Reinforcement Learning towards AI-driven Control

Mengdi Wang
Center for Statistics and Machine Learning

PRINCETON UNIVERSITY

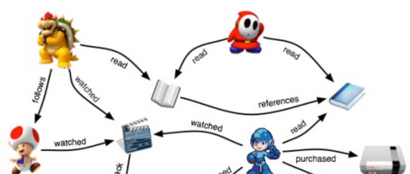
Google DeepMind

Reinforcement learning happening now

Game AI

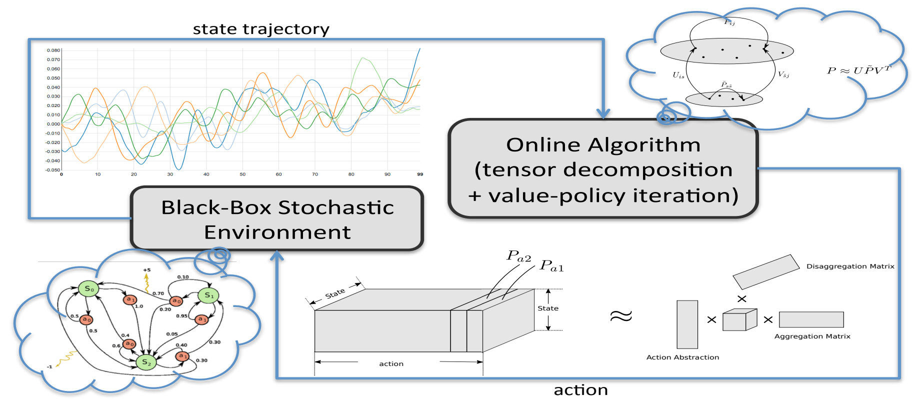
Atari, Breakout, Poker, Battle Zone, Go, Starcraft, …

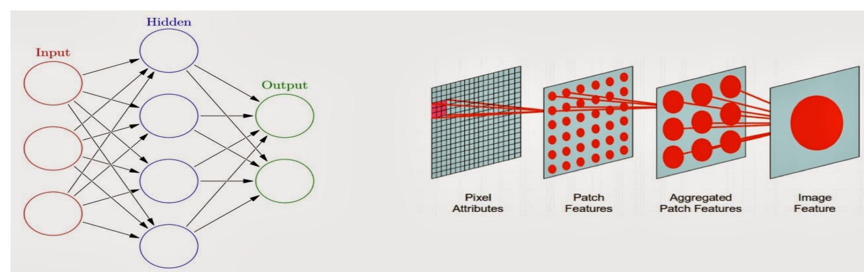
Robots, self-driving cars, e-commerce



Core technologies for Reinforcement Learning

1. Dimension Reduction and Feature Engineering from Machine Learning
2. Deep Policy Network and Deep Value Network
3. Fast training, asynchronous parallel computation
4. Real-time data collection and live experiment
5. CV+NLP+Sensing+Attention Networks
6. State representation learning from high-dimensional data

