

Bayesian Deep Learning

We all have two lives. The second one starts when we realize that we only have one. --- Tom Hiddleston

[博客园](#) [首页](#) [联系](#) [管理](#)

[DQN] What is Deep Reinforcement Learning

已经成为DL中专门的一派，高大上的样子



Boston Dynamics

昵称：[郝一二三](#)

园龄：[6年4个月](#)

粉丝：[41](#)

关注：[5](#)

[+加关注](#)

<	2017年10月						>
日	一	二	三	四	五	六	
24	25	26	27	28	29	30	
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	6	<u>7</u>	
<u>8</u>	<u>9</u>	10	<u>11</u>	12	13	<u>14</u>	
<u>15</u>	16	17	<u>18</u>	<u>19</u>	<u>20</u>	21	
<u>22</u>	23	24	25	26	27	28	
29	30	31	1	2	3	4	

Intro:

[MIT 6.S191 Lecture 6: Deep Reinforcement Learning](#)

Course:

[CS 294: Deep Reinforcement Learning](#)

随笔分类 (284)

~ARM(14)
~Kernel API(11)
[Algo\(10\)](#)
[AR\(10\)](#)
[Bayes\(30\)](#)
[C\(14\)](#)
[C++\(22\)](#)
[CV\(31\)](#)



Jan 18: Introduction and course overview (Levine, Finn, Schulman)

- [Slides: Levine](#)
- [Slides: Finn](#)
- [Slides: Schulman](#)
- [Video](#)

Why deep reinforcement learning?

- Deep = can process complex sensory input
 - ↳ ...and also compute really complex functions
- Reinforcement learning = can choose complex actions

OpenAI

[OpenAI](#) 2016年6月21日宣布了其目标，包括制造“通用”机器人和使用自然语言的聊天机器人。

近期发展



Q-learning

V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, et al. ***“Playing Atari with Deep Reinforcement Learning”***. (2013).

IR(30)
ML(36)
NN(45)
PM(7)
Python(13)
Thought(10)
浅尝辄止(1)

MyLink

Bayesian
cplusplus.com
Linux Kernel
MIT
Stanford
同行

积分与排名

积分 - 61796
排名 - 5123

最新评论

1. [Re:嵌入式读书列表 - 参考](#)

大三，努力学习当中，我肯定要把我认为的好书学完

--sty01

2. [Re:\[Pycharm\] Debugger in Pycharm](#)

写的这么仔细的pycharm教程真少啊

--RammstainX

3. [Re:\[Artoolkit\] ARSimpleNativeCarsProj for Multi Markers Tracking](#)

加油哈博主，你写的都是好东西。我刚开始学ARToolKit，一头雾水。。。看了你写的感觉很好，学了好多东西。这一篇要是能再细致点就好了。。另外您介意不介意留个联系方式，一起学习交流。

policy gradients

J. Schulman, S. Levine, P. Moritz, M. I. Jordan, and P. Abbeel. "Trust Region Policy Optimization". (2015);

V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, et al. "Asynchronous methods for deep reinforcement learning". (2016).

DAGGER

X. Guo, S. Singh, H. Lee, R. L. Lewis, and X. Wang. "*Deep learning for real-time Atari game play using offline Monte-Carlo tree search planning*". NIPS. 2014.

guided policy search

S. Levine, C. Finn, T. Darrell, and P. Abbeel. "*End-to-end training of deep visuomotor policies*". (2015).

policy gradients

J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel. "High-dimensional continuous control using generalized advantage estimation". (2015).

**Finally, AlphaGo 的 四大技术**

supervised learning + policy gradients + value functions + Monte-Carlo tree search

Deep Q Network

问题：RE有没有deep，能如何？

回答：Google's DeepMind published its famous paper [Playing Atari with Deep Reinforcement Learning](#), in which they introduced a new algorithm called **Deep Q Network** (DQN for short) in 2013. It demonstrated how an AI agent can learn to play games by just observing the screen without any prior information about those games(无信息先验?). The result turned out to be pretty impressive.

This paper opened the era of what is called '*deep reinforcement learning*', *a mix of deep learning and reinforcement learning*.

