码农的专栏

个人资料

代码学习者

访问： 2085069次
积分： 28702
等级： BLOG 7
排名： 第183名

原创： 80篇 转载： 463篇
译文： 0篇 评论： 23条

文章搜索

DQN 从入门到放弃1 DQN与增强学习

2017-04-19 16:39 6900人阅读 评论(0)

分类： RL（33）

目录(?) [+]

1 前言

深度增强学习Deep Reinforcement Learning是将**深度学习**与增强学习结合起来从而实现从Perception感知到Action动作的端对端学习End-to-End Learning的一种全新的**算法**。简单的说，就是利用然后通过深度神经网络，直接输出动作，中间没有hand-crafted engineering的现真正完全自主的学习一种甚至多种技能的潜力。

虽然将深度学习和增强学习结合的想法在几年前就有人尝试，但真正成功的开的Playing Atari with Deep Reinforcement Learning一文，在该文中第一次提出名称，并且提出DQN（Deep Q-Network）算法，实现从纯图像输入完全通过DeepMind在Nature上发表了改进版的DQN文章**Human-level Control throug**引起了广泛的关注，Deep Reinfocement Learning 从此成为深度学习领域的前

而Hinton，Bengio及Lecun三位大神在Nature上发表的Deep Learning综述一文最后也将Deep Reinforcement Learning作为未来Deep Learning的发展方向。引用一下原文的说法：

> We expect much of the future progress in vision to come from systems that are trained end-to-end and combine ConvNets with RNNs that use reinforcement learning to decide where to look.

从上面的原文可见三位大神对于Deep Reinforcement Learning的期待。而显然这一年来的发展没有让大家失望，AlphaGo横空出世，将进一步推动Deep Reinforcement Learning的发展。

Deep Reinforcement Learning的重要关键在于其具备真正实现AI的潜力，它使得计算机能够完全通过自
项任务，甚至超过人类的水平。也因此，DeepMind很早受到了Google等企业的关注。DeepMind 50多人
2014年就被Google以4亿美元的价格收购。而15年12月份刚刚由Elon Musk牵头成立的OpenAI，则一开
10亿美元的投资，而OpenAI中的好几位成员都来自UC Berkerley的Pieter Abbeel团队。

Pieter Abbeel团队紧随DeepMind之后，采用基于引导式监督学习直接实现了机器人的End-to-End学习，
引起了大量的媒体报道和广泛关注。去年的NIPS 2015 更是由Pieter Abbeel及DeepMind的David Silver联合组织了
Deep Reinforcement Learning workshop。可以说，目前在Deep Reinforcem
中在DeepMind和UC Berkerley团队。

为了研究Deep Reinforcement Learning，DQN的学习是首当其冲的。只有真
Reinforcement Learning入门。要理解并掌握DQN算法，需要增强学习和深度
开始接触DQN，但由于对基础知识掌握不全，导致竟然花了近1年的时间才真
栏从今天开始推出 **DQN 从入门到放弃 系列**文章，意在通过对增强学习，深度
Tensorflow的代码实现，使大家能够扎实地从零开始理解DQN，入门Deep Re
以一周一篇的速度更新。另外要说明的一点是DQN已被Google申请专利，因

文章分类

TX1 (6)
ELK (5)
ubuntu (43)
ffmpeg (1)
hbase (1)
torch (8)
windows (7)
DL (106)
opencv (8)
ML (15)
NLP (39)
image (2)
centos (1)
face (10)
GAN (9)
RL (34)
iOS (5)
python (15)
lua (2)
caffe (4)
C++ (2)
ROS (3)
algorithm (2)
pytorch (34)
paper reading (76)
tensorflow (2)
automatic driving (3)

2 预备条件

虽然说是从零开始，但是DQN毕竟也还属于深度学习领域的前沿算法，为了理解本系列的文章，知友们还是需要有一定的基础：

- 一定的概率论和线性代数基础（数学基础）
- 一定的Python编程基础（编程基础，后面的代码实现将完全基于Tensorflow实现）

考虑到目前理解深度学习的知友肯定比理解增强学习的知友多，并且专栏也在同步翻译CS231N的内容，章计划用极短的篇幅来介绍DQN所使用的深度学习知识，而用更多的篇幅介绍增强学习的知识。

