Open Time Optimization

Using Reinforcement Learning

Shubham Kumar

Shubham Bhardwaj

Bhoopendra Sharma

Department of Statistics

Banaras Hindu University

Symposium for Innovative & Applied Statistics (SIAS)

28 June 2024

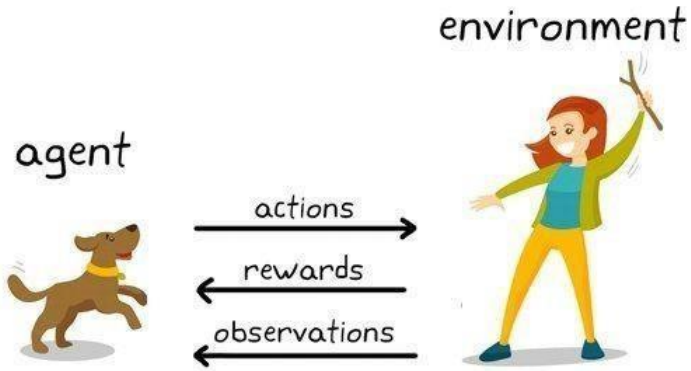
Abstract :

In today's digital landscape, optimizing the timing of customer interactions is crucial for maximizing engagement and revenue. This study uses Reinforcement Learning (RL) to find the best times for various customer interactions, such as email openings and app notifications. RL, a subset of machine learning, involves an agent learning optimal behaviors by interacting with its environment to maximize rewards. By applying RL algorithms like Thompson Sampling, we dynamically adjust our strategies based on customer behavior and interactions. This approach balances exploring new strategies and exploiting known successful times, providing a robust and adaptive method for optimizing open times across multiple channels.

Reinforcement Learning

Reinforcement Learning (RL) is the science of decision making. It is about learning the optimal behavior in an environment to obtain maximum reward. This optimal behavior is learned through interactions with the environment and observations of how it responds, similar to children exploring the world around them and learning the actions that help them achieve a goal.

In the absence of a supervisor, the learner must independently discover the sequence of actions that maximize the reward. This discovery process is akin to a **trial-and-error** search. The quality of actions is measured by not just the immediate reward they return, but also the **delayed reward** they might fetch.

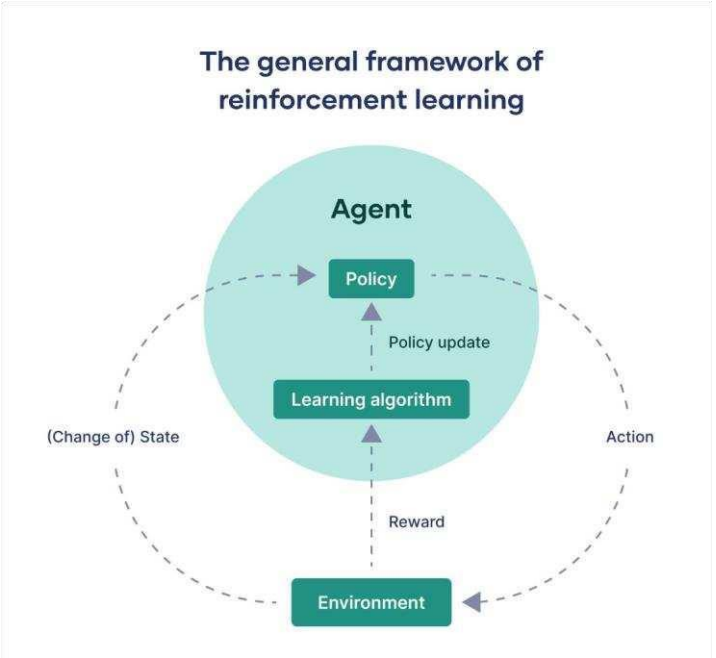


Elements of RL

The main elements of an RL system are:

1. The agent or the learner
2. The environment the agent interacts with
3. The policy that the agent follows to take actions
4. The reward signal that the agent observes upon taking actions

Real world problems: Robotics, Games and Autonomous driving



Challenge

A challenge that is unique to RL algorithms is the trade-off between **exploration and exploitation**. An RL agent must strike a careful balance between exploiting its past experience and exploring the unknown states of the environment. The right balance would lead the agent to discover the optimal policy that yields maximum reward. If the agent continues to exploit the past experience only, it is likely to get stuck in a local minima and produce a sub-optimal policy. On the other hand, if the agent continues to explore without exploiting, it might never find a good policy.

Solution

Greedy Decision: Estimate the model from historical data and maximize the immediate rewards. Could get stuck in a local minima as we are not exploring.

Thompson Sampling: Estimate the model from historical data and keep updating the parameters as you get the rewards and repeat the process. Experimentally we will be close to optimal solution. This is also our use case.

