafsky. 2015. Learning multi-faceted representations of individuals from heterogeneous evidence using neural networks. *CoRR*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, pages 3111–3119, USA. Curran Associates Inc.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124. Association for Computational Linguistics.

- Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, volume 1, pages 116–125.
- Ashish Vaswani, Yonatan Bisk, Kenji Sagae, and Ryan Musa. 2016. Supertagging with lstms. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*, pages 232–237.
- Marilyn A Walker, Rob Abbott, and Joseph King. 2012a. A corpus for research on deliberation and debate. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*.
- Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012b. Stance classification using dialogic properties of persuasion. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*, pages 592–596. Association for Computational Linguistics.