备案 (//beian.aliyun.com/) 登录 (//account.aliyun.com/login/login.htm?oauth_callback=https%3A%2F%2Fyq.aliyun.com%2Farticles%2F58011%3Futm_content%3Dm_5395%3F%26\t/t/acsou

控制台 (//home.coasthe.co



登录 (https://account.aliyun.com/login/login.htm?

from_type=当可以多oau博客Il/ankchttps%3尚客「%系tyq.aliwntenm/Actintiflation of 58性播音版的158性形象 158性形象 1

from_type=yqclub&oauth_callback=https%3A%2F%2Fyg.aliyun.com%2Farticles%2F58011%3Futm_content%3Dm_5395%3F%26utm_source%3Dqq%26do%3Dlogin) 论坛 (https://bbs.aliyun.com) 掌上云栖 (https://www.aliyun.com/app)

云栖社区 > 博客列表 (/articles) > 正文

机器学习理论研究方法探讨

talkless (/users/1934266582757530) 2016-07-25 22:40:52 浏览4048 评论3

机器学习 (/tags/type_blog-tagid_406/)



talkless (/users/19342

关注

2篇文章 (/users/193426 关注

本文译自: http://www.alexirpan.com/2016/07/17/ml-sleep.html (http://www.alexirpan.com/2016/07/17/ml-sleep.html), 原文标题《The Machine Learning Casino》,译文供您参考。

Google专家深度解读VR射摄、后期处理技术 (/articles/55956)

一、机器学习理论研究

机器学习是一种能使计算机半自主地对数据进行分析,并从中学习经验的算法。

机器学习理论的研究听上去就好像做这么一件事,让我们的研究人员去深刻探索,提高计算机学习数据集的效率的方法。



(https://promotion.aliyun.com/n

免费3个月?! 阿里云正式推战器免费使用服务, 助推企业IT:

事实上,机器学习理论的研究可以在我们的生活的每时每刻找到影子。就好像你在不断的进行一场赌博,只不过你的筹码是你的时间。你把你的时间花费在了一些有启发性的事物上去,而这场赌博的最终大奖看上去就是未来光明的前途。当然,即使你付出了大量的时间,你也可能遇到你无法想象而又无法理解的失败。

二、机器学习中理论与实践的关系

尽管机器学习是建立在统计学的基础之上的,但事实上机器学习也可以被更多的说成是一门实践性的科学。

然而谈到到实践性,这并不意味着这是一门轻视基础理论的学科。在机器学习里会遇到众多理论问题。例如,赌博机问题, 凸优化与非凸优化,图形建模,以及信息理论的问题,当中涉及到了许多不同的学科领域。如果你要你深究机器学习,你可以在这些问题里花花时间。

这里面内容太深太广,对于普通的开发者来说,学完全部的内容并不是一件容易的事情。现实就是,人们都希望自己的算法是有保障的,但是不是每个人都愿意花时间去深究为什么他们的算法是有保障的。

好吧,在当下在APP开发火爆的环境下,人工智能的宣传满天飞。但是人们往往首先关心的只是最终的产品,然后在开发进度中逐渐才会关注基础理论的重要性。一般的开发者往往都是会从一些有启发的事情中获得灵感,通常来说,这些有启发的算法能对既已出现的情况做出相应的反应,但是这些算法仅停留在面向过程的层面,基本没有更深的内涵。

三、实验方法

那么我们想想, 我们到底是怎样发现这些有启发性的事物的呢? 这里介绍一个方法:

- 一、提出假设。
- 二、设计一个实验来检验这一假设。
- 三、进行实验并分析结果。
- 四、细化假设并改进实验设计。
- 五、重复以上流程, 直到最终得出结论

在机器学习的领域,提出假设是有前途的算法,进行实验便能检验这些算法的正确性。

重点是什么?就是做做实验,直到最后你的理论被最终证明有效。

好, 回过来, 接下来我和大家分享一些有用的东西。

四、机器学习中的理论是否已经非常成熟了呢

当前的这门艺术的状态还是概率性的并且高度个性化的。经验上来说,概率化的方法现在已经能很好的在大数据领域等领域应用。

个性化意味着你能尝试调试不同参数并观察他们之间的差别,你有没有尝试调整过超参数?有没有尝试过清洗你的数据?或者使用一种不同的最优算法?让你的模型变得更简洁?或者说让你的模型更加复杂化?玩过Batch Normaliztion算法吗?校正过非线性失真的图像吗?我们希望能探索出一些新的方法来打破常规来解决一些实际问题。然而不幸的是,我们目前的工作还没能力达到那个程度。神经网络确实是非常神奇的,但是你一旦用它玩完那些简单的问题后,你会发现其实还有很多问题还不能解决。

