**A Survey on Reinforcement Learning for Recommender Systems**

**大体目录：**

第一节引入以及大体介绍。

第二节介绍了RL的背景，定义了相关的概念，列出了常用的方法。

第三节给出了基于rl的推荐方法的标准定义。

第四节全面回顾了为推荐系统开发的RL算法。

第五节讨论了在推荐系统中应用RL所面临的挑战和相应的解决方案。

第六节讨论了基于rl的推荐系统的各种限制和潜在的研究方向。

第七节总结了本研究。

**一、背景介绍**

**（1）应用场景**

在实践中，基于rl的推荐系统已经被应用到许多特定的场景中，如电子商务，电子学习，电影推荐，音乐推荐，新闻推荐，工作技能推荐，医疗保健，能量优化。

RL在推荐系统上的体现：

主要是：RL使推荐代理能够不断地与环境(例如，用户和/或记录的数据)交互，以学习最佳推荐策略。

**（2）传统算法：**

早期的推荐研究主要集中在开发基于内容和基于协同过滤的方法。矩阵分解是传统推荐方法中最具代表性的方法之一。

**（3）近年来的算法：**

近年来，由于深度学习的快速发展，各种神经推荐方法被开发出来。

**（4）现有算法的缺点：**

现有的推荐方法往往忽略了用户与推荐模型之间的交互。它们不能有效地捕捉到用户的及时反馈来更新推荐模型，往往导致推荐结果不理想。

**二、强化学习相关知识**

**（1）概述**

与有监督学习和无监督学习不同，RL专注于目标导向学习，即在与环境交互时最大化agent获得的总回报。试错和延迟奖励是区分RL与其他类型机器学习方法的两个最重要的特征。

**（2）RL的分类**

RL算法可以分为两大类:无模型算法和基于模型的算法。无模型算法直接学习策略，不需要任何转换函数的模型，而基于模型的算法使用学习或预先确定的模型来学习策略。

根据agent的动作，RL算法可分为价值功能方法、策略搜索方法和Actor-Critic方法。

**（3）马尔可夫决策过程**

接下来需要学习，并未了解清楚！

**（4）三种算法的介绍**

Value-function Approaches

Policy Search Methods

Actor-Critic Algorithms

**三、实验框架**

存在的问题：

社交推荐可以从社交网络中提取有价值的关系。然而，直接利用社交关系可能会降低推荐性能，因为在构建社交网络的过程中，由于随机性，用户通常会与社交邻居分享不同的偏好。

本文提出的框架：

本文提出了基于强化学习(Reinforcement Learning, SRRL)的社会推荐框架，用于识别目标用户的可靠社会关系。

SRRL框架由两部分组成:一个agent采样用户可靠的社会关系并将其传递给环境，而环境根据采样的社会关系生成推荐，并向agent提供奖励以学习最优策略。SRRL通过REINFORCE算法进行端到端训练，动态识别与目标用户有相似偏好的可靠好友。特别是，SRRL自适应采样的社交朋友，以提高推荐质量与用户反馈，因为奖励总是实时的。

**四、实验研究**

RL的任务可以分解为基本组件或一系列子任务，这大大降低了学习任务的复杂性。

这篇文章文主要针对推荐系统中的多智能体强化学习(Multi-Agent Reinforcement Learning, MARL)、HRL和有监督强化学习(Supervised Reinforcement Learning, SRL)进行研究。

**(1)Multi-agent Reinforcement Learning**

提出了一种利用竞争性马尔马尔模型来模拟学术竞争的方法。，多个agent(作者)通过学习优化的推荐轨迹来相互竞争。此外，将采用完全竞争任务的基于市场的推荐系统进行了扩展，改进的基于市场的推荐模型促使所有的agent将自己的推荐划分为不同的内部质量级别，并采用玻尔兹曼探索策略由推荐agent完成这些任务。

**(2)Hierarchical Reinforcement Learning**

如果代理决定修改用户概要文件(即一个高级任务)，它允许高级任务调用一个低级任务来删除相关的噪声课程。然而，这个HRL方法假设用户的首选项是静态的。

为了提高推荐的自适应性和准确性，我们开发了动态注意和层次强化学习(Dynamic Attention and hierarchical Reinforcement Learning, DARL)框架，自动跟踪用户在每次交互中偏好的变化。这两种方法都采用REINFORCE算法对高层和低层的策略函数进行优化。

**五、总结**

这篇文章对基于RL的21个推荐系统进行了全面回顾，使用了三大类RL(即价值功能、政策搜索和角色-评论家)，涵盖了五种典型的推荐场景，并为一些特定场景重构了一般框架，如交互式推荐、会话推荐，基于KG的可解释推荐。然后，系统分析了在推荐系统中应用RL所面临的挑战，包括环境构建、先验知识、奖励功能的定义、学习偏差和任务构建。为了促进该领域的进一步发展，还讨论了RL的理论问题，分析了现有方法的局限性，并提出了未来可能的发展方向。