Online Resource Allocation with Non-Stationary Customers

Xiaoyue Zhang, Hanzhang Qin, Mabel C. Chou

National University of Singapore

Overview

We proposed the Unified Learning-while-Earning Algorithm based on Contextual Bandit with Knapsacks to address resource allocation under nonstationary environments.

- Leveraged contextual information for informed decisions
- Balanced exploration and exploitation for rapidly changing customer preferences
- Overcame limitations of stationary environments
- Achieved regret bound of $\tilde{O}(\sqrt{n|\Theta|T})$ under near-i.i.d. arrivals
- Provided sublinear regret and constant competitive ratio under nonstationary arrivals
- Compared ULwE Algorithm with ALG_{IID} and ALG_{ADV} in experiments, ULwE Algorithm consistently outperformed other algorithms

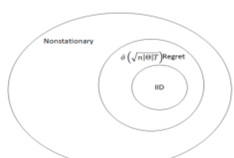
Problem Setting

Objective: Maximize the total reward obtained from customer clicks or purchases while considering resource budget constraints.

- n resources with initial budgets: c_i
- Customer arrival over time horizon: $[0, T]$
- When a customer arrives, we will observe the context information x^t
- Reject or assign a resource j to the customer
- Probability function $f_j(x^t, \theta)$: the likelihood of customer purchasing resource j
- Purchase: the system gets a reward r_j and consumes 1 unit from capacity

Assumption:

- Distribution: $P(x^t = x^{(l)}) = \mu_l^t$ for customer type $l \in 1, 2, \ldots, L$
- For every customer type $x^{(l)}$, calculate the expected number of arrivals over the time horizon: $\lambda_l = \sum_{t=1}^{T} \mu_l^t$
- All μ_l^t values are unknown, but λ_l is provided based on historical data.



ICML International Conference On Machine Learning

Algorithm 1 Unified Learning-while-Earning (ULwE)

Input: c, T, L
Initialize $\Omega_i^0 = \Theta_i$ for all $i \in [n]$
for $t = 1$ **to** T **do**
  if switch = FALSE **then**
    $I^t = \text{ALG}_{\text{LP}}(c, T, L, \Omega_i^{t-1})$
    Check if conditions for switching are met.
    if any condition is violated **then**
      switch = TRUE.
    end if
  else
    $I^t = \text{ALG}_{\text{ADV}}(c, T, L, \Omega_i^{t-1})$
  end if
  Check if conditions for updating Ω_i are met.
  if any condition is violated **then**
    Remove $\bar{\theta}^t$ from $\Omega_{I^t}^t$
  else
    Set $\Omega_i^t = \Omega_i^{t-1}$ for all $i \in [n]$
  end if
end for

Algorithm 2 ALG_{LP} Protocol

Input: c, T, L
Initialize $\Omega_i^0 = \Theta_i$ for all $i \in [n]$
for $t = 1$ **to** T **do**
  Solve LP to maximize revenue and obtain $\bar{s}^t, \bar{\gamma}^t$:

$$\max_{s_{ij}, i \in [n], j \in [L]} \sum_{i \in [n]} r_i \sum_{j \in [L]} \lambda_j s_{ij} \bar{f}_i(x^{(j)}, \Omega_i^{t-1})$$

$$\text{s.t.} \sum_{j \in [L]} \lambda_j s_{ij} \bar{f}_i(x^{(j)}, \Omega_i^{t-1}) \le c_i, \ \forall i \in [n]$$

$$\sum_{i \in [n]} s_{ij} = 1, \ \forall j \in [L]$$

$$s_{ij} \ge 0, \ \forall i \in [n], j \in [L].$$

  Select resource i with probability $\bar{s}_{iJt}$
end for

Algorithm 3 ALG_{ADV} Protocol

Input: c, T, L
Initialize $\Omega_i^0 = \Theta_i$ for all $i \in [n]$
for $t = 1$ **to** T **do**
  Observe the context x^t of the new arrival in period t
  Select

$$I^t = \arg\max_{i \in [n]} r_i^t \bar{f}_i(x^t, \Omega_i^{t-1})$$

  where: $r_i^t = r_i \times (1 - \Psi(N_i^{t-1}/c_i))$
end for

Main Contribution

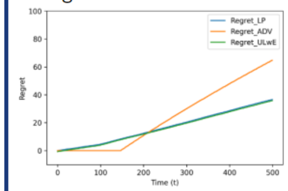
Regret bound:

