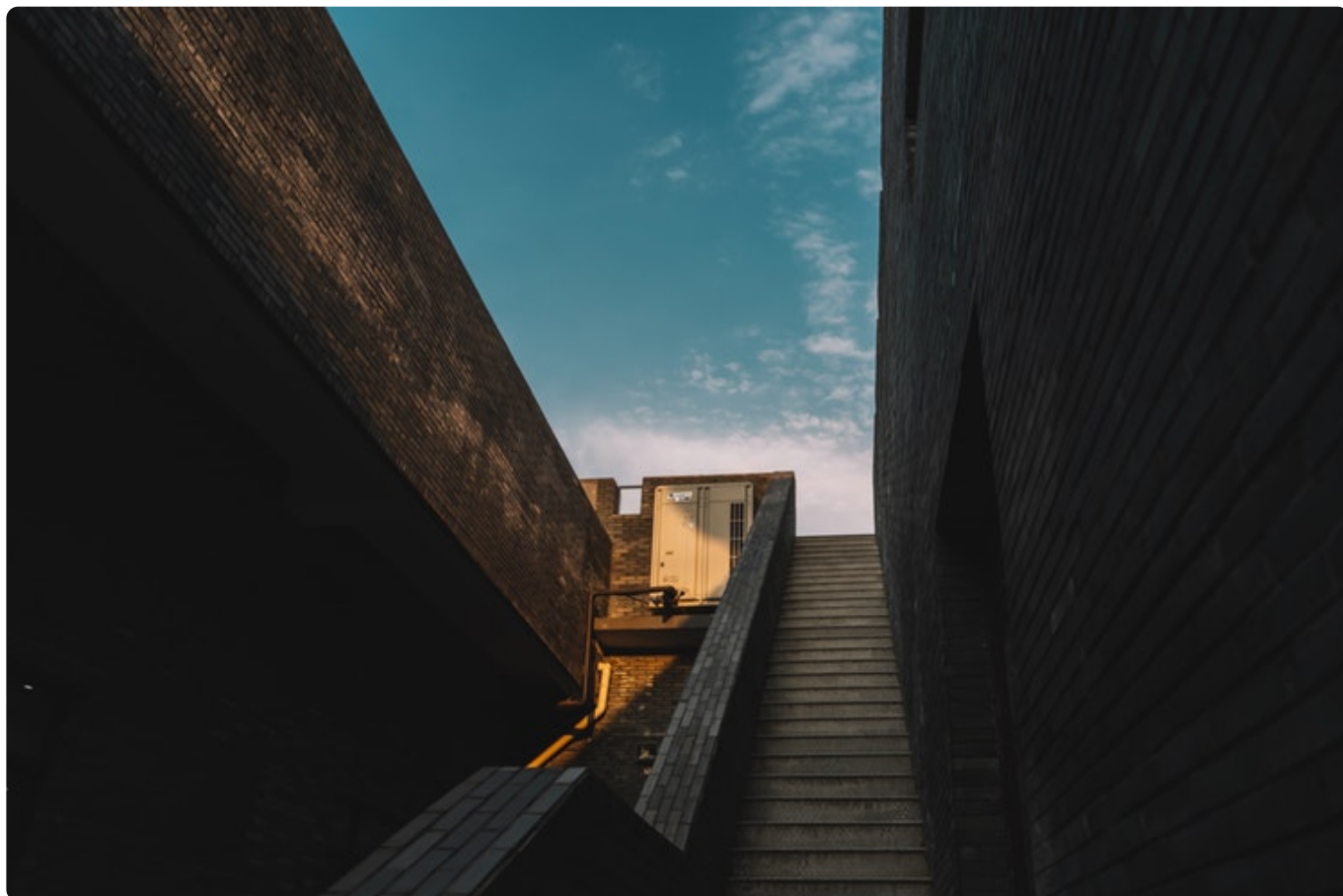


第206讲 | 邵浩：人工智能新技术如何快速发现及落地（下）

2019-04-17 狗尾草科技合伙人，人工智能研究院院长邵浩

技术领导力300讲

[进入课程 >](#)



讲述：李学良

时长 09:29 大小 8.70M



你好，我是狗尾草科技合伙人，人工智能研究院院长邵浩。在上篇文章中我们聊了一些人工智能技术在落地过程中的难点痛点。今天，我们接着这个话题，继续聊聊在新技术层出不穷的情况下，技术管理者如何快速发现可应用的新技术，并将其落地到实际产品中。

