Reinforcement Learning in Healthcare: A Survey

Han Liang Wee Eric

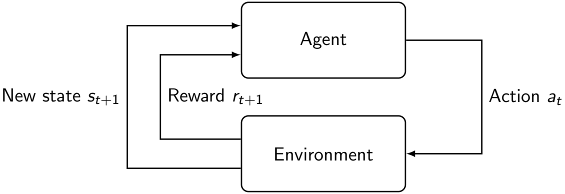
Slides at https://eric-han.com/rl-healthcare.pdf
Other papers –
Combining Kernel and Model Based Learning for HIV Therapy Selection

16 Oct 2020

NUS National University of Singapore

Reinforcement Learning

Introduction



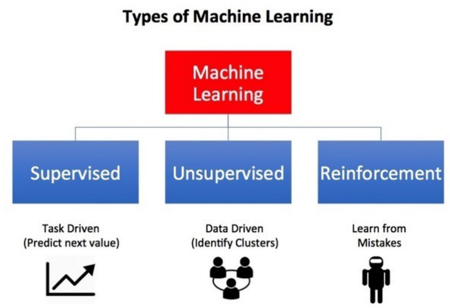
Agents perform action in an environment and observe reaction to maximize cumulative reward.

- **Agent**: What is the optimal policy?
- **Environment**: How states transition?
- **Actions**: What can the agent do?
- **Reward function**: What reward will the agent get?
- **Observation**: What can the agent see?

[1] Img credit - pierrelux

Reinforcement Learning

Machine Learning



- **Supervised**: Data with labels
- **Unsupervised**: Data without labels
- **Reinforcement**: Learning from examples – environment

Reinforcement Learning

Success Stories – How we got here

Dota2
(OpenAI 2019)

Alpha Go
(DeepMind 2016)

Control – Robotics, etc
Various papers on control

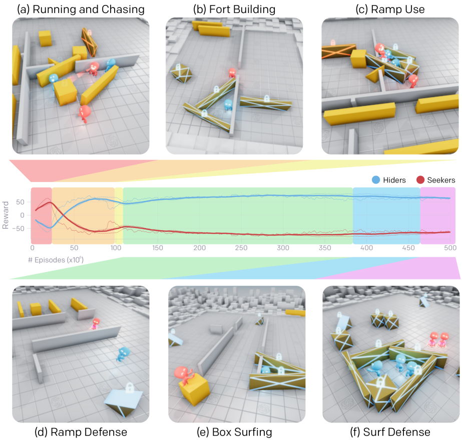
Atari
(DeepMind 2013)

TD-Gammon
(Gerry 1992)

Toy Examples – Cart-Pole
(Michie & Chambers 1968)

Reinforcement Learning

State of the Art - Emergent Tool Use From Multi-Agent Autocurricula (ICLR 2020)



[1] Youtube: Multi-Agent Hide and Seek

Motivation

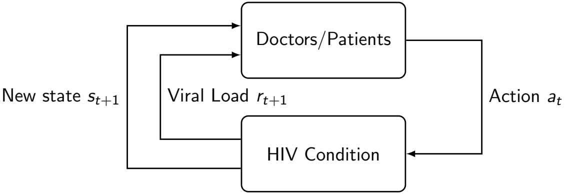
Problems in Healthcare x RL

Properties of Reinforcement Learning (RL) that is attractive for healthcare:

- **No need for labels** – Based on observations from the environment
- **Difficult to model mathematically** – Difficult or impossible to model accurately about a health condition
- **Policy based from trail and error** – Consult textbook and past patient experiences and go with what works best
- **Sequence of actions** – Medicine dosage, Strain, Treatment Type, Operations, Consultations, etc. . .
- **Long-term effectiveness** – Long-term effects of drugs/actions

Motivation

Can Reinforcement Learning help doctors make better decisions?



Want to maximize cumulative patient care/health.