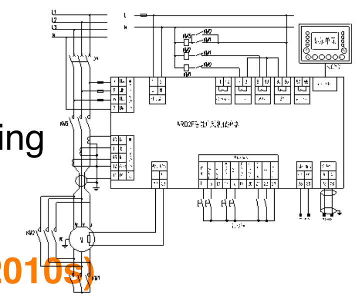


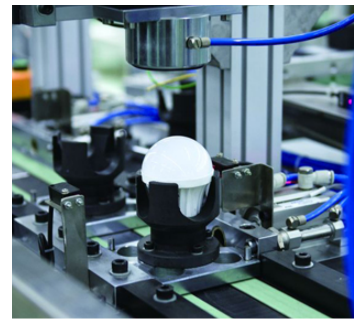


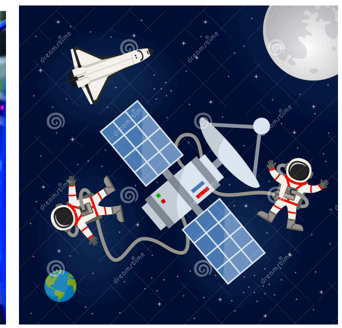
Brief History of AI

Control (1950-1990)
Feedback control for **known** physical systems

Applications: robotics, automation, aeroengineering

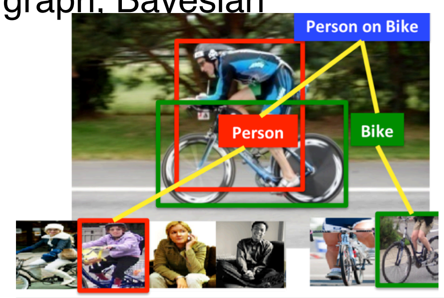


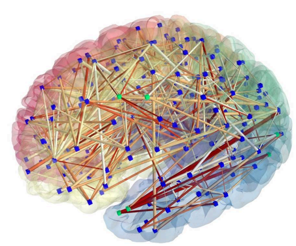




Machine Learning and Data Science (1990 – 2010s)
Find **static mapping** relations from big data

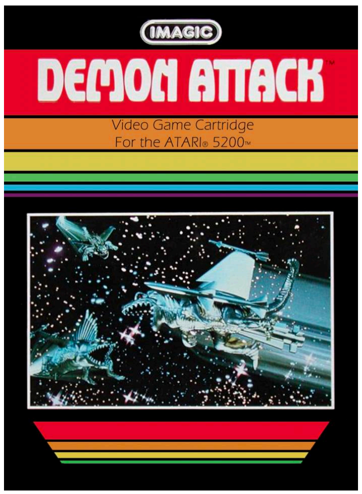
Applications: image recognition, natural language processing, translation, knowledge graph, Bayesian random fields



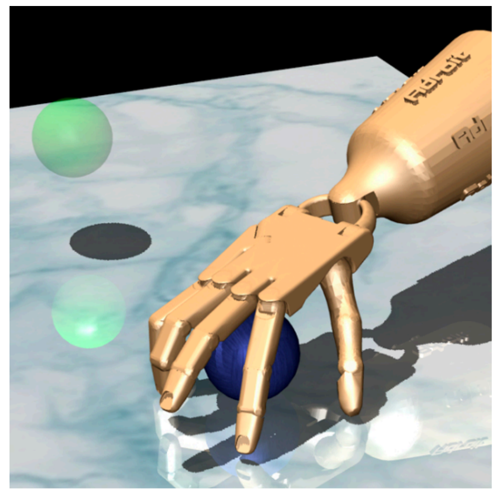


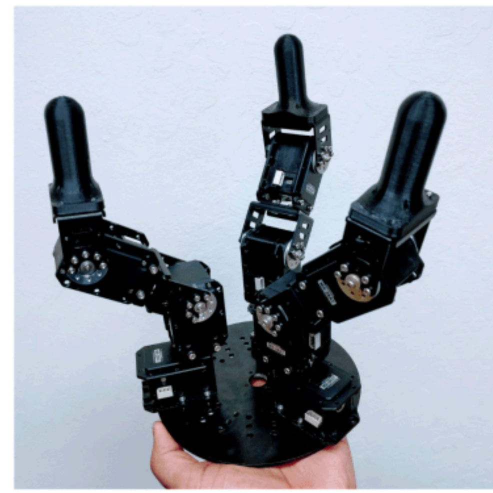
What's next - Deep Reinforcement Learning - Learn to control in **unknown dynamical environment**

How efficient is RL now?
- still heavily relying on simulation and brute force



Game AI means infinite data
AI Training time: ~ 2 GPU hrs
What it means for human： 32 days

