Dynamic Content Recommendation Considering Cache State and Preference Using Reinforcement Learning

Shota Ochiai, Noriaki Kamiyama (Ritsumeikan Univ., Japan)

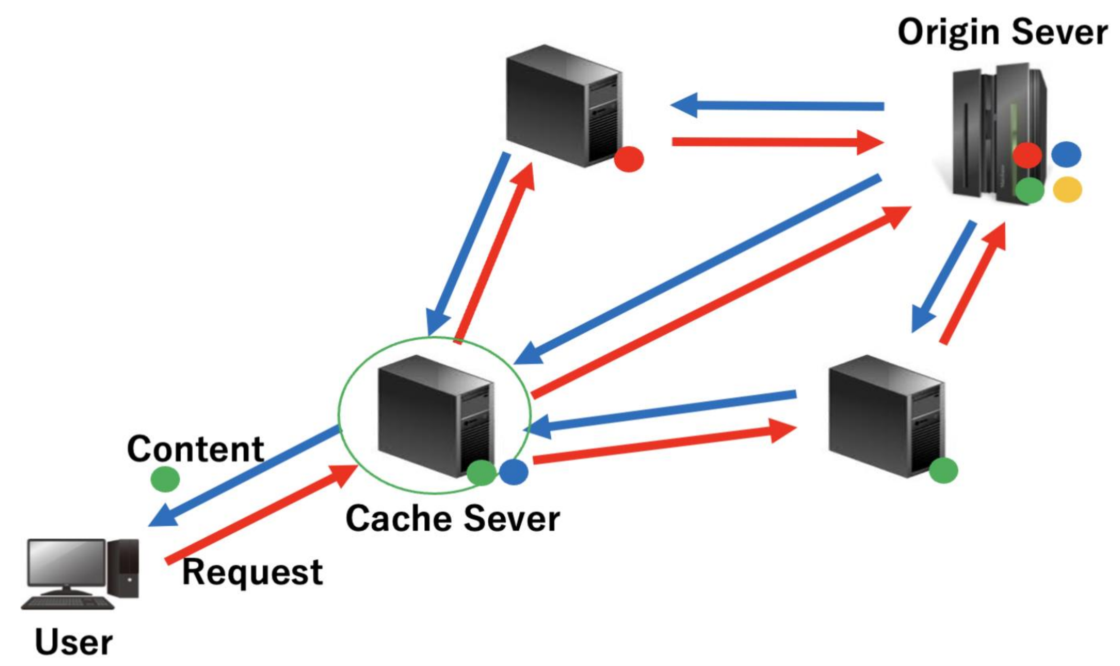
1. Introduction

■ The Need for Recommendation System (RS)

- Widespread use of services with massive content
- Difficulty for users to find content matching their preferences

■ Mitigating Latency

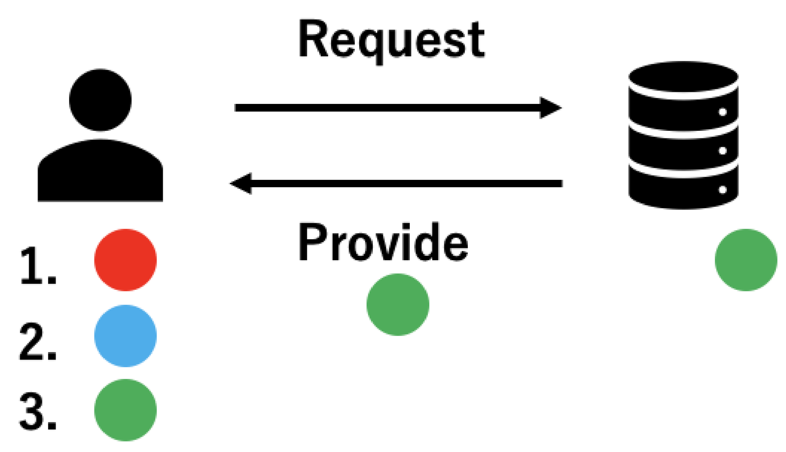
- Delivery latency caused by recommendation increases user churn
- CDN is generally used to mitigate this latency



