

057 | 人工智能领域知识体系更新周期只有5~6年，数据科学家如何培养？

2018-02-12 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:40 大小 3.52M



在上一期的分享里，我们聊了数据科学家团队管理的一个重要步骤，那就是如何来衡量数据科学家或者人工智能工程师在团队中的业绩，我们重点讲了如何看待数据科学家团队的价值和数据科学家评定的一些误区。

今天，我们来聊另一个数据科学家团队的高级话题，那就是数据科学家的培养的问题。

为什么要培养数据科学家

为什么要培养数据科学家？这个问题看上似乎是显而易见的，但实际上，如果不了解数据科学家或者人工智能团队的一个重要性质，你很可能无法很好地运营这样一个团队。究竟是什么

么性质这么重要呢？那就是数据科学家或者人工智能工程师有强烈的持续学习和不断更新自我的需要，这是数据科学家培养的一个非常重要的理念。

那么，数据科学家为什么需要不断学习？简单来说，是因为数据科学家所需要的技能和知识处在一个快速变化的环境中。如果数据科学家不能对这些快速变化的技能和知识加以学习，就很可能被迅速淘汰。

我们这里所说的技能有**知识性的技能**也有实际的**工具性质的技能**。

从知识性的来看，机器学习和人工智能技术每隔一段时间就会有一些重要的发展，了解和掌握这些更新的技术需要一定的门槛。因此，持续学习是为了能够迈过这些门槛。从过去的经验来看，每一次这样的重要发展所带来的新门槛都不可避免地让一些工程师和数据科学家落伍。

比如，在过去不到 20 年的时间里，机器学习就经历了“支持向量机”（Support Vector Machine）、“概率图模型”（Probabilistic Graphical Model）以及“深度学习”（Deep Learning）这三股大的思潮。**也就是平均 5~6 年，数据科学家和人工智能工程师就需要面对一些完全不同的建模思想和工具**。更不要说，在这些大的思潮之下，每年出现的新模型也是层数不穷。这还没有提及应用的领域，比如推荐系统、搜索、广告系统、计算机视觉、自然语言处理等等。如果不能在这些领域知识的快速变化中取得主动，很可能就无法胜任未来的工作。

在实际工具技能层面则更是日新月异。比如近日如火如荼的深度学习框架 TensorFlow 仅有 3 年多的历史，五六年前根本就不存在。而如今借助机器学习迅速崛起的编程语言 Python 在五六年前也没有近日的火爆。而在支持向量机年代非常受欢迎的 LibSVM 和 SVMlight 工具，可能今天已经很少听到。知识框架的变化相比，工具技能层面的变化更加琐碎，更加细节，这也为人工智能科学家提出了更高的挑战。

那么，在知识结构和工具技能都快速变化的情况下，团队的负责人就需要针对这样的特点进行有远见的管理安排。

第一，需要为学习这些技能和知识提供时间。任何数据科学家现有的知识体系都不能保证永不过时。事实上，就像我们刚才提到的，现在每 5~6 年就有一个比较大的知识体系更新，这个更新速度在未来还有可能会更快。那么，花费了非常大的代价招聘来的整个团队就有可

能面临着短时间内过时的危机，所以，要能够利用平时的时间，把持续学习的内容安排进团队的日常运作中，可以有效降低团队遭遇知识鸿沟（Gap）的风险。

第二，需要团队里的资深数据科学家能够战略性的挖掘接下来有可能进入主流视野的技术，从而早作准备。尽管这可能显得有一点过于超前，但是对于大多数的互联网或者高科技企业来说，技术实力上的领先无疑是最强大的生产力。因此，在日常的安排中，如果在团队人手富裕的情况下，能够有一些数据科学家专注比较“面向未来”的技术，从而为今后的技术积累以及整个团队的“技术纵深”打下基础。

其实，谷歌的 DeepMind 或者 Facebook 的人工智能研究院都有着这样的性质。这些机构研发的技术未必能够马上应用到这些公司的主流产品上，但是这些技术有可能让这些公司或者团队能够在未来 3~5 年内有一个比较舒适的纵深，这些公司的其他团队需要做的，就是沿着这个纵深前进。

