

有趣的机器学习

-  Python基础 ▼
-  机器学习 ▼
-  数据处理 ▼
-  其他 ▼

什么是 Sarsa (Reinforcement Learning 强化学习)



切换到 优酷 视频

(Chrome无法播放优酷? 网址框输入"chrome://settings/content/", 勾选允许 Flash Player. 实在不行? 请 [点击这里](#))

« 上一个

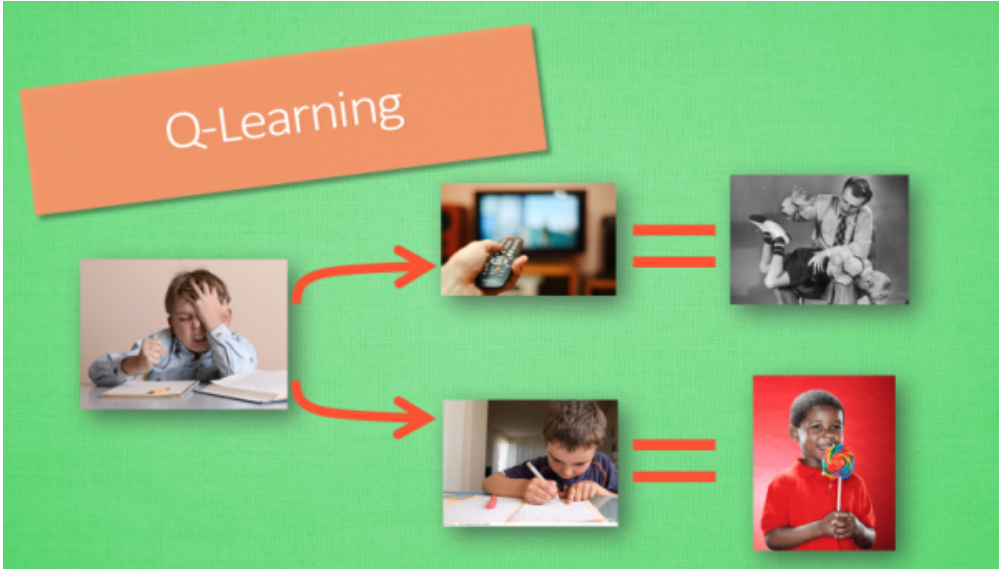
下一个 »

Sarsa

- 学习资料:
 - 强化学习教程
 - 强化学习模拟程序
 - Q-learning 简介视频
 - Sarsa Python 教程
 - 学习书籍 [Reinforcement learning: An introduction](http://ufal.mff.cuni.cz/~straka/courses/npfl114/2016/sutton-bookdraft2016sep.pdf)

今天我们会来说说强化学习中一个和 Q learning 类似的算法, 叫做 Sarsa.

注: 本文不会涉及数学推导. 大家可以在很多其他地方找到优秀的数学推导文章.



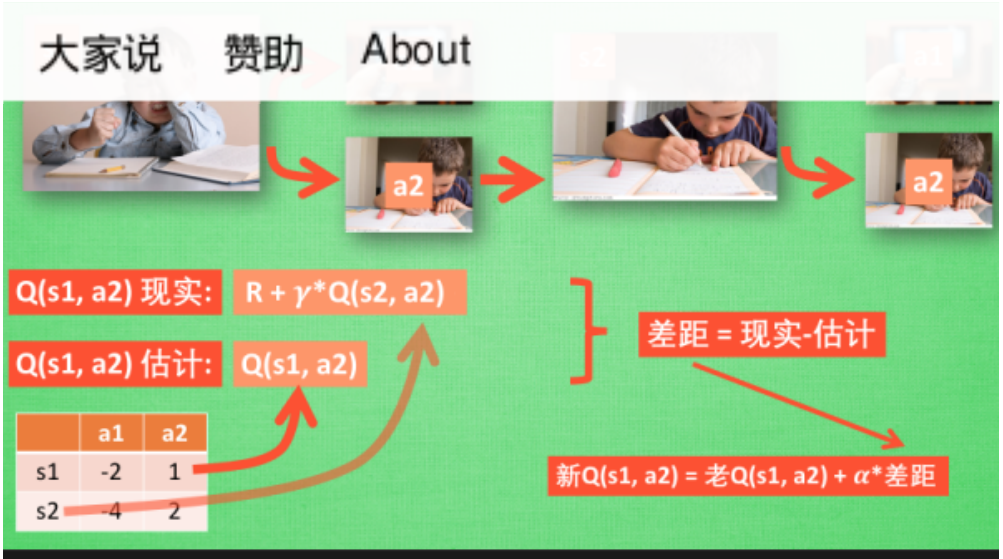
在强化学习中 Sarsa 和 Q learning 及其类似, 这节内容会基于之前我们所讲的 Q learning. 所以还不熟悉 Q learning 的朋友们, 请前往我制作的 Q learning 简介 (知乎专栏). 我们会对比 Q learning, 来看看 Sarsa 是特殊在哪些方面. 和上次一样, 我们还是使用写作业和看电视这个例子. 没写完作业去看电视被打, 写完了作业有糖吃.

