Deep Active Learning for Dialogue Generation

Nabiha Asghar†, Pascal Poupart†, Xin Jiang‡, Hang Li‡
† Cheriton School of Computer Science, University of Waterloo, Canada. (nasghar,ppoupart)@uwaterloo.ca
‡ Noah’s Ark Lab, Huawei Technologies, Hong Kong. (jiang.xin,hangli.hl)@huawei.com

Objectives

- Augment a trained Seq2Seq [1] model with online Active Learning.
- Leverage a novel human-in-the-loop feedback mechanism to generate more coherent, interesting, relevant and engaging responses.
- Use active learning as a form of reinforcement: eliminate the need for hand-crafted reward functions.

State of the Art

LSTM Encoder-Decoder architectures like Seq2Seq [1] are known to generate linguistically robust but dull, short and generic responses (Figure 2).

Training Pipeline

We go through three training phases for the Seq2Seq model (one encoder layer and one decoder layer with 300 hidden LSTM units each):

Phase 1: Offline SL. Generic dataset (300K pairs) of movie dialogues.

Phase 2: Offline SL. Fine-tuning on a second dataset (8K pairs) of human-Jabberwacky bot chatlogs.

Phase 3: Online AL with a human trainer.

Heuristic Response Generation

We use hamming-diverse Beam Search [2] to generate $K$ responses at each turn. The beams are diversified by maximizing an objective that consists of a standard sequence likelihood term and a dissimilarity metric between the beams.

Beam Search: Likely to produce almost identical beams like “I don’t care!” and “I don’t care.”

Diverse Beam Search: Likely to produce beams like “I don’t care!” and “Who cares?”

User Study

Metrics like BLEU, ROUGE, WER and NIST are suitable for machine translation evaluation, but not for response quality evaluation in dialogue.

Online Training: One human trained the model with 200 prompts of his choice.

Test Prompts: We randomly selected 100 of those and linguistically rephrased them. Thus, “How’s it going” was altered to “How are you doing?”, “I hate you.” to “I don’t like you!”, etc.

Test Pairs: We collected the responses of three models SL1, SL2 and SL2+oAL to the test prompts.

Evaluation: We asked 5 human judges to rate the test pairs on 4 axes: Syntactic Coherence, Relevance to Prompt, Interestingness and User Engagement.

Experimental Evaluation

References
