Alexa Project Technical Architecture

William Guss, Phillip Kuznetzov, James Bartlett, Piyush Patell
Machine Learning at Berkeley
{wguss, philkuz, james.bartlett, PIYUSH PO PO W}@berkeley.edu

1 Introduction

In accordance with our vision, in this document, we will propose a radically different architectural approach to learning generative models of human-level conversation. The primary motivation behind the following architecture is that conversation can be modeled as a game in which two players attempt to maximize a mutuakl engangement reward under a minimum information theoretic constraint. In the framework of reinforcement learning, we can view the Social Bot as an agent, π, hereafter referred to as the conversationalist, whose environment is the conversation itself. An additional constraint on π is that of one-shot learning, the ability to work with and engage details pertinent to the conversation, drawing from a model free source of external information, is essential to achieving human-level communication. Under this paradigm, we develop the architecture of the Alexa Social Bot by proposing a novel procedure for semantic information internalization via document embeddings in the regime of deep reinforcement learning with memory augmented metalearning.

2 Model

3 Training

With the core architectural components of the model developed, we draw from a state of the art development in RL, inverse reinforcement learning, to impart natural and expert level conversational skill on the Social Bot.

The principle idea behind inverse reinforcement learning is that behavioral cloning is limited to a biased distribution of only expert examples; that is, if the social bots model is trained to exactly replicate a human level conversationalist response to different dialogues, then its ability to extrapolate beyond that known set of dialogues is unstable due to the severe nonlinearity of the model itself. In answer to this, inverse reinforcement learning attempts to infer the reward function of the agent being cloned. Knowing the reward function which an expert policy maximizes across a variety of states allows that policy to be cloned in a robust and dynamical fashion independent of the particular states in which it has acted; the cloned policy learns not about the particular examples to which it is tuned, but about the general principles which govern the expert policy.

In the context of the social bot, learning to communicate by understanding the forces which govern human level conversation as opposed to mimicking the conversation itself has immediate and fundamental benefits which have not yet been utilized in natural language processing. Therefore we propose the the following training regime for our model.