

腾讯 AI Lab 主任张潼博士：机器学习里的优化问题

雷锋网

雷锋网 AI 科技评论按，日前，在由上海财经大学交叉科学研究院（RIIS）主办，杉数科技有限公司协办的「现代运筹学发展讨论会」上，腾讯 AI Lab（腾讯人工智能实验室）主任张潼博士发表了精彩演说。作为机器学习领域的知名学者，他开场表示，机器学习和运筹优化有很多的共同问题。接下来，他详细介绍了机器学习领域中优化的进展，以及大家比较感兴趣的研究课题。在演讲的最后，他表示，运筹优化的研究人员可以和机器学习研究人员多多交流，大家共同合作，共同促进。



以下为他的发言内容，雷锋网 AI 科技评论做了不改变原意的编辑整理，张潼博士对此文做了校正与确认，同时也感谢上财助理教授邓琪博士对本文提出的宝贵意见。

很感谢葛冬冬邀请我过来，今天是叶荫宇老师的生日，非常高兴能在这里与大家进行探讨。我的主要研究方向是机器学习，现在机器学习领域存在

很多的优化问题，目前来讲，有些优化的研究可能是机器学习领域自己在推动，我将为大家介绍这方面的研究进展，也希望大家以后能更好的去合作，推动发展。

机器学习里的优化相对比较窄，这其中考虑的问题主要是跟数据有关。主要有如下三种数据：

第一种是统计分布上独立的数据，这种问题有一个求和或者求期望的结构，监督学习和非监督学习里都存在这种结构。第二种是类似于 graphical model（图模型）的数据，在这里我们比较关心 graph（图）的结构，这种数据中也存在较多的求和结构。第三种是序列数据，在这种数据中，最根本的结构还是求和。

因为很多问题是以统计期望的形式表达，所以随机优化是大家比较感兴趣的研究方向。

随机优化在五六十年代就出现了，它属于传统优化的范畴。目前有一些与随机优化相关的书籍，也有专门的科研人员在进行这方面的研究。

实际上，在机器学习领域，大家用的也基本上是随机优化，很少有人用确定性优化。此外，最近的一些进展利用 variance reduction 证明了随机优化存在着更好的收敛率，这也是大家使用随机优化的原因之一。

现在我先为大家介绍一阶随机优化，这一领域的相关研究比较多。目前在这一领域，大家比较感兴趣的方向可能是非凸优化。最近有很多计算机理论学家在做非凸优化方面的研究，不过我还没见到做优化的人在这一领域进行研究。非凸优化中之前大家引用得比较多的是 Nesterov 写的 cubic 牛顿法，这个工作基于牛顿法，在非凸情况下取得了一些很好的成果。现在，一些研究人员以 cubic 牛顿法作为基础，但研究得更加深入了。

另一块大家可能研究的方向是二阶或三阶优化，这也跟随机有关，这里的研究包括怎么去做一些特殊设计的采样过程。这一块主要是机器学习领域在进行研究，但我知道目前也有一些做优化的人参与进来了。

还有一块大家可能比较感兴趣的研究跟加速有关。这一领域比较早的研究方法有 momentum 算法，Heavy Ball 优化算法，后来 Nesterov 在凸优化问题中严格分析过的加速算法。现在大家感兴趣的可能是非凸问题上的

加速。如何在非凸情况下做加速？最近也有一些相关文献。实际上，在非凸情况下做不了加速，但最新的文章表明，在执行算法过程中，我们可以检测凸性，一旦遇到凸性情况，就立即转换。

当大家在用 momentum 方法时，条件不合适的情况下应该怎么做？这里可能更偏理论。momentum 法在实际中也有应用，深度学习里的 Adam 算法就用到了它的思想，同时也用了其他一些 scaling 的方式。

此外，在加速问题上，还有一个我个人比较感兴趣的研究问题。加速本身是确定性的，在随机优化上用起来不是很好。随机上只能通过加速增加 mini-batch 规模，所以大家也在研究如何在随机上更好地去加速，不过这个问题在随机的 mini-batch 算法下到现在都没什么特别好的结果。一些人可能知道，把加速方法和随机算法叠加起来提升收敛速度，中间要经过一个过程，这个过程需要用个 deterministic 形式作为转换才行。这里是不是有更好的方式，现在我也不是很确定。