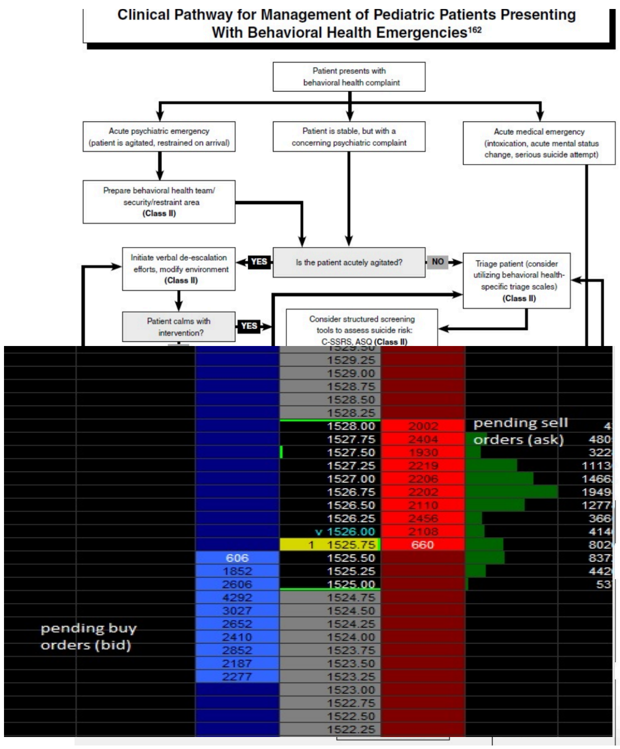


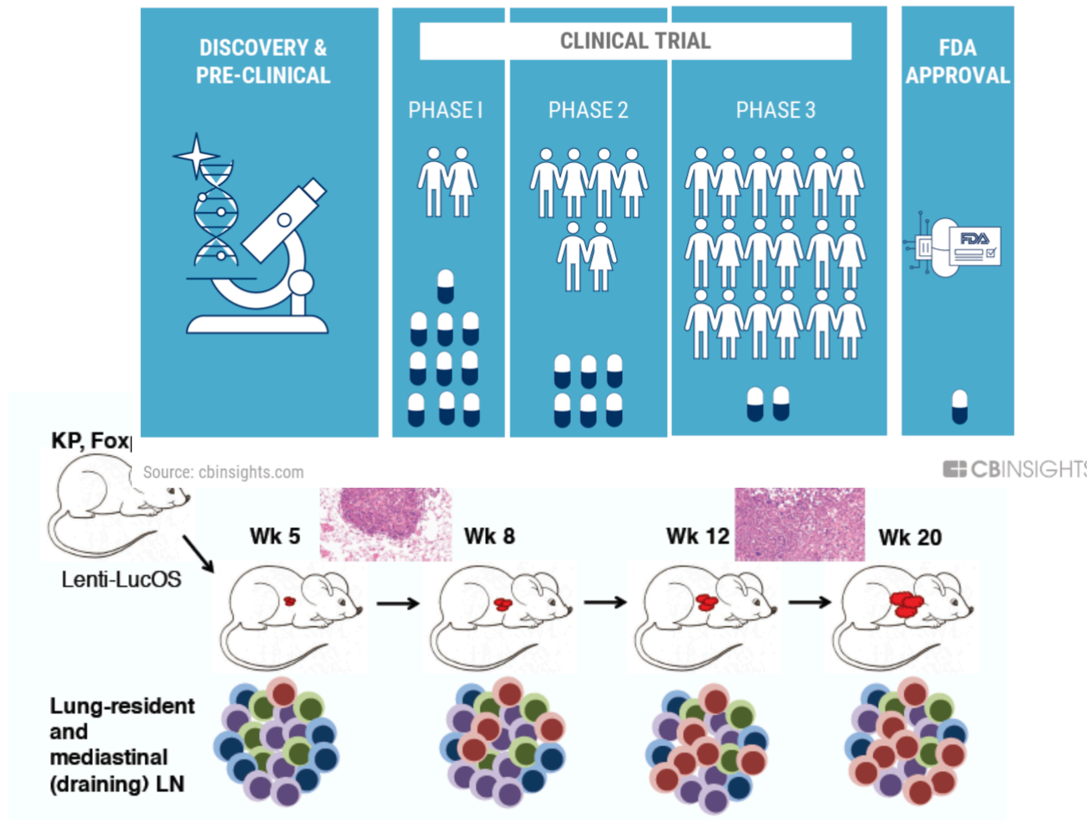


What AI Training time means for human: XXXX years

Lack of generalizability \ sample inefficiency

What if the data/trial is limited and costly

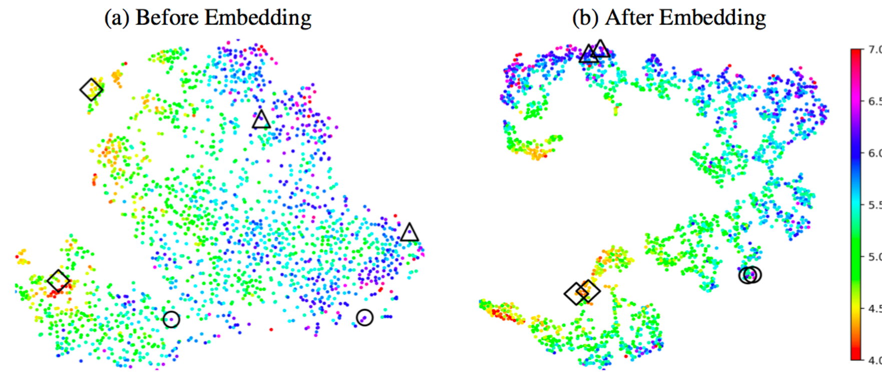




