

Adversarial Learning for Neural Dialogue Generation

1 Main Idea

The authors formulate this dialogue model as a reinforcement learning problem. The network used is a Generative Adversarial Network. The discriminator objective is the same as a Turing test predictor i.e. classifies whether the dialogue response is human or machine-generated. The goal is to improve the generator to the point where the discriminator has trouble distinguishing between human and machine-generated responses [1].

2 Method

- The generator network is a neural seq2seq model, and the discriminator is similar to a Turing test evaluator.
- The generation task is not formulated as a NMT task. Instead, it tries to maximize the likelihood of a response $y = \{y_1, y_2 \dots y_T\}$ given a history of previous sentences x .
- The generator defines the policy by which each word of the output sentence y is generated using a softmax over the space of the vocabulary.
- The discriminator uses a hierarchical neural autoencoder [2] to generate a vector representation of an entire sequence of conversation i.e. $\{x, y\}$. This vector representation is then fed into a binary classifier which predicts whether the sentences were human- or machine-generated.
- The generator is trained to maximize the expected reward of the generated utterance using the REINFORCE algorithm. [3].
- The vanilla REINFORCE model doesn't assign rewards to each generated word, and rather assigns equal reward to all the tokens within a predicted sequence of words.

- However, for partially decoded sequences, the discriminator must also be capable of generating classifications for partial sequences. Two methods are proposed to solve this:
 - Using a Monte-Carlo search to decode $N(= 5)$ top candidate sentences given a partial sequences and using the discriminator average of the 5 complete sequences to predict the classification for the partial sequence.
 - Training the discriminator to directly also be able to classify partial sequences.
 The Monte-Carlo search strategy was found to be more effective.
- Teacher forcing is used to essentially short-circuit the distance between the generator and the true sequence.
- The generative model is trained using seq2seq [4] and an attention mechanism [5]. The discriminator is also pretrained using part of the training data and generating sequences by beam-search and sampling.
- Intuitively, low accuracy of a reasonably well trained discriminator would imply that the quality of generated sentences have improved significantly.

3 Observations

- The authors report that the responses generated by their system are more interactive, interesting, and non-repetitive. It'd be interesting to see how they quantify this. UPDATE: The source for this claim is human evaluations, which of course, could be subjective.
- It's also observed that the system yielded better results when the context i.e. the x preceding utterances were limited to 2.
- The hierarchical neural model is the architecture of choice for the discriminator (evaluator).

References

- [1] Jiwei Li, Will Monroe, Tianlin Shi, Alan Ritter, and Dan Jurafsky. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*, 2017.
- [2] Jiwei Li, Minh-Thang Luong, and Dan Jurafsky. A hierarchical neural autoencoder for paragraphs and documents. *arXiv preprint arXiv:1506.01057*, 2015.
- [3] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

- [4] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- [5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.