- Near-i.i.d. arrivals scenario: $\tilde{O}(\sqrt{n|\Theta|T})$
- Nonstationary arrivals scenario: sublinear regret with a constant competitive ratio

$$\text{OPT} \le \left(1 + \frac{(1 + \min_{i \in [n]} c_i)\left(1 - e^{-1/\min_{i \in [n]} c_i}\right)}{1 - 1/e}\right) \mathbb{E}[\text{ALG}] + \tilde{O}(\sqrt{n|\Theta|T})$$

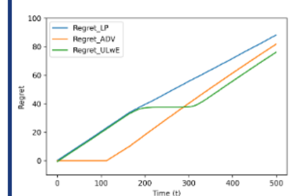
Numerical Studies

Regret over Time under Near-IID Arrivals

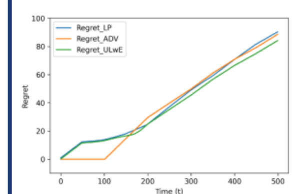


Period	Customer A Resource 1	Customer A Resource 2	Customer B Resource 1	Customer B Resource 2	Regret
LP Algorithm & ULwE Algorithm					
100	0	66	12	22	4.6
200	0	126	38	36	12.4
300	0	174	71	55	20.3
400	0	235	100	65	28.5
500	0	291	132	77	36.6
ADV Algorithm					
100	0	66	0	34	0
200	19	107	10	64	10.7
300	46	128	35	91	30.0
400	76	159	53	112	47.7
500	99	192	76	133	64.7

Regret over Time under Adversarial Arrivals



Period	Customer A Resource 1	Customer A Resource 2	Customer B Resource 1	Customer B Resource 2	Regret
LP Algorithm					
100	0	13	77	10	20.5
200	0	35	141	24	39.5
300	0	75	193	32	55.8
400	0	120	241	39	71.6
500	0	153	250	77	88.1
ADV Algorithm					
100	0	13	0	87	0
200	14	21	21	144	18.0
300	44	31	66	159	40.1
400	79	44	106	174	61.3
500	100	53	150	197	81.8
ULwE Algorithm					
100	0	13	77	10	20.5
200	0	35	121	44	37.1
300	0	75	121	104	37.8
400	27	93	155	125	55.7
500	51	102	199	148	76.1



- Adjust estimates of λ_l every $h = 50$ periods.
- Initial data: 500 customers to estimate initial λ_l.
- Update formula every h intervals:

$$\hat{\lambda}_l = \frac{T}{t} \cdot \sum_{\tau=1}^{t} I_{\{\text{customer at } \tau \text{ is type } l\}}$$

- **Benchmark:**

$$J^D(c, t) = \max_{s_{ij}, i \in [n], j \in [L]} \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{\tau=1}^{t} r_i s_{ij}^{\tau} f_i(x^j, \theta^*)$$

$$\text{s.t.} \sum_{j=1}^{L} \sum_{\tau=1}^{t} \mu_j{}^{\tau} s_{ij}^{\tau} f_i(x^j, \theta^*) \le c_i, \forall i \in [n]$$

$$\sum_{i=1}^{n} \sum_{\tau=1}^{t} s_{ij}^{\tau} = \sum_{\tau=1}^{t} \mu_j{}^{\tau}, \forall j \in [L]$$

$$s_{ij}^{\tau} \ge 0, \forall i \in [n], \forall j \in [L]$$