--左右-

4. Re:[OpenCV] Face Detection

慧眼科技研发总监李汉曦

--郝一二三

5. Re:Git: Basics

谢谢你提的建议，《Git权威指南》已买！

--基本群众

阅读排行榜

1. [Pycharm] Interpreter setting in Pycharm (6874)
2. Communication - 03.RILC(4573)
3. 内核回调 之一"读"到底(4110)
4. 嵌入式读书列表 - 参考(3426)
5. 自旋锁，旋啊旋(2486)

评论排行榜

1. A Game idea...(18)
2. Linux读书片言(17)
3. 嵌入式读书列表 - 参考(12)
4. 挂个文件系统(7)
5. 当c遇到OOA，初探(6)

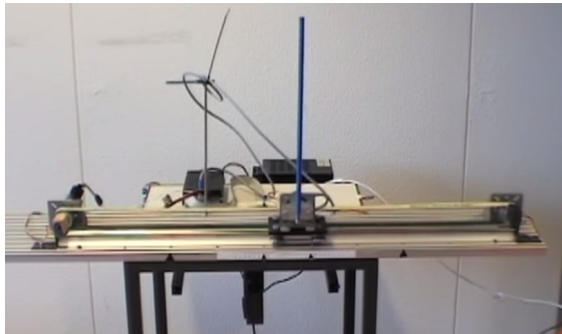
Then, 通过实践了解这个牛牛的网络：[Deep Q Learning with Keras and Gym](#)

外加一个有良心的国内博客：[用Tensorflow基于Deep Q Learning DQN 玩Flappy Bird](#)（课外阅读）

Cartpole Game 简介

CartPole is one of the simplest environments in **OpenAI gym** (a game simulator).

当然了，有经费，也可以这么搞个真玩意。



As you can see in the animation from the top, the goal of CartPole is to balance a pole connected with one joint on top of a moving cart.

Instead of pixel information, there are 4 kinds of information given by the state, such as angle of the pole and position of the cart.

An agent can move the cart by performing **a series of actions of 0 or 1** to the cart, pushing it left or right.

Gym makes interacting with the game environment really simple.

```
next_state, reward, done, info = env.step(action)
```

学习的输入参数，要具体问题具体分析。

As we discussed above, action can be either 0 or 1.

If we pass those numbers, env, which represents the game environment, will emit the results. done is a boolean value **telling whether the game ended or not**.

The **old state information** paired with action and **next_state** and **reward** is the information we need for training the agent.

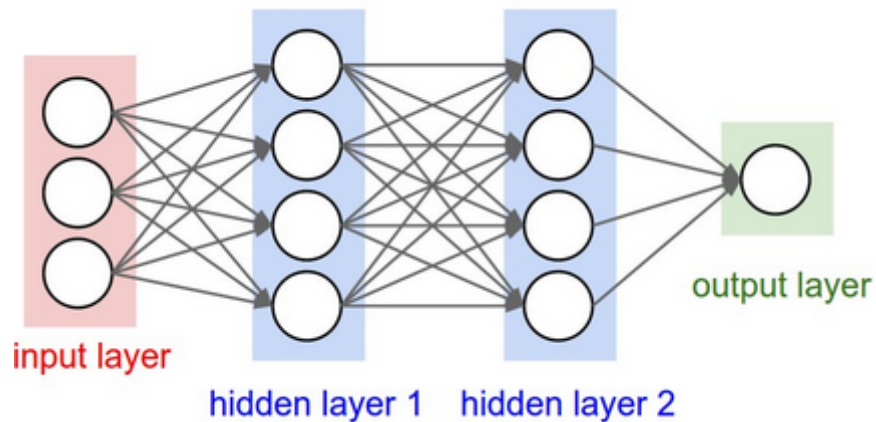
Implementing Simple Neural Network using Keras

This post is not about deep learning or neural net. So we will consider neural net as just a black box algorithm.

An algorithm that learns on the pairs of example input and output data, detects some kind of patterns, and **predicts** the output based on an unseen input data.

But we should understand which part is the neural net in the DQN algorithm.

DQN 算法中哪里涉及神经网络



Note that the neural net we are going to use is similar to the diagram above.

We will have one input layer that **receives 4 information** and 3 hidden layers. 输入层

But we are going to have 2 nodes in the output layer since there are two buttons (0 and 1) for the game.

Keras makes it really simple to implement basic neural network.

The code below creates an empty neural net model.

activation, loss and optimizer are the parameters that define the characteristics of the neural network, but we are not going to discuss it here.

