Hybrid Neural Attention for Agreement/Disagreement Inference in Online Debates

Di Chen[1], Jiachen Du[1], Lidong Bing[2], Ruifeng Xu[1]*
[1]Department of Computer Science, Harbin Institute of Technology (Shenzhen), China
[2]Tencent AI Lab, Shenzhen, China
chandichn@gmail.com, dujiachen@stmail.hitsz.edu.cn
lyndonbing@tencent.com, xuruifeng@hit.edu.cn

Abstract

Inferring the agreement/disagreement relation in debates, especially in online debates, is one of the fundamental tasks in argumentation mining. The expressions of agreement/disagreement usually rely on argumentative expressions in text as well as interactions between participants in debates. Previous works usually lack the capability of jointly modeling these two factors. To alleviate this problem, this paper proposes a hybrid neural attention model which combines self and cross attention mechanism to locate salient part from textual context and interaction between users. Experimental results on three (dis)agreement inference datasets show that our model outperforms the state-of-the-art models.

1 Introduction

The rise of various discussion forums and online debate platforms, has given users a lot of opportunities to express themselves and argue with each other. The online argumentation and discussion are always initiated and evolved by expressions of agreement or disagreement of participants. Inferring the agreement/disagreement in online debates is crucial for many other tasks in broader analysis of social media and argumentation mining, such as stance identification (Somasundaran and Wiebe, 2010), claim/argument extraction (Hidey et al., 2017) and persuasion analysis (Tan et al., 2016).

It is observed that the expression of agreement/disagreement in debates can be decomposed into two factors: 1) the self-expression of claims and 2) argumentative expressions to interact with other participants. To illustrate this observation, we show some examples in Figure 1, which is one of quote-response pair (Q-R pair) in 4forum online debate website. The response expressed disagreement with the claim in quote text. The mark *?* at the end of sentence carries strong emotion of authors, while the phrase *why doesn't He answer* refers to the claims of *God IS GOOD* and express the refutation to it.

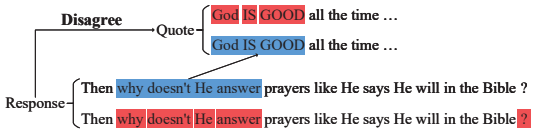

Figure 1: Sampled Q-R Pair with topic of evolution where the words colored red deliver crucial meaning of the text itself, while the words colored blue clarify the interactive relation between users.

Previous works on agreement/disagreement inference mainly focus on exploiting features to model the semantic information which only reveals author's self-expression. (Rosenthal and McKeown, 2015; Menini and Tonelli, 2016). These existing models treat agreement/disagreement inference as a ordinary sentiment classification problem and ignore the interactions between participants in the discussion. In order to jointly leverage the semantic information of the text and interactions between Q-R pairs, we regard the (dis)agreement inference as a special case of Natural Language Inference (NLI) (Rocktäschel et al., 2016), and propose a hybrid neural attention model to this problem. The proposed model consists of two kinds of attention: 1) **self attention** locates salient parts in text of quote and response, and 2) **cross attention** captures the interactive argumentations between Q-R pairs. The fusion of self and cross attention model is capable of jointly modeling the two important factors of inferring (dis)agreement in debates.

The main contributions of this paper are: (1) We propose a neural attention model for (dis)agreement inference which converts this

*Corresponding author.

Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 665–670
Brussels, Belgium, October 31 - November 4, 2018.

problem to a natural language inference task. The proposed model incorporates self and cross attention mechanism, jointly capturing significant part for current context and extracting interactive relations between Q-R pairs. (2) Experimental results on three datasets show that the proposed model significantly improves performance (measured by $F1$ score and accuracy) of state-of-the-art models by 1% on average. The visualization of extracted attention demonstrates different attention mechanism works effectively in different aspect for (dis)agreement inference.

