

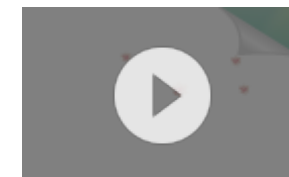


sina 新浪博客

故事：我在你的夜里

[新浪首页](#)[登录](#)[注册](#)

FelixLiu的博客

<http://blog.sina.com.cn/u/1959357162> [订阅] [手机订阅][首页](#) [博文目录](#) [图片](#) [关于我](#)

正文

字体大小：大 中 小

读"Reinforcement learning for Online Optimization of Banner Format and Delivery"

(2012-04-16 22:22:30)

转 载 ▼

标签： [在线学习](#) [多臂赌博机](#) [mdp](#) [计算广告学](#) [it](#) 分类： [计算广告学](#)

FelixLiu

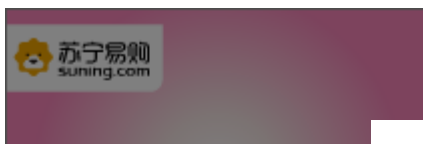
微博

这篇论文主要讲了两种在线学习的模型——马尔科夫决策过程(Markov Decision Processes)和多臂赌博机问题模型(Multi-Armed Bandit Problems)

通常验证两种不同形式的广告效果，可以通过A/B test的方式，通过实验，选择效果好的进行广告投放。这便是最原始的尝试-错误(trial-and-error)模型。对这个过程进行模型化，就可以利用在线学习模型来解决改问题。

本篇文章要解决的问题场景是：对于要投放的条幅(banner)广告，如何确定1)最恰当的广告形式 2)最合适的广告展现时间 3) 广告展现的最佳序列。这三个问题都可以抽象出相同的元素：状态(Stats),动作(actions)和回报(rewards)。在某个状态下，选择某个动作，可以跳转到另外一个状态，同时伴随回报。通过抽象出这三个元素，可以应用下面两个在线学习算法：

1. 马尔科夫决策过程(MDP)

[写留言](#)
[加关注](#)
[博客地图](#) / world map


相关博文

[上帝不小心打翻了调色板](#)
[呼毕勒汗](#)
[2小时不间断爆笑他们如何做到](#)
[老丑](#)
[《鸿门宴》重点字词与句子翻译学](#)
[高原的风中山](#)
[读"Reinforcement learning for Online Optimization of Banner Format and Delivery"_FelixLiu_新浪博客](#)

MDP可以定义为五元组(S; A; T; P_t; R_t)，其中S是状态的集合，A是动作的集合，T是时间序列；P_t是状态见转移的概率矩阵，R_t是状态转移过程中的汇报。在这个框架下，寻找最优策略就变成了寻找最大的累计回报问题。由于不知道P_t和R_t，因此，文章提出了一个Q-学习算法。通过不断的迭代，随机选出状态，动作，计算路径上的累加值，最后选出最大值即刻。

2. 多臂赌博机问题模型(MABP)

MABP相对于MDP要简单得多，而且容易操作和实现。该模型名字源于一个有多个手臂的赌博机，摇动其中某个手臂会有一定得到回报的概率，因此，需要决定的是每次摇哪个手臂。最常用的算法是最佳置信限界法(Upper Confidence Bound)：每次对每只可摇的手臂计算一个分值，该分值考虑已经知道的该手臂的回报和该手臂尝试的次数。公式如下： $p_j = x_j + \sqrt{2 * \ln(n) / n_j}$ ，其中， x_j 是手臂j的平均回报，n是已经摇动手臂的总数， n_j 是对第j个手臂摇动的次数。通过该公式可以知道，当一个手臂的回报很大(x_j 很大)时， p_j 有可能大；当一个手臂被摇动的次数少时(n_j 小)， p_j 有可能大。因此该公式兼顾了已知回报和未知的探测。

该类算法也很难通过实验来验证效果，因此文章中举了两个case study来对MDP和MABP进行说明。其中对MABP的验证主要是统计了再n轮迭代之后选择最优情况所占总情况的比例。

这篇文章算是计算广告中在线学习的基础文章了，文章中列举的方法也是学术上比较常规的方法。MABP模型很常用，尤其对于MABP的求解可以通过动态epsilon的算法，个人觉得可以得到更好的结果。

0

喜欢

0

赠金笔

[分享：](#)
[阅读\(896\)](#) | [评论 \(0\)](#) | [收藏\(0\)](#) | [转载\(1\)](#) | [喜欢▼](#) | [打印](#) | [举报](#)
[已投稿到：](#) [排行榜](#)
[前一篇：读"Click-Through Rate Estimation for Rare Events in Online Advertising"](#)
[后一篇：读"How Effective is Targeted Advertising?"的摘要和介绍](#)

儋州马拉松和黄埔马拉松

用户350958426

泛舟国王湖，聆听德国最美湖泊的

沙漠玫瑰_



平壤新制裁，也获俄罗斯

录

段子（那不是荡妇）

复深沉

看鹿晗和关晓彤的感情

闻你掏了

假牙的种类

不容易啊



百分15年回报容易

流水白菜

十月8日分析：节后还有反弹走

阿丁1970

更多>>

评论







重要提示：警惕虚假中奖信息

[发评论]

做第一个评论者吧！


抢沙发>>

发评论




更多>>


坚持不买你




股市



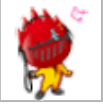
陈水旺



被裁了



贴你



登录名：

密码：

[找回密码](#) [注册](#) ☒ 记住登录状态

☐ 评论并转载此博文

发评论

以上网友发言只代表其个人观点，不代表新浪网的观点或立场。

< 前一篇

后一篇 >

读"Click-Through Rate Estimation for Rare Events in Online Advertising"

读"How Effective is Targeted Advertising?"的摘要和介绍



机多少钱

假牙的种类

¥1560.00

淘广告

推荐博文

[携程“年坑100亿”折射互联网](#)

[网盘衰落之后，为何这款私人云盘](#)

[短期调整或将仍有反复](#)

[人工智能与隐私保护](#)

[从“一人户”到“复式洗”，消费](#)

台湾科技挣扎，人祸大于天灾？

收入份额=市场份额，虎嗅想干什

传奇的谢幕，谈岩田聪和他的任天



轻松做之——培根香葱花

最惊艳流行的婚礼蛋

查看更多>>

假牙的种类



新浪BLOG意见反馈留言板 不良信息反馈 电话：4006900000 提示音后按1键（按当地市话标准计费） 欢迎批评指正

[新浪简介](#) | [About Sina](#) | [广告服务](#) | [联系我们](#) | [招聘信息](#) | [网站律师](#) | [SINA English](#) | [会员注册](#) | [产品答疑](#)

Copyright © 1996 - 2017 SINA Corporation, All Rights Reserved

新浪公司 版权所有