



科研选题

科研选题

大家之前选择的题目都非常好，相信大部分问题也都是同学们感兴趣的。

但是请大家注意一个问题：你们能不能找到数据完成自己的科研题目呢？

科研选题

对于本课程来说，我们希望大家在结课时基于数据分析完成一篇学术论文。这里，我们可以换一种思路：

之前，我们是先有问题，再根据问题找可以回答问题的数据。
但现在，我们可以先找到数据，再根据数据找到有意思的问题。

比如，拿到数据后，我们可以看看哪些变量可以当做Y，哪些变量可以当做X，并根据数据进行简单的分析。

这样有助于同学们快速完成一篇学术论文！

一些有趣的问题

今天，我们讲给大家将一些大数据时代有趣的现象和有趣的问题。这些问题都是基于最顶尖的学术研究，为我们带来别样的启发，也告诉我们，有时候事情并不像我们想象的那么简单。

1. 企业收集的数据越多越好吗？
2. 如何做一名成功的红娘？
3. 数字企业应该打击盗版吗？
4. 特斯拉为什么要公开自己的专利？
5. 为什么我们害怕搭乘飞机？



数据保护：消费者能做什么



这是神马？



这是一个手机摇步器!

为什么有人愿意花钱买手机摇步器?



Matthew Brennan @mbrennanchina · May 14, 2019

Chinese phone cradle for boosting your phone's daily step count. Some insurance companies in China allow people who consistently reach a certain daily step count to get discounted health insurance premiums.



Lê Nguyễn Hoàng (Science4All) @le_science4all · May 14, 2019

Replying to @mbrennanchina

Aka adversarial poisoning attack on data-driven algorithms...



Graeme Douglas ✓ @graemedouglas · May 14, 2019

Replying to @mbrennanchina and @BBHLabs

Not just China - Vitality does this in the U.K.



用户与企业的对抗

我们之前已经讲到，企业会使用种种手段从用户手中获得数据，并利用这些数据对消费者实行“价格歧视”：比如，网络平台Orbitz会向苹果用户收取一个比安卓用户更高的价格。

虽然企业非常狡诈，但消费者并不是任人宰割：很多时候，消费者可以通过自己的手段对抗企业的大数据算法，满足自己的利益！

我们刚才所说的手机计步器就是用户对抗大数据算法的一个例子。

问题

除了刚才的手机摇步器的例子，你还能想到哪些用户战胜算法的例子？

你自己用过哪些手段对抗企业的大数据算法？

战胜算法举例

很多网站根据根据用户的IP地址向不同的用户展示不同的价格：一个纽约用户看到的价格很可能和一个孟买用户看到的价格是不同的。用户可以很容易的通过VPN和代理服务器修改自己的IP地址，使得自己可以获得更理想的价格。

战胜算法举例

还有很多企业根据你的设备为不同的用户定价。如果你用的是最新版的苹果手机或者其他的高端手机品牌，恭喜你，你将享受一个更高的价格。相反，如果你用的是大众品牌手机，那么你的价格更低。

但是，使用高端手机的你，一定不肯付出更高的价格。那么，你可以使用下面的插入程序(Plug-in): User Agent Switcher.

User-Agent Switcher for Chrome

Offered by: google.com

★★★★★ 2,462 | [Developer Tools](#) | 👤 1,000,000+ users

 By Google

战胜算法举例

打车软件Uber会根据乘客的目的地为不同的乘客进行个性化定价：即使路况，历程，举例相似，如果你的目的地是高档小区，那么你的价格更高；如果你的目的地是贫民窟，那么你的价格更低。那么Uber用户可以做什么呢？

聪明的用户想到了战胜Uber算法的方法：我们在打车时可以先输入一个假的目的地（贫民窟），当车开到一半的时候，在中途变换目的地，可以使我们享受更低的价格。

此外，一些Uber用户还会在没事的时候使用打车软件，然后拒绝软件的报价。做的次数多了，软件会认为用户的购买力不强，会在未来给用户降价。

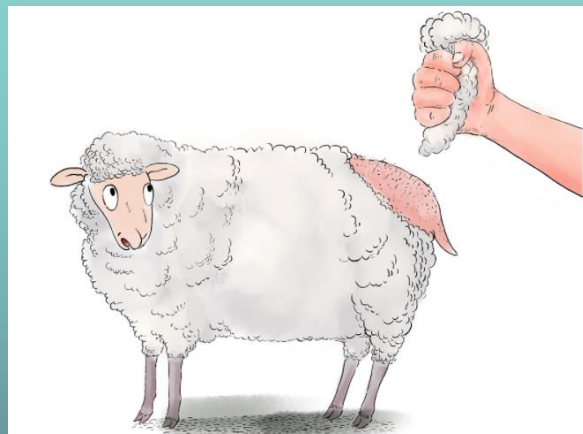
Home › Blog › All › 12 ways people trick algorithms for personal gain