2 Related Work

With the development of social forums, works on (dis)agreement inference have shifted to online debate. Abbott et al. (2011) utilize word-based and dependencies-based features to recognize disagreement in Internet Argument Corpus (IAC) (Walker et al., 2012). Rosenthal and McKeown (2015) present a new corpus derived from participant information, Agreement by Create Debaters (ABCD), and investigate new features for conversational structure. Further, Menini and Tonelli (2016) develop a SVM classifier to detect disagreement, relying on three aspects including sentiment-based, semantic and surface features extracted from both whole text and topic-related part. However, the performances of all these models highly depend on the quality of hand-crafted features. And these representations cannot reflect the interaction between quote and response.

In other NLP tasks, the end-to-end deep learning approaches with attention mechanism have shown impressive results. The attention mechanism is proposed by Bahdanau et al. (2014) in machine translation for selecting alignment between original words and foreign words before translation. For Document Classification, Yang et al. (2016) apply a hierarchical attention from word-level to sentence-level with learnable context vector. In Natural Language Inference (NLI), Liu et al. (2016) construct an inner-attention with mean pooling vector to seize important part from text itself. Hao et al. (2017) propose an cross attention modeling mutual influence between question and answer for Question Answering (QA). But there is no neural attention model incorporating both contextual and interactive information in the scenario of (dis)agreement inference.

3 Model

The overall architecture of our model is shown in Figure 2, comprising two parallel bi-directional LSTM (Hochreiter and Schmidhuber, 1997) networks as quote and response encoder and two attention components that respectively extract self and cross attention.

3.1 Quote and Response Encoder

A quote of length T is denoted as $[q_1, q_2, \cdots, q_T]$, where $q_t \in \mathbb{R}^{d_e}$ is the d_e-dimensional representation of the t-th word in the text sequence. Similarly, the corresponding response can be represented as $[r_1, r_2, \cdots, r_T]$, which shares the same vector space with quote. To model the dependence relation of text sequence, we leverage bi-directional LSTM (BiLSTM) to encode quote and response. The BiLSTM consists of a forward $\overrightarrow{LSTM}$ which reads the text from x_1 to x_T and a backward $\overleftarrow{LSTM}$ which reads from x_T to x_1:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(x_t), t \in [1, T] \quad (1)$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(x_t), t \in [T, 1] \quad (2)$$

Through concatenation, we obtain the representation of each time step $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \in \mathbb{R}^{2d}$ which integrates the information around x_t. The quote and response are encoded as $h_Q = [\overrightarrow{h_Q}; \overleftarrow{h_Q}] \in \mathbb{R}^{T \times 2d}$ and $h_R = [\overrightarrow{h_R}; \overleftarrow{h_R}] \in \mathbb{R}^{T \times 2d}$ respectively.

3.2 Attention Component

After encoding the implicit word semantics, we acquire the representation of both quote and response.

Self Attention

The first source taken into consideration should be the text sequence itself, i.e. the attention from quote to quote itself and that from response to response itself. When issuing an opinion, people tend to center on several keywords which convey the main idea. Thus in some sense, self attention is a kind of dependency parsing that drives the model to focus on salient parts of the context. Here, for quote $h_Q = [h_Q^1, h_Q^2, \cdots, h_Q^T]$, self attention generates signal s_Q^t by:

$$s_Q^t = \frac{\exp[\delta(h_Q^t)]}{\sum_{i=1}^{T} \exp[\delta(h_Q^i)]} \quad (3)$$

