Does Social Pressure Drive Persuasion in Online Fora?

Ayush Jain
Indian Institute of Technology, Kanpur
ayushj@iitk.ac.in

Shashank Srivastava
UNC Chapel Hill
ssrivastava@cs.unc.edu

Abstract

Online forums such as ChangeMyView have been explored to research aspects of persuasion and argumentative quality in language. While previous research has focused on arguments between a view-holder and a persuader, we explore the premise that apart from the merits of arguments, persuasion is influenced by the ambient social community. We hypothesize that comments from the rest of the community can either affirm the original view or implicitly exert pressure to change it. We develop a structured model to capture the ambient community's sentiment towards the discussion and its effect on persuasion. Our experiments show that social features themselves are significantly predictive of persuasion (even without looking at the actual content of discussion), with performance comparable to some earlier approaches that use content features. Combining community and content features leads to an overall performance of 78.5% on the persuasion prediction task. Our analyses suggest that the effect of social pressure is comparable to the difference between persuasive and non-persuasive language strategies in driving persuasion and that social pressure might be a causal factor for persuasion.

1 Introduction

Recent interest in analyzing online discussion forums has led to growing body of work on exploring aspects such as argument quality and persuasion in comments, posts, and tweets (Habernal and Gurevych, 2016; Toledo et al., 2019). In particular, Tan et al. (2016)'s seminal work used linguistic features of comments and their interplay with a view-holder's stance to predict persuasion. Priniski and Horne (2018) investigate differences in persuasion between sociomoral vs non-sociomoral topics. Other work has explored the role of grammatical (Khazaei et al., 2017) and argument structure (Li et al., 2020; Wei et al., 2016; Ji et al., 2018) in persuasive language. Finally, Lukin et al. (2017) and Wang et al. (2019) study the role of prior beliefs and personality in influencing persuasion.

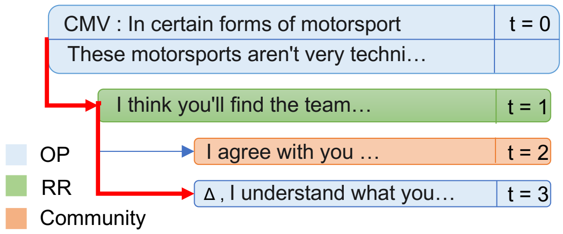
CMV : In certain forms of motorsport
These motorsports aren't very techni…
t = 0
I think you'll find the team…
t = 1
I agree with you …
t = 2
Δ , I understand what you…
t = 3
OP
RR
Community

Figure 1: A sample discussion in CMV. The red path indicates the OP-RR interaction. We explore the premise that comments from the ambient social community influence whether the OP is persuaded by a RR.

In this work, we investigate the ties between social pressure and persuasion. Social influence and peer pressure are highly pervasive in online discussions (Hui and Buchegger, 2009; Huffaker, 2010). Thus, apart from the merits of a discussion, peer pressure can be a strong force that can make one susceptible to change (Cascio et al., 2015; Pruksachatkun et al., 2019). Wei et al. (2016) explore social influence in the context of viewing the attention a comment receives by using discussion-tree-based features from the community. We focus on the task of predicting the persuasiveness in the ChangeMyView (CMV) subreddit by also incorporating content from the community's comments. Towards this, we design a structured model, which can dynamically model the evolution of community opinion and sentiment towards a discussion.

Our experiments show that social features are surprisingly effective at predicting persuasion. In fact, strong predictive performance can be achieved using community-based features even without looking at the language content of a persuasion-seeking post. In §2, we introduce terminology and describe the data and task. §3 describes our method and features. In §4, we present the results and qualitative discussion on the observations. We also present

Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 9201–9208
November 7–11, 2021. ©2021 Association for Computational Linguistics

some analysis suggesting that social pressure might be a causal factor for persuasion.

2 Dataset and Task Definition

ChangeMyView (r/changemyview) is an online discussion group where an Original Poster (OP) posts a view on some topic, and other people try to change the OP's opinion. Users who initiate a new discussion with the OP are referred to as Root Repliers (RR). Other community members can post their comments at any point in the discussion, leading to discussion trees rooted at the RR's post. If the OP is eventually persuaded, the OP indicates this by providing a Δ to the RR. Figure 1 illustrates the structure of a sample discussion from CMV.