原来如此结合，有机会实现下，目前不着急。

马里奥AI实现 简介

Ref: http://www.cnblogs.com/Leo_wl/p/5852010.html

基于NEAT算法的马里奥AI实现

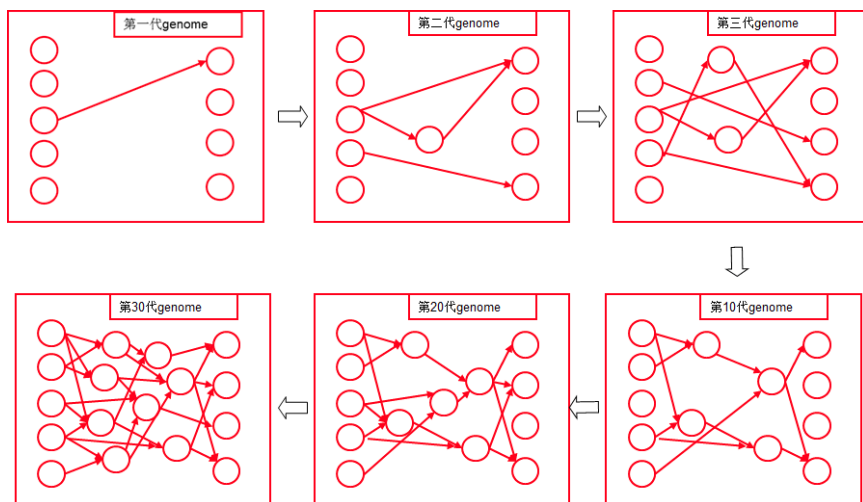
所谓NEAT算法即通过增强拓扑的进化神经网络（Evolving Neural Networks through Augmenting Topologies），算法不同于我们之前讨论的传统神经网络，

- 它不仅会训练和修改网络的权值，
- 同时会修改网络的拓扑结构，包括新增节点和删除节点等操作。

NEAT算法几个核心的概念是：

- 基因：网络中的连接
- 基因组：基因的集合
- 物种：一批具有相似性基因组的集合
- Fitness：有点类似于增强学习中的reward函数
- generation：进行一组训练的基因组集合，每一代训练结束后，会根据fitness淘汰基因组，并且通过无性繁殖和有性繁殖来新增新的基因组
- 基因变异：发生在新生成基因组的过程中，可能会出现改变网络的权重，增加突出连接或者神经元，也有可能禁用突触或者启用突触

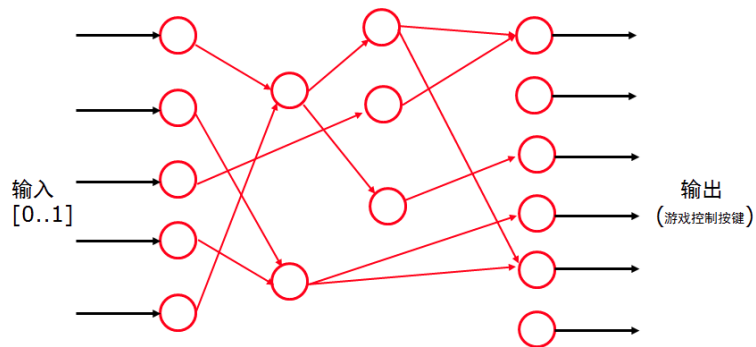
下图我们展示了算法从最开始简单的神经网络，一直训练到后期的网络



利用NEAT算法实现马里奥通关的基本思想便是：

利用上面NEAT算法的基本观点，从游戏内存中获取实时的游戏数据，判断马里奥是否死亡、计算Fitness值、判断马里奥是否通关等，从而将这些**作为神经网络的输入**，

最后**输出对马里奥的操作**，包括上下左右跳跃等操作，如下图：



大多数该算法实现马里奥的智能通关都依赖于模拟器，运用lua语言编写相应脚本，获取游戏数据并操作马里奥。

Ref: [NeuroEvolution with MarI/O](#)。

但，更加复杂的游戏，没有确切的状态指标作为输入，便涉及到了CNN图像识别。

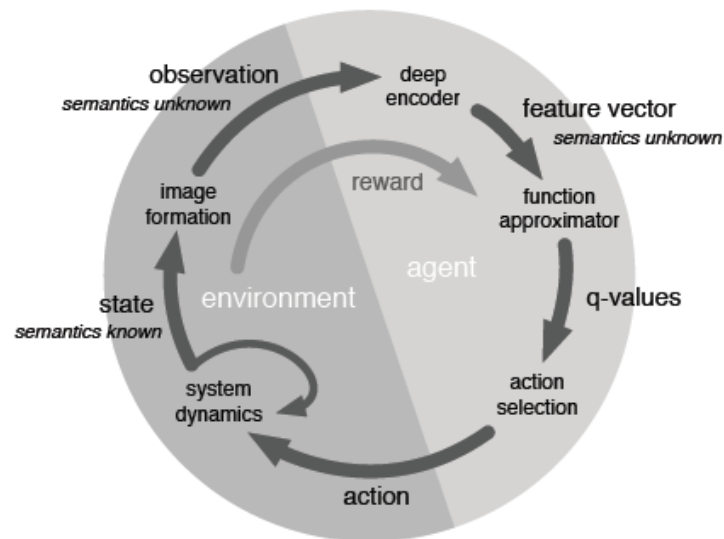
这里的CNN与上述的网络不同，一个是识别图像**产生状态**（input的值）。一个是**处理**input值。