新技术层出不穷，如何发现新技术

如果有关关注 arxiv 网站的读者会感觉到，算法在近年来的迭代速度非常快。以语言模型（Language Model）预训练方法为例，代表性方法有 Transformer，ELMo，Open AI GPT，BERT 以及最新的 GPT2。其中，Transformer 于 2017 年 6 月被提出。ELMo 的发表时间是 2018 年 2 月，刷新了当时所有的 SOTA（State Of The Art）结果。不到 4 个

月，Open AI 在 2018 年 6 月，基于 Transformer 发布了 GPT¹方法，刷新了 9 个 SOTA 结果。又过了 4 个月，横空出世的 BERT 又刷新了 11 个 SOTA 结果。到 2019 年 2 月，Open AI 最新发布的 GPT2，包含 15 亿参数，刷新了 11 项任务的 SOTA 结果。

因此，可能我们还在尝试验证一个算法的时候，另一个新算法已经刷新了原有算法的指标。那我们如何能够保持对技术的敏感度呢？以下是我的几个建议：

首先，优秀论文（英文为主）是必须要跟进的，尤其是阅读算法的原文，对于复现和修改算法细节非常有必要。其中，arxiv (<https://arxiv.org/>) 是一个非常重要的平台，虽然没有同行评审，会带来良莠不齐的问题，但鉴于论文更新速度快，很多重要的成果都会在此进行预发布。一般来说，优秀的论文引用量都比较多，英文苦手也不用担心翻译问题，因为网络上会有很多业内高手对好论文进行解读，通过中文解读可以初步了解算法的性能和工程化的可能性。

其次，高质量的微信公众号也是了解新技术的一个快速通道。通用类公众号比如机器之心、将门创投，细分领域比如 paperweekly（就是带你读论文）等，都是快速接触新概念新技术的途径（注：我没有收取广告费 LOL）。

然后，对于工程化项目而言，GitHub (<https://github.com>) 是一个不可或缺的代码源，工程师经常戏称 GitHub 是全球最大的同性交友网站。在这里，可以直接获取大量优秀算法的开源代码，还有经典论文的算法复现，可以在 issue 中提问和解答细节问题，还可以参与完善算法。

最后，要找机会参与圈内的活动。参加国内的一些圈内会议，不仅可以了解目前的技术进展，还可以和来自高校和企业界的朋友互通有无。尤其是一些大型会议，对公司的招聘和技术宣传也都是非常好的途径。机器学习、计算机视觉、语音技术、自然语言处理、知识图谱等都有专门的会议，这里就不再一一列出了。有兴趣的读者可以关注中国计算机学会的学术会议排名来进行选择。

新技术如何从理论到落地

在上一篇文章中，我们讨论了人工智能技术落地难的问题，接下来，我希望和你简单讨论下如何将优秀的技术进行落地的话题。

首先，从业务的角度来看，需要审视公司的主要竞争力是什么，不能为了用新技术而偏离了公司的核心价值。因为新技术通常是锦上添花，而不是雪中送炭的。在业务中，需要根据自

己的产品做选择，用了人工智能技术，能不能给产品带来显著的提升，公司能否负担起人力成本和数据成本，甚至包括深度学习所需的硬件成本？

举一个简单的例子，机器学习中有个领域是分类（Classification），在深度学习还未兴起的时候，针对不同的数据进行分类，需要根据数据的分布特点，选择不同的适用算法，例如 SVM，LR，决策树等。即便是在今天，传统的算法未必就比深度学习方法要差。我有一次面试一名头部大厂的高级算法工程师，他整个过程中只谈 LR 的方方面面，对于其他的算法不关心也不熟悉，好像是 LR 弄好了就能打遍天下的感觉。不可否认，这些传统分类算法在大量问题上都有优秀的表现，但有了效率更高的模型，也有必要去测试并判断它们能否用于实际产品中。

比如 FastText 这个算法，发布之初就得到了广泛支持，在自然语言分类问题上，不仅速度快，性能提升也相当明显。因此，经过一整套测试流程之后，我们将其作为了主分类模型。后来又出现了 BERT，同样经过测试之后，我们发现其性能超越了 FastText 百分之一到百分之二。但这时候，由于其时间消耗远大于 FastText，并不能作为主模型来用，我们仅将其作为并行处理中的一种参考方法，与若干传统机器学习方法一起来做 stacking（Stacking 是一种集成学习方法，感兴趣读者可以阅读其他资料进行了解）。

因此，算法好不好用，能不能落地，需要综合考虑各方面因素。比如说在做文本匹配的时候，有大量深度学习方法可以使用，理论性能也都是 SOTA 水准，但根据产品需求来看，有时候 96% 的准确率和 95% 的准确率，产品体验并没有多大差别，但投入的研发成本却要高出许多，反而用 ES（ElasticSearch）自带的 BM25 算法又快又好，那就完全没有必要使用深度学习方法了。

