

有趣的机器学习

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

什么是 Sarsa(lambda) (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 »

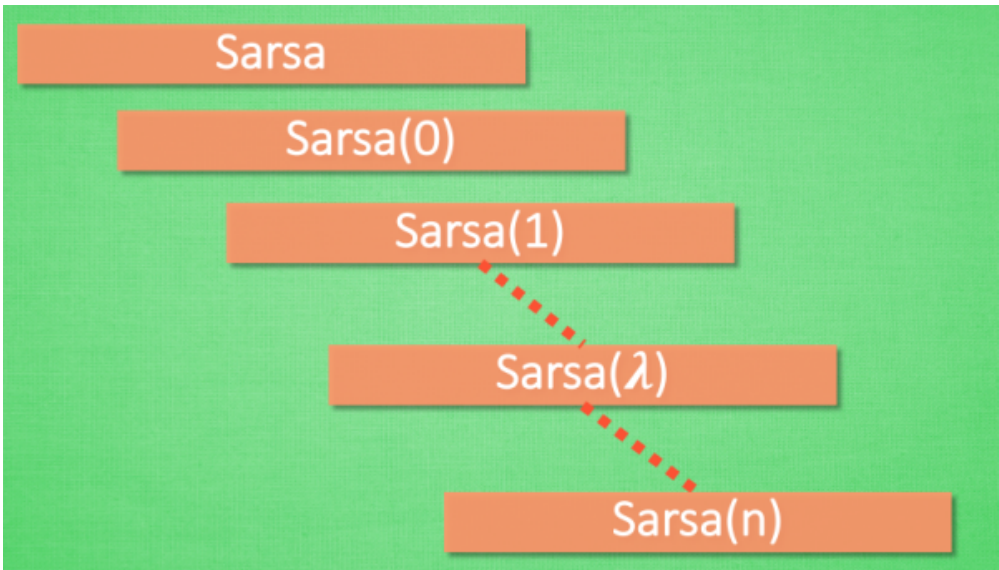
Sarsa(lambda)

- 学习资料:
 - 强化学习教程
 - 强化学习模拟程序
 - Sarsa 简介视频
 - Sarsa(lambda) 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基于 Sarsa 的一种提速方法, 叫做 sarsa-lambda.

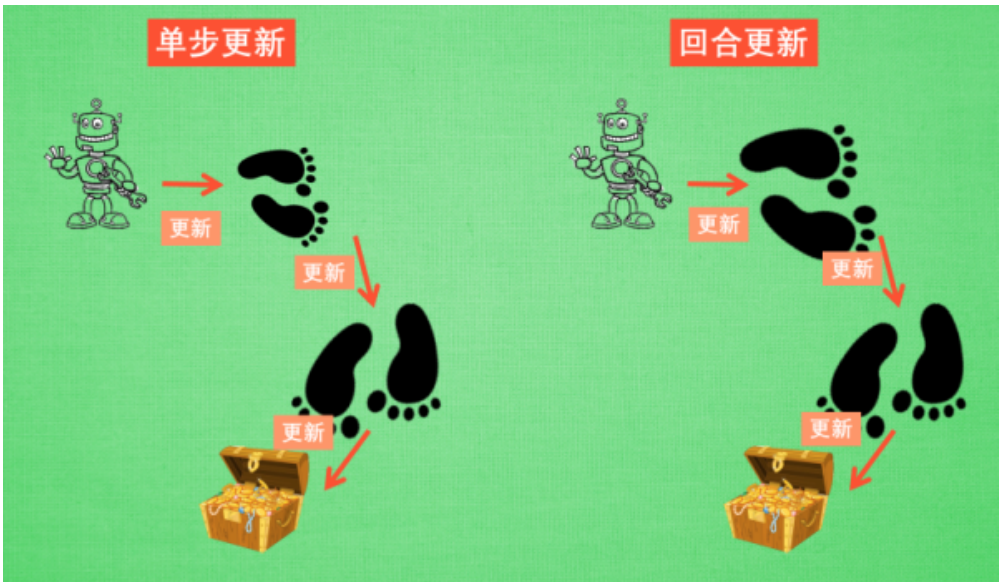
注: 本文不会涉及数学推导. 大家可以在很多其他地方找到优秀的数学推导文章.

