RL Real-Time Bidder

Venkata Chintapalli
AI Insight Fellow

Ad Technology Market

User visits website

Website alerts ad marketplace

Companies bid on ad slots

Highest bid wins

Company gets paid IF user acts

Ad Technology Market

User visits website

Website alerts ad marketplace

Companies bid on ad slots

Highest bid wins

Company gets paid IF user acts

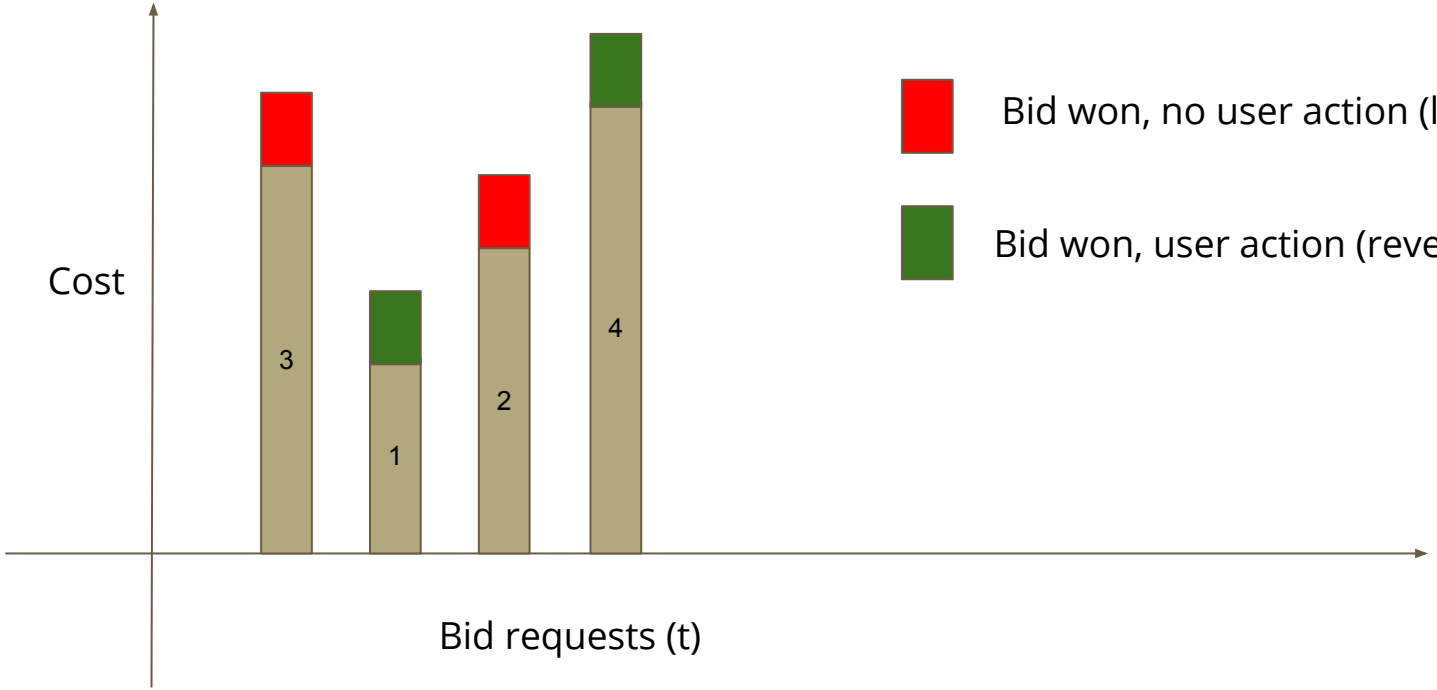
Ad Campaigns are run with limited budget

Improve Ad spending strategy

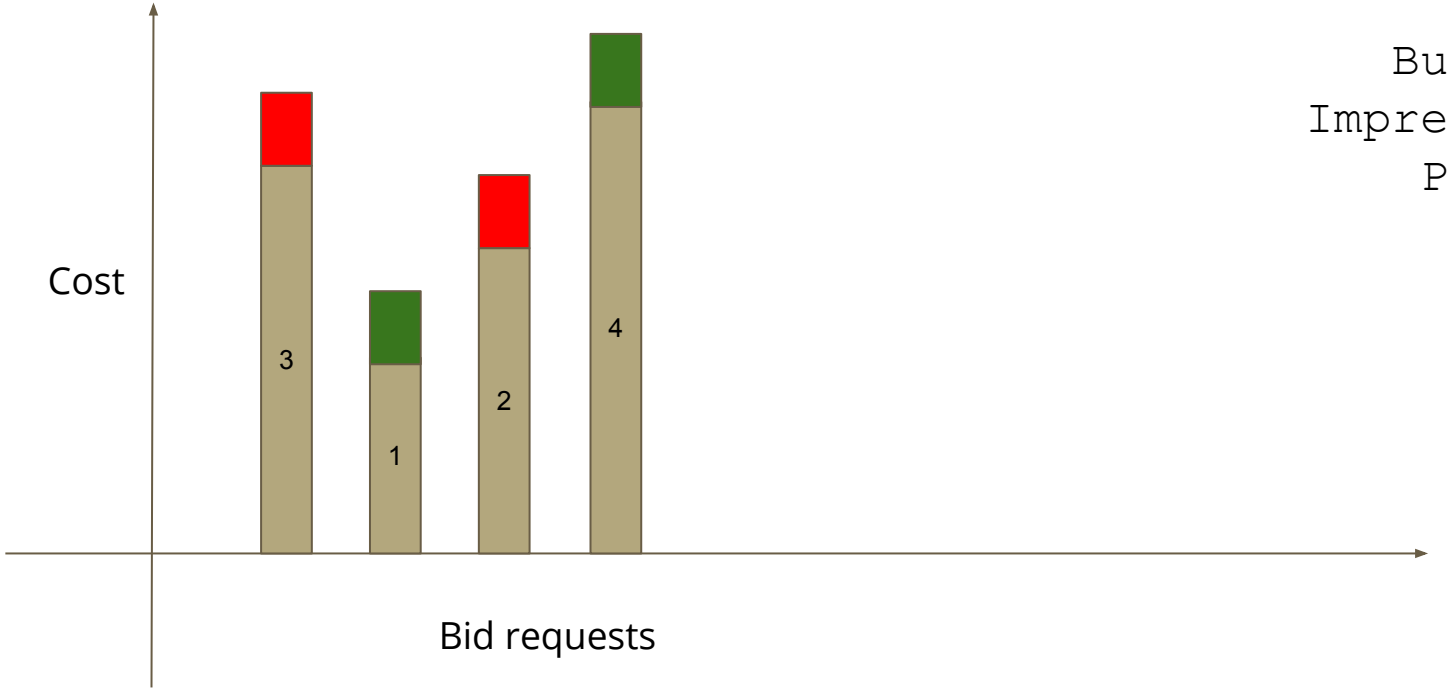
RL BIDDER

Maximize the total value of Ad impressions under a constrained budget.