基于Deep Q-learning的马里奥AI实现

NEAT算法是相对提出较早的算法，在2013年大名鼎鼎的DeepMind提出了一种深度增强学习的算法，该算法主要结合了我们上面讨论的CNN和Q-Learning两种算法。

DeepMind的研究人员将该算法应用在Atari游戏机中的多种小游戏中进行AI通关。

其基本算法核心便是我们之前介绍的CNN和增强学习的Q-Learning，游戏智能通关的基本流程如下图：

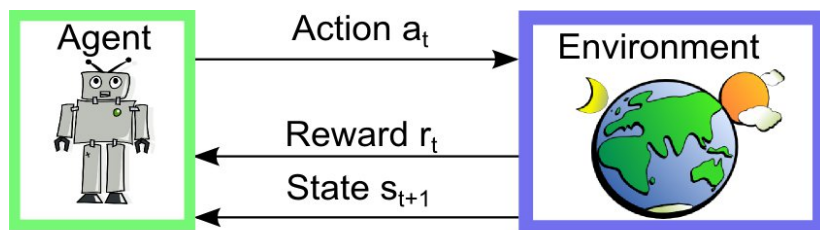


do {

- 利用**CNN**来**识别**游戏总马里奥的**状态**，
- 并利用**增强学习算法**做出**动作选择**，
- 然后根据新的返回状态和历史状态来**计算**reward函数从而反馈给Q函数进行迭代，

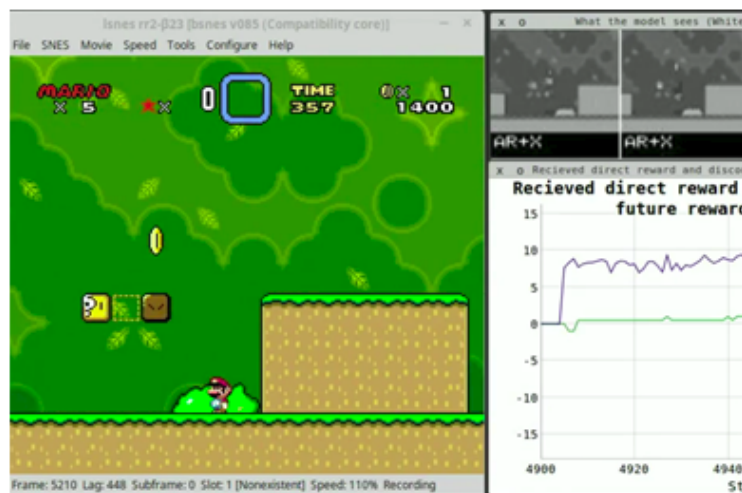
} while(不断的训练直到游戏能够通关)

研究人员在训练了一个游戏后，将**相同的参数**用在别的游戏中发现**也是适用的**，说明该算法具有一定的普遍性。**神奇！**



Reinforcement Learning Setup

而同样的方法，将DRL应用在马里奥上，github上有一个开源的实现方式：[aleju/mario-ai](https://github.com/aleju/mario-ai)
其最终的实现效果图如下：



我们发现在CNN识别过程中，每4帧图像，才会进行一次CNN识别，这是识别速率的问题，图中曲线反映了直接回报函数和简介回报函数。

Q-learning 基础 (应该涉及到Bayes loss function)

Link: [另起一篇](#)

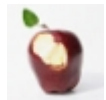
DEEP REINFORCEMENT LEARNING: AN OVERVIEW

From: <https://arxiv.org/pdf/1701.07274.pdf>

中文简述 : <http://www.jiqizhixin.com/article/2208>

有必要总结下later.

分类: [NN](#)



[郝一二三](#)

[关注 - 5](#)

[粉丝 - 41](#)

[+加关注](#)

0

0

posted @ 2017-05-12 08:23 [郝一二三](#) 阅读(51) 评论(0) [编辑](#) [收藏](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问](#)网站首页。

【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库

【活动】腾讯云 【云+校园】套餐全新升级

【推荐】报表开发有捷径：快速设计轻松集成，数据可视化和交互



最新IT新闻:

- [绕过ofo小黄车 摩拜为何把长剑挥向滴滴？](#)
 - [Wine 2.0.3稳定版发布](#)
 - [自媒体曝马云50亿豪宅：阿里愤怒各自索赔100万](#)
 - [苹果中国上线iPhone X福利计划：购买官方服务爽翻](#)
 - [传言将要赴美上市的B站 正在逐步解决他的盈利问题](#)
- » [更多新闻...](#)



最新知识库文章:

- [实用VPC虚拟私有云设计原则](#)
 - [如何阅读计算机科学类的书](#)
 - [Google 及其云智慧](#)
 - [做到这一点，你也可以成为优秀的程序员](#)
 - [写给立志做码农的大学生](#)
- » [更多知识库文章...](#)