McGill Reasoning & Learning Lab: Research Overview

Ryan Lowe

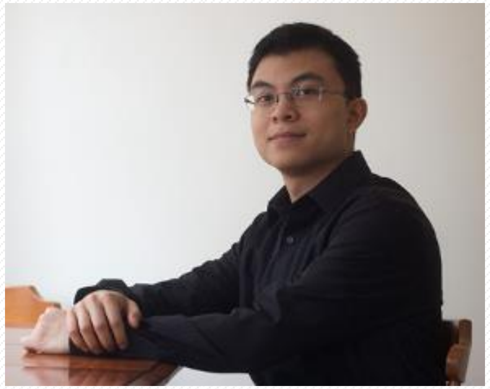
McGill University


McGill
COMPUTER
SCIENCE

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

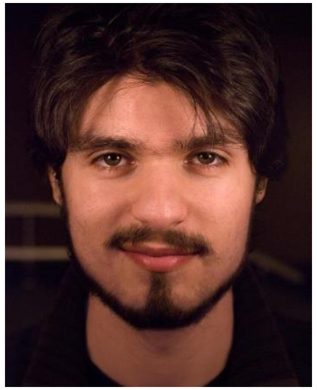
Deep RL

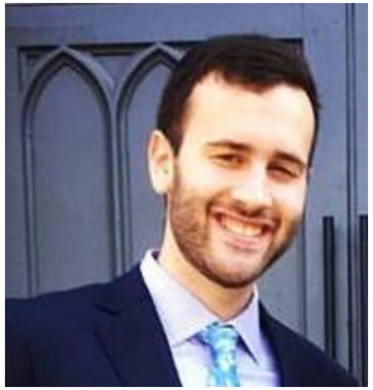
- Common-sense reasoning in NLP
- Natural language generation
- Automatic summarization of fiction
- Task-oriented dialogue systems
- Chatbots
- Dialogue evaluation
- Differential privacy
- Predicting movement of monkey populations
- Automatic sleep staging with EEG
- Seizure prediction with EEG
- Extubation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

Dialogue

Deep Reinforcement Learning






Options

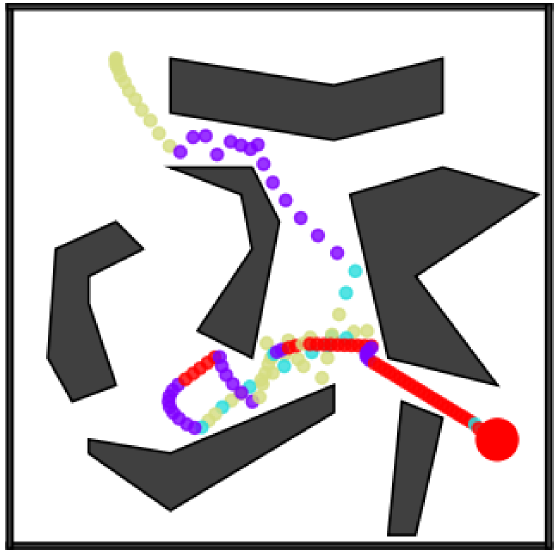
- AIs will need to learn and plan at multiple levels of temporal abstraction

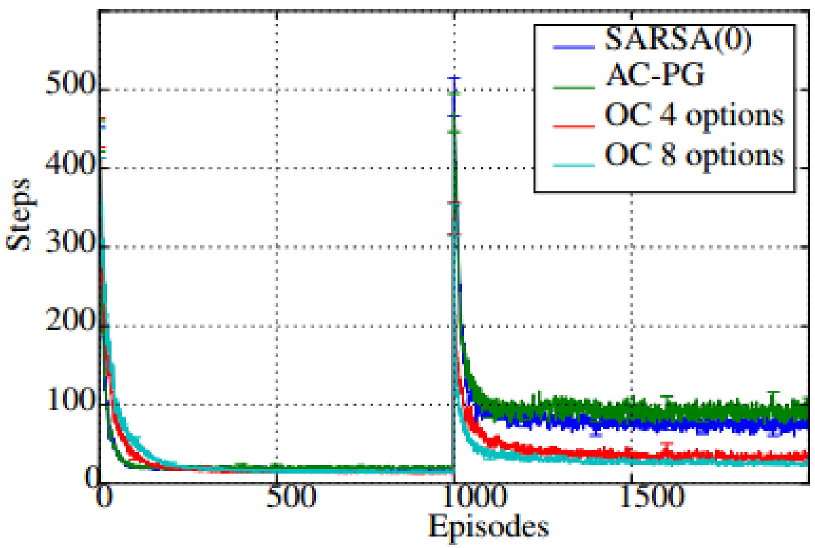
- Options are a (minimal) way to formalize temporal abstraction in reinforcement learning