- **Agent**: Doctors/Patients
- **Environment**: The health problem; ie. HIV
- **Actions**: ie. Which drug to prescribe now?
- **Reward function**: Maximize long-term patient care
- **Observation**: ie. Viral load, mortality

Reinforcement Learning

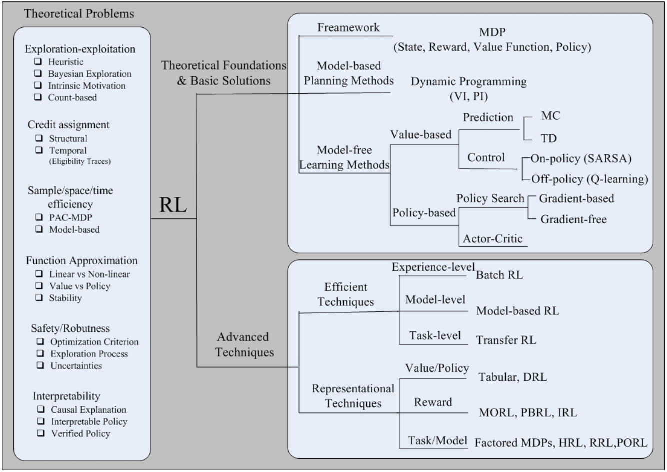
Markov Decision Process

Formally, RL can be formalized by Markov Decision Process (MDP) – $M = (S, A, P, R, \gamma)$:

- S – Finite state space, $s_t \in S$ at time t
- A – Actions avaliable to the agent, $a_t \in A$ at time t
- $P(s, a, s') : S \times A \times S \to [0, 1]$ – Markovian transition function when the agent transits from s to s' after action a
- $R : S \times A$ – Immediate reward after taking action a in state s.
- $0 \leq \gamma \leq 1$ – Discount factor, weighting future rewards vs current rewards.

Reinforcement Learning

Summary of methods



Reinforcement Learning

Q-Learning

Updating the 'goodness' of every state – Q-Table, at time t the agent selects a_t, entering state s_{t+1} and observes r_t [also, DNN]

$$Q^{new}(s_t,a_t) \leftarrow \underbrace{Q(s_t,a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Bigg(\underbrace{\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1},a)}_{\text{estimate of optimal future value}}}_{\text{new value (temporal difference target)}} - \underbrace{Q(s_t,a_t)}_{\text{old value}} \Bigg)}^{\text{temporal difference}}$$

Use Q-Table to decide what actions should be taken – Epsilon Greedy, with probability ϵ take random action, otherwise: [EvE]

$$a_t = \arg\max_{a \in A} Q(s_t, a)$$

Large Q-Table? Use NN to approx. Q-Table $\rightarrow$ DQN

Reinforcement Learning

Various considerations for Healthcare

Replaceable 'parts' in RL that can be changed, concerning healthcare:

- **Model-based vs Model-free** – Textbook approach to treatment or trial-and-error experience
- **Value-based vs Policy-based** – Health metric to pick policy or hard-policy making (ie. always resuscitate)
- **On-policy vs. Off-Policy** – Action derived from optimal policy or not (ie. trying new drugs or not)

Applications in Healthcare

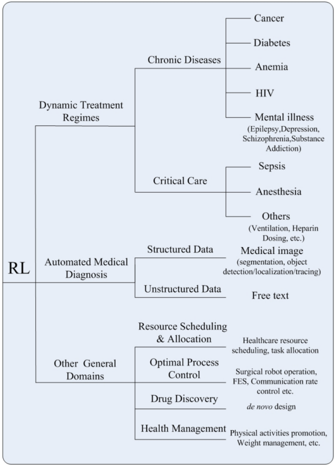
Rapid achievements in RL:

- Generalization
- Representation
- Efficiency

Allowed the use of RL in decision making (mainly discrete)

