一. 马尔可夫链

马尔科夫链是一个数学对象，包含一系列状态以及状态之间的转移概率，如果每个状态转移到其他状态的概率只与当前状态相关，那么这个状态链就称为马尔科夫链。马尔可夫预测的基本方法就是利用状态之间的转移概率矩阵预测事件发生的状态及其发展变化趋势。

(1)特点

马尔科夫链的优势：马尔科夫链在概念上非常直观且易于实现，因为它们不需要使用任何高级的数学概念，是一种概率建模和数据分析的经典方法。

马尔科夫链的劣势：马尔可夫预测法的基本要求是状态转移概率矩阵必须具有一定的稳定性。因此，必须具有足够多的统计数据，才能保证预测的精度与准确性。且马尔科夫链只考虑当前的状态，不考虑之前状态的信息，马尔科夫的无记忆性通常使它们无法成功预测某些潜在会发生的趋势，

(2)应用场景

马尔科夫链是一种非常常见且相对简单的统计随机过程，从文本生成到金融建模，它们在许多不同领域都得到了应用。

马尔科夫链不需要考虑外界因素的影响，只需要利用历史数据或通过状态转移概率的计算得到的数据来预测将来变化的程度，因此在自然灾害中应用前景良好。以及在经济预测中，不需要连续的、大量的历史数据，只需最近一段时间的数据，结合马尔科夫链可以很好的预测下一段时间的状况。

除此之外，马尔科夫链在其他领域的应用还有很多，如教育领域、通过解码字符序列并识别最可能的语言来识别句子的语言，以及在银行的不良资产的管理、企业管理、生存环境演变等科学研究和生产生活中都有广泛应用。

二. 动态规划

动态规划可以将一个复杂问题分为一系列简单的子问题，一旦解决了这些简单的子问题，再将这些子问题的解结合起来就变成复杂问题的解了，并且同时将它们的解保存起来，如果下

次遇到相同的子问题，就不用重新计算了。动态规划本质上就是一种环境模型已知的规划方法。

(1)特点

动态规划的优势：在已知状态转换概率和回报函数的情况下，不需要与环境的交互，直接通过策略迭代或值迭代方法得到最优策略，计算效率高。

动态规划的劣势：但实际情况下，环境的状态转换概率通常是未知的，因此该方法并不实际可行。

(2)应用场景

动态规划适用于环境模型已知的情形。一般地，动态规划需要求解的问题具有 2 个核心特征：最优化的子结构属性、重叠的子问题集。MDP 满足动态规划的性质，因此，可以用动态规划方法来求解 MDP 问题。

三. 蒙特卡洛

蒙特卡洛方法是以概率和统计的理论、方法为基础的一种数值计算方法，将所求解的问题同一定的概率模型相联系，用计算机实现统计模拟或抽样，以获得问题的近似解，故又称随机抽样法或统计试验法。

(1)特点

蒙特卡洛的优势：在不知道状态转换概率的情况下，通过经验平均去估计状态的期望值函数，经验也即是采样或实验，利用当前策略进行很多次试验，每次试验都是从任意的初始状态开始直到终止状态，当采样的次数足够的多（保证每一个可能的状态-动作都能被采样到）时，就可以最大程度的逼近状态的真实期望值函数。它直接从与环境的交互中进行学习，不需要环境的动态模型，可以利用仿真或者采样模型。

蒙特卡洛的劣势：虽然蒙特卡洛方法可以在不知道状态转移概率矩阵的前提下，灵活地求解强化学习问题，但是蒙特卡洛方法需要所有的采样序列都是完整的状态序列。如果我们没有完整的状态序列就无法用蒙特卡洛方法求解。此外蒙特卡洛方法的高方差依然存在。

(2)应用场景

在解决实际问题中，我们通常不太容易获得环境的准确模型，相对而言，获得采样数据通常比较容易实现，蒙特卡洛分析就是通过采样的方法去估计状态的期望值函数，不需要知道状态转换概率，更符合实际情况，

蒙特卡洛是一种模拟统计方法，如果问题可以描述成某种统计量的形式，那么就可以用蒙特卡洛方法来解决。蒙特卡洛方法的应用场景很多，横跨物理、金融、计算机。拿计算机科学来举例，自然语言处理中的 LDA 模型，hinton 较早提出的深度学习模型 DBN 都用到了蒙特卡洛方法。虽然蒙特卡洛方法可以应用在很多场合，但求的是近似解，在模拟样本数越大的情况下，越接近与真实值，但样本数增加会带来计算量的大幅上升。

四. 时序差分

时序差分结合了动态规划的思想和蒙特卡洛的采样，基于已得到的其他状态的估计值来更新当前状态的价值函数。

(1)特点

时序差分的优势：时序差分结合了动态规划和蒙特卡洛方法，并兼具两种算法的优点。结合动态规划的思想，可以实现单步更新，提升效率；结合蒙特卡洛的采样，可以避免对状态转换概率的依赖，通过采样估计状态的期望值函数。时序差分不需要环境的动态模型，直接从经验经历中学习；也不需要等到最终的结果才更新模型，它可以基于其他估计值来更新估计值。

时序差分的劣势：因为 TD target 是估计值，估计是有误差的，这就会导致更新得到 value 是有偏差的，很难做到无偏估计。

(2)应用场景

时序差分则结合动态规划的自举思想和蒙特卡洛的采样思想，使得其边采样边自举，不仅加快了学习的速度，也能适应诸多场景。时序差分算法已经被广泛应用于动态系统、机器人控制及其他需要进行系统控制的领域。