Following previous work, we consider the OP's original post and subsequent comments made along the OP-RR discussion path (but, crucially, also including the community comments) to predict persuasion. We construct a dataset that builds on previous data compiled by Tan et al. (2016) by following a strategy similar to Hidey and McKeown (2018). Our data consists of positive and negative instances of discussion trees (initiated by RR's but also contains comments from the rest of the community) that end in a Δ being awarded by the OP (or not, for negative examples). We remove all posts that receive less than 50 comments to ensure sufficient discussion. For discussions that receive a Δ, we only consider comments until the Δ was received, and for the unsuccessful attempts, we consider the discussion till the last comment from the OP. Positive examples consist of instances where a Δ is received, we store the corresponding path in the discussion tree along with community comments in the subtrees of path nodes. We sample negative examples by following a path (not leading to a Δ) in a depth-first search of the discussion tree such that the path length is similar to those of the positive examples. Table 1 shows statistics of the data.

3 Method

Given a post from the OP and a reply from the RR, one can extract features from both to learn a predictive model (similar to Tan et al. (2016)). To explore the social community's role, this can be extended to also include features based on other parts of the discussion tree (i.e. posts other than from the OP and the RR). However, such an approach would be insensitive to the sequential structure of arguments in the discussion, and the evolution of social sentiment towards the OP and the RR. Therefore, we model this task of persuasion using a *structured model* that considers the discussion's temporal structure. To model social dynamics and the pressure that it exerts, we look at comments from the rest of the community and use features based on these to model the community's influence.

	Positive	Negative
# Train	5199	5199
# Test	1203	1203
Mean Discussion Tree Size	8.3	5.2
Mean Sequence Length	3.9	3.2

Table 1: Dataset statistics for persuasion prediction.

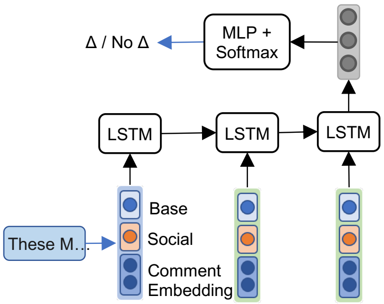

Figure 2: Our structured model incorporates sequential structure in both content-based (base and embedding) and social features to predict persuasion.

We extract the sequence of interactions between the OP and the RR along the path that leads to a Δ (or the longest discussion path) as described earlier. The comments on this path are ordered by time. This sequence of comments serves as an input to our model. At each step in this sequence, we maintain a social vector representing the overall state of the social community at that point, modeled using comments from the community seen till the current time. Our *structured model*, shown in Figure 2, consists of an LSTM network. At each step, the model takes as input the content of an OP/RR comment (represented with base-features, or an embedding representation, as described next) along with a vector of social features (see Figure 2). The final hidden state of the RNN is fed to a feed-forward classifier, which predicts a Δ/no Δ.

Next, we describe three sets of features that are computed at each step, which we use for our task.

Base features: Base features use the content of a comment to provide information about its provenance, agreement, and argument qualities. We performed n-gram analysis on comments that were successful/unsuccessful and observed that words[1] that express agreement or contrast had significantly different proportions.[2] Some of these features draw inspiration from Hidey and McKeown (2018), who model interactions through frames sensitive to agreement/disagreement expressed in comments.

1. Author Type: Indicating if the author of a comment/post is the OP, the RR or Community.
2. # Contrast words: Count of words that are used to express contrast selected from a set of 13 words (“but", “however", etc.).
3. # Agreement words: Count of words that express agreement selected from a set of 10 words.
4. # Disagreement words: Count of words that express disagreement from a set of 3 words.
5. # Question words: Count of question words selected from a set of 8 WH words.
6. # Links: Number of hyperlinks in the comment.
7. # Words: Number of words in the comment.

Comment Embedding: We experiment with two embedding representations of comments: (1) Word averaged 200-D GloVe embeddings (Pennington et al., 2014), and (2) Sentence-BERT embeddings of comments (Reimers and Gurevych, 2019).

Social Features: Social features model the state of the social community and community sentiment towards OP and RR. For every comment by someone other than OP or RR, we see if it responds to the OP or the RR (by looking at the parent comments in the discussion tree). Based on the comment's textual features and the target, we update the social vector. When combining social vectors with other features in the structured models at a current step, we aggregate social features across the multiple community posts which occur before the present step. We use two categories of social features:

1. Global sentiment towards OP(RR): Running average of sentiment intensity in comments written in response to OP(RR) using VADER (Hutto and Gilbert, 2014).
2. # People in support (or against) the OP (RR): Set of four features gauging community support/opposition for the OP (RR) by considering community posts at a certain step. If the VADER sentiment intensity of a comment is above(below) a threshold of $+0.1(-0.1)$, the person is said to be in support(against) of the parent comment. This threshold is selected to avoid non-informative comments.

