



## 码农的专栏

[目录视图](#)[摘要视图](#)[RSS 订阅](#)

### 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

[赠书 | AI专栏 \(AI圣经!《深度学习》中文版\)](#) [评论送书 | 机器学习、Java虚拟机、微信开发](#)

## 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，就是利用深度神经网络，直接输出动作，中间没有hand-crafted engineering的过程，实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，但真正成功的开始是2013年David Silver等人发表的**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神经网络输出动作。DeepMind在Nature上发表了改进版的DQN文章**Human-level Control through Deep Reinforcement Learning**，引起了广泛的关注，Deep Reinforcement Learning 从此成为深度学习领域的热门话题。

关闭





## 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)

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顶任务，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，则一开始就获得了10亿美元的投资，而OpenAI中的好几位成员都来自UC Berkeley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 Reinforcement Learning领域，最活跃的团队是在DeepMind和UC Berkeley团队。

为了研究Deep Reinforcement Learning，DQN的学习是首当其冲的。只有真正理解了DQN，才能开始接触DQN，但由于对基础知识掌握不全，导致竟然花了近1年的时间才真正入门。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，深度学习，Tensorflow的代码实现，使大家能够扎实地从零开始理解DQN，入门Deep Reinforcement Learning。将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，因此目前只能作为学术研究使用。



FPGA (1)  
RNN (5)  
leetcode\_python (9)  
android (2)  
hadoop (2)  
mahout (2)  
openstack (1)  
oracle (1)  
security (4)  
SLAM (17)  
cocos2D-x (1)  
caption (14)  
tracking (5)  
vSLAM (10)  
IOT (2)  
CV (3)  
speech (20)  
object detection (3)  
VAE (2)  
Adreno (2)  
OpenCL (1)  
Snapdragon (6)  
Raspberry Pi (10)  
robot (1)  
system (0)  
segmentation (1)  
linux (2)  
Financial (1)  
VQA (4)  
cuda (1)  
Blockchain (13)  
video (6)

## 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。比如机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拿起手机这个任务。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Observation**。就是Agent执行了动作与环境进行交互后，环境会发生变化，变化的好与坏就如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快的。接下来这里用了Observation观察一词而不是环境那是因为Agent不一定能拿到整个环境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

关闭



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ImageNet (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**动作Action**之间存在映射关系，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，比如  $\pi(a|s)$  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的映射为一个**策略Policy**，一般用  $\pi$  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，使用随机的状态、动作和反馈：

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

关闭





- [TX1 安装 ROS Indigo](#) (2)
- [pytorch学习笔记（八）：](#) (2)
- [Jetson TX1 开发教程（4](#) (2)
- [opencv 仿射变换 根据眼](#) (2)
- [\[iOS\]iOS结合OpenCV做](#) (2)
- [MTCNN训练整理](#) (2)
- [深度学习中的激活函数导](#) (1)
- [MAT: A Multimodal Atten](#) (1)

### 推荐文章

- \* CSDN日报20170725——《新的开始，从研究生到入职亚马逊》
- \* 深入剖析基于并发AQS的重入锁(ReentrantLock)及其Condition实现原理
- \* Android版本的"Wannacry"文件加密病毒样本分析(附带锁机)
- \* 工作与生活真的可以平衡吗？
- \* 《Real-Time Rendering 3rd》提炼总结——高级着色：BRDF及相关技术
- \* 《三体》读后思考-泰勒展开/维度打击/黑暗森林

### 最新评论

- MAT: A Multimodal Attentive Traqq\_39582061: 可以请问一下这篇论文发表的期刊吗？我这里搜不到。。
- [iOS]iOS结合OpenCV做视频流代码学习者: 哪3个C++问题
- [iOS]iOS结合OpenCV做视频流

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！

顶 踩  
2 0

上一篇 [Deep Reinforcement Learning: Pong from Pixels](#)

下一篇 [论文引介 | NMT with Conditional Sequence Generative Adversarial Nets](#)

### 相关文章推荐

- Deep Reinforcement Learning: Pong from Pixels
- 深度增强学习之Policy Gradient方法1
- 深度增强学习David Silver（七）——Policy Gradi...
- 用Tensorflow基于Deep Q Learning DQN 玩Flappy...
- Deep Reinforcement Learning 基础知识（DQN方...
- DQN 原理（二）
- DQN
- double dqn repo
- DQN从入门到放
- 强化学习系列<4

关闭



Await\_Xpf: xcode 中都弄好之后报3个 c++ 的问题，遇到过吗？794778062希望能与您沟通交流一...

MTCNN训练整理

huangbo1221: 您好，能参考一下修改后的softmax\_loss\_layer和euclidean\_loss\_laye...

Jetson TX1 开发教程（4）--Ten: 代码学习者: @u013768935:是的

Jetson TX1 开发教程（4）--Ten: face13138: 博主，在tx2上运行上述例程的话，过程是先安装jetpack，然后配置caffe吗？

pytorch学习笔记（八）：PyTorch: 代码学习者: 可以啊

pytorch学习笔记（八）：PyTorch: Lucio: 博主您好！请教您：我不用pyTorch在Torch中能实现可视化每一层的输出吗？

机器学习&深度学习资料汇总（含noseew: 不要以为蓝色的就是链接，有可能就是字体是蓝色的

opencv 仿射变换 根据眼睛坐标: a\_cainiaoTxl: 您好，我是初学者，最近才开始接触到人脸识别



电脑租赁



人工智能机器人



橡树湾二手房



无地址注册公司



奥特莱斯



显卡排行

猜你在找

【直播】机器学习&深度学习系统实战（唐宇迪）

【直播回放】深度学习基础与TensorFlow实践（王琛）

【直播】机器学习之凸优化（马博士）

【直播】机器学习之概率与统计推断（冒教授）

【直播】TensorFlow实战进阶（智亮）

【直播】Kaggle 神器：XGBoost 从基础到实战（冒教授）

【直播】计算机视觉原理及实战（屈教授）

【直播】机器学习之矩阵（黄博士）

【直播】机器学习之数学基础

【直播】深度学习30天系统实训（唐宇迪）

查看评论

暂无评论

发表评论

用户名： haijunz

评论内容：



Empty text area for user comment.

提交

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

关闭





## 日本八日游



[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

网站客服

杂志客服

微博客服

webmaster@csdn.net

400-660-0108

| 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 | 江苏乐知

司

京 ICP 证 09002463 号 | Copyright © 1999-2017, CSDN.NET, All Rights Reserved



## 电脑租赁



关闭

**无基因 不精准  
唯云端 更高效**

华大基因借云计算承载大量且复杂的  
计算需求完成基因测序

英特尔、Intel是英特尔公司在美国和其他国家的商标。\*其他的名称和品牌可能是其他所有者的商标。