除了从一个团队以及数据科学家本身的不断更新换代的这个思路来看待培训以外，还有一个方面，那就是绝大多数公司和团队的数据科学家都不可能是在招聘的时候就已经是最一流的数据科学家或者人工智能工程师了。

你往往只能招聘到博士毕业生、硕士毕业生。他们的知识面和技能在刚进入公司的时候还非常稚嫩。对于一些博士毕业生而言，以前做的一些研究都是在一个非常窄的领域，还没有形成一个完整的体系。对于一些硕士毕业生而言，很可能完全没有真正接触过现实的问题，之前的学习主要是课堂项目。因此，**对于团队中的年轻成员，学习和培养就成为了一个非常必要的环节，让他们能够真正融入到工业级人工智能解决方案的研究和部署中。**

全方位的培养计划

刚才我们简单聊了聊从技术层面培养数据科学家的一个思路。其实，我们之前反复强调的一个思想就是，**人工智能团队并不是单独存在的。**一个能够真正运转并且能够为公司或者组织带来价值的人工智能团队一定是整个组织中的一个有机部分，并且能够为公司和组织带来数据驱动的持续决策的能力。因此，在这样的一个目标下，数据科学家的培养不应该仅仅是技术层面上的，还应该是更加**全方位的**。

如果说简单一点，对于一个数据科学家的全方位培养中，很重要的一条，那就是**团队协作的能力，特别是跨团队的组织、协调和沟通的能力。**我们在之前的分享中已经提到过，数据科学家的工作需要和数据工程、设计师、前端工程师、后端工程师、产品经理等角色的人员打交道。而在这个过程中，任何一个环节的沟通出了问题，都有可能造成项目的失败。因此，

有没有聆听的能力、有没有表达的能力、有没有了解需求的能力、有没有分清主次的能力等等，这些软实力就成为了数据科学家培养计划中的一个重点。

从过程上来说，一个新入职的数据科学家的核心目标还是从技术上慢慢从学生或者初级工程师逐渐成熟起来。最开始，年轻的数据科学家应该“多听”、“多看”、“多想”，但“少发表意见”。从和资深的员工一起参与项目开始，逐渐学习怎么和其他的部门一起工作，甚至从熟悉其他部门的词汇、语言入手。

最后，我想说的是，除了团队之间的沟通能力以外，**数据科学家上台演讲的能力也很重要。**能够把自己的解决方案说清楚，能够用通俗的语言来解释复杂的问题，能够不使用数学符号依然可以把解决方案的主要思想梳理明白并且能够传递出足够多的信息，这些都是数据科学家进阶必不可少的技能。

小结

今天我们分析了数据科学家或者人工智能工程师团队的培养问题。进行一个简单的总结：第一，我们讲了数据科学家为什么需要培养。在主要的技术技能培养的道路上，有什么样的情况；第二，我们详细梳理了数据科学家全方位培养中的重点是什么。希望这些内容能给你一些启发和借鉴。

最后，给你留一个思考题，怎么能够把数据科学家的持续学习纳入绩效考核呢？或者到底应不应该纳入考核？

欢迎你给我留言，和我一起讨论。

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 056 | 成为“香饽饽”的数据科学家，如何衡量他们的工作呢？

下一篇 058 | 数据科学家团队组织架构：水平还是垂直，这是个问题

精选留言 (2)

 写留言



gurudk

2018-02-14

 1

我想数据科学家主要需要具备两个能力，一个是扎实的技术基础，一个就是商业价值的感知能力



林彦

2018-02-12



一个软件工程和数据科学都能给公司创造价值的团队，如果团队有一定的自主安排时间。我觉得留下10%~20%的时间定期做一些分享(一定要上台讲)；做一些特定话题的小组讨论，小组中每个人可以发言，指定不同的组长，限制单个人的最长发言时间。根据参与人员的评价，专业人员的评价作为考核和未来沟通改进的标准之一。

...

展开 ✓