

内参特刊 | 和你聊聊每个人都关心的人工智能热点话题

2018-01-27 洪亮劼

AI技术内参

[进入课程 >](#)



受极客邦科技邀请，1 月份我回国参加 AICon，1 月 13 日晚做了一场直播，直播的主题是“人工智能 20 问”，编辑们收集整理了一些大家感兴趣的有关人工智能的话题，我在直播中分享了我自己这些年来的一些经验和心得。这场直播的文字精简版已经发布在 InfoQ 的公众号上，今天我也和你在专栏里分享，希望其中的某些话题能够对你有所启发，也欢迎你留言和我讨论。

0:00



暖场篇

Q：您本硕博学的都是计算机专业，这个专业其实有很多方向，您为什么会选择人工智能这样一个比较新的领域？

A：我到美国读博的时候，开始并没有想去学习人工智能，而是想做一些操作系统及软件的研发。到了美国以后，我的导师认为社交媒体的数据挖掘、机器学习应用比较前沿，希望有一个相关课题，当时把新的课题给了我。我从这个课题入手，接触了机器学习，接触了人工智能，然后慢慢地走上了人工智能的发展方向。

Q：您怎么定义人工智能？

A：我认为人工智能有两个方面非常重要。一是人工智能是**数据驱动的**，也就是说一个人工智能产品一定得是有数据支撑的；另一方面人工智能是一个**持续的决策过程**，比如说AlphaGo、无人驾驶等，都是利用机器学习，以及其他算法进行持续决策的一个过程。**在人工智能这个定义中，数据驱动和持续决策应该说是非常重要的两个元素。**

落地篇

Q：您能不能详细介绍一下目前在 Etsy 有哪些人工智能的技术是实实在在落地了的？落地的场景是什么？取得了什么样的效果？

A：Etsy 人工智能应用主要是围绕三块在做。

的搜索，不管是大一点的电商，如亚马逊、京东、阿里，还是小一点的电商，真正能够做到比较完美的搜索体验是非常困难的。比如一部分用户的购买欲望很强烈，那么我们如何能够帮助他找到最合适的商品；也有一部分用户，并没有特别强的购买意愿，这个时候他希望通过搜索一些关键字来获取灵感，从你的平台上能得到一种逛街的体验，所以这个时候，搜索带来的东西就未见得一定是他要买的，如何能够通过搜索给他带来逛街的体验是比较困难的。这块搜索是我们现在一直在做的，并且用不同的模型来进行落地尝试。

那么第二个方向就是**推荐系统**。电商的推荐也是一个老生常谈的话题了，但在一些新的场景下如何做？比如 Etsy 上有很多的商品是手工艺品，数量可能很少，那么在给用户推荐的时候，必须考虑到它的件数是有限的，这个地方也是有一些难点。

第三个当然就是我们的**广告平台**。如何能够通过人工智能帮助广告提升，广告是一个买卖的双方系统，如何帮助卖家实现好的宣传效果，这个也需要通过算法来更新的。

Q：目前已落地的一些人工智能技术，起到的作用和我们传统的一些办法相比，效果明显吗？

A：人工智能在搜索广告和推荐这块效果已经比较明朗了，肯定比传统的方法做得好，但是它的提升度在不同的情况下区别还蛮大的。**我们如果能够让效果最大化，必须跟产品界面，还有前端等融合，形成一个统一的产品，才能让算法得到的结果以最好的形式展现出来。**

Q：您认为人工智能落地遇到的最大的一个挑战，或者困难是什么？

A：**现在人工智能落地最难的并不是算法本身，最难的是和场景的对接。**算法本身可能是一个数学描述，是个模型，但是真实的场景跟模型中间还有一段距离。**这个距离需要数据科学家进行“翻译”，把我们的场景和算法衔接起来，需要我们的工程师，能够去根据新翻译的场景调整算法。**所以不是所有的经典的算法你都能够马上利用，而是需要一些更改，这种更改必须是基于你的场景，你去理解这个场景，并且能够进行合理的翻译，这实际上是最困难的。

观点篇

A: 我刚才已经谈到了, 数据科学家和人工智能工程师职能有重叠, 但也略有不同。打个比方, 就像你踢一场足球赛, 需要有 11 个队员, 有前锋, 有中场, 有后卫, 有守门员, 一个复杂的人工智能产品需要有不同的角色, 这些角色都需要有一定人工智能机器学习背景, 比如说前端工程师, 一些设计的人员, 一些产品经理都需要人工智能背景。**那么从这个角度看这个问题, 我觉得反而是大家的机会更多了, 大家可以根据自己的背景, 根据自己的喜好, 寻找适合自己的角色。**

Q: 人工智能会在哪个行业爆发?

