CS 194-129 Deep Neural Networks
Spring 2018 Lewis Lecture 19

1 Reading

Learning Cooperative Visual Dialog Agents

As *goal-driven* training for VQA and dialog agents, the authors pose a cooperative QA "image guessing" game between two agents, where a Q-agent asks an A-agent questions so that it can identify the correct image within a collection of images. The agents' policies are learned end-to-end (from pixels to multi-agent multi-round dialog to game reward) using deep RL.

The end product: a pair of agents that can understand the contents of an image and communicate that understanding in natural language (in *visually grounded dialog*).

End-to-End Learning for Negotiation Dialogues

In this case, agents with different goals (*reward functions*) attempt to agree on a common decision (*make a deal*) via natural language dialogue, despite not knowing each other's reward functions. The agents must be able to understand, plan, and generate dialogue in order to achieve their goals.

Given a dataset of natural language negotiations between two humans, agents can first learn to negotiate by maximizing the likelihood of human actions (in a supervised, end-to-end fashion). They can further be improved by training to maximize reward with self-play RL (using **dialogue rollouts** to estimate the expected reward of utterances).

2 Lecture

In a negotiation setting, an agent must generate language that is not only understandable, but also helps the agent achieve its goals. Generally, all of the agents' goals won't be (simultaneously) fully achievable, so the agents must come to some mutually acceptable compromise. Additionally, there is usually no simple way to achieve a goal, meaning interesting strategies can arise (e.g. strategic withholding of information).

Problem Setup

There are two agents, each with its own reward function. They converse via dialogue until a deal is agreed upon. Finally each agent independently selects a deal, and receive a reward from the environment.

A simple task: object division. Here, the agents are shown the same objects but different values for each. A value represents an object's worth to an agent. The goal is to come to an agreement as to how to split the objects, despite not knowing the other agent's reward function.

Models

We'll define end-to-end models, in which we input language and try to spit out some new language. This avoids the need for dialogue annotations ("which phase are we in?") and allows for easy multitasking (by combining datasets).

Baseline Model

As a baseline, we can (1) compress our dialogue into a sequence of tokens, (2) train a seq2seq model to predict tokens given the value function, and (3) train an extra classifier to predict the final deal (which/how many items the agent receives). We can do this from each agent's perspective. The model will be trained to maximize the likelihood of human-human dialogues, and will be decoded by sampling messages.

The problem is that the negotiator is too friendly; it tends to agree even if offered a very poor deal. This is because RNNs are prone to generating very generic responses. The model doesn't really understand what's going on here, and kind of squishes all of the different modes into these nonspecific utterances (which tend to indicate agreement).

Another problem is that the model can never go beyond what humans do.

Goal-Based Training

So instead we'll train to maximize the reward the agent receives. First we'll train according to the baseline model, and then we'll have the agent hold lots of dialogues with itself. The [normalized] reward it receives can be backpropagated using `REINFORCE`.

Note: we'll need a "likelihood of human dialogue" loss in addition to a reward-based loss. Otherwise the agents will start communicating in nonsense (at least to actual humans). We can interleave RL updates (*maximize reward*) with supervised learning updates (*generate humanlike dialogue*).

Other issues:

- The agents can be unwilling to compromise (because then the other agent would get a better reward).
- The model is sensitive to hyperparameters, and therefore difficult to tune/train.

To resolve these problems, we can try to train models that are able to anticipate reactions and guess how the dialogue is going to go. Such models should theoretically be able to negotiate better. In other words, we'll roll out how the dialogue is going to proceed from a certain point, predicting complete dialogue sequences which end with a received reward.

We'll have each agent choose the option that leads to the highest expected reward.

Another issue: word similarity does not equate to semantic similarity. Sentences with similar words might have very different meanings (e.g. "you take" versus "I take"). We can fix this by disentangling the reasoning step from the language step. First of all, we'll convert our RNN dialogue model into a hierarchical RNN (with a separate RNN to encode each sentence in isolation, and where each sentence is fed into a higher-level RNN). This shortens the dependencies between turns, s.t. the model can more easily access information from many tokens ago. We can also introduce latent variables. Instead of directly generating a sentence, the model will first generate a discrete latent variable z (which represents the meaning of the next message), then use z to generate said message. This factorizes the problem into "what to say" (reasoning) and "how to say it" (language).

Thus, during RL, we can fix all of the parameters that correspond to translating z into natural language, and only fine-tune the parameters that affect which z we're going to generate. This means the model can't actually diverge from human language, but *can* choose what it's going to say. So RL now decides *what* to say, but not *how* to say it.

References

[1] Abhishek Das, Satwik Kottur, Jos M. F. Moura, Stefan Lee, Dhruv Batra. Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. *arXiv preprint arXiv:1703.06585* (2017).

[2] Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, Dhruv Batra. Deal or No Deal? End-to-End Learning for Negotiation Dialogues. *arXiv preprint arXiv:1706.05125* (2017).