Sarsa(n)

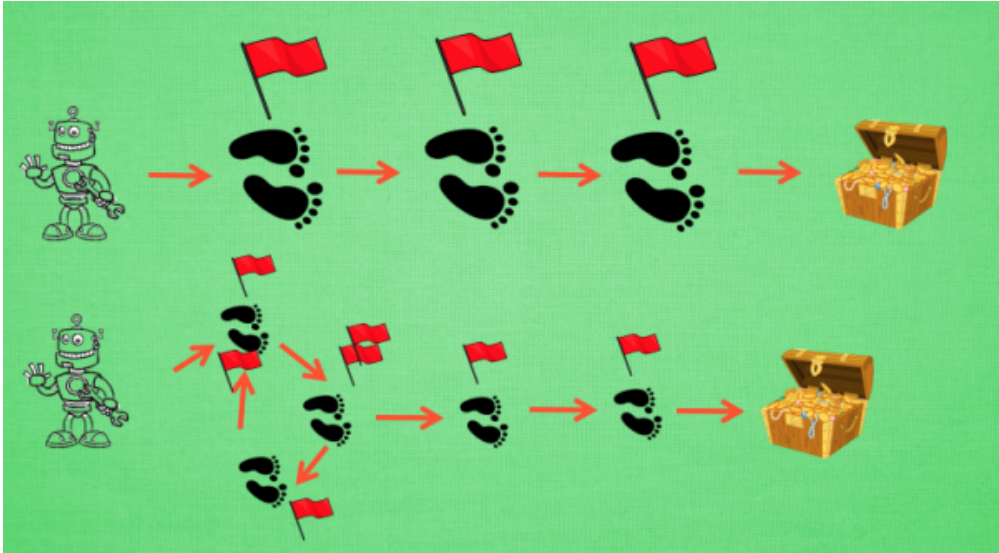


通过上个视频的介绍, 我们知道这个 Sarsa 的算法是一种在线学习法, on-policy. 但是这个 lambda 到底是什么. 其实吧, Sarsa 是一种单步更新法, 在环境中每走一步, 更新一次自己的行为准则, 我们可以在这样的 Sarsa 后面打一个括号, 说他是 Sarsa(0), 因为他等走完这一步以后直接更新行为准则. 如果延续这种想法, 走完这一步, 再走一步, 然后再更新, 我们可以叫他 Sarsa(1). 同理, 如果等待回合完毕我们一次性再更新呢, 比如这回合我们走了 n 步, 那我们就叫 Sarsa(n). 为了统一这样的流程, 我们就有了一个 lambda 值来代替我们想要选择的步数, 这也就是 Sarsa(lambda) 的由来. 我们看看最极端的两个例子, 对比单步更新和回合更新, 看看回合更新的优势在哪里.