Sarsa 决策



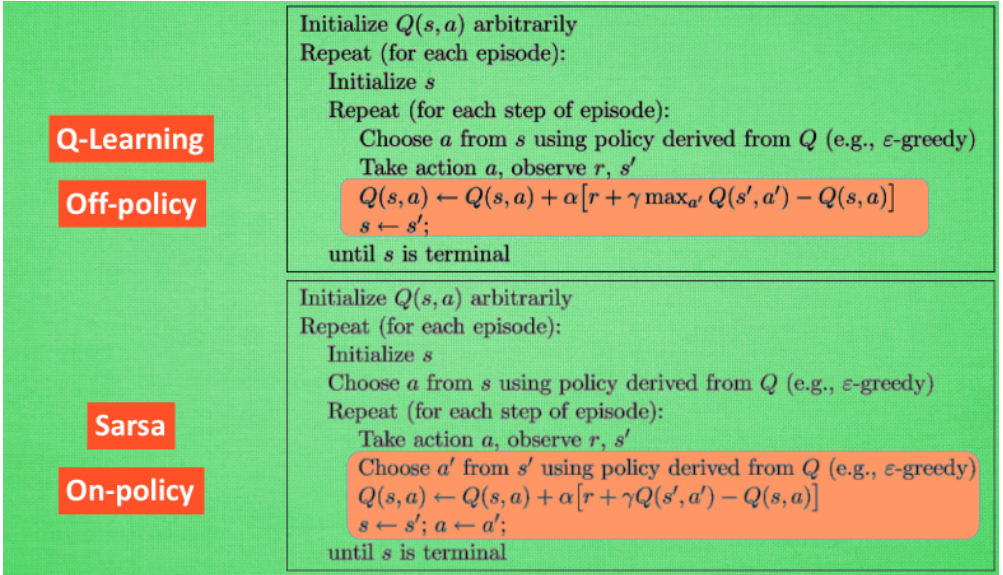
Sarsa 的决策部分和 Q-learning 一模一样, 因为我们使用的是 Q 表的形式决策, 所以我们会在 Q 表中挑选值较大的动作值施加在环境中来换取奖惩. 但是不同的地方在于 Sarsa 的更新方式是不一样的.

