Dual Learning for Machine Translation

Abstract

While neural machine translation (NMT) is making good progress in the past two years, tens of millions of bilingual sentence pairs are needed for its training. However, human labeling is very costly. To tackle this training data bottleneck, we develop a dual-learning mechanism, which can enable an NMT system to automatically learn from unlabeled data through a dual-learning game. This mechanism is inspired by the following observation: any machine translation task has a dual task, e.g., English-to-French translation (primal) versus French-to-English translation (dual); the primal and dual tasks can form a closed loop, and generate informative feedback signals to train the translation models, even if without the involvement of a human labeler. In the dual-learning mechanism, we use one agent to represent the model for the primal task and the other agent to represent the model for the dual task, then ask them to teach each other through a reinforcement learning process. Based on the feedback signals generated during this process (e.g., the language model likelihood of the output of a model, and the reconstruction error of the original sentence after the primal and dual translations), we can iteratively update the two models until convergence (e.g., using the policy gradient methods). We call the corresponding approach to neural machine translation dual-NMT. Experiments show that dual-NMT works very well on English French translation; especially, by learning from monolingual data (with 10% bilingual data for warm start), it achieves a comparable accuracy to NMT trained from the full bilingual data for the French-to-English translation task.

1 Introduction

Specifically, the dual-learning mechanism for MT can be described as the following two-agent communication game.

1. The first agent, who only understands language A, sends a message in language A to the second agent through a noisy channel, which converts the message from language A to language B using a translation model.

2. The second agent, who only understands language B, receives the translated message in language B. She checks the message and notifies the first agent whether it is a natural sentence in language B (note that the second agent may not be able to verify the correctness of the translation since the original message is invisible to her). Then she sends the received message back to the first agent through another noisy channel, which converts the received message from language B back to language A using another translation model.

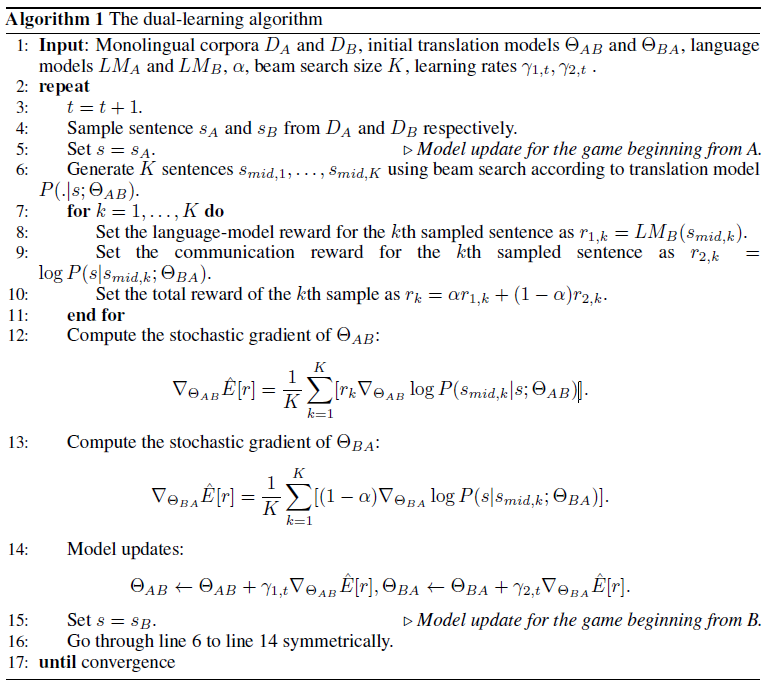
3. After receiving the message from the second agent, the first agent checks it and notifies the second agent whether the message she receives is consistent with her original message. Through the feedback, both agents will know whether the two communication channels (and thus the two translation models) perform well and can improve them accordingly.

4. The game can also be started from the second agent with an original message in language B, and then the two agents will go through a symmetric process and improve the two channels (translation models) according to the feedback.

2 Background: Neural Machine Translation

Neural machine translation systems are typically implemented with a Recurrent Neural Network (RNN) based encoder-decoder framework.

3 Dual Learning for Neural Machine Translation



4 Experiments

5 Extensions

Actually, many AI tasks are naturally in dual form, for example, speech recognition versus text to speech, image caption versus image generation, question answering versus question generation

(e.g., Jeopardy!), search (matching queries to documents) versus keyword extraction (extracting

keywords/queries for documents), so on and so forth.

Therefore, if more than two associated tasks can form a closed loop, we can apply our technology to improve the model in each task from unlabeled data.

6 Future Work