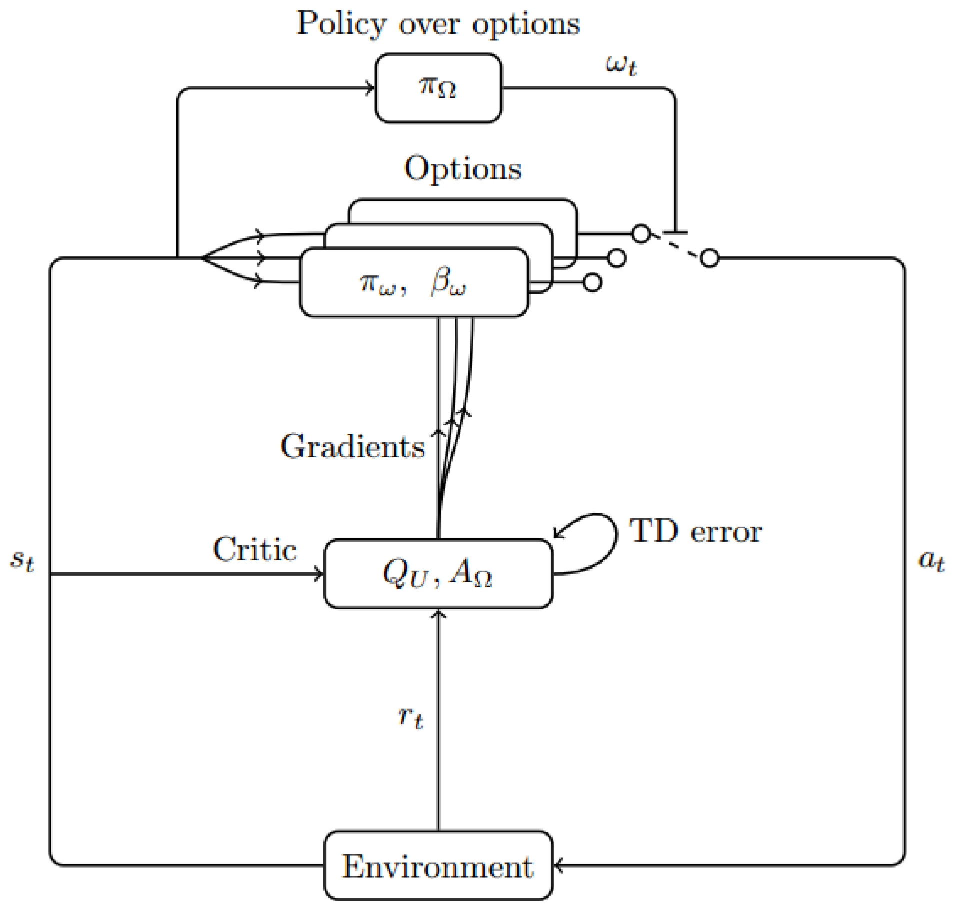
- When planning, first choose an option (high-level plan), then execute the option (low-level details)

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







Bacon, P. L., Harb, J. & Precup, D. (2016). The option-critic architecture. *Submitted to AAAI*.

Deep Option-Critic

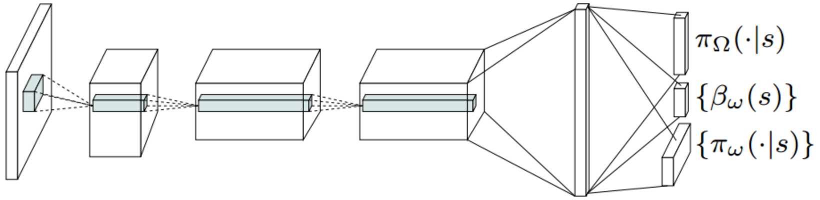


Figure 4: Deep neural network architecture. A concatenation of the last 4 images is fed through the convolutional layers, producing a dense representation shared across intra-option policies, termination functions and policy over options.

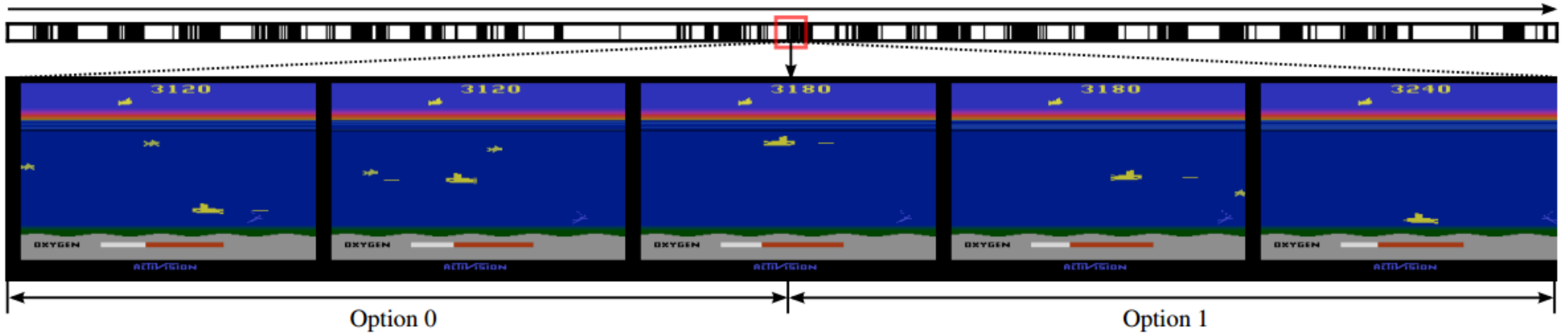
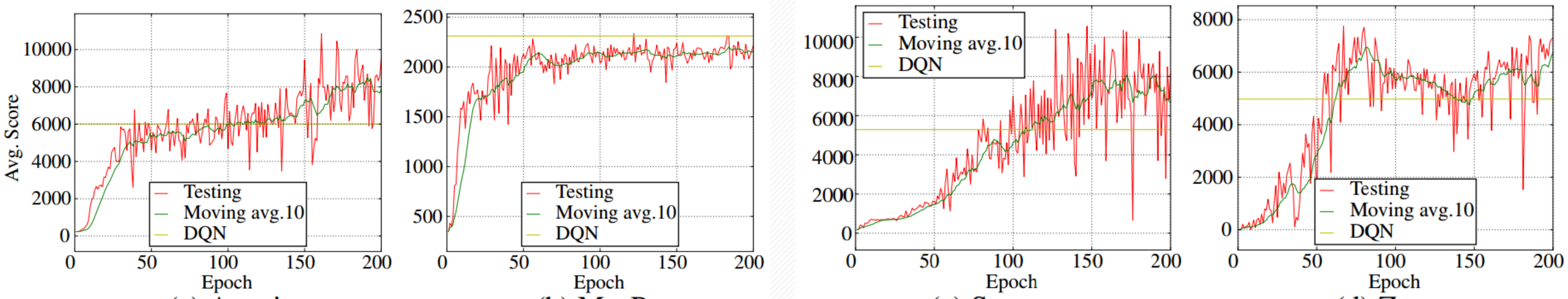


