

《人工智能》研究报告

**年级：大三**

**班级：32**

**姓名：汪杰、马鹤立**

**学号：23190815、21190911**

**2021年12月13日**

**计算机学院**

1. 强化学习的学习任务和目标

强化学习面临两类任务：一类是非顺序型任务；另一类是顺序型任务。

在非顺序型任务中，当agent学习环境状态空间的映射是，agent的动作会瞬时得到环境奖赏值，而不影响后记的状态和动作。而在顺序型任务中，agent采用的动作可能影响未来的状态和未来的奖赏报酬，在这种情况下，agent需要在更长的时间周期内与环境交互，估计当前动作对未来状态的影响。因此agent的学习设计到时间信度分配问题，即agent在采用一个动作后得到的奖赏值，如何分配到过去每个行为动作上。当前的研究主要集中在顺序型任务。

深度

1. 强化学习分类依据
2. 分类方法
3. 几个强化学习的代表性开源实验环境及数据集链接
4. 改进方法

1.引入注意力机制。

注意力模型（AM）最初被用于机器翻译[Bahdanaetal,2014]，现在已成为神经网络领域的一个重要概念。在人工智能（Artificial Intelligence，AI）领域，注意力已成为神经网络结构的重要组成部分，并在自然语言处理、统计学习、语音和计算机等领域有着大量的应用。

在bert模型中，注意力机制使得模型能够知道不同句子之间的相关性，极大的提升了对文章的分析能力。

在强化学习模型中，针对一些问题，我们也可以引入注意力机制，使得模型知悉不同数据间的联系，以此来强化模型的训练效果。

1. state-action-value

在强化学习中，学习过程是慢且不稳定的。可以通过ensemble learning的方法，类似于监督学习的bagging approach。我们可以通过对Q-learning的state-action-value 函数的action部分增加其稳定性，从而达到增加其性能的效果。

1. 引入rnn模型。

深度强化学习(DRL)也取得了多项突破，最引人注目的是围棋和Atari游戏。每个游戏都是在不同游戏阶段具有不同状态的环境。代理消耗状态并通过遵循某些游戏策略发出动作来做出响应。反过来，环境提供奖励（以及下一个游戏状态）作为对代理的反馈，以优化其策略。