

MSI Webinar:

Ensembling Experiments to Optimize Interventions Along Customer Journey: A Reinforcement Learning Approach

May 9, 2023 | Virtual | 12:00 pm – 12:30 pm EDT

Speaker:

Yicheng Song – *Assistant Professor, Carlson School of Management, University of Minnesota*

Overview:

In this MSI Webinar, Yicheng Song (University of Minnesota) discussed the integration of multiple experiments using the reinforcement learning (RL) method to address scenarios that cannot be addressed by "standalone randomized experiments." In his presentation, Song gave an introduction to reinforcement learning in an example using video games and the board game Go. The computer program AlphaGo was specifically highlighted as an exceptional example of reinforcement learning, with the ability to observe a human expert opponent to learn over time and outperform that expert opponent based on the opponent's own technique. In addition to examples using gaming, Song noted applications of the reinforcement learning model that are currently used in business. A challenge for the reinforcement learning model is the exploration-exploitation dilemma: balance or trade-off in exploiting existing knowledge to obtain a reward or explore new actions to create better decisions.

In the second part of his presentation, Song examined a study he conducted in partnership with Tianshu Sun, Associate Professor of Data Sciences and Operations at the University of Southern California's Marshall School of Business. To do this they integrated historical data from earlier experiments with the reinforcement learning (RL) model to see if RL can improve an intervention policy along the customer journey. Song and Sun proposed a revised Bayesian Recurrent Q Network (BRQN) model which leveraged both recent and historical data. This allowed them to study two complementary research questions: (1) how to ensemble historical experiments to optimize a sequence of interventions along the customer journey, and (2) how to utilize the RL model to guide the design of future experiments to balance exploitation and exploration.

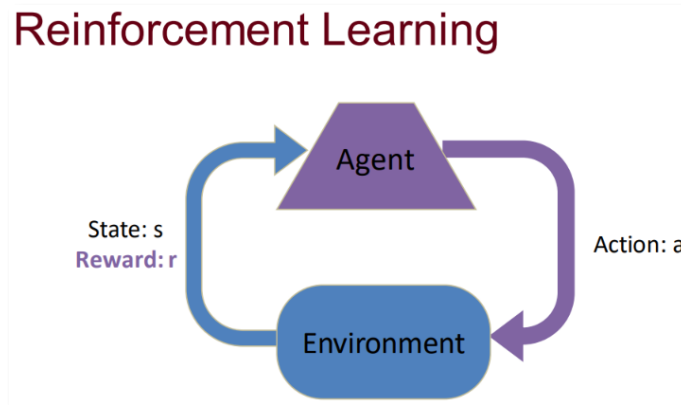
Takeaways

Introduction to Reinforcement Learning



- **Two main components of reinforcement learning are the agent and the environment**, with the agent interacting with the environment by observing the state of the environment directly. Based on this observation the agent will make a

decision regarding action. Once the environment receives the decision on the action it will send a reward to the agent.

- **The key idea of reinforcement learning** is that **the agent utility receives feedback in the form of rewards and** is defined by the reward function. Therefore, the agent must act to maximize the cumulative rewards it gains from the environment.



- **The basic components in RL are the agent, environment and reward function.**
 - **In terms of applications of RL in video games** the agent is the controller or input device (human control), the environment is the video game itself and the reward is the increase in score when "killing enemies."
 - In response, Reinforcement Learning is going to train the model on "optimal control" which can outperform many human beings on these video game tasks.





	Agent	Environment	Reward Function
Video Game			Get 20 scores when killing an enemy

- **In an example, the product of the reinforcement learning model is the computer program AlphaGo.** In this setting this environment includes the human champion of the board game Go. The reward function is the rule of Go.
 - In this case, **reinforced learning based on actions from the human champion of the board game Go**, AlphaGo can defeat the champion opponent, despite being the champion of the game Go.

	Agent	Environment	Reward Function
Go			The rule of GO

- In the above settings it is noted that there is **no control over the environment or the reward function**. However, **there is control over the "reinforced learning agent"** (actor) which is being trained to learn optimal control.

Basic Components




	Actor	Environment	Reward Function
Video Game			Get 20 scores when killing an enemy
Go			The rule of GO

We cannot control


- **The goal of the reinforcement learning model** is to observe and learn this control policy to **"maximize the cumulative reward."**

Example: Playing Video Game

Start with observation s_1 Observation s_2 Observation s_3

After many turns



This is an **episode**.

Total reward:

$$R = \sum_{t=1}^T r_t$$

Action a_T

Obtain reward r_T

- **A challenge** for the reinforcement learning model **is the exploration-exploitation dilemma.**
 - There is a challenge in the balance or trade-off in exploiting existing knowledge to obtain a reward or explore new actions to create better decisions.