Figure 6: Up/down specialization in the solution found by option-critic when learning with 2 options in Seaquest. The top bar shows a trajectory in the game, with "white" representing a segment during which option 1 was active and "black" for option 2.



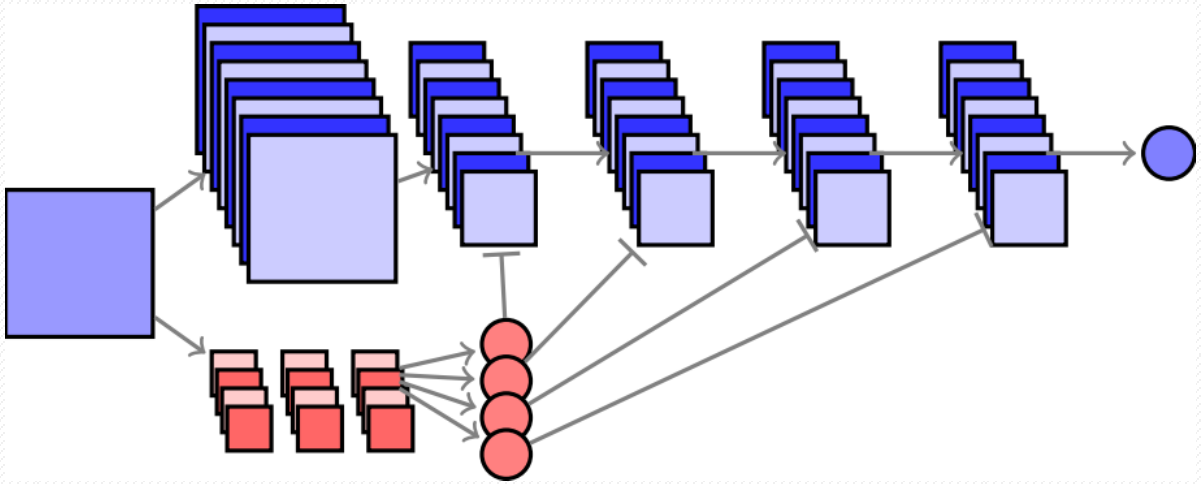
(a) Asterix (b) Ms. Pacman (c) Seaquest (d) Zaxxon