12 ways people trick algorithms for personal gain

How to Outsmart Algorithms and Take Control of Your Information Diet

Money › Spend & Save

How to beat the retailers' algorithm



企业利用大数据“割韭菜”，而用户则用自己的方法“薅羊毛”
割韭菜与薅羊毛是天生的一对矛盾



我是大熊猫宝宝！

启示

为什么用户需要开发种种手段战胜算法？

因为企业用大数据方法为消费者画像，并利用大数据为消费者进行个性化定价。而没有用户愿意接受高价。

什么时候用户更愿意战胜算法？

当企业大量收集用户数据的时候。当企业的数据越多，就越有能力对消费者进行个性化定价，也就意味着用户有更大的动机采取种种手段战胜算法。

数据越多越好吗？

我们的直觉认为，数据往往是越多越好。现实生活中，企业也往往用尽手段采集用户的数据，并通过数据剥削他们的消费者。但数据真的是越多越好吗？

其实不尽然。收集更多的数据意味着：

1. 企业可以分析更多的消费者
2. 用户一旦知道企业收集很多数据，他们更加担心企业对剥削自己，因此更有动力采取手段篡改自己的数据，因此企业的数据质量反而更低了！

企业需要平衡数据的数量和数据的质量！更多的数据意味着更劣质的数据，对企业并不一定意味着是一件好事。

对现实的启发

现在很多互联网金融机构大量收集用户的数据并以此作为为用户发放贷款的依据。你在网上的言论，你在社交媒体的互动，你的购物行为，你的资金流水都会影响你的贷款状态。很多用户为了获得贷款，会花大量精力“洗白”自己，也就是我们说的篡改数据。

与其逼着用户篡改数据，企业不妨后退一步，少收集一些数据，用户也不必花大量的时间精力去洗白自己。这样，我们会得到双赢的局面！



红娘的烦恼

问题

假如你是一名职业红娘（或者媒人），你觉得怎么才能赚到钱？



婚恋网站的生意

传统上，媒婆通过自己的经验撮合男女，为男性或女性挑选合适的对象，并从中获利。媒婆的成功与否取决于自己有多少资源，还有自己能不能准确判断适合客户的伴侣。

前一种资源，我们可以理解为媒婆手中的“数据”
后一种资源，我们可以理解为媒婆手中的“算法”，即数据分析的能力

那婚恋网站做的又是什么样的生意呢？

婚恋网站的生意

婚恋网站一手有大量的数据（数万名的适龄男女），因此数据量远超传统媒婆。

同时，几乎所有主流的婚恋网站都通过数据挖掘技术分析如何才能最高效的匹配男女。例如，男女身高差多大婚配成功的几率最大？男女收入相差多大婚配成功的概率最大？男女年龄相差多大婚配的成功几率最大？哪些性格最相配？哪些家庭背景最相配？等等等等。

因为婚恋网站有大量的数据，因此分析数据，建立算法不成问题。

婚恋网站的生意

我们的直觉是，当婚恋网站的算法越准确，那么用户更容易在婚恋网站实现配对，这样婚恋网站的效益越高，用户的满意度也越高。

但事实是这样吗？

国内婚恋网站为什么这么不靠谱



haichuan



婚恋网站的生意

其实这是婚恋网站刻意为之的：它们不是没有技术实现更好的匹配，而是好的匹配不利于互联网站的发展。

更好的匹配意味着用户能够更快的找到自己理想的配偶。而一旦一个用户找到了理想的配偶，这个用户就会离开婚恋网站，不再进行消费。因此，婚恋网站将永久性的失去这个用户。

相反，如果婚恋网站的算法不够准确，那么用户需要更多的时间才能够找到理想的配偶。用户逗留的时间越长，意味着更多的消费和更高的利润。这就是“红娘的烦恼” (Matchmakers' Dilemma).

Matchmaking is a complex process that requires considerable expertise. Matchmakers in various industries often advertise proprietary technologies that presumably help users find an ideal match in a short time. However, matchmakers may have incentives to provide suboptimal matchmaking services so that users remain clients longer and pay more fees. This article considers a matchmaking market with network effects and strategic consumers and analyzes under what conditions matchmakers would offer more effective versus less effective matchmaking services. The authors find that stronger pricing power paradoxically leads to lower technology provision when consumers have high valuation for the matchmaking service. Moreover, network effects typically encourage matchmakers to retain the users in the market to create positive externalities, which can result in less precise matchmaking. In addition, consumer patience prompts competing matchmakers to implement ineffective technology. In two extensions, the authors explore asymmetric two-sided markets and discuss the impacts of alternative pricing schemes on technology provision.