Bidding Terminology

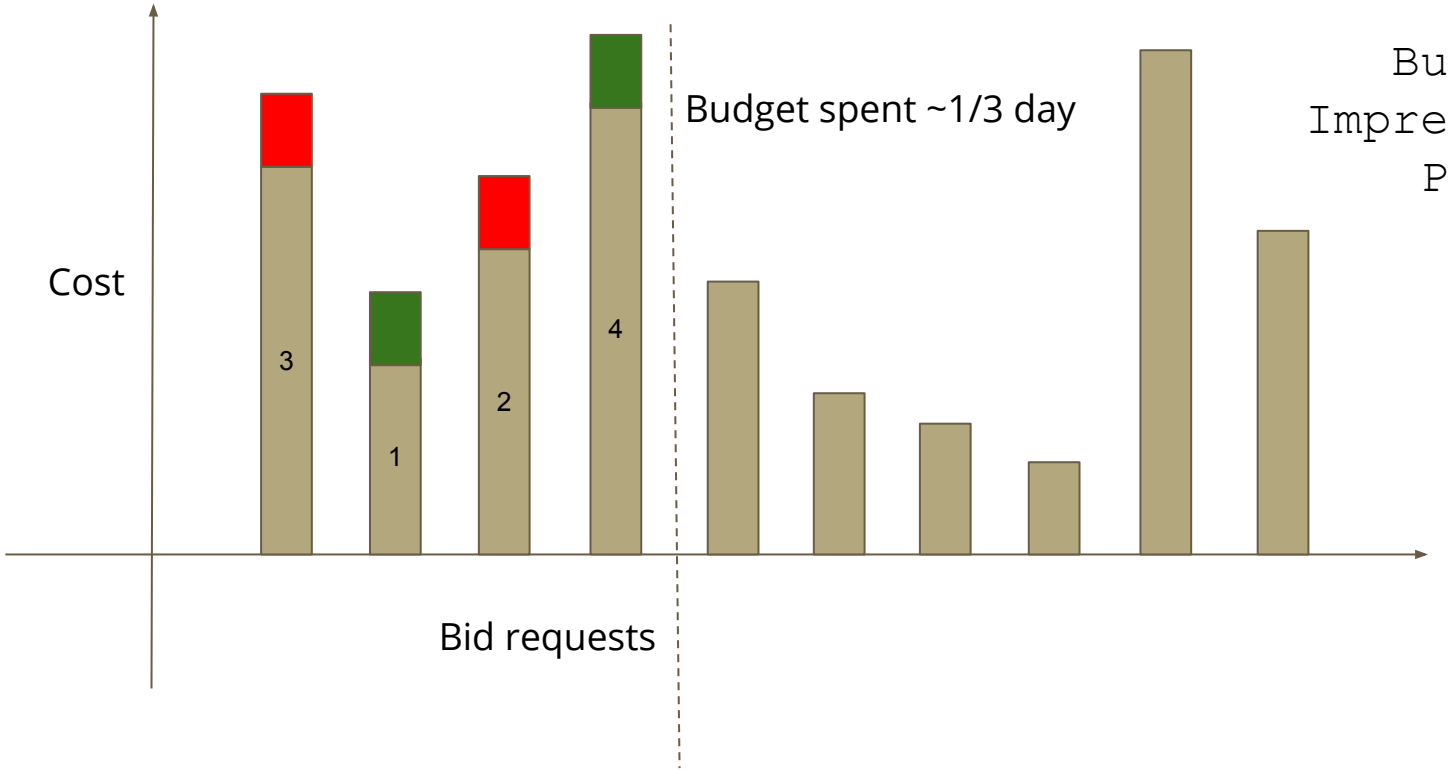


Naive Bidding Strategy

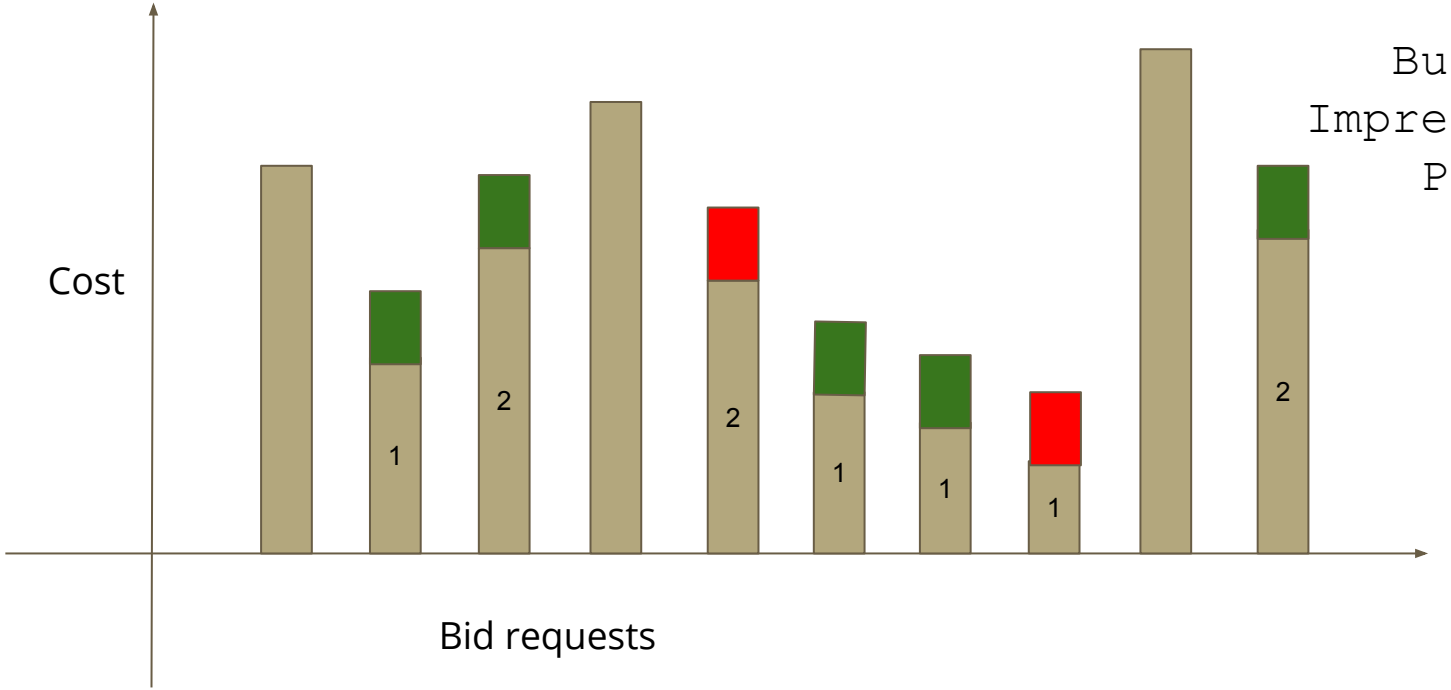


Budget $10
Impressions 4
Profit $2

Naive Bidding Strategy

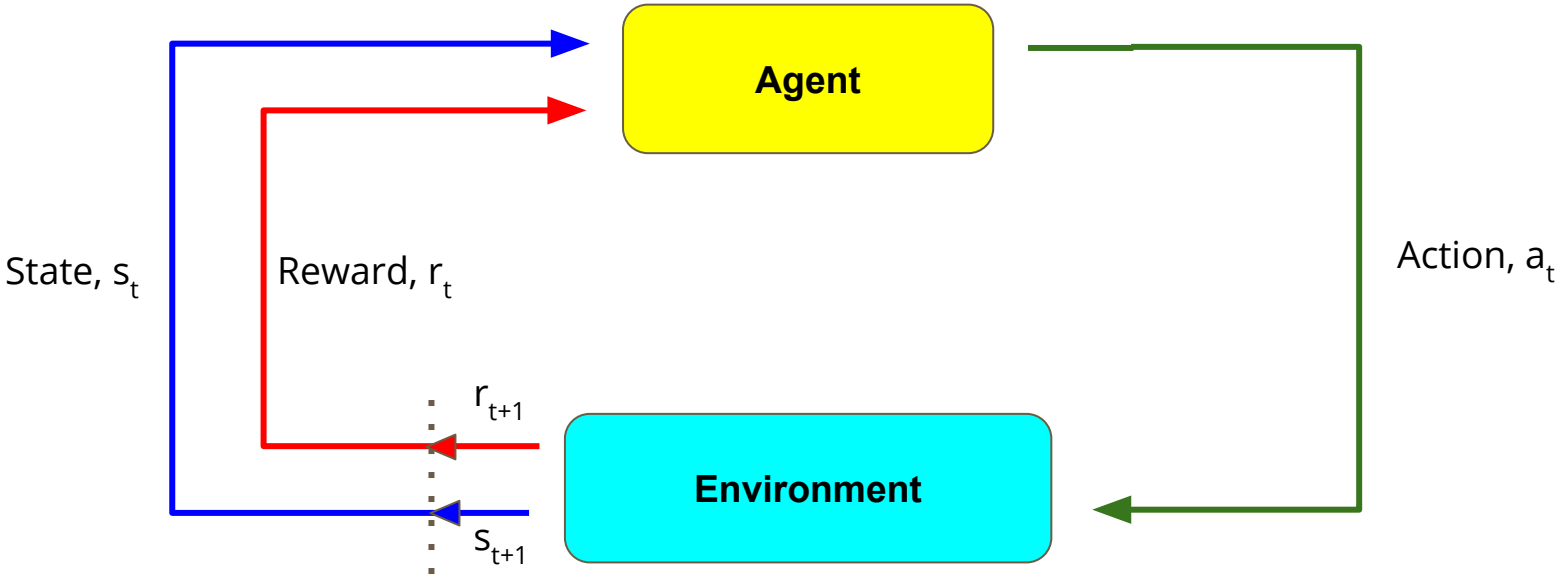


Optimal Bidding Strategy



Budget $10