Problem Statement

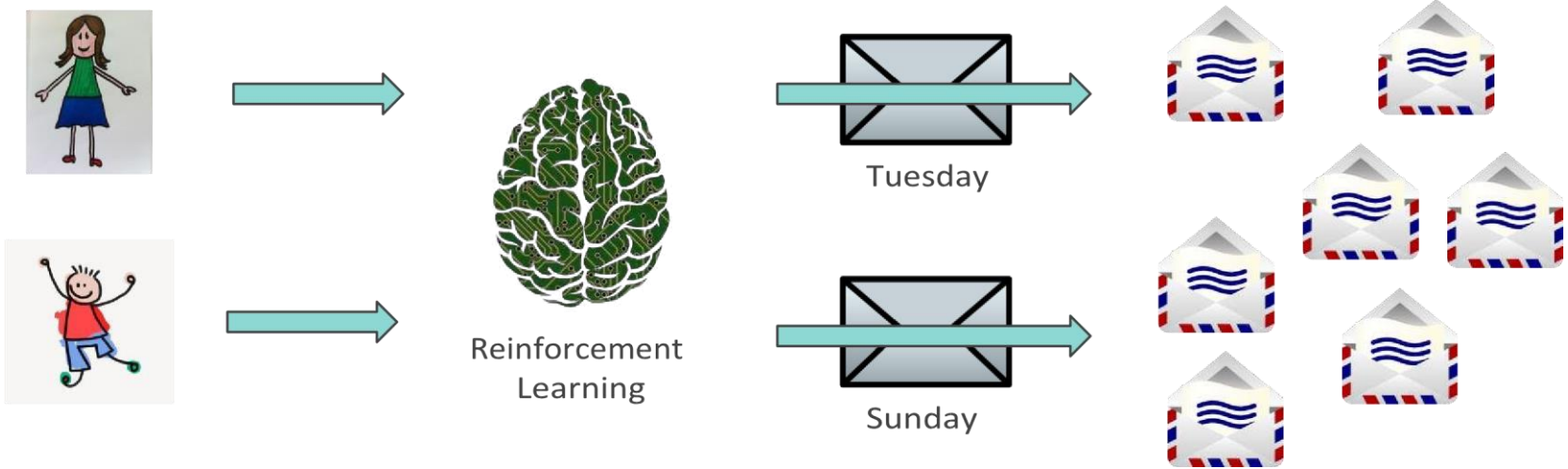
Context: You are a merchant and would like to show your best offers to your customers at the right time through emails, in order to maximize your revenue.

Problem: If you don’t send them email at the right time, your email will be stacked lower in the inbox and will probably be missed.

Objective :

In Open time optimization we are using Reinforcement Learning to predict the best day for a customer to open the email from a merchant.

Goal is to increase the open rate:



Data Preparation:

	OPEN	IGNORED	TOTAL
Sunday	2	1	3
Monday	1	1	2
Tuesday	2	1	3
Wednesday	3	4	7
Thursday	1	1	2
Friday	2	5	7
Saturday	3	1	4
		Total	28

Sun_1	Sun_2	Mon_1	Mon_2	Tue_1	Tue_2	Wed_1	Wed_2	Thu_1	Thu_2	Fri_1	Fri_2	Sat_1	Sat_2
2	1	1	1	2	1	3	4	1	1	2	5	3	1

RL in Action :

Sun_1	Sun_2	Mon_1	Mon_2	Tue_1	Tue_2	Wed_1	Wed_2	Thu_1	Thu_2	Fri_1	Fri_2	Sat_1	Sat_2
2	1	1	1	2	1	3	4	1	1	2	5	3	1

Parameters:

Beta(2, 1)	Beta(1, 1)	Beta(2, 1)	Beta(3, 4)	Beta(1, 1)	Beta(2, 5)	Beta(3, 1)

Probability:

0.3683	0.4528	0.1730	0.6526	0.6646	0.2385	0.1748

Optimal Days:

Thursday	Wednesday	Monday	Sunday	Friday	Saturday	Tuesday

Updating Parameters :

Thursday

	Open	Ignored
Sunday	2	1
Monday	1	1
Tuesday	2	1
Wednesday	3	4
Thursday	1 + 1	1
Friday	2	5
Saturday	3	1

Thursday

	Open	Ignored
Sunday	2	1
Monday	1	1
Tuesday	2	1
Wednesday	3	4
Thursday	1	1 + 1
Friday	2	5
Saturday	3	1

Conclusions :

Our research shows that Reinforcement Learning (RL) using Thompson Sampling effectively optimizes the timing of customer interactions across various digital touchpoints, including emails, app notifications, SMS, and social media. This adaptive method improves personalization, scalability, and engagement metrics, leading to higher revenue and customer satisfaction.

Future work will integrate advanced RL algorithms and additional metrics, such as click-through and conversion rates, to enhance interaction relevance. Optimizing open times for different customer segments based on demographic or behavioral data will further refine targeting strategies.