Research in the RL Lab

Doina Precup
- Reinforcement learning
- Deep learning
  - Generative models
  - Deep RL
- Health applications

Joelle Pineau
- Reinforcement learning
- Deep learning
  - Dialogue systems
  - Deep RL
- Robotics
- Health applications

Jackie CK Cheung
- Natural language processing
  - Natural language generation
  - Automatic summarization
- Common-sense reasoning

Prakash Panangaden
- Semantics of probabilistic systems
- Logic and Computation
- Machine learning
  - Weighted automata
- Quantum mechanics
Research in the RL Lab

• Autonomous robot navigation (SmartWheeler)
• Model-based RL
• Hierarchical RL (options)
• Multitask/transfer learning in deep RL
• Conditional computation
• Real-time machine translation with deep RL
• Deep energy-based causal models
• Deep generative models
• Spectral learning
• Imitation learning
• Pedestrian motion prediction
• Human motor control with RL
• Comparative genomics

• Common-sense reasoning in NLP
• Natural language generation
• Automatic summarization of fiction
• Task-oriented dialogue systems
• Chatbots
• Dialogue evaluation
• Dialogue evaluation
• Differential privacy
• Predicting movement of monkey populations
• Automatic sleep staging with EEG
• Seizure prediction with EEG
• Extubation prediction for infants
• Weighted automata
• etc.
Deep Reinforcement Learning
Options

• AIs will need to learn and plan at multiple levels of temporal abstraction

• Options are a (minimal) way to formalize temporal abstraction in reinforcement learning

• When planning, first choose an option (high-level plan), then execute the option (low-level details)
Option-Critic

- Learns options **automatically**
- Each option is a policy. Options are chosen using a meta-policy (‘policy over options’)
- Options learn to **specialize**
- Options **aid transfer** to related tasks

Deep Option-Critic

Figure 4: Deep neural network architecture. A concatenation of the last 4 images is fed through the convolutional layers, producing a dense representation shared across intra-option policies, termination functions and policy over options.

Figure 6: Up/down specialization in the solution found by option-critic when learning with 2 options in Seaquest. The top bar shows a trajectory in the game, with “white” representing a segment during which option 1 was active and “black” for option 2.
Conditional Computation

• Running large neural networks at test time can be expensive!

• Want to learn an input-dependent dropout

• Different areas of network specialize for different classes

• Beneficial for lower-power devices (e.g. phones)
Conditional Computation

• Learn policy (red units) that drops out certain nodes of a neural network (blue units)

• Can do this for both feed-forward and convolutional networks

Conditional Computation

- Dropout policies are *input-dependent*
- Can achieve up to *5x speed-up* with similar accuracy
- Single hyperparameter controls accuracy/speed trade-off

Figure: Probability distributions of the dropout policy for class 0 (left) and class 1 (right)
Neural Dialogue Systems
End-to-End Dialogue Systems

- A single model trained directly on conversational data

- Uses a single objective function, usually maximum likelihood on next response

- Most of our work uses neural networks to predict the next response. (Ritter et al., 2011; Sordoni et al., 2015; Shang et al., 2015)
Ubuntu Dialogue Corpus

• Large dataset of ~1 million tech support dialogues
• Scraped from Ubuntu IRC channel
• 2-person dialogues extracted from chat stream

Generative Models

- Use RNN to **encode** text into fixed-length vector representation

- Use another RNN to **decode** representation to text

- Can make this hierarchical


The Problem of Generic Responses

• Most models trained to predict most likely next utterance given context

• But some utterances are likely given any context!

• Neural models often generate “I don’t know”, or “I’m not sure” to most contexts

(Li et al., 2016)
Variational Encoder-Decoder

• Augment encoder-decoder with Gaussian latent variable

• Inspired by VAE (Kingma & Welling, 2014)

• When generating first sample latent variable, then use it to condition generation

• Generates longer responses with higher entropy

Evaluating Dialogue Responses

**Context**

Hey, want to go to the movies tonight?

**Generated Response**

Yeah, let’s go see that movie about Turing!

**Ground-truth response**

Nah, I’d rather stay at home, thanks.

- Having humans evaluate is expensive and time-consuming
- Want to evaluate dialogue responses automatically (an automatic Turing test)
Existing Metrics Correlate Poorly with Human Judgement

Goal: (roughly linear correlation)

Reality:

- Asked 25 CS students to rate the quality of dialogue responses on a scale from 1 – 5, on Twitter and Ubuntu datasets
- The scores from the automatic metrics (e.g. BLEU) correlate very poorly or not at all with human scores

Thank you!
References