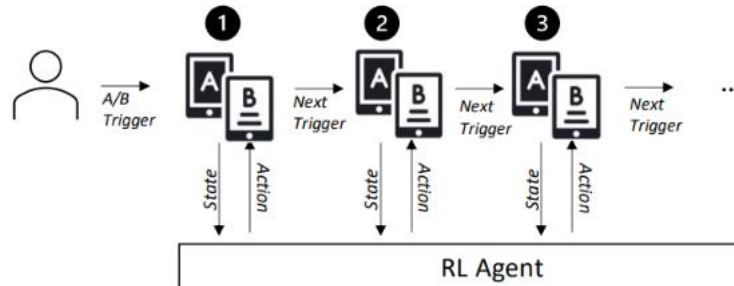
Existing Business Applications of the RL Model

	Environment	Action	Objective
Programmatic Advertisement	Customers	Display ads	Click rate
Recommender System	User	Item recommendation	User engagement
Portfolio Optimization	Stock market	Investment strategy	Asset value
Inventory Control	Warehouse	Replenishment strategy	Balance supply and demand

Reinforcement Learning + Randomized Experiments: Optimal Intervention

- **Randomized experiments (A/B testing) is the "holy grail of causal inference to evaluate various interventions along the consumer journey"** providing unbiased information on causal effects from new website design, price promotion, to system recommendation.
 - One caveat is that stand-alone experiments designed to identify the impact of one specific intervention, fail to consider interdependency.
- The accumulative nature of historical experiments creates a "gold mine" for reinforcement learning models. Historical experiments hold the following attributes:
 - **Diverse:** Maximal 2^N different intervention combinations.
 - **Exogenous:** All interventions are exogenously designed or developed externally.
 - The combination provides feedback from all stages of the consumer journey.
- In an experiment, the researchers **integrated historical data with the reinforcement learning (RL) model** to see if **RL can improve an intervention policy, along the customer journey**, considering key components such as the environment, state, actions and reward.

RL+AB



- **Environment:** Heterogeneous customers
- **State:** Summary of customer state
- **Actions:** Interventions from historical randomized experiments
- **Reward:** Monetary reward from the customer

- **The experiment considered the following research questions:**



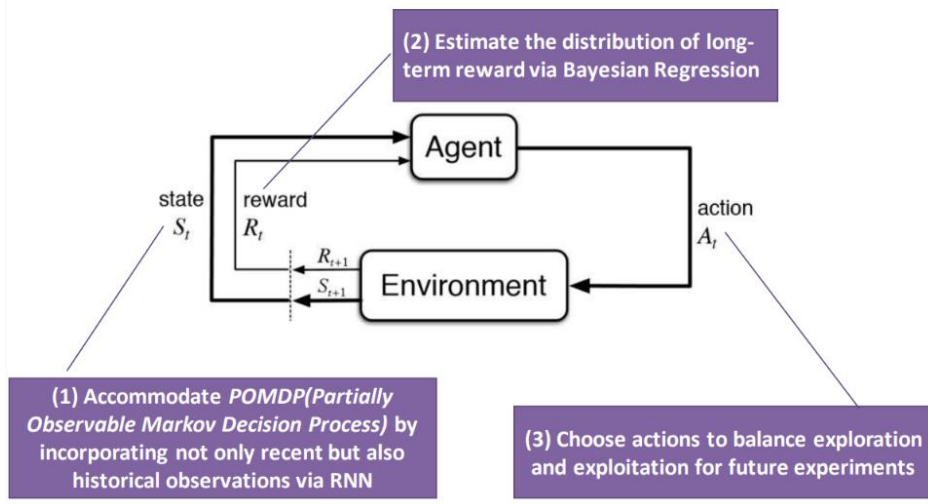
A/B→RL: How ensembling large number of historical experiments to design an algorithm to optimize sequence of interventions along customer journey?



RL→A/B: How to utilize the RL to guide the design of future experiments to balance exploitation and exploration?

- **The widely used classical RL model** (Q-learning) **contains limitations** such as the assumption of the Markov decision process, a single number estimation of reward and no guidance on future experiments.
- **The revised RL model (Bayesian Recurrent Q Network (BRQN) model) incorporated both recent and historical observations**, to accommodate a "partially observable Markov decision process." Using Bayesian regression, they estimate the distribution of long-term reward
 - Additionally, this model allows for the choosing of "actions to balance exploration and exploitation for future experiments." (See the revised model below).

Model Outline



- The proposed **Bayesian Recurrent Q Network (BRQN)** model provides a "**two-way complementarity between RL and A/B** (randomized experiments) facilitating RL to learn "optimal interventions."
 - This creates a more **holistic approach to learning and optimizing interventions** along the customer journey.