

科研求索路之学术界的“工厂模式”

——“自古真情留不住，唯有套路得人心”

微软亚洲研究院机器学习组实习生 赵鉴

我有一个梦想，想课题可以不再费心；

我有一个梦想，做研究能像刷题一样；

我有一个梦想，得成果能像批发一样。

这就是一个刚刚接触学术、即将踏入博士生涯的我的一点“梦想”。科研短短几个月，我忽然看到我的梦有望实现了：**Neural [*] Search with Reinforcement Learning**

- Paper 1: [*] = Architecture
- Paper 2: [*] = Optimizer
- Paper 3: [*] = Activation Function
- Paper 4: [*] = Loss Function...
- Paper 5: [*] = Batch Size...
- Paper 6: [*] = ...

文章已经陆续发表在 ICLR 2017, ICML 2017, ICLR 2018……都是机器学习顶级会议，擦一擦口水……哦不对，前三篇已经发表，后两篇我计划投稿。哦不对，前三篇也没有一篇是我的，我还是看看第四、第五篇有没有希望吧：(

抱着看一看、学一学的心理，我仔细拜读了这几篇文章。在下不才，也谈谈这几篇“孪生”文章的学习心得。

背景

近些年深度学习已经火爆至极了，各个领域都已经广泛使用。对于很多实际问题，它模式识别的精度和之前算法的精度已经是天壤之别。虽然，迄今为止没有人告诉我怎么解释模型，但是这不影响我将它作为研究领域的热情。

作为新手，我必须了解这个领域的关键技术。找了一圈，问了一圈，我发现神经网络对系统参数高度敏感，而深度学习最关键的技能就是：调参数。怎么关键呢？参数调的好才能出文章呀，而调参这件事愁坏了一大批深度学习的研究人员。虽然大家也不知道怎么调，但我相信，有一天我也能掌握这门独门绝技。不知是哪一天，我忽然发现了一个新的机器学习研究方向：Meta Learning ——让机器学会自动调参，想想都激动。

'In this paper, we use a recurrent network to generate the model descriptions of neural networks and train this RNN with reinforcement learning to maximize the expected accuracy of the generated architectures on a validation set.'

这篇文章引领了最近两年神经网络结构搜索这一研究方向的热潮，为神经网络结构的设计提供了新的思路。

论文二

紧接着，在 ICML 2017 上 论文一的作者带来了一篇新的文章：'Neural Optimizer Search with Reinforcement learning' ([*] = Optimizer)。

文章的题目和使用的图表让我分外亲切。这篇文章跟前面那篇文章采用了几乎完全相同的搜索方法，不同的是，搜寻目标从网络结构变成了优化方法，相应地，卷积核的长宽等这类参数变成了操作数、一元函数和二元函数，目标是搜索出一个更好的神经网络优化器。当然，该方法可以控制是否引入优化方法中的另一个重要参数——动量 (Momentum)，也可以通过调整优化函数的复杂程度和搜索组件的数量来控制搜索的空间大小。最终，这篇文章例举了他们该方法新发现的两个优化器 PowerSign、AddSign，以及几种新的 Learning Rate 衰减函数。

思路非常套路，图表也高度相似，这让我对我的“科研梦”充满了希望（如图）。

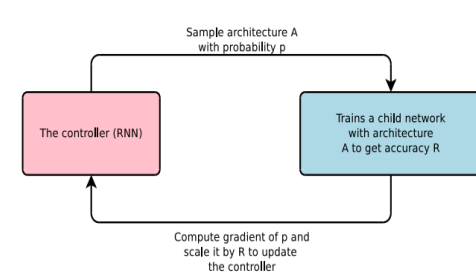


Figure 1: An overview of Neural Architecture Search.