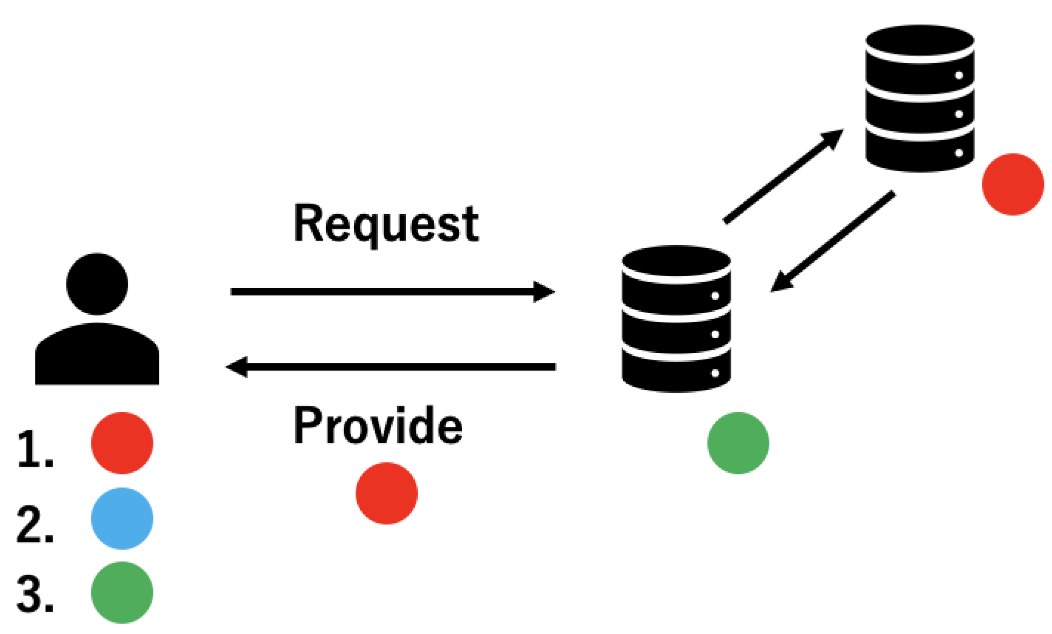
2. Challenges in Recommendation System Using Cache

■ The Trade-off

- Inherent trade-off between improving cache hit rate and recommendation accuracy
- Prioritizing cache state



- Prioritizing preferences



■ Environmental Limitation

- Most existing studies addressing both metrics assume a static environment

3. Purpose of This Work

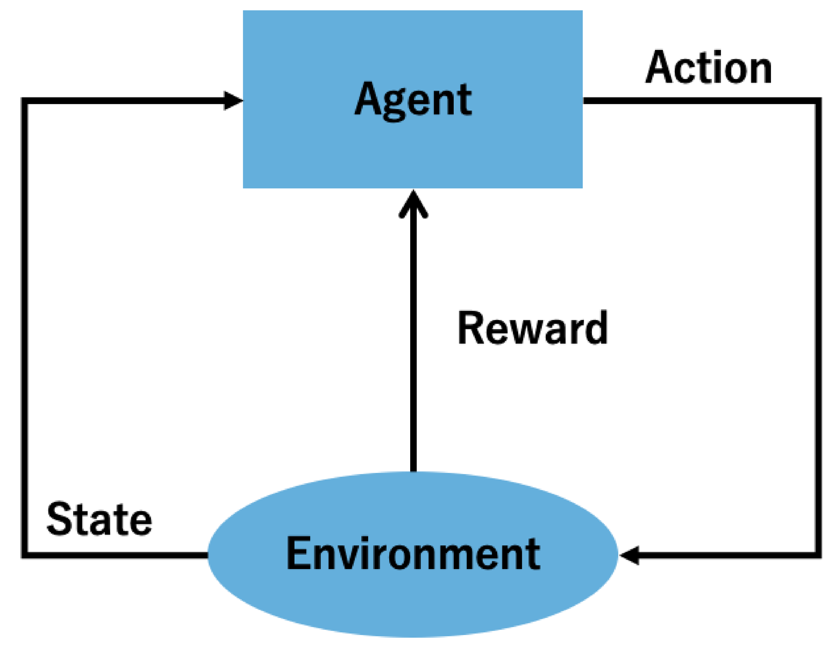
■ Dual Goal in Dynamic Settings

- Target a dynamic environment where user count and content popularity fluctuate
- Aim to simultaneously achieve a high cache hit rate and accurate preference-matching recommendation using Reinforcement Learning

3. Proposed Method

■ Applying Reinforcement Learning (RL)

- Formulation of the recommendation and caching problem for RL



■ Optimization via Reward Function

- Use a reward function designed to improve both cache hit rate and preference-matching performance

■ Flexible Prioritization

- Adjustable weight in the reward function allow flexible tuning of priorities based on specific environmental requirements