Impressions 8
Profit $5

Reinforcement Learning Cycle

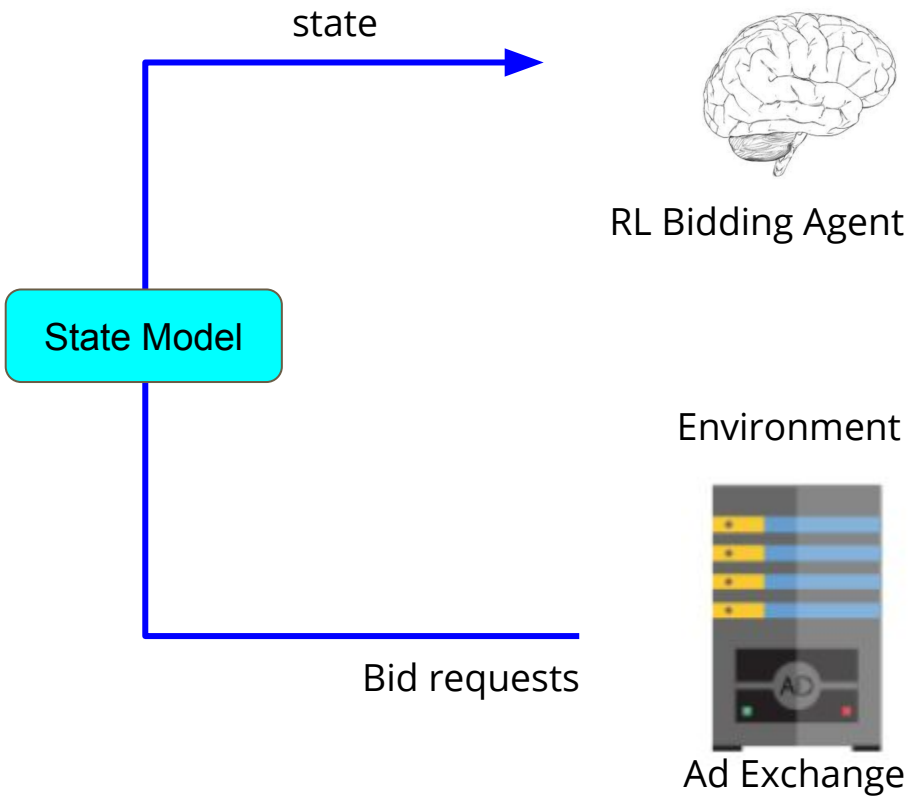


Reinforcement Learning Cycle

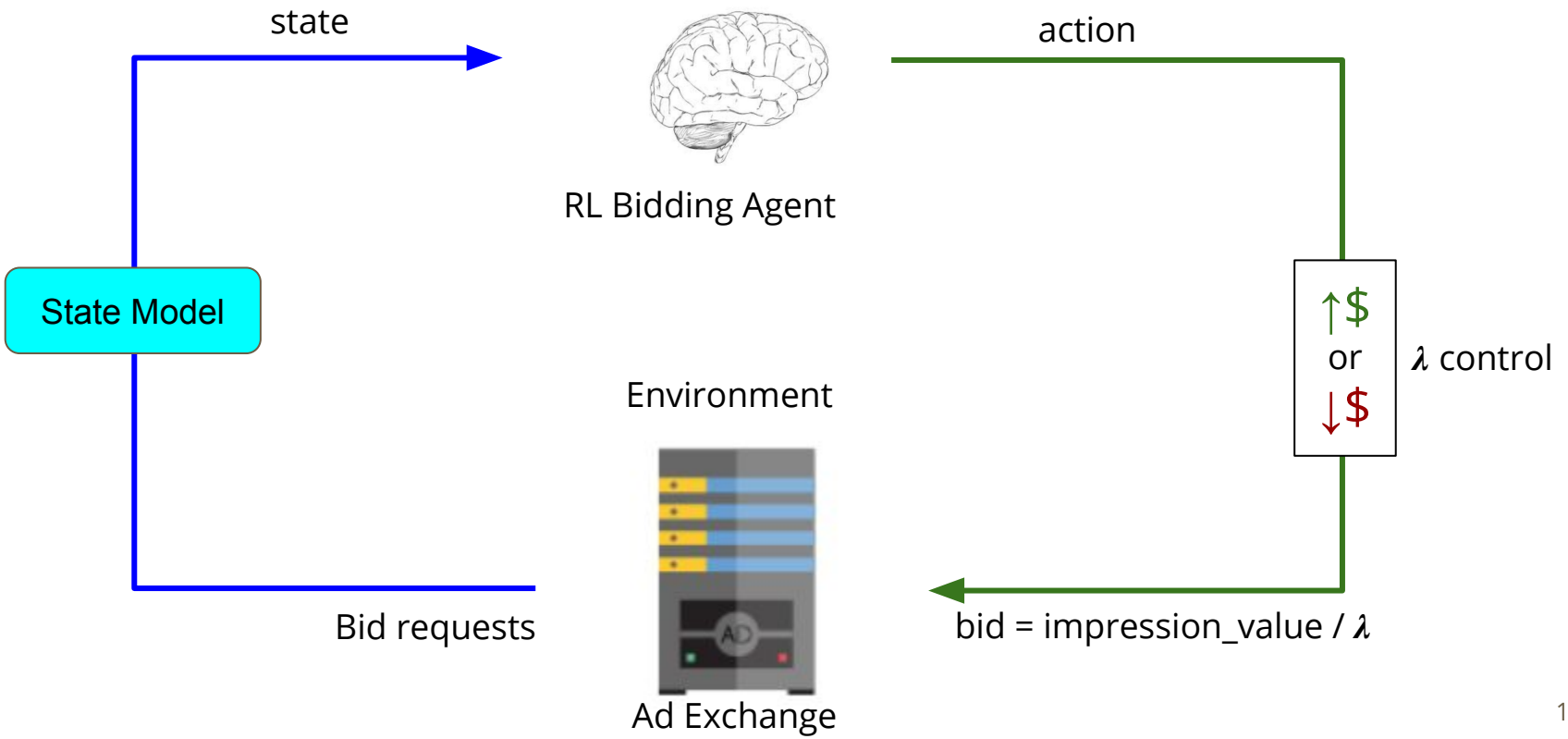
Environment

Ad Exchange

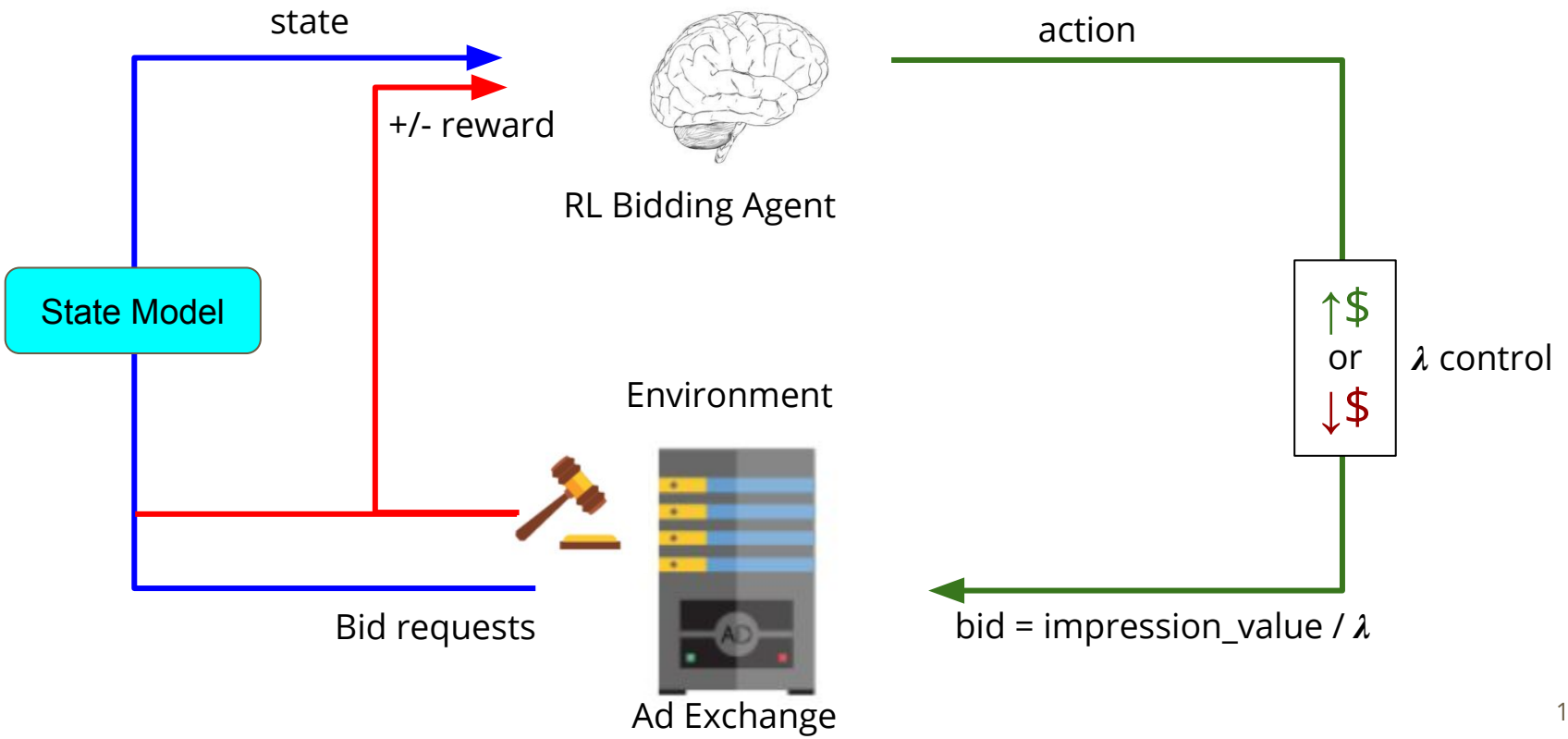
Reinforcement Learning Cycle