大家也对一些特殊结构比较感兴趣，比如说复合损失函数和 proximal 结构（比如稀疏和低秩）。非凸问题中对特殊结构的研究相对来说比凸问题中少一点。

还有一块大家感兴趣的就是超参数的优化，这块的研究实际上更加复杂。最近 learning to optimize（学习优化）这个研究就与超参数优化有关，即优化过程不是你自己推导出来的，是机器学习学出来的，这也比较有意思。这套方法目前还很初级，大家可以这么考虑这个问题——已经有人用机器学习来设计优化算法了。这里的研究更偏实践，因为如果要做分析会比较困难。与此相关的是随机 bandit 算法。在优化里也存在这种算法，和无导数优化相关，即优化过程中不见得用导数去做，这在很多实际问题的系统调参里是非常有用的。

前面是一系列与单核优化相关的研究。另外一块大家感兴趣的方向是大规模分布式和多核优化。即使到现在，很多优化软件包还没有多核的支持。

在这里，第一步可能要多核计算，第二步要做分布式计算，这是实际上的需求。另外一方面，从理论上来说，大家比较感兴趣的是通信和计算的平衡，即在这种结构里，如果有不同的计算单元，在执行算法的过程中需

要进行信息的交换，信息交换了多少，计算时间有多少，以及如何去平衡。这里存在很多相关研究，其中有同步的，也有异步的。大家实际上可能或多或少都会用到异步，目前也有一系列与异步分布式优化相关的研究。此外，现在有些人开始感兴趣的是去中心优化和低精度优化（比如用低精度传输导数），这些方法甚至可以和今后新的低精度硬件结合。这里有另一部分跟芯片相结合的研究是模型压缩，这也是和优化相关的研究范畴。

另外还有一个方向就是以实用为导向的非凸优化算法研究。非凸优化其实除了解决深度神经网络中的问题，还能解决别的一些问题，不过大家可能比较关心的就是深度神经网络问题。很多研究人员会做出一些有效的但没有太多理论基础的 heuristic 算法，包括 batch normalization 算法。另外还有类似 Adam 的算法，在实用上也很流行。它是把两个凸优化的想法结合在了一起：将 momentum 优化算法与 Adaptive Gradient 结合，再做一些参数调整。

另外在理论上，最近也有一些进展。最近有一些年轻的华裔科学家在这一领域做了一些比较前沿的研究，比如说研究怎么去优化逃出鞍点，他们最后得出了一个局部最优解。

现在很多是理论上的工作，比如说对收敛复杂度的研究，把 ε 的负 2 次方变成 ε 的负 1.5 次方、负 1.25 次方等。虽然这些算法目前看起来可能都不太实用，但是在理论上取得了一些进步。我们可以研究这些方法到底能够走得有多快，之后慢慢将它们变得实用化。

另外一块大家感兴趣的研究是算法在一些非凸问题上的全局最优收敛性，在这里需要一些结构信息。目前的研究在某些问题上取得了一些结论。有些是优化算法，但有些可能把算法改变了，用了其他的结构。

还有一块新的大家比较感兴趣的研究领域，是 saddle point 问题。我个人对这个问题也感兴趣，我研究过相关文献。目前在优化上，对这一领域的研究不算太多。现在比较好做的是凸、凹问题，以及对变量是线性，primal 和 dual 之间是 bilinear 作用关系的问题。在这些问题上，现在有了一些成果，包括机器学习领域现在也开始进行这方面的研究了。如果不是 bilinear 的，情况相对来说会更难一些，有些结论就不好推广了。目前，在非凸情况下，或者不是凸、凹的情况下，这一块的结果非常少，连收敛性

都没有好的一般性的结果，很多问题能收敛到什么地步我们也不是很清楚。

从实际上来看，确实会存在这种问题，比如说强化学习的某些形式可以写成一个鞍点问题。包括其他的一些生成模型，例如生成对抗网络，都会出现这种形式，既不是凸的，也不是凹的，但它有一些特殊结构，所以有些对它感兴趣的人会进行专门的研究。**而强化学习本身也是跟优化联系比较紧密的方向，它也跟运筹学研究的马尔科夫决策过程紧密相关。目前这个方向是研究热点，有越来越多的结果出现。**

虽然机器学习领域的问题比较窄，很多传统优化问题都不属于机器学习的研究范畴，但就像我前面说的，这里面也有很多有意思的问题，它们和优化紧密相连。机器学习领域在这些问题上钻研的相对比较深，常常做出了超出优化领域本身的一些理论工作。

希望到时候和大家能多多交流，如果大家感兴趣，可以参与进来，共同研究。

雷锋网 AI 科技评论编辑整理。

(完)