机器学习中会设置一些特征参数,而且这些我们也不会说哪些参数是失败参数。这样的事可能是在其他学科中不会出现的。但是在机器学习中,事情就是这样。而且有时,完全相同的参数设定也不会得到相同的结果。你可以认为我们能做的更好,但是难以重现结果确实是一个问题。

对每一个欣赏机器证明艺术的人来说,这也是机器学习中最气死人的地方。在我们进行超参调优的时候,我们往往并没有带着专门去探索其理论基础的学习动机。但这却是你们必须做的事情,机器学习之美和现实应用有时会脱节。这里有一个段子,是一个计算机视觉教授和他的学生说的:

学生: 这理论没有什么实际应用。

教授: 真的没有用吗, 就算真的没有什么实际应用, 理论本身就已经是艺术。

学生: 你说的我都懂,理论很高端,但却在实际生活中没有用。甚至连Lenna图像也不能很好的处理。

教授: (半开玩笑的说)也许我们处理一百万个图像,可能在这个过程中,我们就会发现机器学习的作用。

学生: 如果连Lenna都不能处理, 就更别说无数个Lenna了。

教授: 哈哈哈, 我也这么认为, 真是倒霉啊。

我能体会他们的苦衷。

在训练了足够的机器学习模型后,你就找到问题的关键,能识别常见的无效模式。当你能正确进行模式识别,你便能开始总结出新的识别方法。然而,这种识别没法达到保证百分百成功的程度。我喜欢开玩笑的说,有一天理论学家会找到比纯粹靠经验的好方法。但是我不认为在当下能很快实现。因为这些理论上的东西真的比较难。

理论上,在监督机器学习中,已经发现了"No Free Lunch"定律。我如果这么说可能会显得不太正式,但是确实没有哪一种算法能在每一个可能的问题里打败另外一种算法。换句话说,永远不会有一个算法统治所有其他算法。当然,如果有人能证伪这一点,那当然是最好不过的事了。

五、如何在实验中获取有价值的结论

我依旧没有解释为什么机器学习的研究能应用到我们的生活当中。

我认为我已经迂回地说过了。机器学习实验的是随机性的。即使是这个领域内的大牛也不能掠过超参调优的过程。

这就让这个领域让人感觉像一个巨大的赌场,你拉动了水果机的拉杆,希望自己能中奖。有的时候,你确实能中彩,但是也不会中奖。 或许凑巧的是,有人告诉你玩的这台水果机已经十年没有开奖了,你也许应该尝试新的水果机。机器学习就好像 民间传说,这里面的窍门和门道人们一代又一代相传。 我们理解了许多知识,但是水果机就是是水果机,这里面不可避免的随机性问题随随便便就能毁了你美好的一天。或者有更糟的情况,这样的"水果机"也许就隐藏在你的工作里,你的股票基金里,你的前途里。

在这样的机器学习的游戏里,要么你很幸运,要么你在尝试了许多次后你必将获得幸运。这里面唯一保证成功的途经就是做这个后者。

那就意味着实验,你需要做海量的实验。其实在进行实验的这段时间你也不需要做什么事情。你在你准备休息的时候运行实验。比如说,你准备去吃中饭。那就开始在计算机上运行你的实验吧。看看你吃完后会不会有什么新的实验结果。又或者说你准备出去玩,那你或许可以整夜的运行你的实验,在第二天再去查看你的实验结果。再比如,你不想在这个周末工作?那好吧,你的电脑可不会介意在周末上班的。我们写好代码,我们便能在无人值守的情况下进行实验,这真的是非常的棒。只要你的代码是有效的,如果你的代码出现了问题,那么好吧你需要修复出现的bug。除去修复代码的时间,你可以每时每刻都在运行你的实验。

计算机的工作就像是一个工作狂一样。你也不需要立即修改你的代码,就像你不需要一次把所有的筹码都买满水果机的下注。我很确信随着时间的增长,你最终会变得十分幸运,即使你错过了一两天,这也不会影响最终的结果。

如果有一种方法能让机器算法能有效。我也不知道会不会真的有这样的算法,我在心里还曾想过那么一秒,我如果在这个满月下献祭一只山羊,会不会就能让这项工作有效了。如果机器学习的算法真的这么简单就能有效了,我想应该有人会想出无数的方法想着怎么去献祭山羊了。