Improve RL efficiency and generalizability

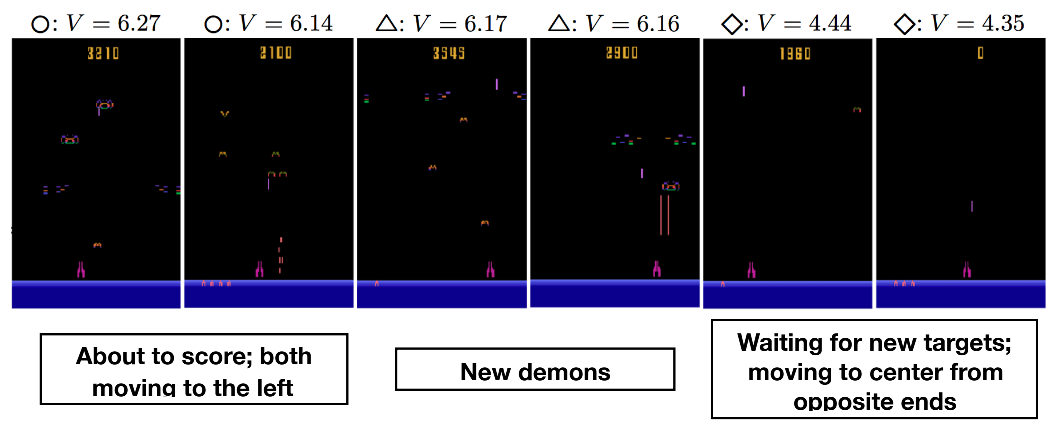
Acceleration and efficiency improvement

- For an100-step planning problem, our algorithms improve sample efficiency by 100,000,000 time （NeurIPS 17, 18, ICML