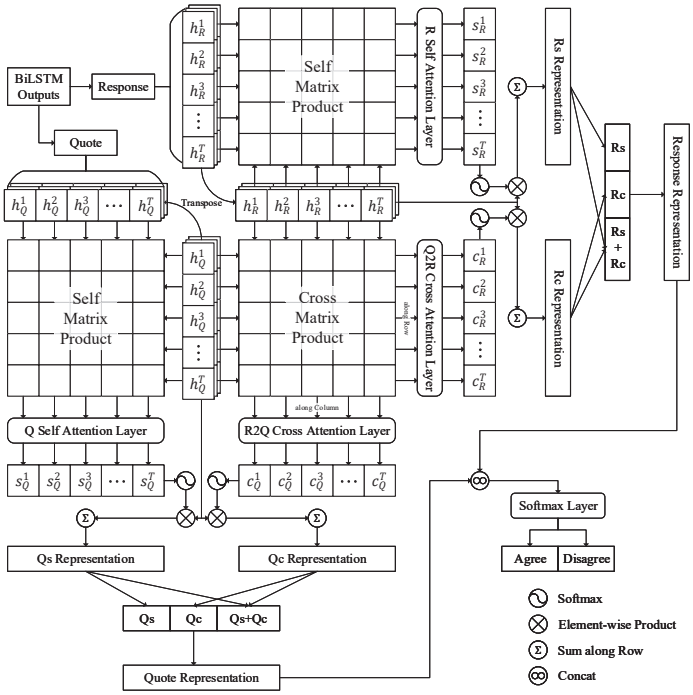

Figure 2: Model Architecture.

where δ is a transformation mapping $2d$-dimensional vector into scalar value, with learnable weight $W_S \in \mathbb{R}^T$ and bias $b_S \in \mathbb{R}$ defined as:

$$\delta(h_Q^t) = \tanh[W_S(h_Q \cdot (h_Q^t)^{\mathrm{T}}) + b_S] \quad (4)$$

Similarly, with another parallel transformation, the self attention signal of response can be calculated as above. Then, we can obtain a more compact representation of quote and response respectively derived from the weighted sum, where $Q_S, R_S \in \mathbb{R}^{2d}$.

$$Q_S = \sum_{t=1}^{T} s_Q^t h_Q^t \quad (5)$$

$$R_S = \sum_{t=1}^{T} s_R^t h_R^t \quad (6)$$

Cross Attention

Another prominent facet comes from the relation between each Q-R pair, i.e the attention from quote to response and that from response to quote. In whether disagreement or agreement cases, both quote and response provides a precise context for each other. The cross attention integrates interactive influence which produces more specific features for (dis)agreement inference.

As discussed above, cross attention c_Q^t, c_R^t for quote and response can be computed by:

$$c_Q^t = \frac{\exp[\gamma(h_R^t)]}{\sum_{i=1}^{T} \exp[\gamma(h_R^i)]} \quad (7)$$

$$c_R^t = \frac{\exp[\gamma'(h_Q^t)]}{\sum_{i=1}^{T} \exp[\gamma'(h_Q^i)]} \quad (8)$$

where γ and γ' are two parallel transformation with learnable weight matrix $W_C, W'_C \in \mathbb{R}^T$ and bias $b_C, b'_C \in \mathbb{R}$ defined as:

$$\gamma(h_R^t) = \tanh[W_C(h_Q \cdot (h_R^t)^{\mathrm{T}}) + b_C] \quad (9)$$

$$\gamma'(h_Q^t) = \tanh[W'_C(h_R \cdot (h_Q^t)^{\mathrm{T}}) + b'_C] \quad (10)$$

The representation of whole sequence $Q_C, R_C \in \mathbb{R}^{2d}$ embracing cross attention signal are:

$$Q_C = \sum_{t=1}^{T} c_Q^t h_Q^t \quad (11)$$

$$R_C = \sum_{t=1}^{T} c_R^t h_R^t \quad (12)$$

Hybrid Attention

In order to cooperate the advantage of self attention and cross attention, we design hybrid attention to get a more specific representation for quote and response:

$$Q = Q_S \oplus Q_C \oplus (Q_S + Q_C) \quad (13)$$

$$R = R_S \oplus R_C \oplus (Q_S + Q_C) \quad (14)$$

where $Q, R \in \mathbb{R}^{6d}$ and $\oplus$ is the vector concatenation operation.

3.3 (Dis)agreement Inference

Finally, the quote representation Q and response representation R are concatenated as a vector v. We use a fully-connected network to project $12d$-dimensional representation into n-dimensional vector space, i.e.

$$y = softmax(W_l v + b_l) \tag{15}$$

where $y \in \mathbb{R}^n$ is predicted probability distribution for (dis)agreement inference, W_l and b_l are parameters of softmax layer.