单步更新, 回合更新对比

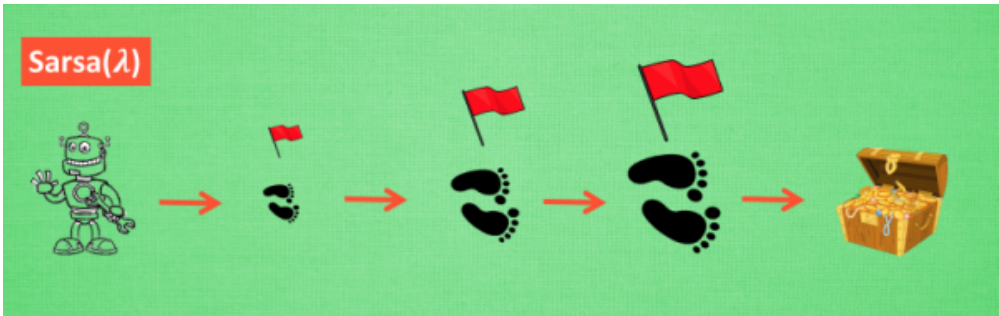


虽然我们每一步都在更新, 但是在没有获取宝藏的时候, 我们现在站着的这一步也没有得到任何更新, 也就是直到获取宝藏时, 我们才为获取到宝藏的上一步更新为: 这一步很好, 和获取宝藏是有关联的, 而之前为了获取宝藏所走的所有步都被认为和获取宝藏没关系. 回合更新虽然我要等到这回合结束, 才开始对本回合所经历的所有步都添加更新, 但是这所有的步都是和宝藏有关系的, 都是为了得到宝藏需要学习的步, 所以每一个脚印在下回合被选则的几率又高了一些. 在这种角度来看, 回合更新似乎会有效率一些.



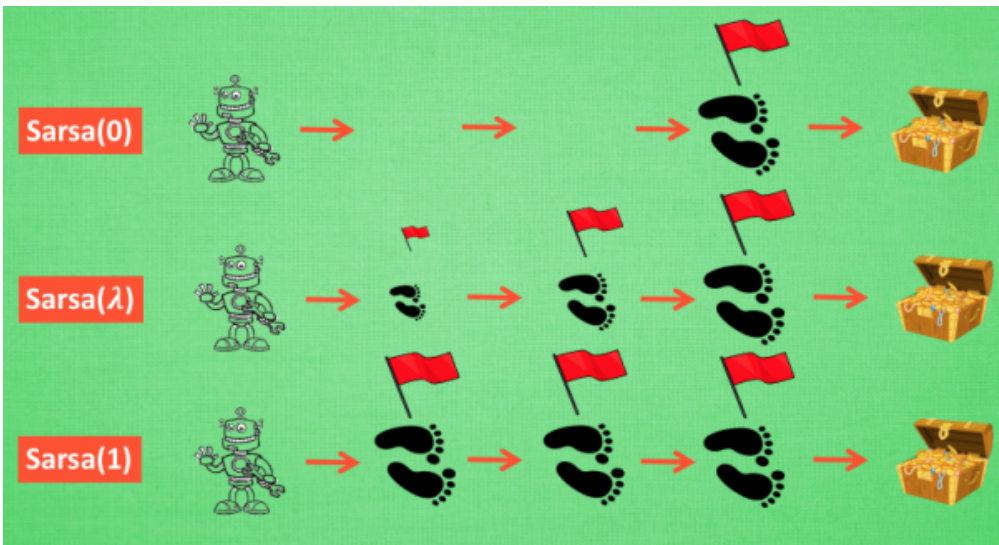
我们看看这种情况，还是使用单步更新的方法在每一步都进行更新，但是同时记下之前的寻宝之路。你可以想像，每走一步，插上一个小旗子，这样我们就能清楚的知道除了最近的一步，找到宝物时还需要更新哪些步了。不过，有时候情况可能没有这么乐观。开始的几次，因为完全没有头绪，我可能在原地打转了很久，然后才找到宝藏，那些重复的脚步真的对我拿到宝藏很有必要吗？答案我们都知道。所以Sarsa(lambda)就来拯救你啦。

Lambda 含义



其实 lambda 就是一个衰变值，他可以让你知道离奖励越远的步可能并不是让你最快拿到奖励的步，所以我们想象我们站在宝藏的位置，回头看看我们走过的寻宝之路，离宝藏越近的脚步越看得清，远处的脚印太渺小，我们都很难看清，那我们就索性记下离宝藏越近的脚步越重要，越需要被好好的更新。和之前我们提到过的 奖励衰减值 gamma 一样，lambda 是脚步衰减值，都是一个在 0 和 1 之间的数。

Lambda 取值



当 lambda 取0，就变成了 Sarsa 的单步更新，当 lambda 取 1，就变成了回合更新，对所有步更新的力度都是一样。当 lambda 在 0 和 1 之间，取值越大，离宝藏越近的步更新力度越大。这样我们就不用受限于单步更新的每次只能更新最近的一步，我们可以更有效率的更新所有相关步了。

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PYTHON

教程 ▼

大家说

赞助

About

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 »

正在载入来必力

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