Sequence to Sequence Generative Argumentative Dialogue Systems with Self Attention

Ademi Adeniji, Nate Lee, and Vincent Liu
{ademi,natelee,vliu15}@stanford.edu
Stanford University Department of Computer Science

Overview

Problem

- Argument mining, a growing field in natural language generation, includes the automatic identification and generation of **argumentative structures** within conversation
- We experiment with various methods for creating a **dialogue agent that can engage in argumentative discourse**

Significance

- Utility in **education and assessment** as well as **business use** for investment decision
- Advances **self-attention/transformer** in argument **NLG/NLU objectives**

Existing Approaches

- Current state-of-the-art generative model: **hierarchical recurrent neural network**, encoding and decoding at one level and updating a conversation-level state at another
  - **Encoder**: Bidirectional GRU encoder w/ **conversation-level RNN memory**
  - **Decoder:** Vanilla RNN
- Model often **misinterprets arguments** or produces **irrelevant responses.**

Data

$$d = [\, p^{(1)}, p^{(2)}, ..., p^{(m)} \,]$$
$$p^{(i)} = [\, w_1^{(i)}, w_2^{(i)}, ..., w_n^{(i)} \,]$$

- **Internet Argument Corpus Dataset-v1:** 11,800 discussions w/ ~390,000 posts total
- **Training instance: discussion, d** (sequence of posts)
- **Gold** instances are offset from train instances
- ***p* is a padded sequence of tokens, *w***

Task

Given a post (w, w/o context), generate an appropriate adversarial argumentative response

Approach

Project Phases

1. LSTM Seq2Seq - model baseline, *context-free* argument generation
2. Pure transformers - *context-free* argument generation
3. Transformer with LSTM Session Memory - *context-rich* argument generation

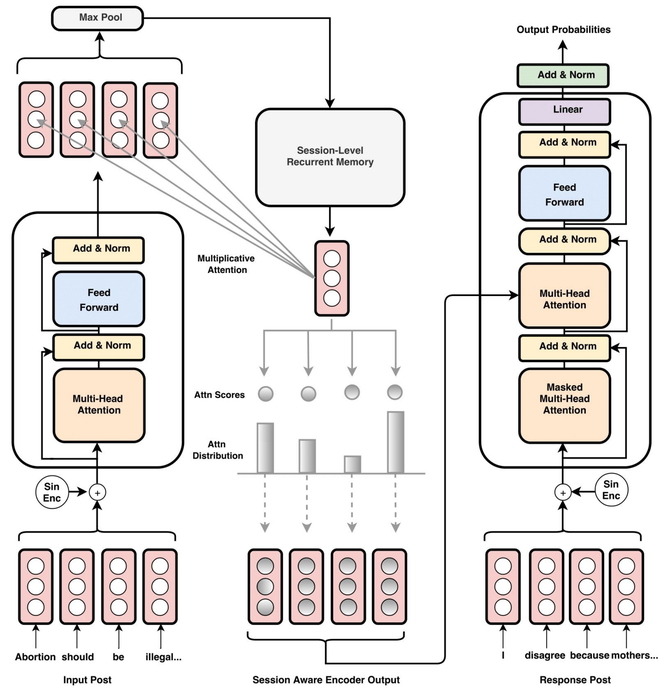

Figure 1. Transformer Model Architecture w/ LSTM. We borrow the Transformer architecture and use an LSTM between the encoder/decoder to encode session level memory.

Additional Tunings

1. Hyperparameter search - layers, dimensions, attention heads, learning rate, vocabulary size, min word count, etc.
2. Pre-training with cross-argumentative embedding objective (**Self-referential**)
3. GloVe embeddings vs. training from scratch
4. <unk> thresholding, vocabulary pruning, etc. (16k size)

Results

LSTM Seq2seq (w/o char decoder)

Query	Output
I think the bible is real and it should be respected	i 'm not sure that you are not a christian. i do n't think it is a matter of the bible . i do n't think it is a matter of a person .
Woman 's rights and suffrage should be upheld under the law	i do n't know what i said , but i do n't know what you are talking about .