A: **我觉得另外一个角度看这个问题更好, 有多少行业能够去利用人工智能?** 就像你刚才提到的, 这个人工智能可能会像空气, 会像水, 会像电一样, 成为一个基础设施, 我很认同, 今年可能会有一些新的契机, 我对这方面也很感兴趣。我们也看到, 像吴恩达, 他想去对传统制造业和人工智能做一些结合, 我觉得这个趋势就是一个很好的切入点, 不是说现在就能够彻底地改变制造, 因为毕竟各行各业有些很深的问题, 你得去了解它, 但是它可以打开一扇门, 让更多的传统行业, 比如像制造业的人来了解人工智能, 我觉得这个可能会是未来的方向。

Q: 现在有一种说法是人工智能就是人工智障, 您怎么看?

A: 今天我们所看到的人工智能, 像 AlphaGo 也好, 或者图像识别的一些突破, 云技术的一些突破, 它的确比 10 年前, 比 20 年前的技术要好的多, 我觉得人工智能的确到了一个好的阶段。同时, 我们也必须要承认, 人工智能不能做的事情也很多, 并且它做绝大多数事情可能不如我们想象的完美。但是我觉得我们也不必给它贴上人工智障的标签, 以表达我们的失望之情。**人工智能目前还处于一个很初级的阶段, 前面还有很长的路走。我自己希望通过做一个传播者和一个教育者, 让更多的人了解人工智能。**

Q: 您觉得人工智能存在泡沫吗?

A: 我个人觉得现在的确是有些泡沫, 比如说媒体的热炒, 媒体可能不是很了解情况, 然后去下一些断言。但是, 我个人感觉现在的泡沫处于一个比较健康的状况, 这个泡沫帮助更多的人了解这个行业, 帮助更多的人投身到这个行业, 帮助更多的社会资源聚焦到人工智能。

质是什么？

A：招聘大概有这么几个方面，一个是对候选人的**基础知识的考察**，比如说机器学习基础、统计背景、计算机的基本系统的掌握。第二点，我们会给一些场景，去考察这个候选人是否有能力把一个现实的场景跟他所学的挂上钩，他**能否用机器学习或人工智能算法作为一种语言去描述这个场景**，这个能力也是很重要的。第三个是**动手的能力**，比如说他是不是会一些最基本的编写代码，或者数据结构等，大概是分这么三块。

Q：您在专栏里面提到过，数据科学家大概有两类，您能具体说说这两类吗？有什么区别呢？

A：在国外，数据科学家是一个比较泛的统称，大概有两大类。

一类是**偏数据分析**，比如我有一个商业的问题，或者产品的问题，我想去挖掘这个数据，看能不能回答我这个问题，它重在分析，不偏重如何提高产品品质。比如说他分析圣诞期间什么东西卖得比较火，明年男生会不会更喜欢某款相机，回答类似的静态问题。

另外一类的数据科学家主要侧重于怎么用机器学习的算法提高我们的搜索、推荐或广告效率，更**偏重于驱动产品**。这两类数据科学家的技能有一些重叠的部分，比如最基本的机器学习、统计背景、编程背景等，但是他们的侧重方向不一样。

Q：是不是硕士学历及以上才能做人工智能？

A：我觉得要分现状和未来两方面。现状是本科生没有全面的一个训练，一些需要高阶技能的部分，可能硕士也没有完整训练过，并不是说，做人工智能就一定要博士生，而是需要的这些技能刚好只有一些博士生满足。但是随着教育和推广的演进，这个门槛会逐渐降低。我跟朋友闲聊的时候经常举一个例子，牛顿发明整个力学体系时候，那会儿全世界可能只有几个人知道，但是现在牛顿力学体系只是高中的一堂课，它的门槛低多了。未来人工智能、机器学习的门槛也会降低，普通人也可以从事相关的工作。

