基于k-最近邻训练的深度Q学习

(Deep Q-learning based on Neural K-nearest Neighbourhood Training)

# 背景：

使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息学习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，即采取行为和产生激励信号之间可能间隔多步，与传统意义上的监督学习不同；（2）深度学习的预设条件是样本数据是独立的，增强学习问题中的样本不独立，特别是相邻序列之间的状态是高度相关的；（3）在增强学习中，样本数据的分布随着采取动作的不同而发生变化，深度学习则假设样本的分布不变；为了缓解样本高度相关和分布变化带来的模型的不稳定的情况，引入经验回访机制，将所产生的样本数据放入memory中，随机选择一定数量的样本进行训练，在一定程度上缓解模型的不稳定性。但是深度学习应用的过程中学习速度慢的弊病：（1）应用于深度神经网络优化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，使得模型收敛速度慢；（2）样本数据的高度不平衡，低激励的样本数据数量往往大大超过高激励的样本数据，使得模型很难学习；为解决上述问题，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。

# 方法：

Q-learning估计：, 成本函数,更新方法

基于K-最近邻训练深度Q学习方法，设置一个KND（K-nearest Neighbourhood Dict）,KND可以查询和写，在通过状态查询得到k-最近邻的激励集合和状态集合，则状态的一个激励，是的第i个最近邻在对应动作的激励，，是的第i个最近邻对应的状态,是向量和向量的和函数。如高斯核函数.

# 结果：

将算法应用在Atari 2600的游戏上，结果表明方法相较于Prioritised Replay和Retrace()回放机制，回放机制KND将相似状态的样本数据合并，得到一个更加合理的激励，基于K-最近邻的回放机制具有更高的样本数据效率，提高了模型的训练效率。