Transformer w/ LSTM

Query	Output
i believe in god	i 'm not sure what you mean by `` god '' .
well then, have you heard of the bible?	i do n't think it 's a good idea .

Transformer w/o LSTM

Query	Output
I think that gun control has the potential to solve a lot of the problems with school shootings	i think that 's a good thing .
I think that abortion should be illegal because its effectively murder	i think that 's a good point . i think that if you are a christian , you are a christian.

Figure 2. Transformer w, w/o LSTM sample argumentation

Table 1. Transformer w/ LSTM validation metrics with tuned parameters

Epoch:	Perplexity	Accuracy	Loss
0	205.88	17.30%	5.33
4	84.21	24.83%	4.43
8	72.10	26.35%	4.28
12	65.65	27.36%	4.18
16	62.40	28.00%	4.13

Analysis

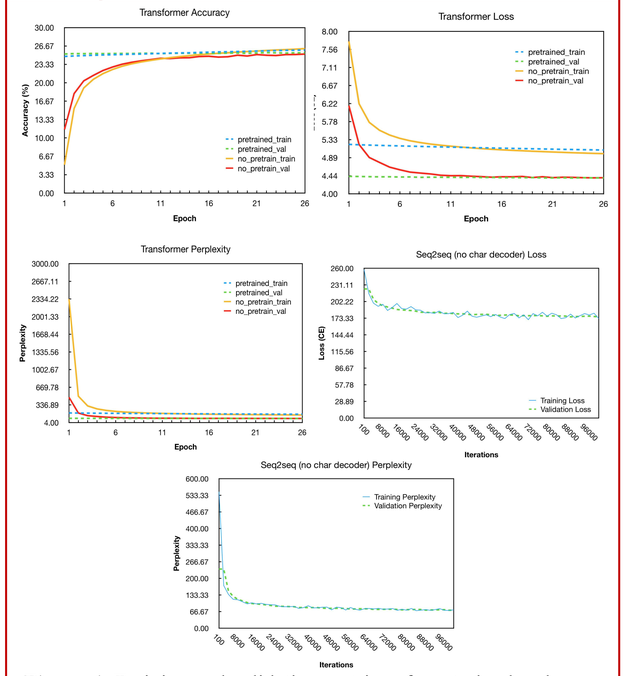

Figure 4. Training and validation metrics of pre-trained and from-scratch Transformer w/ LSTM models and Seq2Seq over 26 epochs.

Conclusion

- From our qualitative results, we conclude that our **dataset is ill-suited for generating more sophisticated language models** typical of advanced argumentative discourse
- Our **extensive hyperparameter** search suggests that our **cross entropy training objective is overly simplistic** for more complex generation tasks. A more involved theoretical formulation of training loss could yield qualitative translation improvements
- We were impressed by the model's **ability to infer the underlying basis** of the human input arguments
- Additionally, the dialogue agent was proficient in establishing a **sufficiently resolute position** on many topics

Future Work

- Less primitive argumentation datasets increases **language model expressivity**
- Fine-tuning on pretrained contextual embeddings (BERT) captures **word relationships** more precisely for better NLG
- More sophisticated **attention mechanisms** may allow for a **more informative signal** for decoding

References

M. Walker, J. F. Tree, P. Anand, R. Abbott, and J. King, "A corpus for research on deliberation and debate," in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, N. C. (Chair), K. Choukri, T. Declerck, M. U. Dog˘an, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds., Istanbul, Turkey: European Language Resources Association (ELRA), May 2012, ISBN: 978-2-9517408-7-7.

Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," no. Nips, 2017, ISSN: 1469-8714. DOI: 10.1017/ S0952523813000308. arXiv: 1706.03762. [Online]. Available: http://arxiv.org/abs/ 1706.03762.

D. Thu Le, C.-T. Nguyen, and K. Anh Nguyen, "Dave the debater: a retrieval-based and generative argumentative dialogue agent," pp. 121–130, 2018.