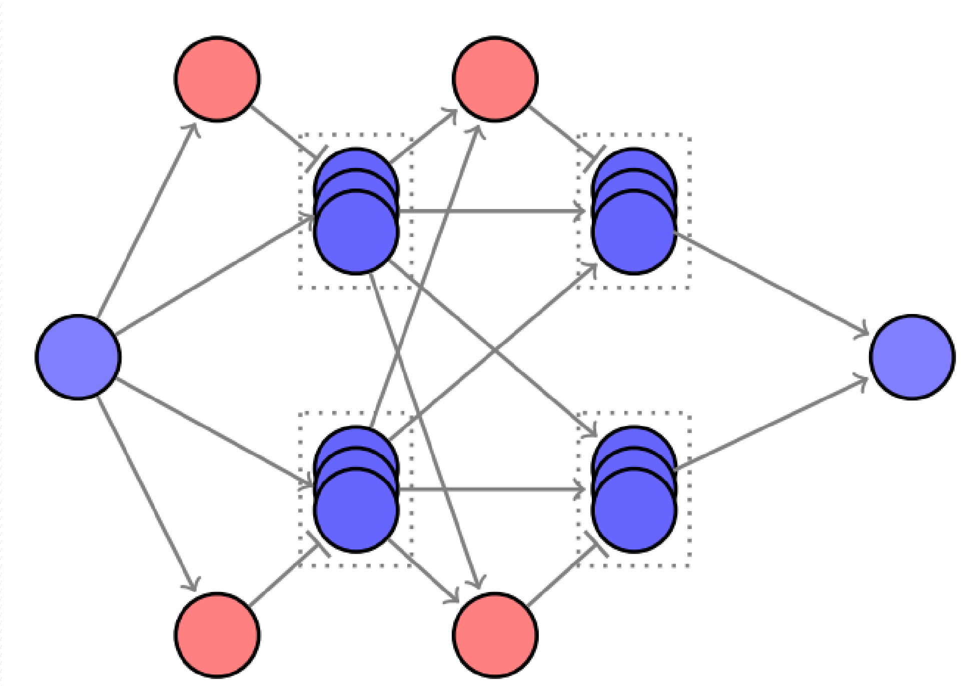
Conditional Computation

- 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

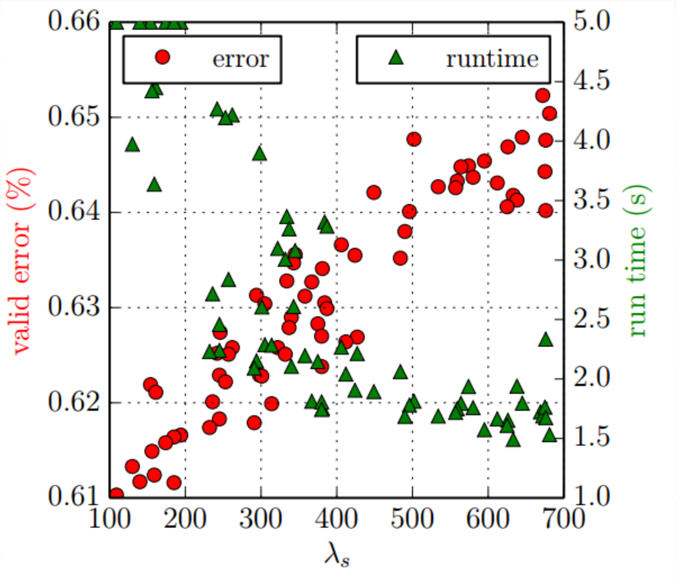




Bengio, E., Bacon, P. L., Lowe, R., Pineau, J., & Precup, D. (2016) Reinforcement learning of conditional computation policies for neural networks. *ICML Workshop on Abstractions in RL*.

Conditional Computation

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



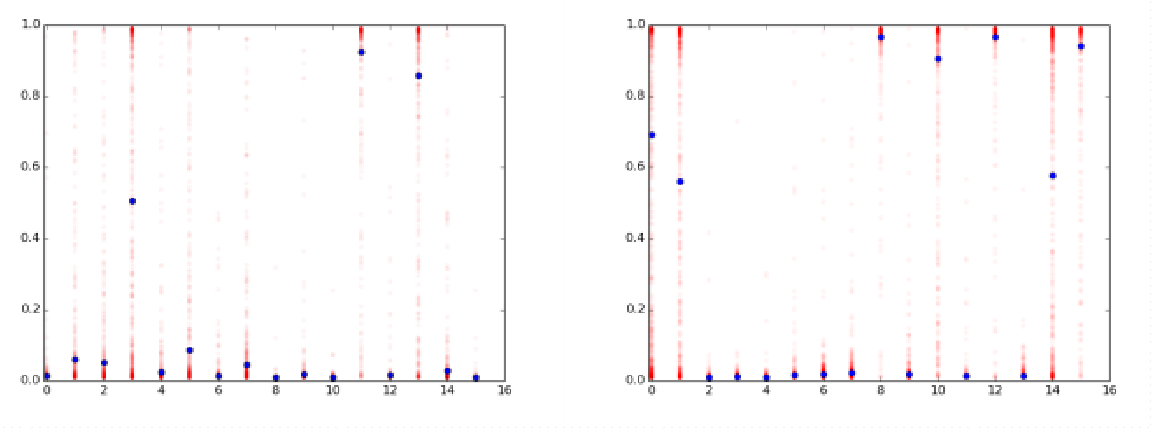
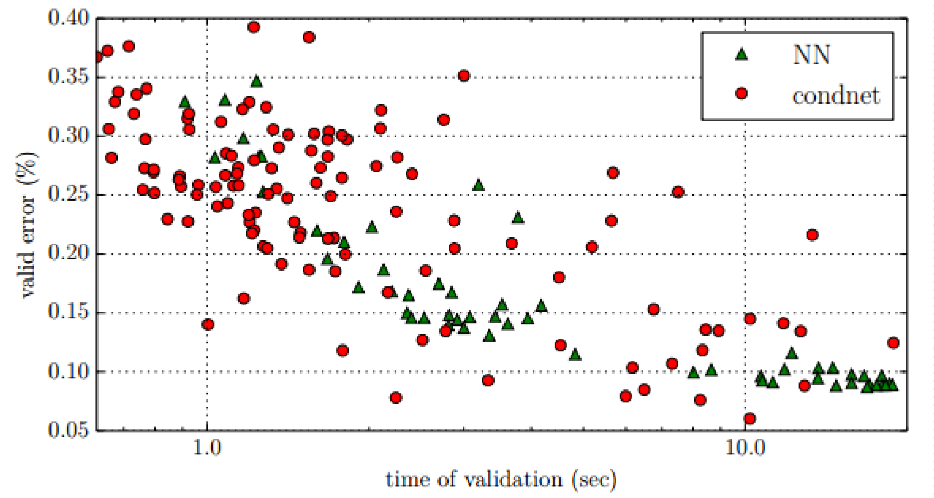


Figure: Probability distributions of the dropout policy for class 0 (left) and class 1 (right)