如果知友们具备以上的基本预备条件，那么我们就可以开始DQN学习之旅了。

接下来本文将介绍增强学习的基础知识。

3 增强学习是什么

在人工**智能**领域，一般用**智能体Agent**来表示一个具备行为能力的物体，比如机器人，无人车，人等等。那么增强学习考虑的问题就是**智能体Agent**和**环境Environment**之间交互的任务。比如
械臂周围的物体包括手机就是环境，机械臂通过外部的比如摄像头来感知环境
起手机这个任务。再举玩游戏的例子，比如我们玩极品飞车游戏，我们只看到
作（键盘操作）来控制车的运动。

那么，不管是什么样的任务，都包含了一系列的**动作Action**,**观察Observatio**
就是Agent执行了动作与环境进行交互后，环境会发生变化，变化的好与坏就
果机械臂离手机变近了，那么Reward就应该是正的，如果玩赛车游戏赛车越
的。接下来这里用了Observation观察一词而不是环境那是因为Agent不一定能
的摄像头就只能得到某个特定角度的画面。因此，只能用Observation来表示A

wordpress (0)
latex (1)
boost (1)

文章存档

2017年07月 (11)
2017年06月 (73)
2017年05月 (54)
2017年04月 (237)
2017年03月 (77)

展开

阅读排行

linux pytorch 安装 (9473)
SLAM算法解析：抓住视 (8354)
PyTorch中文文档 (8229)
GAN学习指南：从原理入 (8104)
CNN浅析和历年ImageNe (8073)
MTCNN训练整理 (8047)
带你搞懂朴素贝叶斯分类 (8043)
Adversarial Nets Papers (7965)
elasticsearch 5.0 版本安 (7933)
CentOS安装nvidia显卡驱 (7814)

评论排行

MTCNN训练数据整理 (4)
linux pytorch 安装 (2)

上面这张图（来自David Silver的课程ppt）可以很清楚的看到整个交互过程。事实上，这就是人与环境交互的一种模型化表示。在每个时间点time-step Agent都会从可以选择的动作集合A中选择一个动作 a_t 执行.这个动作集合可以是连续的比如机器人的控制也可以是离散的比如游戏中的几个按键。动作集合的数量将直接影响整个任务的求解难度，因此DeepMind才从玩最简单的游戏做起，DQN算法（不考虑其变种）仅适用于离散输出问题。

那么知道了整个过程，任务的目标就出来了，那就是要能获取尽可能多的Reward。没有目标，控制也就无从谈起，因此，获取Reward就是一个量化的标准，Reward越多，就表示执行得越好。每个时间片，Agent都是根据当前的观察来确定下一步的动作。观察Observation的集合就作为Agent的所处的**状态State**，因此，**状态State**和**动**存在映射关系，也就是一个state可以对应一个action，或者对应不同动作的概率（常常用概率来表示，概就是最值得执行的动作）。状态与动作的关系其实就是输入与输出的关系，而状态State到动作Action的**映**为一个**策略Policy，**一般用 π 表示，也就是需要找到以下关系：

$$a = \pi(s)$$

或者

$$\pi(a|s)$$

其中a是action，s是state。第一种是一一对应的表示，第二种是概率的表示。

> 增强学习的任务就是找到一个最优的策略Policy从而使Reward最多。

我们一开始并不知道最优的策略是什么，因此往往从随机的策略开始，使用随**机**列的状态,动作和反馈：

$$\{s_1, a_1, r_1, s_2, a_2, r_2, \ldots s_t, a_t, r_t\}$$

这就是一系列的**样本Sample**。增强学习的算法就是需要根据这些样本来改进Policy，从而使得得到的样本中的Reward更好。由于这种让Reward越来越好的特性，所以这种算法就叫做增强学习Reinforcement Learning。

4 What's Next?

在下一篇文章中，笔者将和大家分享MDP马尔科夫决策过程的知识，这是构建增强学习算法的基础。敬请关注！

上一篇 Deep Reinforcement Learning: Pong from Pixels
下一篇 论文引介 | NMT with Conditional Sequence Generative Adversarial Nets





查看评论

暂无评论

发表评论

用 户 名： haijunz

评论内容：

提交