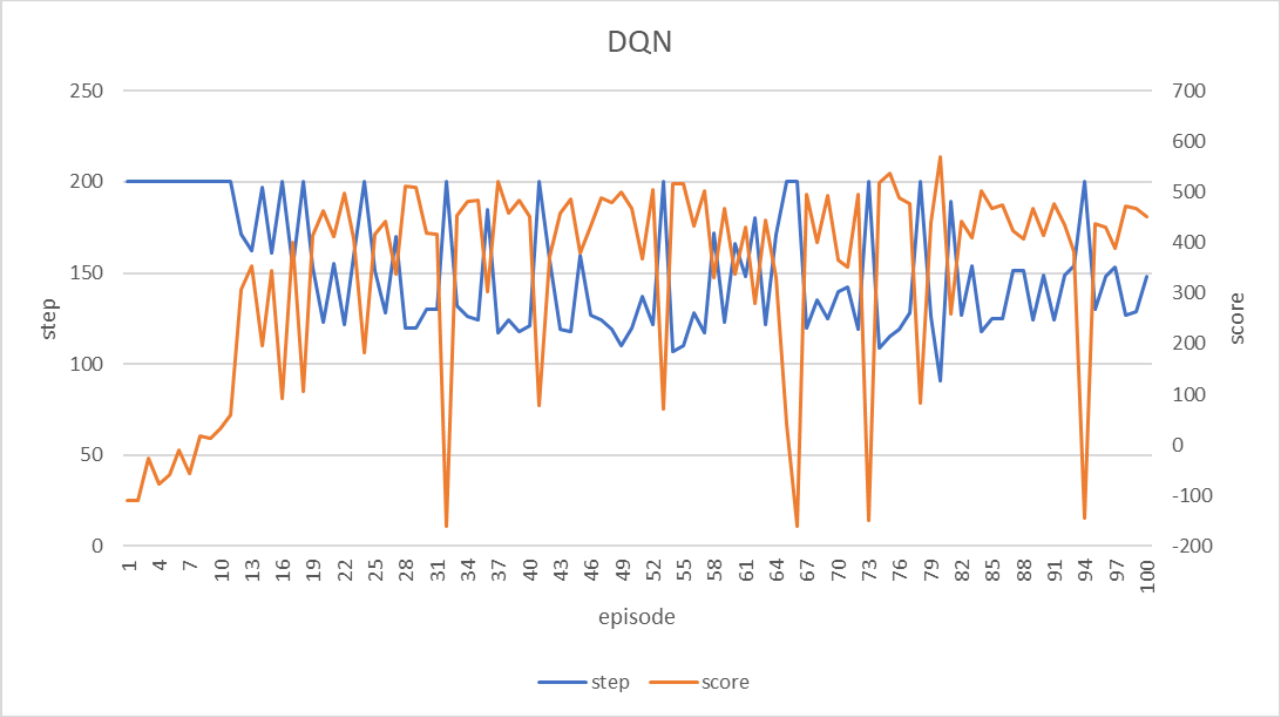
本题状态空间包括小车的x坐标以及小车的速度，原始reward恒定为-1。由于随机选取action时，小车难以到达终点，导致模型训练停滞，因此我们修改了reward计算方法以加速模型训练。

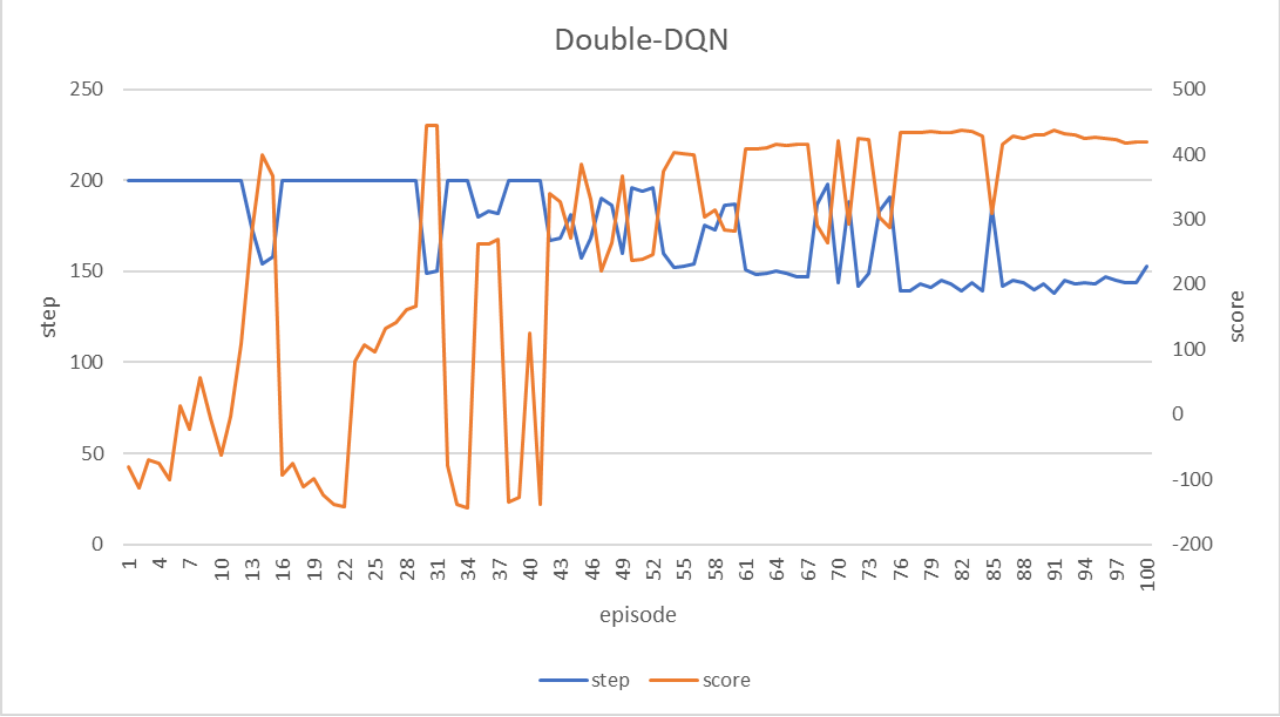
我们使用的reward如下：

$$R = x^2 + 50|v| - 1$$

小车距离起点越远、速度越快，获得的奖励越多。此外，为了鼓励更快到达终点，当到达终点时，将额外获得$100 + 5 \times (200 - \texttt{step})$的奖励。

我使用了DQN和Double-DQN进行训练。训练结果如下，其中蓝色曲线代表小车到达终点所需要的步数，橙色曲线代表该轮获得的总reward。





可以看到，Double-DQN开始时的性能提升要慢于DQN，这可能是因为Double-DQN需要让两个模型轮流训练，导致单个模型提升较慢。在训练后半段，可以明显看出Double-DQN的波动小于DQN，证明Double-DQN确实能提高模型的稳定性。