

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 \rightarrow

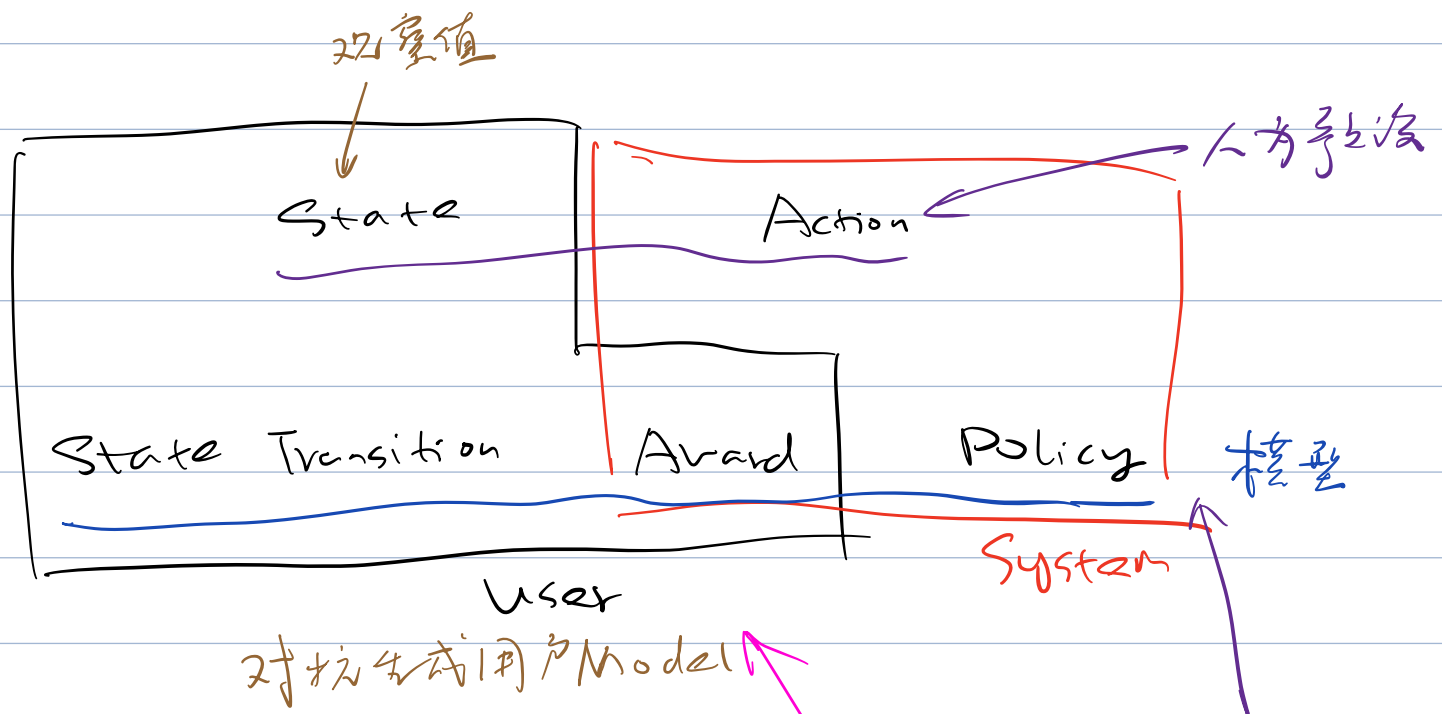
1. RL Loss function \rightarrow reward (标签不一致)

收敛性差 \leftarrow 标签不一致 / 少.

2. 在长期价值的评估

3. Q值: 量化评估 action 的优越性

reward + next state
evaluation



状态转移函数:

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

假设用户模型

GAN 作为人类点击概率

环境奖励:

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

场景没有很多数据

$$CTR \uparrow \rightarrow Reward \uparrow$$

Cascading Q-networks

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

reward, ctr