学习篇

Q：如何入门和学习人工智能？

的。

Q：如何提高看书的效率？

A：这个需要在实践中去学，如果需要某个知识点，要去通读这个知识点相关的内容，相同的内容多看一些不同的观点，把它搞懂搞透。有的讲的比较粗的，需要去其他地方找到比较细的论证，有的比较细可能没有抽象到一个高度，在其他地方反而能找到，要互补，这样看书的方式是最高效的。

Q：您前面提到除了看书，看要看国际会议论文。如果让您给大家推荐三个顶尖国际会议，您推荐哪三个？

A：我个人觉得，工业界的朋友可以关注这几个会议。

第一个是**KDD**（Conference on Knowledge Discovery and Data Mining，数据挖掘与知识发现大会）。这个会议有个工业组，论文包括一些已经落地的项目经验，或者将要上线的经验，并且在开会的过程中，会跟工业界有互动的环节。你去 KDD 可以看到很多人在翻译现实中的场景，这对工业界的朋友来说是很有帮助的。

另外一个**WSDM**（International Conference on Web Search and Data Mining，网络搜索与数据挖掘国际会议）。这个会议缘起于雅虎、谷歌等硅谷公司，认为传统的学术会议不能满足工业界需求，最近几年，工业界的不少解决方案和新问题都是在这个会议上提出的。

第三个是**CVPR**（Conference on Computer Vision and Pattern Recognition，国际计算机视觉与模式识别会议）。这个会议可以帮助很多初学者拓宽视野，让大家知道计算机视觉是一个非常广的领域，深度学习只是其中的一部分。

Q：外行看热闹，内行看门道，有很多报道只关注谁获得了最佳论文奖，谁获得了终身成就奖，除此之外，您觉得对于这些顶级会议我们还应该关注什么？

注搜索、推荐等，我一次性去了解这次会议在该领域大概有多少论文，有些什么新的东西，这就极大的缩小了关注范围，节省了时间。另一个需要去了解这个会议上有哪些新的想法，这些也许跟你现实的工作不直接相关，但是对于你拓宽视野很有帮助。

Q：那针对阅读论文，我们应该怎么入手呢，您的经验是怎样的？

A：学习人工智能方面的知识，特别是读论文，很多时候大家觉得难以入手，因为有太多的论文，一个比较高效的方法是**先读文献综述**。文献综述总结了当前的发展状况，能帮你解决至少两个问题，第一个问题是，了解相关术语，你用这些术语能查到更多相关的东西，如果你连术语都不知道，可能连这些论文查都查不出来，这些综述能帮助你熟悉这个领域的语言，和它里面的一些子问题。另外一个，能够起到文献综述作用的是**国际的顶级会议**，这些会议会邀请学者做分享，分享往往会有一个 PPT，这些 PPT 会归纳最近几年的发展和观点。通过综述和讲座，你就已经对这个领域有了一个概念性的理解，然后你再去读三到五篇该领域、子领域的经典论文，这是我个人去体会过相对比较高效的一个方法。

Q：小白如何入门人工智能？

A：如果你想真正投身到人工智能，我个人建议，还要相对系统地去学习整个知识体系，比如刚才我们聊过的，机器学习、统计、系统概念、最基本的编程的能力等，还有一些工具，比如 Python 等，磨刀不误砍柴工，但这肯定需要一个过程。

Q：目前有个说法，Python 是机器学习第一语言，一般推荐大家先学习 Python，您怎么看？

A：Python 的确是一个比较好的机器语言学习工具，并且现在围绕 Python 也产生了不少生态系统，使得 Python 比较适合机器学习、数据分析，或者是一些场景。但这里我必须强调一下，**在一个比较完整的端到端的人工智能应用中，特别是工业级的应用中，往往是一个比较异构化的环境**。所谓异构化是指不是单一的语言，或者单一的工具就能够完全的把数据转化成你想要的模型，并且能够用这个模型去推动我们的产品，这里面可能有 Python 的部分，但可能还有 Java 的部分，有 C++ 的部分等，如果你真想要有一个进阶，成为一个更加全面的工程师，或者是科学家，就不能只掌握 Python。


AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 050 | 如何将“深度强化学习”应用到视觉问答系统？

下一篇 051 | 精读2017年NIPS最佳研究论文之一：如何解决非凸优化问题？

精选留言 (1)

 写留言



林彦

2018-01-27

 3

请问老师您推荐的3个会议里文献综述类文章除了review, survey，还有哪些常见的关键字或其他检索方法？谢谢