Neural Dialogue Systems





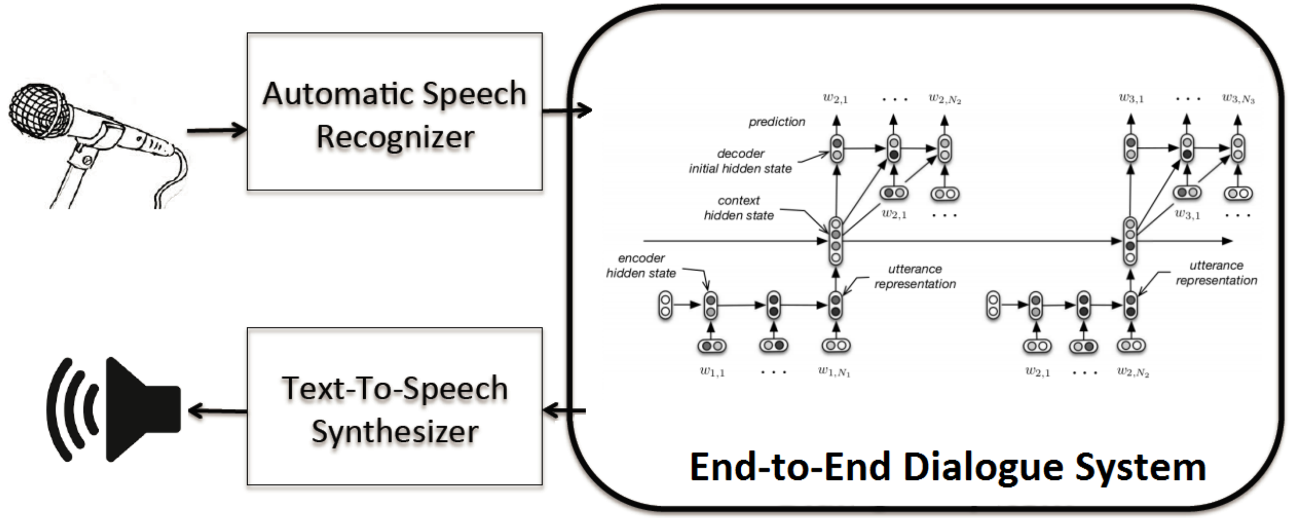




End-to-End Dialogue Systems

- A single model trained directly on conversational data

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



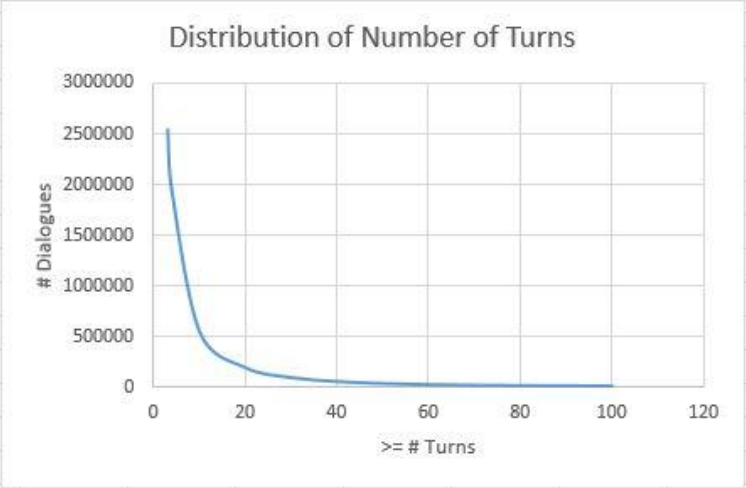
- Most of our work uses neural networks to predict the next response. (Ritter et al., 2011; Sordoni et al., 2015; Shang et al., 2015)

Ubuntu Dialogue Corpus

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

ubuntuaddicted	what's my ip?	[02:59]
DF3D2	k11: so I reinstalled fglrx manually, and startx just keeps saying "no protocol specified"	[02:59]
nahtnam	ubuntuaddicted: Are you in europe?	[03:00]
xtpeeps	Anyone can introduce me some interest channel of irc:p THX	[03:00]
timwis	hey guys, just did a fresh install on a Lenovo yoga to Pro, and I'm getting Wi-Fi is disabled by hardware switch. Any idea how to resolve?	[03:01]
DF3D2	k11: and time out in locking the .Xauthority file	[03:01]
Bashing-om	DF3D2: Before you rebooted, did you do -> sudo amdconfig --initial <- ??	[03:01]
timwis	this article suggests I modify ideapad-laptop.c but it doesn't seem to exist on the filesystem http://billauer.co.il/blog/2014/08/linux-ubuntu-yoga-hardware-blocked-wireless-lan/	[03:01]
xangua	!alis | xtpeeps	[03:01]
ubottu	xtpeeps: alis is a services bot that can help you find channels. Read "/msg alis help list" . For more help or questions relating to alis, please join #freenode. Example usage: /msg alis list #ubuntu* or /msg alis list *http*	[03:01]
DF3D2	Bashing-om: yes	[03:01]
ubuntuaddicted	nahtnam, no, why?	[03:01]
DF3D2	Bashing-om: I also did rm -r ~/.Xauthority as I saw suggested on the web, didn't help	[03:02]
cfhowlett	timwis, yep. only took me 3 years to learn. hit the windows wifi switch but experiment with combinations: ctrl F2 does it on my DELL in ubuntu. In windows: f2	[03:02]
cfhowlett	timwis, ctrl, alt, shift and super keys are all candidates	[03:03]
timwis	that article actually suggests that with the Lenovo laptops there's a problem beyond that	[03:04]
timwis	what is the super key?	[03:04]
cryptodan	the windows key	[03:04]
cfhowlett	timwis, aka "windows" key	[03:04]
timwis	ah! super indeed	[03:04]
somsip	timwis: windows key, or mod key, between left ctrl and left alt usually	[03:04]

Sender	Recipient	Utterance
Old		I dont run graphical ubuntu, I run ubuntu server.
bur[n]er	Old	you can use "ps ax" and "kill (PID#)"
kuja	Taru	Haha sucker.
Taru	Kuja	?
kuja	Taru	Anyways, you made the changes right?
Taru	Kuja	Yes.
kuja	Taru	Then from the terminal type: sudo apt-get update
Taru	Kuja	I did.