同样的，对于感知技术，目前技术的成熟度也非常高。比如说人脸识别，如果和产品不是重度耦合，直接使用大厂的 API 或 SDK 就行了。但如果是自己的核心产品需要搭配人脸识别模块，就可以考虑在已有成熟的开源的算法上进行二次开发，甚至搭配自己的芯片，从而打造自己的核心竞争力。（注：别忘了确认算法的商用条件，以免陷入版权纠纷）

其他需要注意的点

1. 技术和产品的权衡

正如前文提到的技术成熟度曲线，有些技术，在现阶段的确没有达到人类的期望值。这个时候，一个可选的解决方案是通过产品设计来补偿。

比如说大量“弱智”的聊天机器人产品，可以利用产品设计，来弥补技术缺陷。那该如何去做呢？关键在于，我们要让用户不要重点关注其技术表现，而是对产品的体验有一种惊艳感，发出类似“天哪，这样都可以”的赞叹。比如我们狗尾草科技提出的聊天机器人的虚拟生命形态，这个概念之前日本的 Gatebox 公司也提出过。基于这个概念诞生的琥珀·虚颜，就是我们推出的一款结合了 AI+AR+IP 以及 GAVE 引擎（Gowild AI Virtual Engine）的虚拟生命产品，它搭载 holoera 硬件平台及 360°全息投影，创造了一个有情感、可养成、可进化的虚拟存在，但这种存在又可以和周边世界进行多模态真实互动，并针对用户行为习惯形成不同的性格体系。同时，人物还可以换成二次元角色和真实的明星，进一步提升用户体验和粘性。

有一句话说的好，技术不够产品来凑，产品不够运营来凑。用户接受才是硬道理，技术可以加分，但过硬的还是产品本身。

2. 小公司和大公司

我接触过一些创业公司，一开始就期望做一个人工智能开放平台，虽然在前两年拿到不少融资，但最近大多数都销声匿迹。做平台型的事情，没有大量的人员和资金支持，是无法实现的。我们可以看一下国内的人工智能研究院，且不说百度腾讯的换帅事件，很多中小型公司的研究院也由于没有办法进行工业级产出，导致了大量资本投入换来的只是 Demo 和论文。

所以，对于中小型公司而言，需要做的事情是借助利好，顺势而为。同时精耕细作一个细分领域，在大公司无暇顾及的垂直行业杀出一条血路。无论是后期被收购还是能够独立壮大，都是比较好的结果。切忌大而全，什么都想做，做自己擅长的才是最重要的。

写在最后

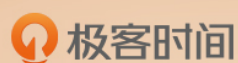
大家现在都在谈人工智能技术，而且很多人都会把人工智能和 AlphaGo 以及深度学习划上等号。其实人工智能涵盖的学科范围是非常广泛的，包括心理学、神经科学、哲学、认知科学等等。我们目前看到的大量成果都只是深度学习和大数据的化学反应。而且，大量的人工智能应用还都是人工 + 智能，离真正的认知智能差距甚远。如何利用技术赋能产品，得到用户和资本的认可，才是最重要的。

参考文献：

1. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018). Improving language understanding by generative pre-training

作者简介

邵浩，TGO 鲲鹏会会员，日本国立九州大学工学博士。现任上海瓦歌智能科技有限公司总经理，深圳狗尾草智能科技有限公司合伙人，人工智能研究院院长，带领团队打造了聊天机器人产品“公子小白”及 AI 虚拟生命产品“琥珀·虚颜”的交互引擎。中国中文信息学会青年工作委员会委员，中国计算机学会 YOCSEF 上海学术委员会委员。研究方向为人工智能，共发表论文 40 余篇，出版了业内第一本聊天机器人著作，主持多项国家级及省部级项目，曾在联合国、WTO、亚利桑那州立大学、香港城市大学等任访问学者。



技术领导力 300讲

每个技术人都应该知道的管理心经

梁宁 / 著名产品人

张雪峰 / 饿了么CTO

陈皓 左耳朵耗子 / 知名创业者

许式伟 / 七牛云创始人兼CEO

李大学 / 前京东CTO

汤峥嵘 / turtorABC COO

右军 / 蚂蚁金服

程浩 / 迅雷创始人

邵浩 / 瓦歌CTO

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 第205讲 | 邵浩：人工智能新技术如何快速发现及落地（上）

下一篇 第207讲 | 许良：科创板来了，我该怎么办？

精选留言

💬 写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。