不过好事不会像山羊献祭这么简单的,因我也不想做这种没有什么意义的事情。

六、结语

在这里,你可能会想知道,为什么我有时会对机器学习感到有些厌烦。

事实上,所有的废话也会有一定的价值,实验的运行中有许多令人激动的事情在不断发生,因此,我对机器学习中出现的问题也越来越能容忍,我对此也感到习以为常了。和理论相比,运行试验就是浪费时间,但这是令人激动的垃圾时间。

我认识的计算机理论科学的朋友们可能会认为我已经疯了, 我正做的这些实验简直让我近乎疯狂! 哦, 好吧还有什么新理论吗。

我们的这些工作,处于获取真理的边缘。如果这样繁复的工作就是我们必须要付出的代价,我必然会毫不犹豫地会选择付出。 出。

"标枪远远领先于她,并移动的更快。殖民者还有很多许多时间来进行活动。当她到达天狼星的时候也许会在那儿发生什么,也许会十分友好。就 算真的遇到不友好的事情,她也能到时候再随机应变。"

——摘自小说《RA》

作者: alexirpan 译者: 语林

更多详细内容,请您点击这里:阅读原文 (http://www.alexirpan.com/2016/07/17/ml-sleep.html)



【云栖快讯】云栖聚能聊,开启技术交流新体验!云栖社区特邀请到封神、义从、戒空、沧龙、坤健、刘光利等众位技术专家,与广大开发者 进入技术话题交流,更有专家打赏礼品等你抢哦~ 详情请点击 (https://yq.aliyun.com/activity/120)

评论文章(3)

凸(15) ☆(0)



(http://service.weibo. title=%E6%9C%BA%E5% sleep.html&url=https:

上一篇: Google专家深度解读VR影片拍摄、后期处理技术 (/articles/55956)

相关文章

数据-缓存漫谈 (/articles/57754)

优云软件数据专家最佳实践:数据挖掘与运维分析 (/articles/56834)

一文读懂机器学习, 大数据/自然语言处理/算法全有了...... (/articles/382...

网友评论

模掌618 (/users/1651667003421861)

2016-07-29 11:00:20

这个感觉很套路

(/users/1651667003421861) 1F

(来自社区APP) (https://promotion.aliyun.com/ntms/mobile.html)

160 **F**0



韩韩天下 (/users/1423360508351147)

2016-08-03 17:23:10

(/users/1423360508351147) 2F

看了文章,但是还有好多疑问,机器学习到底是不是谷歌宣传的那样,完全是以数据为基础,只要有足够 量的数据,机器学习就能发现数据本身存在的规律,第二,机器学习是不是不应该被人干预,让是自己去 学习,找规律,得出参数,机器学习的终极是不是完全拜托人的干预,自己做事情,这些问题没看到正面

(来自社区APP) (https://promotion.aliyun.com/ntms/mobile.html)

160 **F**0



Iollipopcc (/users/1285202453684677)

2016-08-12 05:07:49

调参就是一门玄学。

(/users/1285202453684677) 3F

(来自社区APP) (https://promotion.aliyun.com/ntms/mobile.html)

160 ₽0

登录后可评论,请登录 (https://account.aliyun.com/login/login.htm?

评论

搜索

■ 中国站



售前咨询

) pe=yqclub&oauth_callback=https%3A%2F%2Fyq.aliyun.com%2Farticles%2F58011%3Futm_content%3Dm_5395%3F%26utm_source%3Dqq%26do%3Dlogin)

全方位的购买咨询 (https://www.aliyun.com/support/95187) 精准的配置推荐 (https://www.aliyun.com/support/95187)

友情链接 (//www.aliyun.com/links)

阿里巴巴集团 (http://www.alibabagroup.com/cn/global/home) 淘宝网 (//www.taobao.com/) 天猫 (//www.tmall.com/)

聚划算 (//ju.taobao.com/) 全球速卖通 (http://www.aliexpress.com/) 阿里巴巴国际交易市场 (http://www.alibaba.com/)

1688 (http://www.1688.com/) 阿里妈妈 (http://www.alimama.com/index.htm) 阿里旅行去啊 (//www.alitrip.com/) 阿里云计算 (http://www.aliyun.com/) YunOS (http://www.yunc阿里通信 (//aliqin.tmall.com/) 万网 (http://wanwang.aliyun.com/) 高德 (http://www.autonavi.com/) UC (http://www.uc.cn/) 友盟 (http://www.umeng.com/) 虾米 (http://www.x天天动听 (http://www.ttpod.com/) 来往 (http://www.laiwang.com/) 钉钉 (http://www.dingtalk.com/?lwfrom=20150205111943449) 支付宝 (https://www.alipay.com/)

© 2009-2016 Aliyun.com 版权所有 ICP证: 浙B2-20080101





(mailto