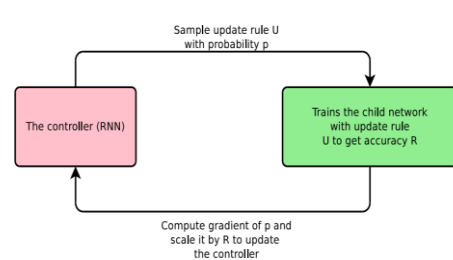
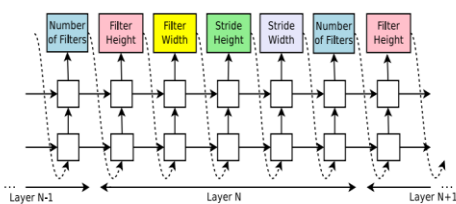
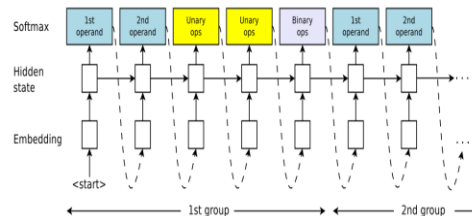


Figure 1. An overview of Neural Optimizer Search.



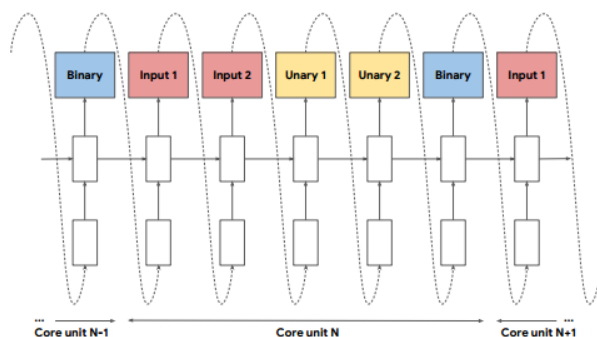
论文图表对比

[*] = Architecture vs. [*] = Optimizer

论文三

再接再厉，作者又在 ICLR 2018 上发表了新的工作，文章名 ‘Searching For Activation Functions’ ([*] = Activation Function)。

虽然这次在标题上没有了 Reinforcement Learning 字样，但文章依旧是熟悉的方法、熟悉的套路以及熟悉的图表。这一次，文章将搜索优化器表达式的方法应用在了搜索激活函数表达式上。文章中，作者称该方法找到了新的激活函数 $f(x) = x \operatorname{sigmoid}(\beta x)$ ，而这个结果也在多个网络结构上表现出了更高的准确率。最后，在这篇文章里我们又再次看到了那个熟悉的示意图。



行动

虽然曾经有人告诉我，可解释性、泛化性对学术价值很重要。但作为一名小博士生，发现一个可以发表在顶级会议上的课题方向，总是像“老鼠遇到了奶酪”。上述三篇文章都是采用强化学习框架，使用 RNN 作为生成器，通过在特定任务上训练神经网络，将验证集上的性能作为 Reward 反馈给 RNN，进行不断搜索。需要变化的只是搜索内容从网络结构，到优化器，再到激活函数。沿着这个思路，我完全可以去搜索损失函数 (Loss Function)、批处理大小 (Batch Size) 等等。只要拥有足够多的计算资源，我们就可以扩大搜索空间，找到比目前更好的模型！换一个目标投一篇文章，这下文章有着落啦！说干就干，我打开我的笔记本，迅速敲下了代码，运行！趁代码在跑，我开启多进程开始写文章。忍住内心的激动，我开始憧憬文章列表长到一页纸放不下，想象着被当作大神簇拥着……

“梦”醒

等等，我再次打开我的小破本，看看模型训练那里纹丝不动的进度条，我看了一眼系统配置，我想，我用它几十年也训练不出来……我默默关掉了程序，删掉了 TeX 编辑器里写的 “Neural Loss Function Search with Re...”……

俗话说得好，识时务者为俊杰。俗话又说了，撑死胆大的，饿死胆小的。我想我是时候换一个梦了。灵光一闪，机智如我，一口茶的时间已经拟好了我新的星辰大海：

Adaptive [*] Search with Explainable [*].

关灯，睡觉。