# Support/Against is a count-based feature that treats a passionate supporter or someone in slight agreement as the same, whereas Global sentiment is finer-grained in weighing the intensity of emotion expressed.
While here we use VADER to capture the sentiment intensity, other sentiment analysis tools such as SenticNet (Cambria et al., 2020) and TextBlob (Loria, 2018) may be used for similar analyses.

4 Experiments and Results

In this section, we evaluate the performance of our approach on the persuasion prediction task. We first perform a comparative analysis of our structured model variants using different feature families and their combinations. Next, to understand the role of the evolution of community social features, we compare our structured models with unstructured models that use the same sets of features. Further, we discuss some analysis that indicates a causal relationship between social pressure and persuasion. Finally, we consider counterfactual explanations of the learned models, which indicate the highly significant role of social features[3].

Feature Ablation: Table 2 shows the test set performance of our structured model for different combinations of feature families. Tan et al. (2016) report accuracies of around 60% using linguistic features extracted from the root comment and computing its interplay with OP's post. We note that our structured model that only uses social features as input achieves similar accuracy. This is a surprisingly strong result since this approach never looks at comments authored by either OP or RR. We also note that both embedding-based representations achieve much better performance than manually defined Base features. Adding social features to GloVe and BERT leads to the highest performance on this task. In particular, BERT+Social achieve an accuracy of 78.1%. We also individually ablated social feature categories and found global sentiment and support/against features to be about equally discriminative individually (details in Appendix).

Learned Social Vectors: We experiment with a fully neural approach that learns representation of social vectors through an LSTM network (this is in addition to existing LSTM, which takes in OP/RR feature embeddings). For each community

[1] The exact word sets are described in the Appendix.

[2] For example, the bi-gram "disagree with" is observed nearly 3 times more in unsuccessful attempts at persuasion.

[3] The code will be available at `https://github.com/ayushjain9501/cmv-soc`

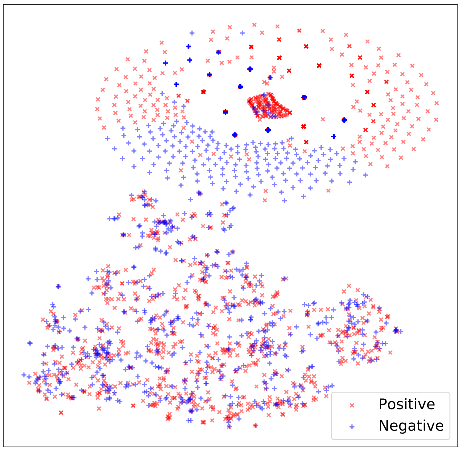

Figure 3: t-SNE reduction of the learned social vectors. Blue/red points represent Δ/no Δ examples.

Model	Test (%)	F1 Score
Social	58.43	0.45
Base	64.50	0.63
GloVe	75.1	0.74
BERT	75.93	0.75
Base + Social	65.71	0.64
GloVe + Social	76.52	0.76
BERT + Social	**78.09**	**0.77**
BERT + Base + Social	76.22	0.75
BERT + Social LSTM(Learned)	**78.51**	**0.78**

Table 2: Test set performance of structured models.

comment, we use the network to update the social embedding by feeding in the concatenation of the comment and its parent's BERT embedding. The new hidden state reflects the updated social vector. With this model, we get the overall best performing model with accuracy of 78.5%. Figure 3 shows the t-SNE(van der Maaten and Hinton, 2008) reduction of the learned social vectors, showing clear clustering that segregates the Δ cases from the No$-\Delta$ cases. This substantiates an underlying relation between the community comments and persuasion.

Role of Structure: To clearly delineate the role of structure, we also explore unstructured models (Logistic Regression and Multi-layer perceptron) trained on the averaged feature embedding of all the comments and the final social vector. Table 3 shows these results (reporting best model in each case). We note that the structured models perform significantly better than unstructured variants in every case. In particular, the difference in performance when only using social features indicates that the sequential nature of evolving social vectors carries more information (58.43%) than just the aggregated social vector over all comments (56.23%).