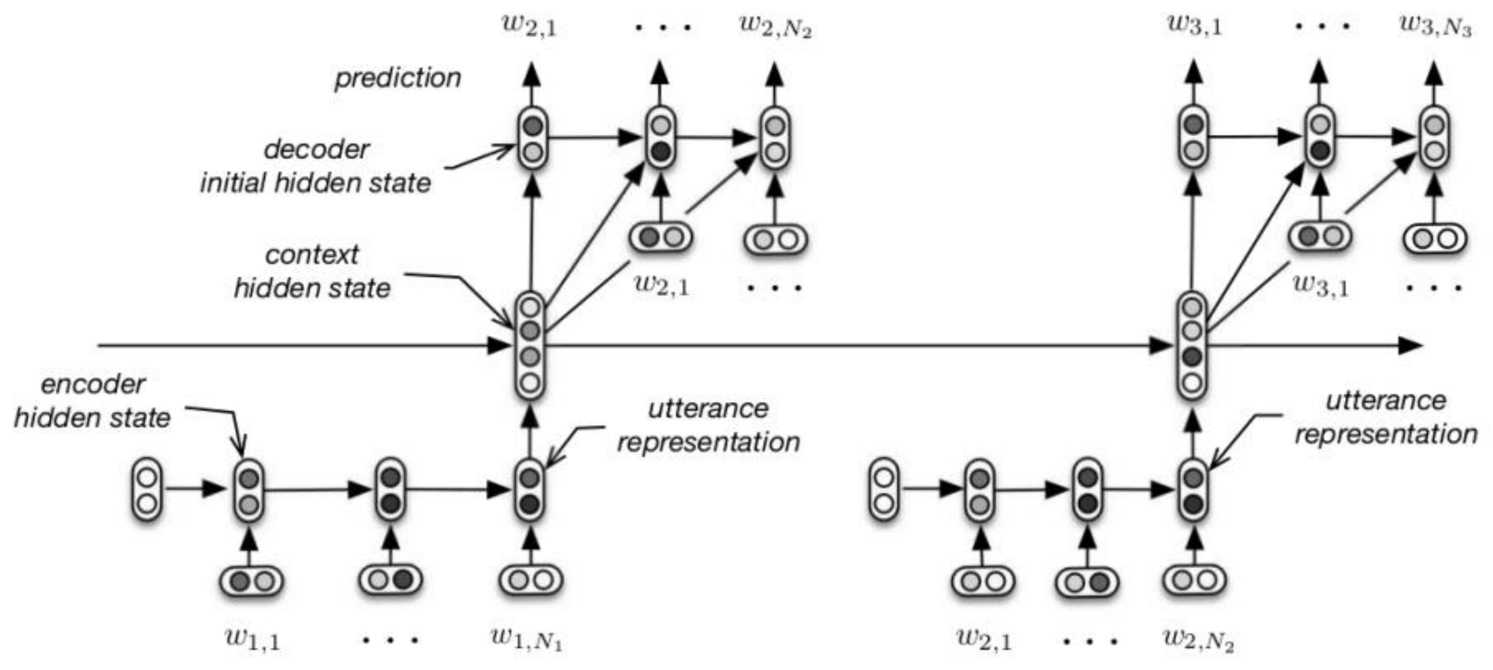
Lowe*, Pow*, Serban, Pineau. "The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems." *SIGDIAL*, 2015.

Generative Models

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

- Use another RNN to decode representation to text

- Can make this hierarchical



Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." *EMNLP* 2014.

Serban, Sordoni, Bengio, Courville, Pineau. "Building End-to-End Dialogue Systems using Generative Hierarchical Neural Network Models" *AAAI*, 2015.