In a supervised learning framework, we train our model in an end-to-end way. Given a set of training data $\{(Q_i, R_i), y_i\}$, let $\widehat{y_i}$ denote the predicted probability distribution, the goal of training is to minimize the cross-entropy loss:

$$loss = -\sum_i \sum_j y_i^j \log \widehat{y}_i^j \tag{16}$$

where i is the index of quote-response pair, j is the index of class and y_i is the ground truth of corresponding pair.

4 Experiment and Results

As prior work, we concentrate on direct disagreement and agreement between quote-response (Q-R) pairs. Specifically, in the proposed model, the size of hidden units is 128 and all word embeddings are initialized by GloVe (Pennington et al., 2014) of 300d. Both length of quote and response are set to 64, padded where necessary. Adam is the optimizer of model whose learning rate is $1e-3$, β is $(0.9, 0.999)$, ϵ is $1e-8$ and weight decay is $1e-5$. All models are trained by mini-batch of 32 instances, with 5-fold cross validation.

4.1 Datasets

We conduct experiments on three most commonly-used (dis)agreement inference datasets. Table 1 shows the detail of these datasets.

- *Internet Argument Corpus (IAC)* (Walker et al., 2012) is a corpus crawled from online political debate `4forums.com`. Following prior work, we compute average score for each pair and convert the score into binary labels, with $[-5, -1]$ as disagreement and $[+1, +5]$ as agreement.

Table 1: Detail of Datasets

Dataset	# Disagree	# Agree	# Neutral
IAC	6,157	1,113	-
DP	12,899	11,875	-
ABCD	25,200	13,519	72,683

- *Debatepedia (DP)* (Menini and Tonelli, 2016). DP corpus is crawled from `debatepedia.org`, which is an online encyclopedia of debates.

- *Agreement by Create Debaters (ABCD)* (Rosenthal and McKeown, 2015) is developed from `createdebate.com` with labels of agreement, disagreement and neutral. As the original settings, the comparison experiments are conducted on a balanced training set by downsampling and the full test set.

4.2 Comparison with Baseline Methods

As shown in Table 2, by accuracy and average $F1$-score in percentage, we compare our model with the best performing model of corresponding dataset to our knowledge. These models are reported in (Abbott et al., 2011; Menini and Tonelli, 2016; Rosenthal and McKeown, 2015) as Naive Bayes (NB), JRipχ^2 (ruled based classifier using χ^2 for feature selection), SVM, Maximum Entropy (ME), exploiting a rich suite of features including n-grams, sentiment lexicon and syntax.

We also analyze the contribution of each component in ablation experiment. BiLSTM-sum and BiLSTM-concat refer only sum or concat operation is applied to both self and cross attention respectively. Results show that BiLSTM-hybrid gives the best performance across all datasets regardless of data sizes. For smaller dataset such as IAC, our model outperforms the previous best methods by 8.8%. This outcome is consistent across other larger datasets with a significant improvement of 19.6% on DP. What' s more important, on DP, the length of text is longer than other datasets, so ordinary BiLSTM suffering from gradient vanishing results in the poor performance. It is the hybrid attention that effects. As for ABCD, compared with ME based on textual features, Our BiLSTM-hybrid also gives superior performance of average $F1$ in 3-way inference. Since ABCD is a corpus annotated by meta-thread rules, the ME attaching conversational structure attains the best

Table 2: Experimental results on three datasets.

2-way Inference				3-way Inference	
IAC		DP		ABCD	
Approaches	Accuracy	Approaches	Accuracy	Approaches	Average $F1$
NB	60.3	SVM	74.0	ME	50.8
JRipχ^2	68.2	-	-	ME+structure*	**77.6**
Liu et al., 2016	74.9	Liu et al., 2016	92.4	Liu et al., 2016	76.5
BiLSTM	72.9	BiLSTM	55.2	BiLSTM	71.1
BiLSTM-self	76.1	BiLSTM-self	93.5	BiLSTM-self	75.8
BiLSTM-cross	76.4	BiLSTM-cross	93.3	BiLSTM-cross	75.9
BiLSTM-sum	76.4	BiLSTM-sum	93.3	BiLSTM-sum	76.4
BiLSTM-concat	76.8	BiLSTM-concat	93.5	BiLSTM-concat	76.4
BiLSTM-hybrid	**77.0**	BiLSTM-hybrid	**93.6**	BiLSTM-hybrid	76.9

(*) Conversational structure is a corpus-specific feature.

performance. We think it a corpus-specific feature with weak generalization ability.