Model	Test (%)	F1 Score
Social	56.23	0.40
Base[4]	62.18	0.62
GloVe	68.99	0.69
BERT	72.40	0.72
Base + Social	62.86	0.63
GloVe + Social	69.84	0.70
BERT + Social	**73.08**	**0.73**
BERT + Base + Social	**72.99**	**0.73**

Table 3: Test set performance of unstructured models.

Volume of Discussion v/s Social Features: In constructing the dataset, we try to match the number of back-and-forth interactions between the OP and RR and include all comments from the community in this time frame. The average sequence length for the two classes (Table 1) is comparable, but average the tree size has a significant difference. To delineate the contribution of this factor, we re-ran the previously described analysis by subsampling the comments such that the mean subtree size was also comparable (5.1 vs 4.9). This only led to a small drop in performance (58.4% vs 56.6%), suggesting this factor (volume of social comments) only contributes marginally, and the content of the social comments is more significant.

Interestingly, for the same amount of back-and-forth between the OP and RR, the total amount of discussion is more in positive examples. This supports the hypothesis that community intervention can exert pressure driving persuasion.

Relative Importance of Persuasive Language vs Social Features: Significant previous work has studied the efficacy of content-based features in predicting persuasion (Tan et al., 2016; Khazaei et al., 2017; Li et al., 2020; Wei et al., 2016; Ji et al., 2018). We perform an analysis to quantify the relative effect-sizes of language-usage and social pressure in influencing persuasion. For this experiment, our content-based features consist of BERT representations[5] of the OP's initial statement and RR's opening comment alone, and do not incorporate any information about the subsequent interactions between the OP and the RR. The social features are represented by their averaged value over all the community comments, as in previous experiments. This partition ensures that there is no conflation between content-based and social features as they are computed from non-overlapping parts of the discussions. The results are shown

[4] Similar to the accuracies reported by Tan et al. (2016) when IP was not considered

[5] Since BERT-based representation perform the best among our content-based representations

in Table 4. While the content-based features are more predictive than social features, the difference is not substantial. The joint model that combined the content-based and social features achieves the best performance.

Model	Test (%)	F1 Score
Social	56.3	0.40
Content	59.8	0.59
Content + Social	60.8	0.61

Table 4: Test set performance of models using content-based and social features.

Direction of Association: In addition to social features being correlated with persuasion, we note that this association also has a directional trend. Of all discussions that receive a Δ, in 65% of cases, the Δ is awarded after the discussion had at least half of the comments which it receives eventually ($p < 0.01$, Binomial test). This suggests that it is the social community that pressures the OP to give a Δ, rather than the Δ attracting comments.

Counterfactual analysis: Counterfactual analysis with our learned models reveals a significant role of social features on persuasion. We consider examples predicted as no Δ and replace their social features with those of examples with high percentiles value of $\#$ of people in support of the RR feature (Table 5). Doing this expectedly flips the prediction for many examples, with the trend increasing with the number of people supporting the RR. Figure 4 plots the change in the probability of persuasion for individual examples for the 99%-ile scenario. The probability increases in almost all cases. In particular, the model suggests that a discussion with a 25% chance of persuasion can have a 45% chance instead, if 13 new comments support the RR.

Percentile	% flipped	# Support
$85th$	5.18	2
$95th$	42.2	5
$99th$	55.5	13

Table 5: Percentage of examples that change from no-Δ to Δ when the social vector is replaced by vector with xth percentile value of $\#$ of people in support.

5 Conclusion

We present a model that explores the role of social pressure in affecting persuasion. By incorporating structural information in discussions and simple social features, the approach shows surprisingly high predictive performance. We also perform experiments to quantify the relative effect-sizes of content-based and social features. Our counterfactual analysis indicates that the learned models assign a very high value to social features. Future work can develop more sophisticated methods for extracting social features and perform comprehensive causal analysis, such as with intervention studies.

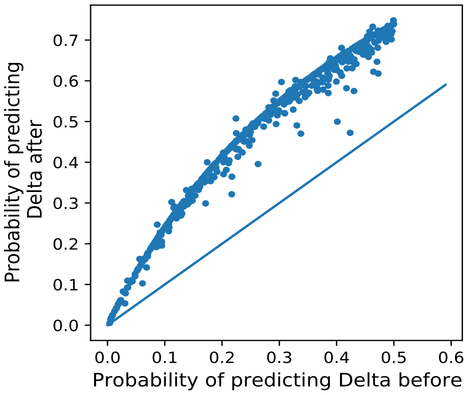

Figure 4: Probability of predicting Δ before and after counterfactually replacing social vector with one corresponding to $99th$ percentile value for $\#$ of people in support. All examples were predicted as no Δ before.