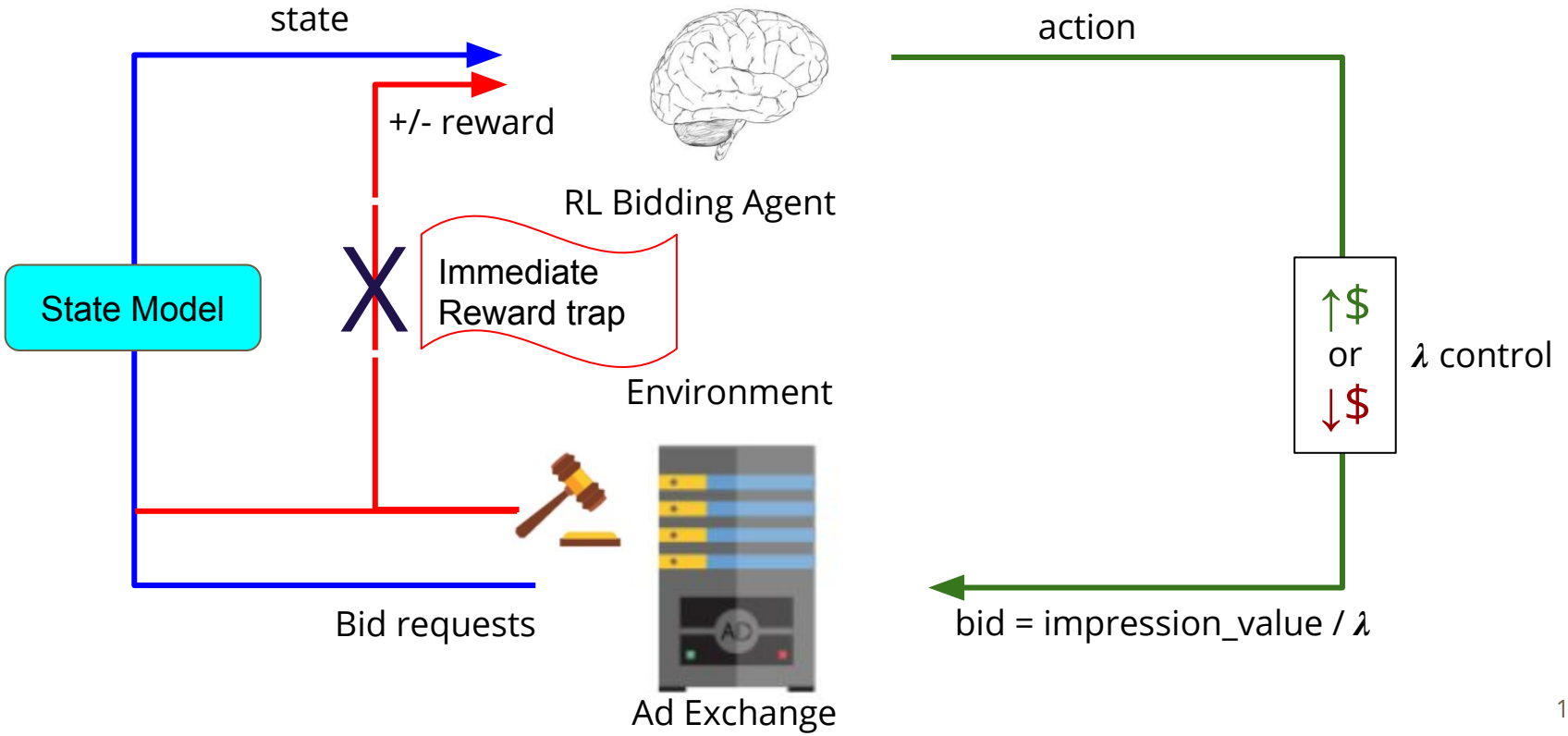
Reinforcement Learning Cycle



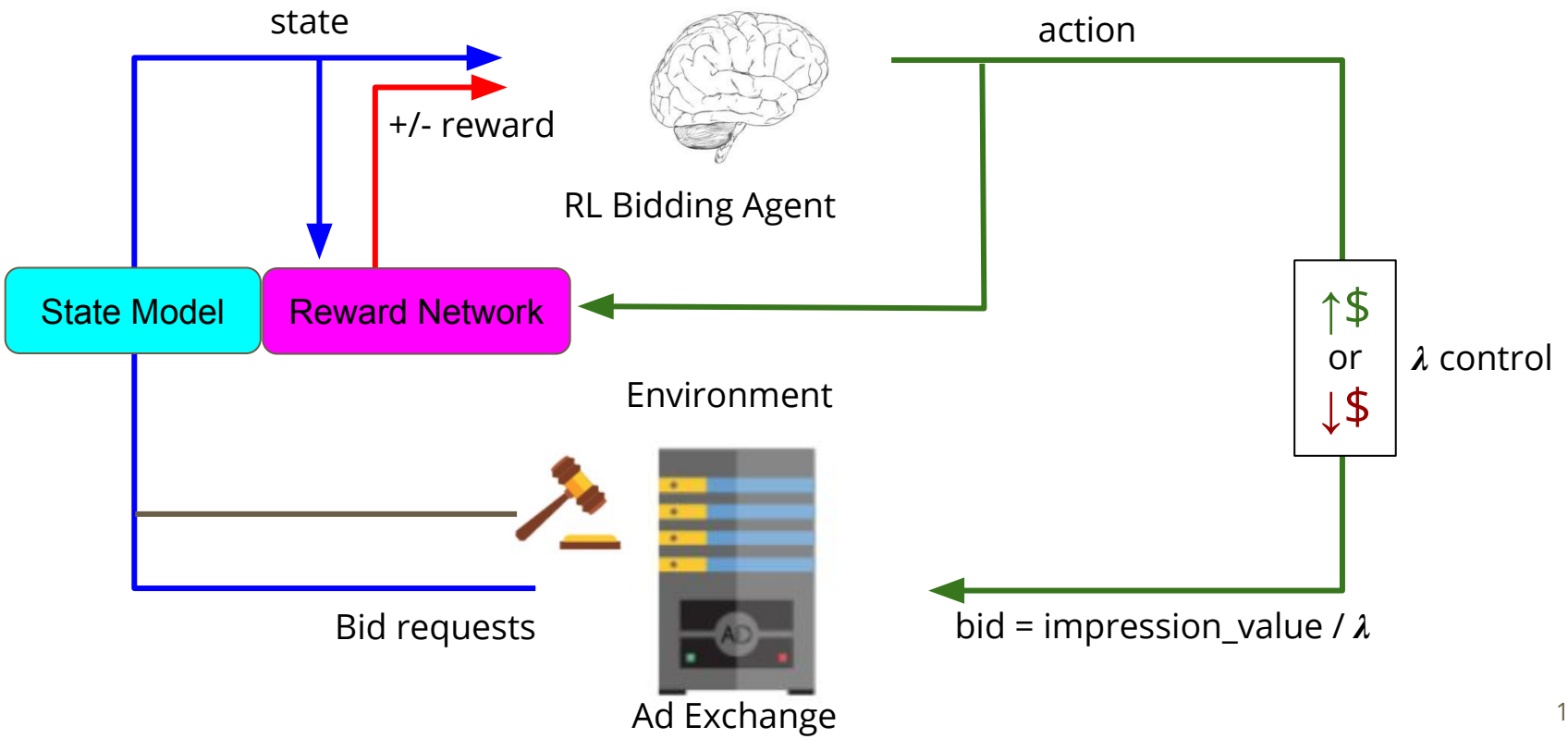
Reinforcement Learning Cycle



Reinforcement Learning Cycle

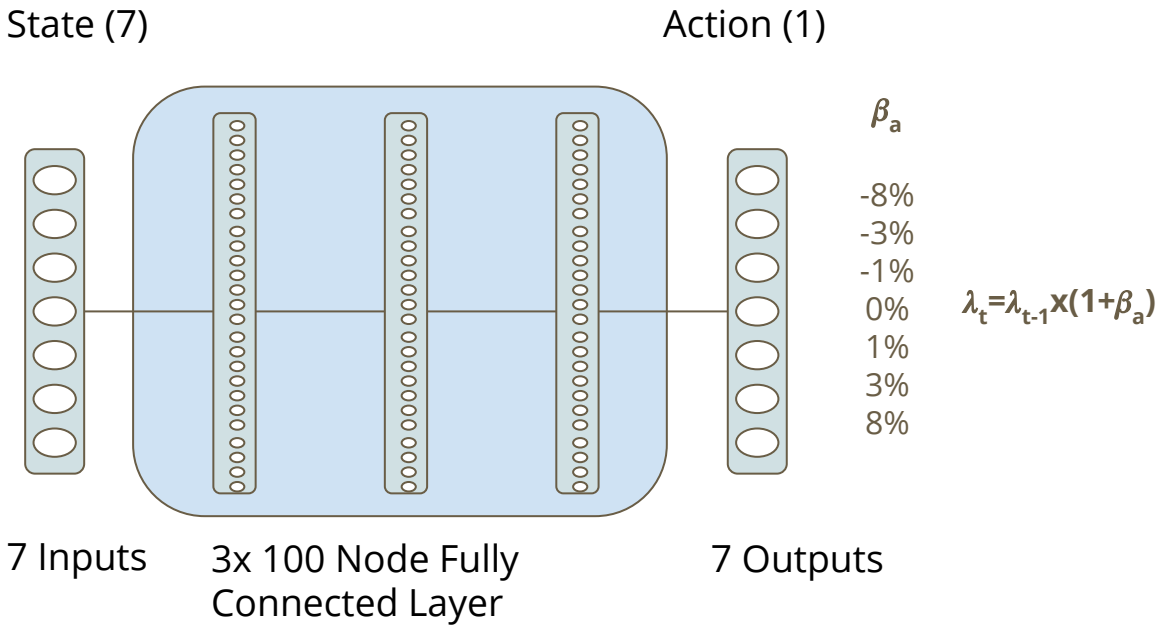


Reinforcement Learning Cycle



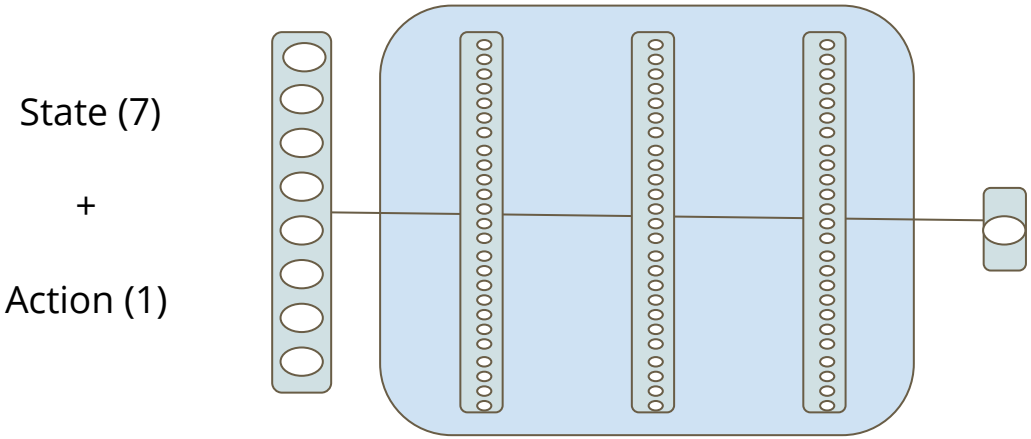
RL Agent DQN

1. **t** - The current time-step
2. **B_t** -Remaining Budget at t
3. **ROL_t**-Regulation Opportunities Left at t
4. **BCR_t**=(B_t-B_{t-1})/B_{t-1} Budget Consumption Rate
5. **CPM_t**=Cost per mille (winning impressions t-1 to t)
6. **WR_t**=winning impressions/total impression opportunities
7. **r_{t-1}**=Total value of winning impressions (clicks or conversions)



Wu, D., Chen, X., Yang, X., Wang, H., Tan, Q., Zhang, X., ... & Gai, K. (2018, October). Budget constrained bidding by model-free reinforcement learning in display advertising.

