

156 | 聊一聊谷歌特立独行的混合型研究

2018-10-01 洪亮劫

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:26 大小 2.95M



上一讲我们介绍了微软研究院发展早期的一段故事，一起讨论了为什么说微软研究院是工业界研究院的楷模。

今天我们来看另外一种“**混合型**”的工业界研究机构模式，聊一聊**谷歌研究院**，一起来讨论这种模式是不是更加适合互联网企业的需求。

研究背景起家的谷歌

谷歌的创立比微软晚了将近 20 年，但两个公司有一些相似的地方，其中之一就是创始人都是中断了学业，投身到创业的浪潮中。不过拉里·佩奇（Larry Page）和谢尔盖·布林（Sergey Brin）当时是在攻读博士学位，关于如何进行网页搜索的最初想法，是从他们的

博士研究课题衍生出来的 [1]。由此可见，谷歌从一开始就和研究类项目有着千丝万缕的联系。

佩奇和布林的论文发表在 1998 年的国际万维网 (WWW) 大会上。这篇论文介绍了 PageRank 算法，在当时这简直就是一个石破天惊的算法。那它和当时其他搜索引擎的关键技术相比，独特之处在哪里呢？计算网页的相关度或者说是重要度完全不依赖文本信息，而仅仅依靠由网页之间关系组成的图，而且能够得到一种非常稳定的排序。或许就是因为这个独特的算法让两个斯坦福大学的年轻人放弃了继续攻读博士学位的想法，转身在硅谷找了一个车库，从而演绎了一个传奇的创业故事。

或许就是因为创始人的背景，我们可以看到谷歌对学术界的最新研究成果有一种特殊的青睐，在谷歌发展的路上，屡屡上演收购案例，收购的很多公司都是因为有一些研究成果而成立的小公司。

举几个例子。同样来自斯坦福的博士毕业生塔哈尔·哈维利瓦拉 (Taher H. Haveliwala)，改进了 PageRank 算法 [2]，他创办的公司在 2003 年的时候被谷歌收购。2010 年，达蒙·霍洛维兹 (Damon Horowitz) 和瑟潘达·卡姆瓦尔 (Sepandar D. Kamvar) 在当年的国际万维网大会上发表了一篇“社交搜索” (Social Search) 的论文，论文标题都跟佩奇和布林当年发表的 PageRank 论文有异曲同工之妙。两个作者所在的公司很快就被谷歌收购了。最近的例子就是我们都熟知的杰弗里·辛顿 (Geoffrey Hinton) 所创立的公司以及位于英国的 DeepMind 公司，也都是因为在深度学习方面的重要贡献被谷歌先后收购。

混合型工业界研究

尽管谷歌对于学术研究有一种天然的亲近，然而在很长一段时间里，谷歌其实并没有真正成立完全独立的基础研究部门。所以很多学术圈的研究人员，还有工业界的研究同仁，都对谷歌产生了一种误解。

2012 年的时候，谷歌的研发总监彼得·诺维格 (Peter Norvig) 发表了一篇文章 [4] 来介绍谷歌的研究模式。

谷歌的研究模式到底是怎样的呢？简单来说，就是**让研究和产品的研发紧密结合起来，而不完全建立独立运行的研究院**。当然，这个模式在收购了 DeepMind 之后算是被打破了。但是在谷歌 20 年的发展历史上，**混合型研究一直是谷歌研发的主流**。

诺维格在文章中解释道，谷歌研究工作的一大重要目的就是**为终端用户带来重要的和实际的好处**。很明显，这一目的和纯粹的学术研究有很大的距离。我们来看微软研究院和雅虎研究院，它们的重要贡献指标就是学术论文发表的质量和数量。

我们再具体来看看这两种道路的差异。

传统的学术研究是这样的：研究人员首先构想一个学术课题，然后在一种比较受限的环境中对这个课题进行研究。这里说的受限的环境，往往是指数据并不是全量数据，而是采样过后的数据，这些数据能够在学者们的笔记本或者小型集群中进行计算。甚至在有的情况下，研究人员会使用完全虚拟的数据。另外，在这种受限的环境中进行研究还会带来这样一个问题，由于开发的代码和软件不需要重复使用，也不需要开发生产环境的代码，所以这些代码质量都相对较低。

那谷歌的研究工作是怎样做的呢？**谷歌研究要求从一开始就使用生产环境来编写代码**。这些代码和普通的产品代码没有区别，也使用和一般产品线代码相同的数据、相同的流程。这样，一旦有了研究成果就可以无缝接入现在的产品线中。这样的要求对于研究者来说，其实是拔高了研究的难度，但是对于研究成果和产品对接来说就将困难降低到了最小。

总结一下，**让工程和研究结合在一起，并且有意模糊这两者的区别，就是谷歌混合型研究的核心思想**。工程师和科学家在同样的项目中工作，大家都面临同样的限制，这样就可以让研究的课题不至于完全天马行空，而是能够落到实处。

混合型研究的思考

我们上面介绍了谷歌的研究背景以及基于此慢慢形成的混合型研究模式。为什么这条道路在谷歌就能够落地实施呢？我想这里面有一个非常重要的先决条件，这是诺维格的文章里没有提到的，那就是大量高素质人才的涌入。

在这些人才中，博士生比比皆是，甚至有很多教授。所以，谷歌的工程团队中有很多博士生担任着普通工程师的角色。说得通俗一点，谷歌用博士生在干硕士生甚至是本科生就可以胜任的工作。一个团队中的工程师和科学家并没有本质的区别，这才使得任何一个科研想法都可以很容易地在工程层面得以实现。

由此我们可以看到，从某种意义上来说，谷歌其实并不需要单独的研究机构，自己工程团队的水平就已经非常出色了。我们从 TensorFlow、语言翻译等知名项目就可以看出来，这些项目都是工程团队达到了很高的研究水平。

当然，在这种模式下，谷歌的基础研究其实是受限的。在收购了 DeepMind 后，谷歌也开始依靠这样单独的研发机构来推动和产品结合得不那么紧密的研究方向。

小结

今天我为你介绍了谷歌与众不同的混合型研究模式。这种模式对工程团队的水平要求比较高，如果没有高水平的工程团队，研究人员和工程师就会产生隔阂，沟通不畅，研发就会有问题。从这个角度来看，建立单独的研究机构或许更能实现很多公司的初衷。

讲到这里，如果我们要借鉴谷歌的这种混合型模式，你觉得挑战是什么？如果有了高水平工程团队这一保障后，你觉得想要成功还有什么挑战？

欢迎你给我留言，我们一起讨论。

参考文献

1. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. Proceedings of the seventh international conference on World Wide Web 7 (WWW7), Philip H. Enslow, Jr. and Allen Ellis (Eds.). Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, 107-117, 1998.
 2. Taher H. Haveliwala. Topic-sensitive PageRank. Proceedings of the 11th international conference on World Wide Web (WWW '02). ACM, New York, NY, USA, 517-526, 2002.
 3. Damon Horowitz and Sepandar D. Kamvar. The anatomy of a large-scale social search engine. Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 431-440, 2010.
 4. Alfred Spector, Peter Norvig, and Slav Petrov. Google' s hybrid approach to research. Commun. ACM 55, 7 (July 2012), 34-37, 2012.
-

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 155 | 微软研究院：工业界研究机构的楷模

下一篇 复盘 4 | 广告系统核心技术模块

精选留言 (2)

写留言



金文

2018-10-02

感谢洪老师这一路来的分享，感谢

展开



金文

2018-10-02

感谢洪老师这一路来的分享，感谢

展开



