RL for RS.

Reinforcement Learning for Recommendation System

Generative Adversarial User Model

Agent Environment

S_t : state

A : action

S_{t+1}, R : state, reward

SL: Loss function →

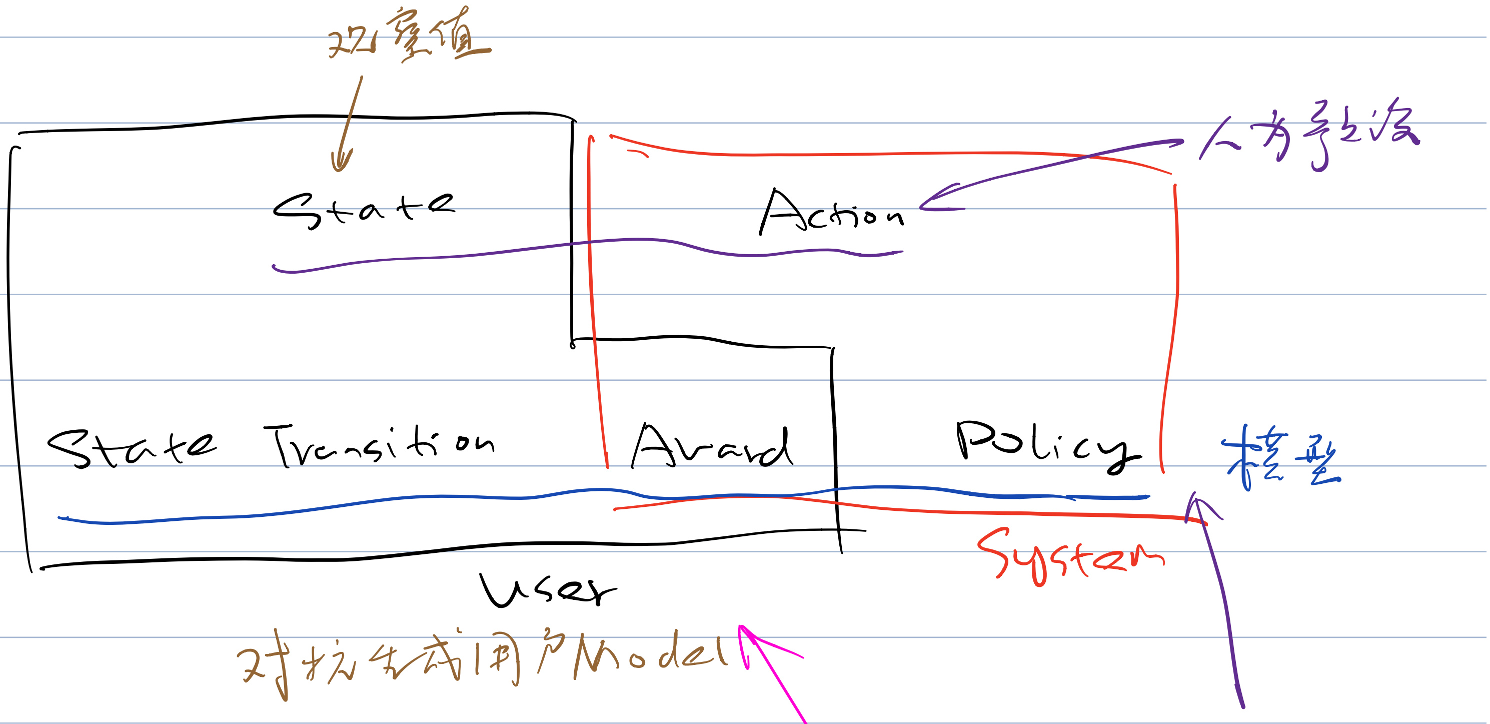
1. RL: Loss function → reward (标签不固定)

收敛性差 ← 标签不固定/少.

2, 在长期价值的评价

3. Q值: 量化评价 action 的优越性

reward + next state evaluation



状态转移函数：

$$P(\cdot \mid s^t, A^t)$$

GAN 仿真人类点击概率

环境奖励：

$$r(s^t, A^t, a^t)$$

场景没有很多数据

CTR ↑ → Reward ↑

Cascading Q-networks

$$y = r(s^t, A^t, a^t) + \gamma Q^k(s^{t+1}, a^*_{1:k}; \theta_k)$$

$\underbrace{\text{reward , ctr}}$