Keywords: matchmaking, competition, customer lifetime, platform

Online Supplement: <http://dx.doi.org/10.1509/jmr.16.0423>

Matchmaker Competition and Technology Provision

核酸检测的生意

假如你是核酸检测机构，你根据核酸检测的次数收费。那么，你会如何开发你的核酸检测包呢？

如果你的核酸检测太不准确，那么没人会用你的服务，你根本赚不到钱。

但如果你的核酸检测太准确，那么很快疫情就会被控制，这样下来，你只能做一次生意，以后再也赚不到钱，也不是一笔好买卖。

因此，对核酸机构来说，一个“凑合”的核酸检测可能是最划算的，即使他们有能力开发一个完美的核酸检测包！

核酸检测的生意

7天7次核酸检测后才确诊 检测试剂怎么了？

“石家庄有人检测6次才呈阳性”，核酸检测是不是不灵了？

大量病例经多次核酸检测才呈阳性，病毒变狡猾了吗？

药厂的生意

药厂开发药物的时候，同样面临上述的抉择。我们以乙肝为例：

如果药厂可以开发一款高质量的乙肝药物，一次性彻底的治愈乙肝，那么病人只会消费一次，药厂只能做一笔买卖。

相反，如果能够开发一种能够控制乙肝但是不能治愈乙肝的药物，那么病人就会成为药厂的“回头客”，反而可以为药厂创造更高的利润！

这也是为什么很多疾病无法得到彻底解决的原因。

2018年4月10日，著名投资银行高盛（Goldman Sachs）发布了一份针对生物技术领域的调查报告，这份名为《基因组革命》的医药产业调查报告却引发了轩然大波。高盛分析师公开表示，快速甚至是一次性治愈患者的商业模式并不利于长期利润，比如说基因疗法。

一般情况下来说，对于制药企业而言，生产药品的根本目的究竟是什么？大家肯定会说治病救人，医者父母心啊？然而实际情况并非如此，在医疗药品行业中，药品的目的是为了治疗疾病不假，但完全治愈疾病以及缓解部分症状却不是制药的核心。因为药厂也是企业，他们也需要盈利，那么盈利的关键就在于自己的药能不能成为一种：可持续的商业模式。

举例来说，“基因疗法、细胞基因工程疗法和基因编辑最具吸引力的因素在于能提供‘一次性治疗方法’，然而，这种疗法在复发性疾病与慢性疾病方面有着非常不同的前景，”高盛分析师Salveen Richter在报告中写道。“虽然这个创想对患者和社会将带来巨大价值，但对于追求可持续商业利润为目标的基因药物开发商来说，这可能是一个挑战。”这个“挑战”就在于，快速甚至一次性治愈患者，医药公司将难以保证自己的利润，因为市场总体趋势将会是萎缩的。

古人的智慧

狡兔死、走狗烹；飞鸟尽、良弓藏

——《史记》

养寇自重

——《资治通鉴》



企业应该打击盗版吗？

盗版产品

数字经济时代，盗版产品越来越容易生产。比如，程序员经常破解各种软件，我们甚至可以在淘宝上以很低的价格获得软件的使用权。而在音乐，游戏，艺术品等领域，盗版更是层出不穷。

那么企业是不是应该尽最大的努力保护自己的知识产权，反击盗版呢？

盗版产品

其实不然。大部分企业对盗版的打击都很宽松，对盗版产品的使用往往是睁一只眼闭一只眼，甚至纵容盗版产品的流通。

你周围有没有朋友使用盗版产品呢？

你周围有没有人因为使用盗版产品被起诉呢？

问题：为什么企业会纵容盗版产品的流行，而不是打击盗版产品？

网络效应

大部分数字产品都具有网络效应(Network effect): 它是指一个产品的用户数量越多, 那么这个产品对用户价值就越大。

用微信的人越多, 那么微信对每个人个价值都更大。如果你没有朋友使用微信, 那么你也大概不会自己使用微信。

玩游戏的人越多, 那么这个游戏对每个玩家的价值也更大。你可以向朋友吹嘘交流自己打游戏的成果。而如果只有你自己玩游戏, 那么你大概也觉得没什么意思。

甚至对于PPT, 只有很多人都用PPT的时候, 它才有价值。如果只有你一个人用PPT, 那么你做了PPT给谁看呢? 同样, 大部分软件都面临这一问题。

网络效应与打击盗版

但是，网络效应和打击盗版有什么关系？

打击盗版对数字产品的价值有两个影响：

1. 如果打击盗版，很多用户被迫使用正版软件，企业将卖出更多的正版软件。
2. 如果打击盗版，很多用户无法接受正版软件的价格，会干脆不用软件或者用其他品牌的软件，这样，用户总量变少了，因为网络效应，软件的价值也会变小，正版软件也卖不出去多少钱。