References

Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 105–114. ACM.

Christopher N Cascio, Christin Scholz, and Emily B Falk. 2015. Social influence and the brain: persuasion, susceptibility to influence and retransmission. *Current Opinion in Behavioral Sciences*, 3:51 – 57. Social behavior.

Ivan Habernal and Iryna Gurevych. 2016. What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1214–1223, Austin, Texas. Association for Computational Linguistics.

Christopher Hidey and Kathleen R. McKeown. 2018. Persuasive influence detection: The role of argument sequencing. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence*

(EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 5173–5180. AAAI Press.

David Huffaker. 2010. Dimensions of leadership and social influence in online communities. *Human Communication Research*, 36(4):593–617.

Pan Hui and Sonja Buchegger. 2009. Groupthink and peer pressure: Social influence in online social network groups. In *2009 International Conference on Advances in Social Network Analysis and Mining*, pages 53–59. IEEE.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Lu Ji, Zhongyu Wei, Xiangkun Hu, Yang Liu, Qi Zhang, and Xuanjing Huang. 2018. Incorporating argument-level interactions for persuasion comments evaluation using co-attention model. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3703–3714, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Taraneh Khazaei, Xiao Lu, and Robert Mercer. 2017. Writing to persuade: Analysis and detection of persuasive discourse. *iConference 2017 Proceedings*.

Jialu Li, Esin Durmus, and Claire Cardie. 2020. Exploring the role of argument structure in online debate persuasion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8905–8912, Online. Association for Computational Linguistics.

Steven Loria. 2018. textblob documentation. *Release 0.15*, 2.

Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753, Valencia, Spain. Association for Computational Linguistics.

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, Doha, Qatar. Association for Computational Linguistics.

John Hunter Priniski and Zachary Horne. 2018. Attitude change on reddit's change my view. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, pages 2279–2284. Cognitive Science Society. 40th Annual Meeting of the Cognitive Science Society, CogSci 2018 ; Conference date: 25-07-2018 Through 28-07-2018.

Yada Pruksachatkun, Sachin R. Pendse, and Amit Sharma. 2019. Moments of change: Analyzing peer-based cognitive support in online mental health forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, page 64. ACM.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 613–624. ACM.

Assaf Toledo, Shai Gretz, Edo Cohen-Karlik, Roni Friedman, Elad Venezian, Dan Lahav, Michal Jacovi, Ranit Aharonov, and Noam Slonim. 2019. Automatic argument quality assessment - new datasets and methods. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5625–5635, Hong Kong, China. Association for Computational Linguistics.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.

Zhongyu Wei, Yang Liu, and Yi Li. 2016. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 195–200, Berlin, Germany. Association for Computational Linguistics.

A Appendix

A.1 Filtering samples on the volume of discussion

We filtered examples based on the number of comments in the discussion to ensure participation from the community. The threshold of 50 is a heuristically-chosen trade-off between community involvement and conversation content on the one hand (tree size and sequence length), and the fraction of training examples that satisfy the criteria on the other. We note that a large fraction of the CMV dataset does not consist of any interactions apart from OP and RR. 50 is a threshold that we found helpful to filter the data. To give an idea, if the threshold is set to 10, around 10% of data is rejected but the average discussion size is only 3 (OP and RR comments)

A.2 Description of Base vector features

We use the count of occurrences words that indicate Agreement, Disagreement, and Contrast as features of our Base vector. We describe the complete set of words used here.

1. **Agreement** : agree, true, right, thanks, correct, alright, good point, never thought, get this, see why
2. **Disagreement** : false, wrong, disagree
3. **Contrast** : no, but, yet, although, despite, however, while, though, spite, whereas, unlike, besides, instead

A.3 Analysing Feature importance in the Social Vector

We ablate the social vector to test if a specific feature is important to predict persuasion. Table 6 shows these results, indicating that global sentiment and support/against features are nearly equally discriminative individually. As mentioned in the paper, the global sentiment features is more fine-grained in capturing the intensity of emotion than the # Support/Against count-based features. This may account for the slight increase in performance.