- Dynamic Treatment Regimes
- Automated Medical Diagnosis
- Others

Case Study: HIV Therapy



Applications in Healthcare

HIV Therapy Selection

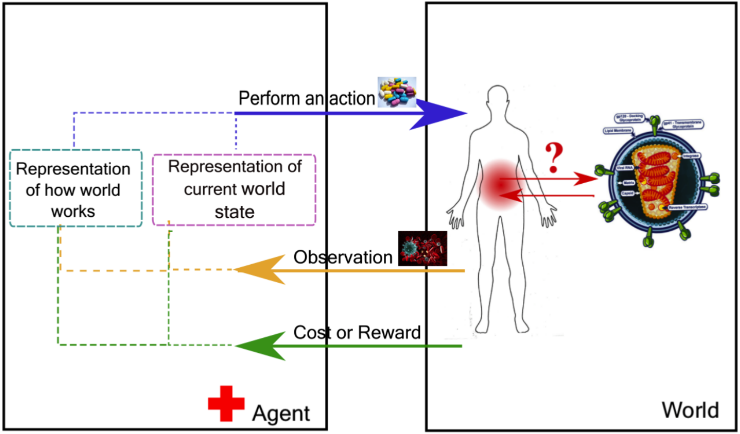
Sequential temporal decision process to choose antiviral drug.
Mixture of experts approach to get the best of both worlds – rely on established human model & RL for treatment exploration.

- Kernel-based history alignment (K-Nearest Neighbour)
- Partially Observable MDP, bayesian reinforcement learning – Learns a distribution over models, then use the model to determine the optimal decision.

In a sense, combing the models allow us to take advantage of each method in different situations.

Applications in Healthcare

HIV Therapy Selection



Healthcare Challenges in RL

Solving some of the theoretical/traditional problems would contribute to wider applicability of RL:

- Exploration-vs-Exploitation: How much exploration?
- Credit Assignment: How should the reward function be?
- Sample/Space/Time Efficiency: How fast can RL be?
- Function Approximation: What underlying modelling?
- Safety/Robustness: Can RL be safe?
- Interpretability: Why the action?

Healthcare Challenges in RL

Exploration Strategies

Exploration strategy is an important component in RL; trading off exploration and exploitation decisions for the agent.

- Most of RL applications in healthcare adopt simple heuristic-based exploration strategies (ie. ϵ-greedy)
- Simple strategies require large sample complexity (which in healthcare can be limited)
- Accounting for the true cost of the action; we can always reset machines but not for a life (safety exploration dilemma).

Healthcare Challenges in RL

Interpretable Strategy Learning

Most ML, including RL lack clear interpretability, unable to reveal the correlation between features and actions.

- No perceivable worse-case or safety guarantees.
- Medical domain requires rigorous validation for safety, correctness and robustness, which is challenging for RL.
- Understanding, the robustness of RL methods in uncertain healthcare settings, is still an open problem.

Conclusion

Future Work

Open problems in RL:

- Interpretable Strategy Learning - Understanding for safety, correctness and robustness.
- Integration of Prior Knowledge - Medical knowledge/models can be used as priors to improve accuracy
- Learning from Small Data - RL is fundamentally highly dependent on large number of training samples.
- Healthcare under Ambient Intelligence - integration into more certain healthcare settings with sensors.
- Future in-vivo Studies - learning from data, without a model.

Current models get around the problem by augmenting RL with another technique in the case of HIV, prior clinical knowledge.

Conclusion

My Thoughts

IBM applied reinforcement learning to the strategy of IBM Watson playing Jeopardy! IBM Watson defeated former Jeopardy! champions in 2011 and was subsequently adapted to healthcare.

Article:
IBM Watson Overpromised and Underdelivered on AI Health

[1] A history of reinforcement learning

Conclusion

My Thoughts

One criticism of the article that I think is not adequately addressed:

The goal of RL is to explore new treatment strategies.

But how can we be sure of its recommendation and evaluation?

- When we haven't tested the new treatment strategy before?
- Are measuring the performance based on what we know?
- Not know how does the new strategy works.

The need for a well defined, robust metric to measure performance.

Are we ready for RL to be used in healthcare?

500×10^6? In my opinion, No.