In addition, we adapt a NLI-oriented model proposed by Liu et al. (2016) as a stronger baseline, which comprises inner-attention with mean pooling. The mean pooling of text encoder is set as the summary representation for inner-attention to seize important part from text itself. It is similar to our self attention but with coarse-grained level from text to word. The results imply that our BiLSTM-hybrid modeling additional interaction with fine-grained attention from word to word performs better.

4.3 Qualitative Analysis

To validate that different attention focuses on different part of text sequence, we visualize the outputs of self attention layer and cross attention layer, with a Q-R pair of disagreement from IAC. As show in Figure 3, darker color indicates larger weight in the corresponding attention vector.

In quote, the self attention selects *good* which is exactly the point that quote wants to argue. Similarly, the self attention selects *?* in response, which indicates a rhetorical mood to show disagreement. On the other hand, even though *why doesn't he answer* in response is endowed less weight from the self attention, the cross attention highlights it and *god* in quote. When inspecting the cross matrix product of this pair, Figure 4 demonstrates that our method is able to model the reference between *god is good* and *why doesn't he answer* in the whole interactive context.

5 Conclusion

In this paper, we propose a hybrid attention based neural network for (dis)agreement inference in debate. The main motivation is to jointly leverage self attention for textual context and cross attention for interactions between users to improve the capability of inference on agreement/disagreement relations. Experimental results show that our model outperforms several strong baselines. Visualization of extracted attention of our model illustrates that our models is effective in capturing the main point from different aspects.

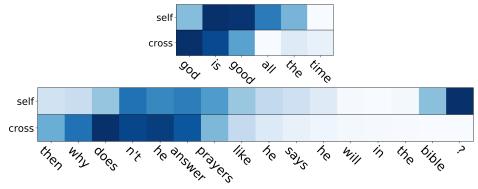

Figure 3: Attention Visualization. The topic is about evolution and the attitude of response is *disagreement*.

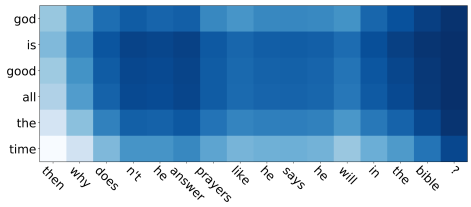

Figure 4: Cross Matrix Product Visualization.

Acknowledgements

This work was supported by the National Natural Science Foundation of China U1636103, 61632011, Key Technologies Research and Development Program of Shenzhen JSGG20170817140856618, Shenzhen Foundational Research Funding 20170307150024907.

References

Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani, and Joseph King.

2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 2–11, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 221–231.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21.

Sepp Hochreiter and Jrgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Yang Liu, Chengjie Sun, Lei Lin, and Xiaolong Wang. 2016. Learning natural language inference using bidirectional LSTM model and inner-attention. *CoRR*, abs/1605.09090.

Stefano Menini and Sara Tonelli. 2016. Agreement and disagreement: Comparison of points of view in the political domain. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2461–2470.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomáš Kočiskỳ, and Phil Blunsom. 2016. Reasoning about entailment with neural attention. In *The 4th International Conference on Learning Representations (ICLR 2016)*.

Sara Rosenthal and Kathy McKeown. 2015. I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions. In *Proceedings of the SIGDIAL 2015 Conference, The 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2-4 September 2015, Prague, Czech Republic*, pages 168–177.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. pages 116–124.

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*, pages 613–624. International World Wide Web Conferences Steering Committee.

Marilyn A Walker, Pranav Anand, Jean E Fox Tree, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *Eighth International Conference on Language Resources & Evaluation*, pages 23–25.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J. Smola, and Eduard H. Hovy. 2016. Hierarchical attention networks for document classification. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 1480–1489.