经验回放使你获得更多吹捧

简介

吹捧是一个汉语词语，意思是吹嘘捧场。出自 《长官意志》。亦有人称为鼓励、吹逼等。

获得他人的吹捧有利于我们心情的愉悦，我们都希望最大化自己得到的吹捧。我们发现，借助深度强化学习中应用的经验回放机制，有利于我们获得更多的吹捧，同时可以显著减少训练时间。

背景

我们发现，人一旦获得别人的吹捧，心情会倍加愉悦，工作起来也会倍加卖力。所以一些心术不正的领导会通过适当的夸奖来进一步压榨员工的价值，在代价几乎为零的情况下得到的员工效率的提升。但我们在此不考虑这种情况，我们的目标是获得尽可能多的吹捧，除去其对身心健康的帮助作用，心情愉悦还能帮助我们缓解脱发。因此我们都希望最大化得到的吹捧。

经验回放是 DQN 的主要做法，其将系统探索环境得到的数据储存起来，然后随机采样样本更新深度神经网络的参数。优点：1. 数据利用率高，因为一个样本被多次使用。2. 连续样本的相关性会使参数更新的方差（variance）比较大，该机制可减少这种相关性。

问题定义

获得吹捧的过程定义为马尔科夫模型(MDP),环境为 agent 表现自己的场景，如组会讲论文，评奖学金展示等。状态 S 代表每一刻环境的剪影。动作集包含 agent 的语调、体态、表情等各个方面。回报为他人的吹捧，由于一般不会很多人同时讲话，所以每个人的单次吹捧等价，训练目标就是最大化在特定场景结束时获得的吹捧次数。

模型

加入经验回放的方法为将自己的表演过程录下来，然后均存放在硬盘中。类似舞蹈演员录下自己的舞蹈过程，通过录像发现自己的不足。Agent 通过随机观察录像找到获得更多吹捧的策略。这样你记不清的表演过程、细节也不用怕忘记了，提高了数据利用率。除此之外，当你生动地解释了一个概念时，有人夸你太

幽默了，有人夸你理解的太对了，这时你就可能陷入混乱，到底是幽默好，还是理解充分好。但如果你有经验回放，你就能观察过往的记录，找到合适的优先级。有些概念幽默更重要，而另一些讲得清楚更重要。

实验

事实证明经验回放机制是有效的，不信可以尝试复现。

结论

多记录生活。