The Problem of Generic Responses

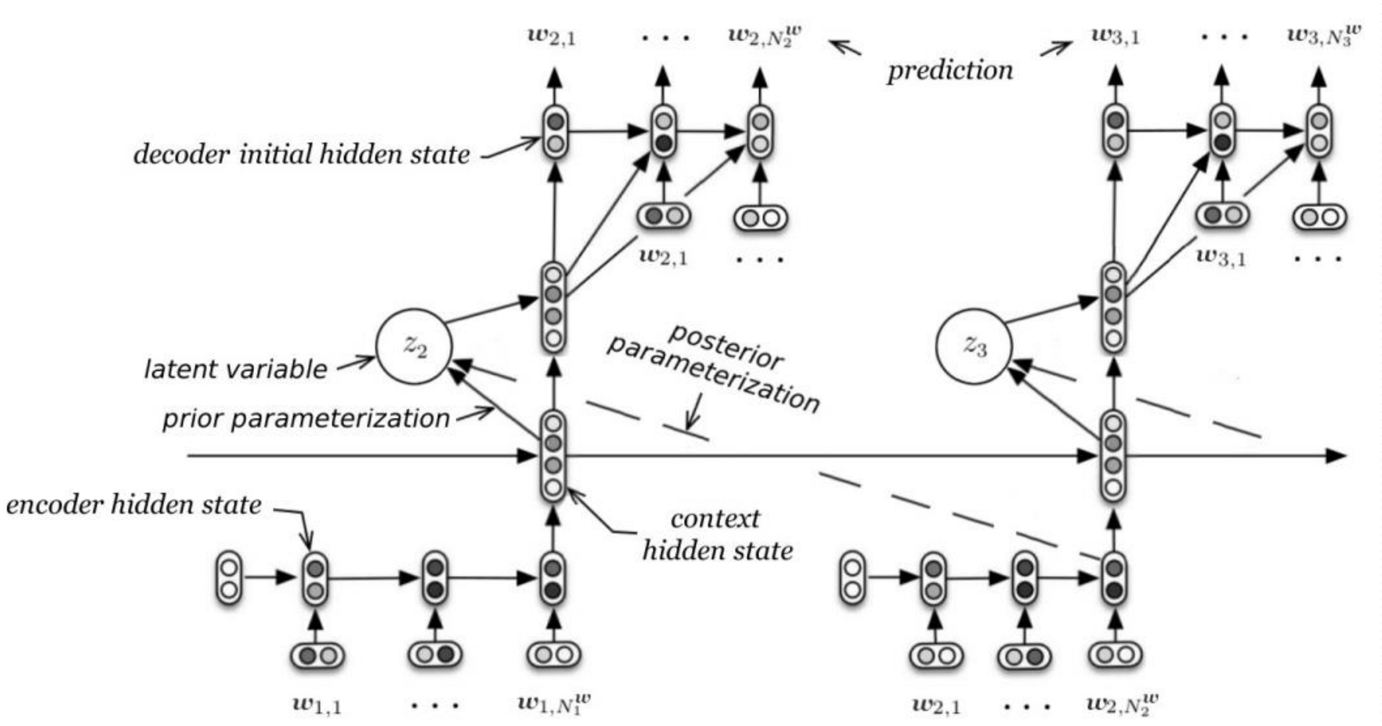
- Most models trained to predict most likely next utterance given context
- But some utterances are likely given any context!
- Neural models often generate "I don't know", or "I'm not sure" to most contexts

Input: What are you doing?		
−0.86	I don't know.	–
−1.03	I don't know!	–
−1.06	Nothing.	–
−1.09	Get out of the way.	–
Input: what is your name?		
−0.91	I don't know.	...
−0.92	I don't know!	–
−0.92	I don't know, sir.	–
−0.97	Oh, my god!	–
Input: How old are you?		
−0.79	I don't know.	...
−1.06	I'm fine.	–
−1.17	I'm all right.	–
−1.17	I'm not sure.	–

(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

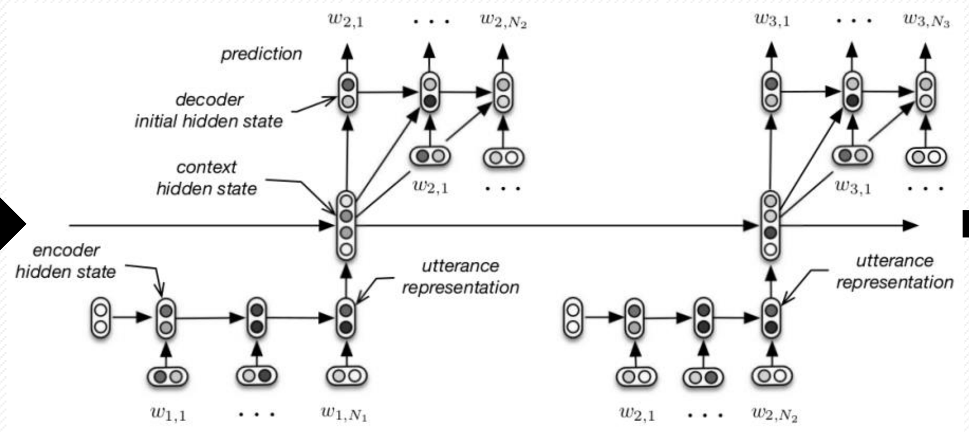


Serban, Sordoni, Lowe, Charlin, Pineau, Courville, Bengio. "A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues." *arXiv:1605.06069*, 2016.

