Generating Text with Deep Reinforcement Learning

Target Task
- Model sequence-to-sequence learning, with length not known a-priori \((y_1, y_2, \ldots, y_n) = f(x_1, x_2, \ldots, x_n)\)
- Important applications: speech recognition, machine translation, question answering

State-of-the-art
- Encoder-decoder Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN)
- Shortcoming: decoding process is essentially a RNN language model, thus a left-to-right decoding schema

Contributions
- Deep Q-Network (DQN) to iteratively decode sequence
- Enable the decoder to first tackle easier portions of the sequence, and then difficult parts
- Decode with BiDirectional LSTMs

DQN for Iterative Decoding
- Represent value function by deep Q-network with weights \(w\)
- Search to modify words in non-exact decoded sentence

\[
Q(s, a, w) = Q^*(s, a)
\]

- Input: the encode sentence \(\langle A, B, C, D \rangle\) and the current decoded sentence \(\langle A', B, C, D \rangle\)
- Output: \(Q(s, a)\) for each word in the sentence (edit or not) plus "no modification needed" action
- Reward: BLEU score between the target and source pair; normalized to -1 (decreased score), 0 (same score), +1 (increased score), compared to the previous iteration

Training DQN
- Iteratively improve value function, based on Bellman Equation
- Objective function: mean-squared error in Q-values
- Optimise objective end-to-end by SGD with gradient

Convergence Issues
- Softmax supervised signal to help the convergence of the DQN
- Pre-train the encoder-decoder LSTMs to generate state representation features and potential actions
- Replay memory: update DQN with sentence different from the current input
- Fixed target Q-network with periodically update

Evaluation Settings
- Regenerate natural sentences
- Randomly select 12000 sentences, with max length of 30, from the Billion Word Corpus. Train with 10000 sentences; test with 1000 seen sentences and 1000 unseen sentences
- Comparison baseline: encoder-decoder LSTM used for machine translation, with beam search of size 1

Predictive Performance
- Performed competitively well when decoding sentences from the training set
- Significantly outperformed the baseline when decoding unseen sentences

Learning Curves
- Both the state generation functions and the DQN were easy to be trained

Examples
- Click here to read more than the New York Times
- Click here to read more from the New York Times

Conclusions
- Deploy a Deep Q-Network (DQN) to embrace an iterative decoding strategy for sequence to sequence learning
- Evaluation using a sentence regeneration task shows promising performance especially when decoding unseen sentences