企业必须平衡上述两点，既不能完全不打击盗版，也不能彻底打击盗版，让盗版软件没有生存空间。

网络效应与打击盗版

现实生活中，企业往往做的更聪明。

对于那些没有什么消费能力的用户，比如学生，发展中国家的用户等，企业彻底纵容他们使用盗版，甚至变相的鼓励他们使用盗版。这也是为什么我们周围没有人因为使用盗版被起诉。

但是，对于那些有消费能力的用户，比如大企业员工，企业往往对他们使用盗版非常敏感。如果你在跨国公司用盗版软件，那么你大概率会被起诉并且赔的倾家荡产。

这种做大对企业是最优的！

微软纵容盗版？你不懂真正的微软，出手刀刀见血

如果让你用一个词来描述一下微软公司，你会想到什么词呢？或许是仁慈？也许吧，可能在大部分人的印象里，微软是一个良心企业，这么多年来，大家都在用人家的产品，操作系统，从XP到WIN7再到现在的WIN10，office办公软件，却极少有人为此花过钱，都是免费下载、安装、再用软件一键激活搞定，微软对此也完全是默许的态度，真的是非常仁慈了。



特斯拉为什么开放专利?

特斯拉的举动

“我们所有的专利都属于你。”这是特斯拉创始人Elon Musk在2014年6月12日发布的一篇官方博客的标题。对于一家技术公司来说，这是个让人印象深刻的表白。

在这篇博客中，Musk表示公司将开放所有专利技术，不会向使用其专利技术的公司提出侵权诉讼。当然，特斯拉对开放的目标有一个模糊的界定—善意使用其专利的公司。

All Our Patent Are Belong To You

Elon Musk, CEO • June 12, 2014

Yesterday, there was a wall of Tesla patents in the lobby of our Palo Alto headquarters. That is no longer the case. They have been removed, in the spirit of the open source movement, for the advancement of electric vehicle technology.

Tesla Motors was created to accelerate the advent of sustainable transport. If we clear a path to the creation of compelling electric vehicles, but then lay intellectual property landmines behind us to inhibit others, we are acting in a manner contrary to that goal. Tesla will not initiate patent lawsuits against anyone who, in good faith, wants to use our technology.

When I started out with my first company, Zip2, I thought patents were a good thing and worked hard to obtain them. And maybe they were good long ago, but too often these days they serve merely to stifle progress, entrench the positions of giant corporations and enrich those in the legal profession, rather than the actual inventors. After Zip2, when I realized that receiving a patent really just meant that you bought a lottery ticket to a lawsuit, I avoided them whenever possible.

特斯拉的考虑



特斯拉开放专利有着种种考虑。但是，网络效应仍然是关键的一点。

当特斯拉开放自己的快速充电专利时，不仅特斯拉，其他玩家也会进入快速充电的市场，这样，市场上的充电桩会迅速变多。而当充电桩变多时，用户也会发现驾驶电动车越来越方便了。当更多的用户进入电动车市场的时候，特斯拉的销量也会增加。

特斯拉的考虑

除了刚才说的，特斯拉还有其他的考虑。

公布专利，还能帮助特斯拉打击竞争对手！

为什么这么说？我们假设有一个创业公司，开发了一款更加先进的电动车，他的充电效率更高。但是，但特斯拉公布了自己的专利后，特斯拉的技术已经占据了整个市场，创业公司难以进入市场获取利润，因此，即使创业公司的技术更先进，也很难在残酷的市场竞争中生存。

蚊帐的例子

我们知道，非洲大陆疟疾横行，严重危及人类的健康。而应对疟疾传播的一个重要的重要手段是使用蚊帐。



蚊帐的例子

后来，一些慈善机构向贫穷的非洲人捐献了一批蚊帐。

但这却带来了意想不到的后果：本来，当地有一些小的公司聘用手工业者生产蚊帐，能够维持自己的生存。但是，当捐献的免费蚊帐达到非洲后，这些小作坊无法继续盈利，继而破产，手工业者失业。