Model	Test (%)	F1 Score
Global Sentiment	**58.97**	**0.55**
# Support/Against	57.73	0.53

Table 6: Ablation analysis of social feature categories using structured model.

A.4 Processing an example in the Structured Model

We describe an example in the dataset containing an OP-RR interaction path, with community comments lying in the subtree of the path nodes. Figure 5 shows an extended positive example in our dataset. We now explain how we process this discussion tree. As mentioned, for positive(negative) examples we only consider comments before the Δ is awarded(last OP comment). We use the discussion path that leads to a Δ(the longest discussion path).

In this case, the sequence corresponding to timesteps $[0, 1, 4, 6]$ serves as a sequence of input. The comments $[2, 3, 5]$ are community comments which are used to update the social vector. We initialize the social vector to be a zero vector. All comments are represented by one of Base features, Glove averaged embedding or Sentence-BERT embedding, based on the model. We traverse across these comments in increasing order of time.

1. Feed the OP comment's feature embedding concatenated with the zero vector to the LSTM cell.
2. Similarly feed the RR's comment.
3. Now, comment 2 is a community comment. We find its parent, which is the RR. We use the sentiment and base features of this comment to update the social vector.
4. Similarly update the social vector for comment 3.
5. Now find the embedding for the RR comment 4, concatenate it with the social vector which contains information from comments 2 and 3 and feed it to the LSTM.
6. Update social vector for comment 5.
7. Concatenate the embedding of comment 6 with social vector to feed.
8. The final hidden state of the LSTM captures the whole discussion.
9. This serves as input to an MLP(1 hidden layer) classifier that predicts Δ/No-Δ.

Note that the social vector keeps changing as more and more community comments are encountered to capture the changing community opinion. In the model where we learn the social embeddings, we concatenate the community comment's and its parent comment's embedding, which is fed to an LSTM that updates the social vector.

A.5 Hyperparameter Tuning and Optimisation

We create an 80/20 Train-Validation split on the Training Data. For all the structured models, we manually tune the hyperparameters to achieve the best validation set accuracy across varying learning rates, LSTM hidden layer size, MLP hidden layer size etc. We train the models for 10 epochs. We measure Cross-Entropy loss on our predictions and use Adam optimizer for gradient updates.

Model	Val (%)	Test (%)	F1 Score	LSTM Hidden Size	MLP Hidden Size
Social	57.39	58.43	0.45	16	16
Base	66.02	64.50	0.63	16	8
GloVe	73.3	75.1	0.74	256	256
BERT	73.8	75.93	0.75	1024	512
Base + Social	65.57	65.71	0.64	16	8
GloVe + Social	72.39	76.52	0.76	256	256
BERT + Social	74.43	**78.09**	**0.77**	1024	512
BERT + Base + Social	77.27	76.22	0.75	1024	512
BERT + Social LSTM(Learned)	77.4	**78.51**	**0.78**	1024	1024

Table 7: Validation and Test Accuracy along with the optimal hyperparameters for the structured models.

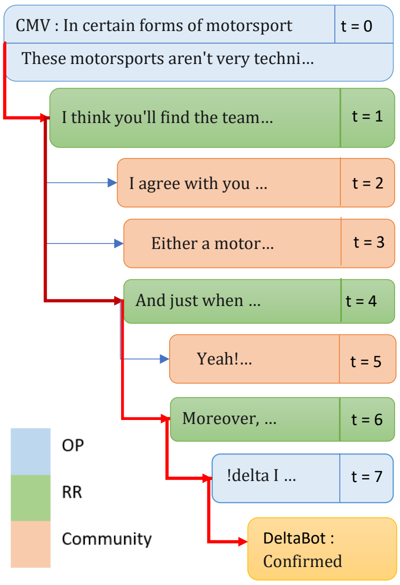

Figure 5: A Positive example in the Dataset. The community comments are used to model the evolving community opinion and are used to predict persuasion.

A.6 Training Resources

All the models were trained on Google Colab using the Tesla K80 GPU. On average, almost all models take less than 1 minute per epoch. The BERT + Social LSTM model takes 4 minutes per epoch on average.

A.7 Accuracy Metrics

We use accuracy and F1-score as our metric. F1 was calculated using scikit-learnhttps://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html.

A.8 Validation Accuracy and Optimal Hyperparameters

Table 7 list the validation accuracy and optimal hyperparameters for the structured models.