RL Agent RewardNet



Wu, D., Chen, X., Yang, X., Wang, H., Tan, Q., Zhang, X., ... & Gai, K. (2018, October). Budget constrained bidding by model-free reinforcement learning in display advertising.

Data

iPinYou public tabular dataset

bidid	timestamp	IP	slotprice	bidprice	payprice	click	...
81aced04baad90f9358aa39a4521cd6f	20130606000104828	115.45.195.*	0	300	51	0	...
...	...	...	...	...	...	...	...

~200 million bid requests spanning across 7 days

Results

Strategy	% Impressions	Profit(Estimate)
Linear Bidder	1.44%	239
RL Bidder RewardNet	1.28%	244

Venkata Chintapalli

Work Experience

NUTANIX

ORACLE

Education

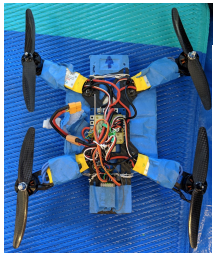
M.S. in Machine Learning

GEORGIA TECH

M.Tech. Electronics Design & Tech

IDT

Volunteer



DESTINATION IMAGINATION

INDIAN INSTITUTE OF SCIENCE

Backup slides

How to optimize ad spending budget

Improve ad targeting

With the same budget

Can the same ad impressions be targeted to the right user?

DQN Algorithm

```
Initialize replay memory D1 to capacity N1
Initialize Q with random weights θ
Initialize Q target with weights θ− = θ
for episode = 1 to K do
    Initialize λ0
    Bid with λ0 according to Eq. (2)
    for t = 1 to T do
        Update RewardNet (8-10 in Algo. 2)
        Observe state st
        Get action at from adaptive □-greedy policy
        Adjust λt−1 to λt
        Bid with λt according to Eq. (2)
        Get rt from RewardNet
        Observe next state st+1
        Store (st, st+1, at, rt ) in D1
        Sample mini-batch of (sj, sj+1, aj, rj ) from D1
        if sj+1 is the terminal state then
            Set yj = rj
        else
            Set yj = rj + γ max a′ Q(sj+1, a′; θ−)
        end
    Perform a gradient descent step on (yj −Q(sj, aj ; θ))2 with respect to θ
    Every C steps reset Q target = Q
    end
    Store data for RewardNet
end
```

RewardNet Algorithm

```
Initialize replay memory D2 to capacity N2
Initialize reward network R with random weights η
Initialize reward dictionary M to capacity N3
for episode = 1 to K do
    Initialize temporary set S
    Set V = 0
    for t = 1 to T do
        if len(D2) > BatchSize then
            Sample mini-batch of (sj, aj, M(sj, aj)) from D2
            Perform a gradient descent step on (R(sj, aj; η) − M(sj, aj))2 with respect to the network parameters η
        Observe state st
        RL agent executes at in the Environment
        Obtain immediate reward rt from the Environment
        Set V = V + rt
        Store pair (st , at ) in S
    end
    for (sj, aj ) in S do
        Set M(sj, aj) = max(M(sj, aj), V )
        Store pair (sj, aj, M(sj, aj)) in D2
        if |M | > N3 then
            Discard old key in M based on LRFU strategy [18]
    end
end
```

Bidding Strategies

Linear Bidding

- Bid Price = MLL * Target Price

RL Bidding

- Bid Price = MLL * Target Price * λt

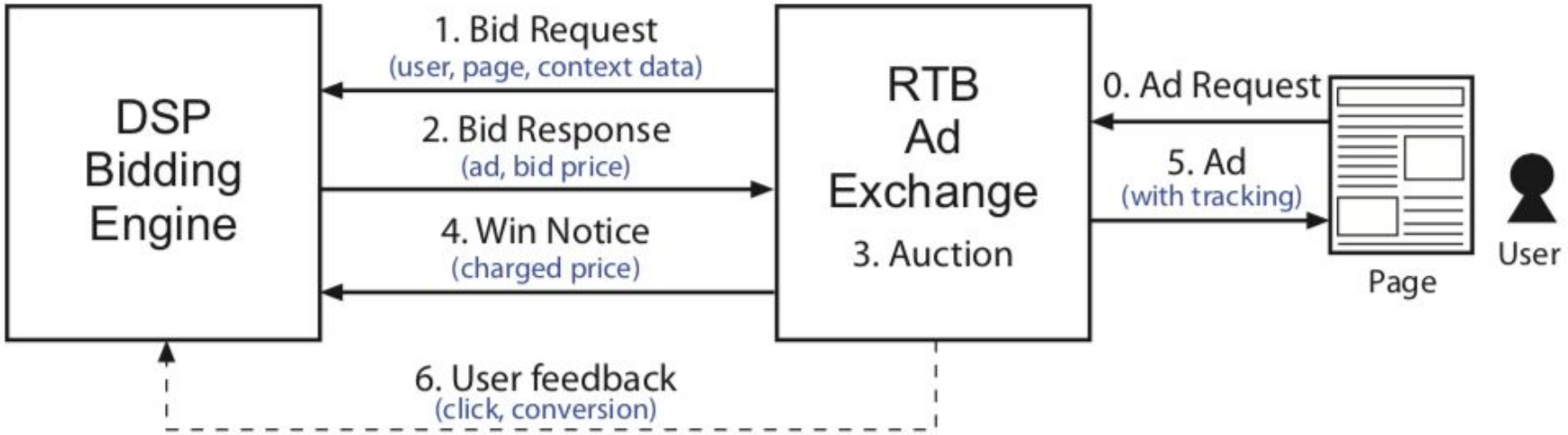
Ad Tech: Budget Constrained Bidding

- **Problem**: Project goal is Optimizing advertisers(also, Display Side Platforms) Budget to provide highest total value(Clicks) of their Ad campaigns
- **Solution**: Replace heuristics and classic control based Real-time bidding(RTB) agent with a Model-free Reinforcement Learning(RL) agent to learn the optimal bidding strategy
-