Sarsa 更新行为准则

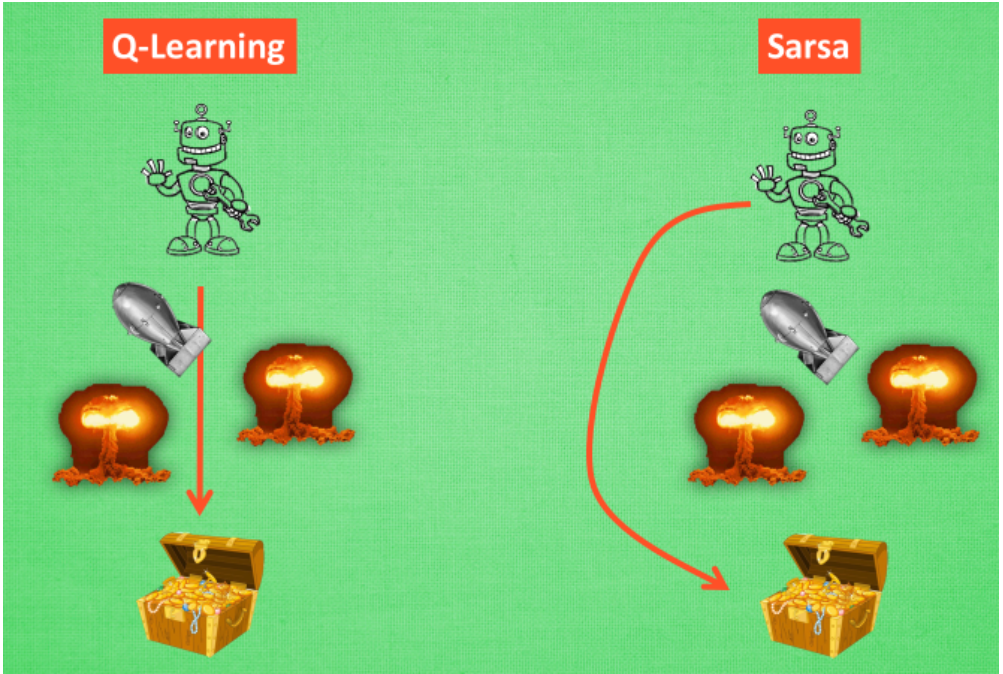


同样, 我们会经历正在写作业的状态 s1, 然后再挑选一个带来最大潜在奖励的动作 a2, 这样我们就到达了 继续写作业状态 s2, 而在这一步, 如果你用的是 Q learning, 你会观看一下在 s2 上选取哪一个动作会带来最大的奖励, 但是在真正要做决定时, 却不一定选取到那个带来最大奖励的动作, Q-learning 在这一步只是估计了一下接下来的动作值. 而 Sarsa 是实践派, 他说到做到, 在 s2 这一步估算的动作也是接下来要做的动作. 所以 Q(s1, a2) 现实的计算值, 我们也会稍稍改动, 去掉 maxQ, 取而代之的是在 s2 上我们实实在在选取的 a2 的 Q 值. 最后像 Q learning 一样, 求出现实和估计的差距 并更新 Q 表里的 Q(s1, a2).

对比 Sarsa 和 Q-learning 算法



从算法来看, 这就是他们两最大的不同之处了. 因为 Sarsa 是说到做到型, 所以我们也叫他 on-policy, 在线学习, 学着自己在做的事情. 而 Q learning 是说到但并不一定做到, 所以它也叫作 Off-policy, 离线学习. 而因为有了 maxQ, Q-learning 也是一个特别勇敢的算法.




为什么说它勇敢呢, 因为 Q learning 机器人 永远都会选择最近的一条通往成功的道路, 不管这条路会有多危险. 而 Sarsa 则是相当保守, 他会选择离危险远远的, 拿到宝藏是次要的, 保住自己的小命才是王道. 这就是使用 Sarsa 方法的不同之处.

如果你觉得这篇文章或视频对你的学习很有帮助, 请你也分享它, 让它能再次帮助到更多的需要学习的人.

莫烦没有正式的经济来源, 如果你也想支持 莫烦Python 并看到更好的教学内容, 请拉倒屏幕最下方, **赞助他一点点**, 作为鼓励他继续开源的动力.

« 上一个

下一个 »

 撰写评论

使用社交网站账户登录

或使用来必力便捷评论 

邮件

写评论

总评论数 5

按时间正序



Gary 2017年5月17日

感觉两种算法并没什么区别啊。。。

00



implicitUnit 2017年5月15日 · 已分享的SNS(1)

莫烦老师,您好,您的博文非常的好,希望您可以附上参考文献,这样可以方便我们进一步学习,不知可以吗?

100



莫烦Python 2017年5月16日

@implicitUnit 我在一些最新研发的技术上面附了一些文献, 我尝试附一下这些文献


00



cuix_Aga 2017年5月2日 · 已分享的SNS(1)

不明白最后那里对于 Q-learning 为什么很勇敢 , Sarsa为什么很保守

100



莫烦Python 2017年5月3日

@cuix_Aga 因为看公式的话, Q-learning 有 max 的操作, 他是理想化了未来的情况, 而 Sarsa 没有这没做, 他是一步步实打实的再更新

00

来必力是？

询问

支持 让教学变得更优秀

点我 赞助 莫烦

关注我的动向:

[Youtube频道](#) [优酷频道](#) [Github](#) [微博](#)

Email: morvanzhou@hotmail.com

© 2016 morvanzhou.github.io. All Rights Reserved