4. Evaluation

■ Performance Results

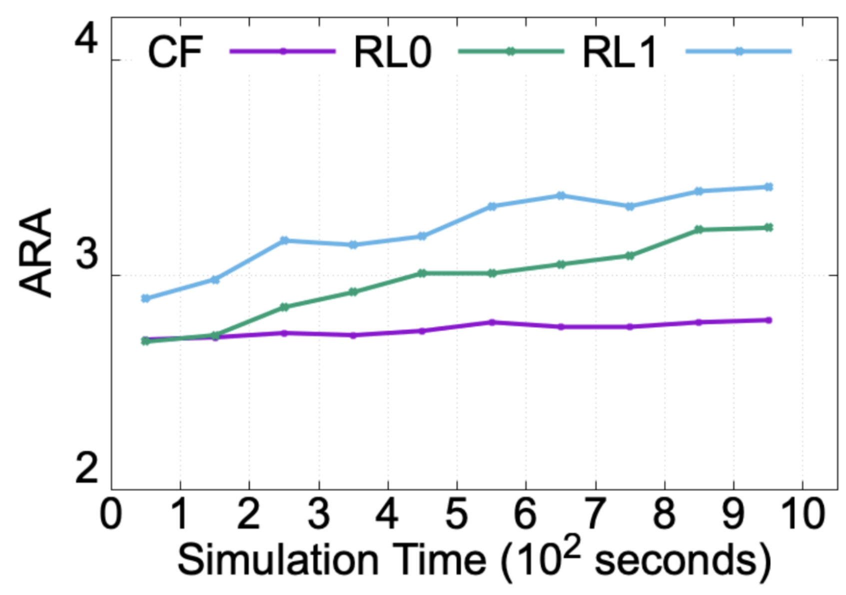
- Achieved superior performance in both cache hit rate and recommendation accuracy compared to conventional Collaborative Filtering (CF)
- CF

Target user

content

	1	2	3	4	5	6	7	8	9	10	Correlation coefficient
X	—	4	2	—	—	—	—	1	1	5	1.000
A	5	3	4	—	—	—	—	1	2	1	0.063
B	—	—	—	1	1	2	4	5	4	0	−0,982
C	2	3	2	1	—	4	3	2	1	4	0.941
D	2	—	—	2	4	5	1	2	5	5	0.500
E	—	5	0	—	5	1	—	1	1	2	0.615
Degree of recommend	2.00	4.00	1.00	1.50	4.50	3.33	2.00	1.66	2.33	3.66	

- ARA (Average Recommendation Accuracy)



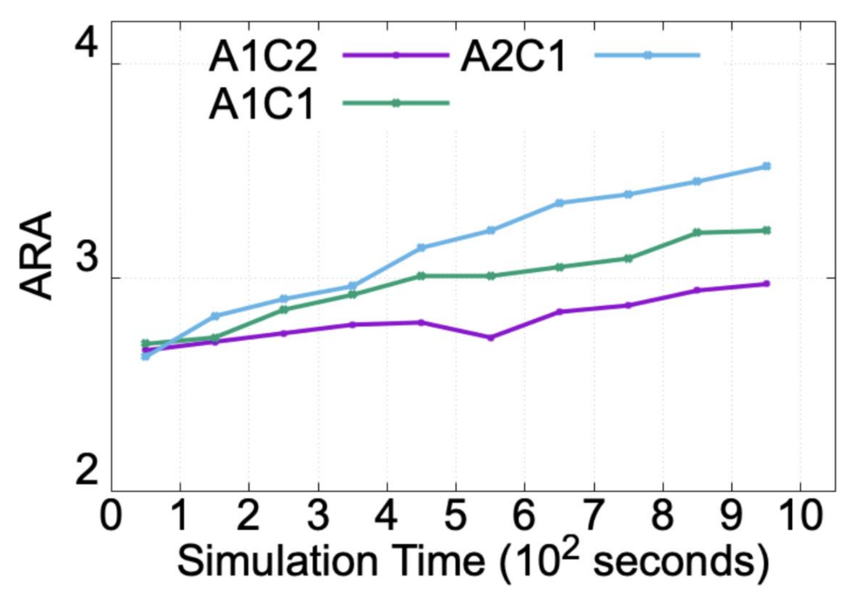
RL0: No Pre-training
RL1: Pre-trained

CF	RL0	RL1
0.0325	0.1153	0.1545

■ Adaptability Check

- Verified that adjusting reward weights enables the system to adapt its focus to environmental needs

A1C2: Cache Priority (Accuracy : Cache = 1 : 2)
A1C1: Balanced (Accuracy : Cache = 1 : 1)
A2C1: Accuracy Priority (Accuracy : Cache = 2 : 1)



A1C2	A1C1	A2C1
0.1723	0.1153	0.0576