Acronyms

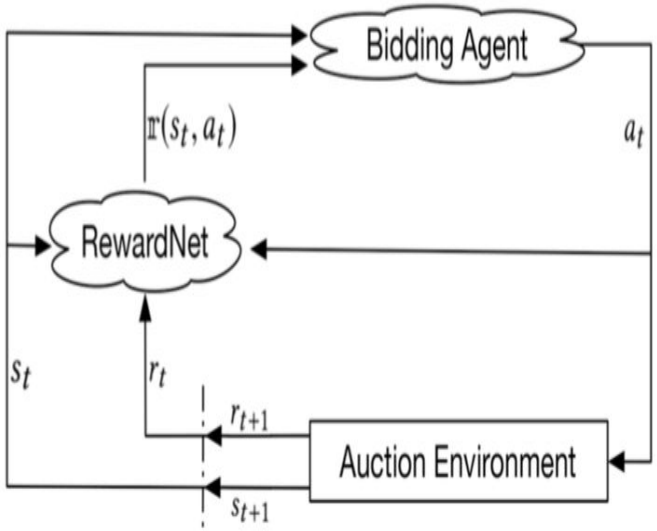
RTB	Real-Time Bidding	MDP	Markov Decision Process
DSP	Display Side Platform	CMDP	Constrained MDP
SSP	Supply Side Platform	RL	Reinforcement Learning
DMP	Data Management Platform	DQN	Deep Q-Network
GD	Guaranteed Delivery	FLB	Fixed Linear Bidding
		BSLB	Budget Smoothed Linear Bidding
		RLB	RL to Bid
		DRLB	Deep RL Bid
		IDFA	Identifier for Advertisers

Ad Tech: Components



Model-free RL agent

- State
  - t - The current time-step
  - B_t -Remaining Budget at t
  - ROL_t-Regulation Opportunities Left at t
  - $BCR_t=(B_t-B_{t-1})/B_{t-1}$ Budget Consumption Rate
  - CPM_t=Cost per mille (winning impressions t-1 to t)
  - WR_t=winning impressions/total impression opportunities
  - r_{t-1}=Total value of winning impressions(clicks or conversions)
- Action
  - $\lambda_t=\lambda_{t-1}x(1+\beta_a)$ where β_a=-8%,-3%,-1%,0%,1%,3%,8%
- Reward
  - $r(s,a) = \underset{e \in E(s,a)}{max} \sum_{t=1}^{T} r_t^{(e)}$
  - E(s,a)=set of existing episodes that the agent took action a at state s and $r^{(e)}_t$ is the original immediate award at t within e



Train & Test Phases

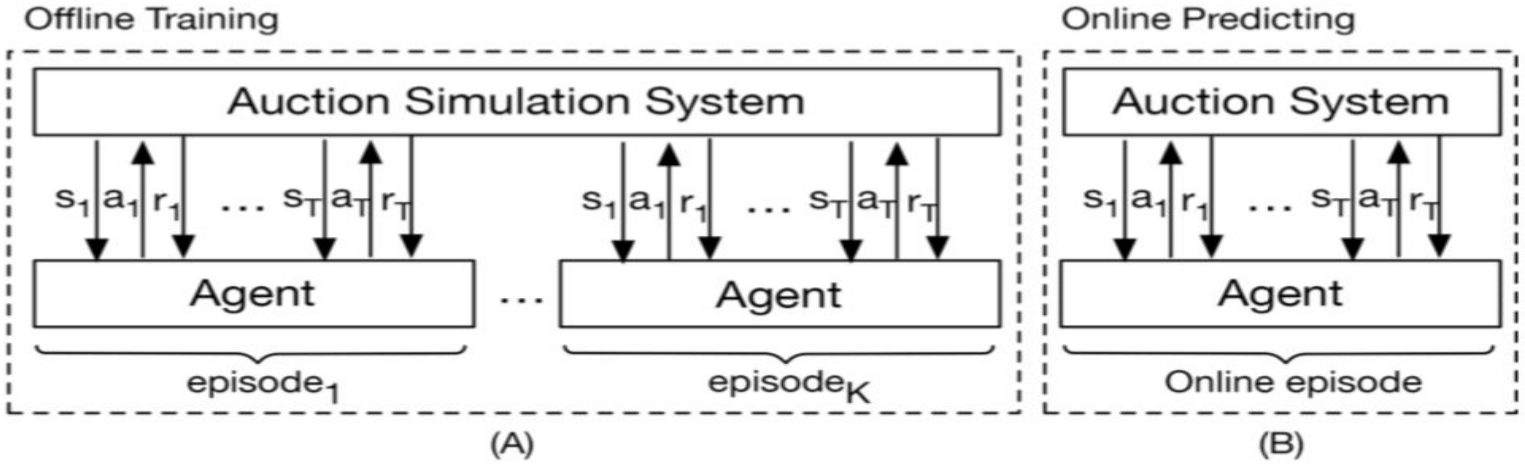


Figure 1: Illustration of λ control process in budget constrained bidding. (A) Agent training process. (B) Agent online predicting process.

Display Advertising (Ad Tech)

Display Advertising digital revenues for FY 2018 surpassed $100 billion dollars for the first time. So for my Insight project, I am using the iPinYou dataset to develop a real-time bidding agent that interacts with the Ad exchanges to maximize the winning impressions with a limited budget per day. To achieve this, I am building an auction simulator and a model-free Reinforcement learning agent that can also be used to solve problems in control, robotics, financial sectors.

Problem

- Problem: Formalized Budget Constrained Bidding as a Knapsack problem.

$$\max \sum_{i=1\ldots N} x_i v_i$$

$$s.t. \sum_{i=1}^{N} x_i c_i \leq B$$

  - xi = 1, if advertiser wins impression i, else 0
  - vi = impression value
  - Ci = bidding an impression associates with a cost

- Solution: The optimal bidding strategy under the second price auction

$$b_i = v_i / \lambda$$

  - bi = advertisers bid according to the impression value vi
  - λ = scaling factor

Second Price Auction

Source: http://www.science4all.org/le-nguyen-hoang/auction-design/

Challenges

- Immediate reward trap for the Budget constrained problem
  - Reward Network handle the budget constraint
- Dealing with very large datasets - bid tables with 200 million entries
  - Used batch processing

References

- Wu, D., Chen, X., Yang, X., Wang, H., Tan, Q., Zhang, X., ... & Gai, K. (2018, October). Budget constrained bidding by model-free reinforcement learning in display advertising. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 1443-1451). ACM.
- Ad Tech Simplified : What is Real Time Bidding, (RTB)?
- IAB internet advertising revenue report, 2018 full year results
- 20 Must Know Digital Advertising Acronyms
- Reinforcement Learning Applications