

基于K最近邻训练的深度Q学习

2018年3月13日

1 问题描述

利用深度神经网络解决增强学习问题时，往往存在着这些问题：（1）模型需要通过激励信号学习，估计agent在一段时间内的总激励，而通常激励信号具有稀疏、有噪声和延迟的特点，即采取动作和产生激励信号之间可能间隔多步，与典型的监督学习不同；（2）深度神经网络的预设条件是样本数据相互独立，增强学习问题中的样本数据不独立，特别是相邻序列产生的状态之间高度相关；（3）增强学习中，样本数据的分布随着采取动作的不同而发生变化，深度神经网络则假设数据样本的分布不变。

为了缓解数据样本高度相关以及分布变化带来的深度神经网络模型不稳定的情况，在模型训练过程中引入经验回放机制，将所产生的样本数据放入memory中，再随机选择一定数量的样本进行训练，能在一定程度上缓解模型的不稳定性。但是深度神经网络应用过程中存在着学习速度慢的弊病：（1）用于优化深度神经网络模型的随机梯度下降方法存在着优化速度慢的特点。随机梯度下降方法需要使用较小的学习率，如果使用较大学习率可能导致模型振荡，不收敛，而较小的学习率则导致模型收敛速度慢；（2）样本数据不平衡，低激励的样本数据数量大大超过高激励的样本数据，使得模型很难学习。为缓解上述问题，本文提出基于K最近邻训练深度Q学习方法，提高样本数据的利用效率，提高模型的训练效率。