而当慈善机构离开之后，非洲既没有了蚊帐，也没有了生产蚊帐的小作坊。经济和健康受到了双重的打击！

特斯拉是不是一样的想法呢？



为什么我们怕坐飞机？

飞机焦虑

2022年3月21日14时38分许，一架东航波音737-800客机在广西壮族自治区梧州市藤县埌南镇莫埌村神塘表附近山林坠毁，机上132人全部遇难。这次空难激发了大家对航空安全的大讨论。



飞机焦虑

然而大量的统计数据表明，其实飞机是我们这个时代最安全的交通工具：

对于每十亿公里的旅行，预期的死亡人数是：

摩托车108.9人，步行54.2人，自行车44.6人，汽车3.1人，轮船2.6人，铁路0.6人，航空0.05人。可见，乘坐飞机往往比搭乘其他交通工具更加安全。

但事实上，往往大家更担心乘坐飞机。

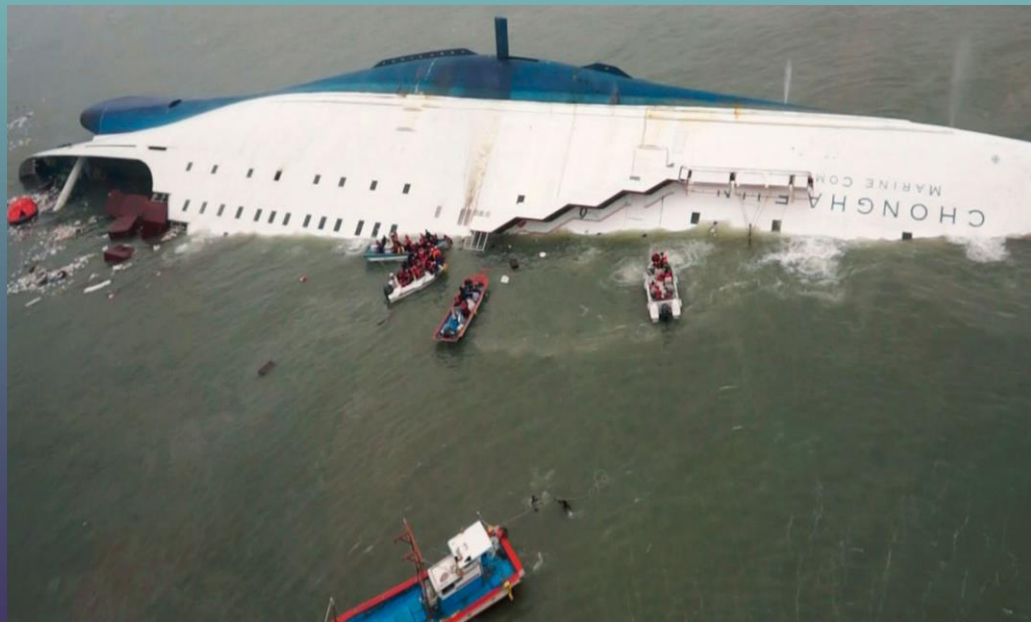
飞机焦虑

其实，其他运输工具也会带来大量的伤亡，例如2011年的温州高铁撞车事故，导致了40名乘客死亡：



飞机焦虑

其实，其他运输工具也会带来大量的伤亡，例如2014年的韩国世越号轮船沉没，更是导致了476名乘客死亡：



飞机焦虑

问题：你觉得是什么原因导致乘客更担心乘坐飞机而不是搭乘其他交通工具？

飞机焦虑

我的解释：请大家看下面这张图

原创 飞机是最安全的出行选择，一旦出事，后果就不堪设想

2022-03-22 12:12



飞机焦虑

我的解释：虽然飞机的事故率不高，但是一旦发生事故，死亡率往往接近100%

这一结论也适用于很多其他情况：

例如，在新冠时代，我们发现Omicron导致的死亡人数和患病人数都比Delta更多，但是大家却更害怕Delta。这是因为得了Delta之后死亡率比较高，但是得了Omicron之后死亡率却没有那么高。

飞机焦虑

我的解释：虽然飞机的事故率不高，但是一旦发生事故，死亡率往往接近100%

这一结论也适用于很多其他情况：

假如有两款新冠疫苗，都会有一些不良反应。

第一种疫苗的不良反应率是百万分之1，但是一旦发生不良反应必死无疑。

第二组疫苗的不良反应率是百万分之15，发生不良反应之后死亡率是10%

一般来讲，大家更愿意接受第二种疫苗而不是第一种。