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

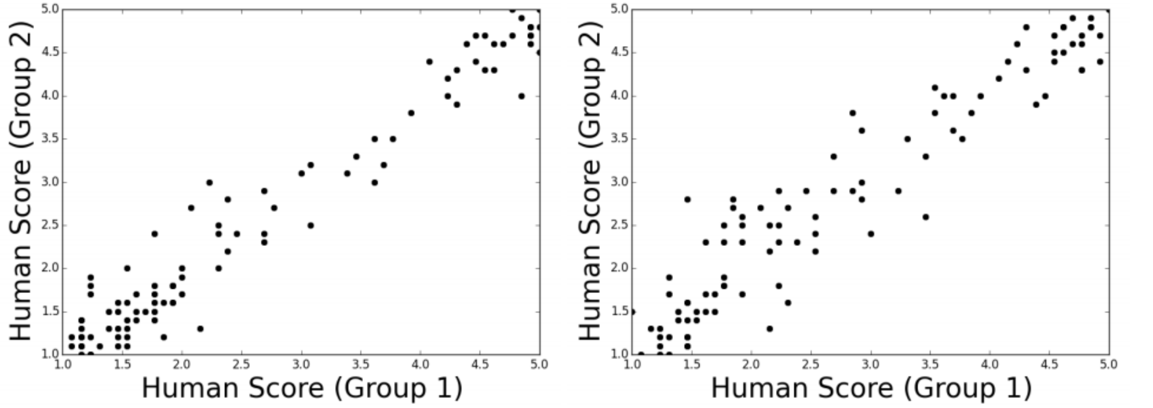
Nah, I'd rather stay at home, thanks.

SCORE

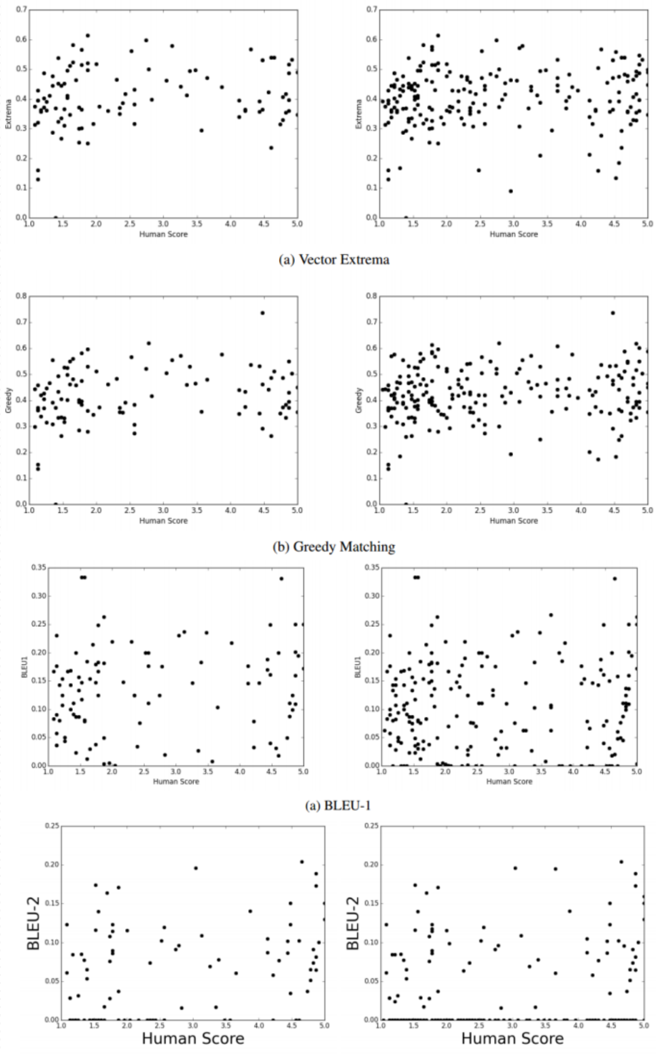
- Having humans evaluate is expensive and time-consuming
- Want to evaluate dialogue responses automatically (an automatic Turing test)

Existing Metrics Correlate Poorly with Human Judgement

Goal: (roughly linear correlation)



Reality:



(a) Vector Extrema

(b) Greedy Matching

(a) BLEU-1

(b) BLEU-2

- 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

Liu*, Lowe*, Serban*, Noseworthy*, Charlin, Pineau. "How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Systems." *EMNLP*, 2016.

Thank you!

References

- Bengio, E., Bacon, P. L., Lowe, R., Pineau, J., & Precup, D. (2016) Reinforcement learning of conditional computation policies for neural networks. *ICML Workshop on Abstractions in RL*.
- Bengio, E., Bacon, P. L., Pineau, J., & Precup, D. (2015). Conditional Computation in Neural Networks for faster models. *ICLR Workshop*.
- Lowe, R., Pow, N., Serban, I., & Pineau, J. (2015). The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. *SIGDIAL*.
- Liu, C. W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., & Pineau, J. (2016). How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *EMNLP*.
- Serban, I. V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., & Bengio, Y. (2016). A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues. *arXiv preprint arXiv:1605.06069*.
- Bacon, P. L., & Precup, D. (2015). The option-critic architecture. In *NIPS Deep Reinforcement Learning Workshop*.
- Dodge, J., Gane, A., Zhang, X., Bordes, A., Chopra, S., Miller, A., ... & Weston, J. (2016). Evaluating prerequisite qualities for learning end-to-end dialog systems. In *ICLR*.
- Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. In *ICLR*. Ritter, A., Cherry, C., & Dolan, W. B. (2011). Data-driven response generation in social media. In *EMNLP*.
- Shang, L., Lu, Z., & Li, H. (2015). Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.
- Sordoni, A., Galley, M., Auli, M., Brockett, C., Ji, Y., Mitchell, M., & Dolan, B. (2015). A neural network approach to context-sensitive generation of conversational responses. In *NAACL-HLT*.
- Precup, D., Sutton, R. S., & Singh, S. (1998, April). Theoretical results on reinforcement learning with temporally abstract options. In *European conference on machine learning* (pp. 382-393). Springer Berlin Heidelberg.