Improve generazability

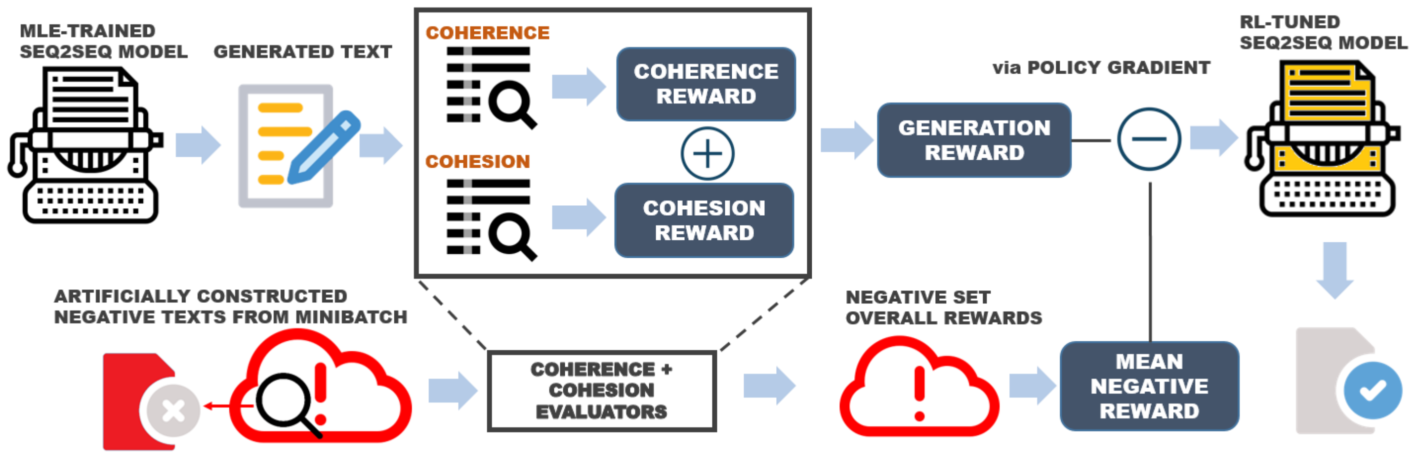
- Dimension reductio in “phase” space - finding latent logic of complex games



Imitation Reinforcement Learning

- Learn from human experts and learn faster （ICML 19)

RL for Writing Long-Form Text



Human evaluations via Amazon Mechanical Turk:

Cohesion					Coherence				
Human judges preferred:					*Human judges preferred:*				
Our Method		Neutral	Comparison		Our Method		Neutral	Comparison	
$G_{\text{MLE+RL}}$	**36.41%**	33.57%	30.50%	G_{MLE}	$G_{\text{MLE+RL}}$	**37.23%**	31.44%	31.80%	G_{MLE}
$G_{\text{MLE+RL}}$	29.91%	30.85%	**39.24%**	Human	$G_{\text{MLE+RL}}$	28.96%	31.32%	**39.72%**	Human

WSCho*,PZhang, YZhang,XLi,MGalley,MWang,JGao. Towards coherent and cohesive long- form text generation. *Narrative Understanding workshop at NAACL*, 2019.

RL for Quantitative Trading

RL in HFT and Monetization
Use reinforcement learning to adapt trading strategies
Why RL?
- Market impact cannot be predicted. Backtesting doesn't work

Solution:
- Learning-while-doing， inverse reinforcement learning

Deep RL for Dynamic Portfolio Optimization
- Classical finance model does not capture dynamic market movements

Why RL?
- Collect data online and adapt strategies

Solution:
- Deep recurrent neural networks for predicting market movements

Y Sun, J Mulvey, M Wang, J Ye. Learning Mean Reversion Trading Strategy with Transaction Costs Using Deep Neural Networks. Quant Finance, to appear

Qi Wu, Shumin Ma, Cheuk Hang Leung, Wei Liu. Understanding Distributional Ambiguity via Non-robust Chance Constraint. 2019

Thank you!

Mengdi Wang's Group

Group Members: Lin Yang, Saeed Ghadimi, Yichen Chen, Galen Cho, Yaqi Duan, Hao Gong, Zheng Yu, Hao Lu, Zachary Moore

Center for Statistics and Machine Learning

Department of Operations Research and Financial Engineering

Department of Computer Science

Princeton University

Princeton, NJ, USA 08540

PRINCETON UNIVERSITY