初探计算机视觉的三个源头、兼谈人工智能｜正本清源

2016-11-27 视觉求索

谈话人：
杨志宏　视觉求索公众号编辑
朱松纯　加州大学洛杉矶分校UCLA统计学和计算机科学教授
Song-Chun Zhu
www.stat.ucla.edu/~sczhu

时间：2016年10月

杨：朱教授，你在计算机视觉领域耕耘20余年，获得很多奖项， 是很资深的研究人员。近年来你又涉足认知科学、机器人和人工智能。受 《视觉求索公众号》编辑部委托，我想与你探讨一下计算机视觉的起源，这个学科是什么时候创建的， 有哪些创始和代表人物。兼谈一下目前热门的人工智能。

朱：好， 我们首先谈一下为什么需要讨论这个问题。 然后， 再来探讨一下计算机视觉的三个重要人物David Marr，King-Sun Fu， Ulf Grenander以及他们的学术思想。我认为他们是这个领域的主要创始人、或者叫有重要贡献的奠基人物。

第一节： 为什么要追溯计算机视觉的源头， 这有什么现实意义?

中国有句很有名的话：“一个民族如果忘记了历史,她也注定将失去未来。” 我认为这句话对一个学科来讲，同样发人深省。我们先来看看现实的状况吧。

首先，假设你当前是一个刚刚进入计算机视觉领域的研究生，很快你会有一种错觉，觉得这个领域好像就是5年前诞生的。 跟踪最新发表的视觉的论文，很少有文章能够引用到5年之前的文献，大部分文献只是2-3年前的，甚至是1年之内的。现在的信息交换比较快，大家都在比一些 Benchmarks,把结果挂到arXiv 网上发布。 很少有一些认真的讨论追溯到10年前，20年前， 或30年前的一些论文，提及当时的一些思想和框架性的东西。现在大家都用同样的方法，只是比拼，你昨天是18.3%的记录（错误率），我今天搞到17.9%了。大家都相当短视，那么研究生毕业以后变成了博士，可能也会带学生做研究，他只知道这几年的历史和流行的方法的话，怎么可能去传承这个学科，让其长期健康发展呢？特别是等当前这一波方法退潮之后，这批人就慢慢失去了根基和源创力。这是一个客观的现象。

其次，还有一个现象是，随着视觉与机器学习结合，再混合到人工智能的这么一个社会关注度很高的领域去以后，目前各种工业界，资本、投资界都往这里面来炒作。所以，你可以在互联网上看到各种推送的文字，什么这个大师，那个什么牛人、达人说得有声有色，一大堆封号。中国是有出“大师”的肥沃的土壤的，特别是在这个万众创新、浮躁的年代。 这些文字在混淆公众的视听。也有的是一些中国的研究人员、研究生， 半懂不懂，写出来一些， 某某梳理机器学习、神经网络和人工智能的历史大事。说得神乎其神。我的大学同学把这种帖子转发给我，让我担忧。

杨：这大多是以学术的名义写的软文，看起来像学术文章，实际上就是带广告性质的，一般都是说创投、创业公司里的人，带着资本的目的，带商业推广性质的。

朱：我甚至不排除有些教授，比如与硅谷结合很紧密的、在IT公司或者风投公司兼职的，有意识地参与、引领这种炒作。

这对我们的年轻学生其实是很致命的，因为他们不了解这背后的动机， 缺乏免疫力。而且现在年轻人和公众都依赖短平快的社交媒体，很少去读专业文献。当公众的思想被这些文字占领了，得出错误的社会性的共识，变成了 false

common sense， 对整个社会， 甚至对学术界，都会产生长久的负面冲击。

这就形成了**新时代的皇帝的新装**。我们需要对这种现象发声， 做一些严肃的探讨。所以，正本清源有着重要的现实意义。

第二节：计算机视觉和人工智能、机器学习的关系

杨：谈到这里，我想先问一下计算机视觉和人工智能是什么关系？还有机器学习这三个东西。

朱：人工智能是在60年代中后期起步的。一直到80年代，翻开它的教科书，就是一些启发式搜索，研究最多的是下棋， 从国际象棋一直到最近的围棋，都是比较抽象的表达。棋盘的位置是有限的、下棋的动作也是有限的， 没有感知和动作执行的不确定性。 所有的问题都变成一个图搜索的问题，教科书上甚至出现了一个通用图搜索算法号称可以解决任何人工智能问题。当时视觉问题还没引起大家重视。我这里有一份1966 年7月 的 MIT AI 实验室的第100号报告（备忘录memo 100），很短，题目叫做“The Summer Vision Project”。这个备忘录的基本意思就是暑假的时候找几个学生构造一个视觉系统。他们当时可能就觉得这个问题基本上是不需要做什么研究的。所以你就一个暑假，几个人一起写个程序，就把它干掉算了。现在说起来，当然是个笑话。

人的大脑皮层的活动， 大约70%是在处理视觉相关信息。视觉就相当于人脑的大门，其它如听觉、触觉、味觉那都是带宽较窄的通道。视觉相当于八车道的高速， 其它感觉是两旁的人行道。如果不能处理视觉信息的话，整个人工智能系统是个空架子，只能做符号推理，比如下棋、定理证明， 没法进入现实世界。所以你刚才问到的人工智能和计算机视觉的关系，视觉，它相当于说芝麻开门。大门就在这里面，这个门打不开，就没法研究真实世界的人工智能。

到80年代，人工智能， 连带机器人研究就跌入了低谷， 所谓的冬天。那个时候，很多实验室都改名字了， 因为拿不到经费了。 客观来说，80年代， 一个微型计算机的它的内存只有640K字节，还不到一兆（1MB一百万字节），我们现在一张图像，随便就是几个兆的大小，它根本无法读入一张图像，还谈什么理解呢？等到我做博士论文的时候（1992-1996），我导师把当时哈佛机器人实验室最好的SUN工作站给我用，也就是32兆字节。我们实验室花了25万美元构建了一个图像采集系统，因为当时没有数字照相机。可以这么说，一直到90年代中期的时候，我们基本上不具备研究视觉这个问题的硬件条件和数据基础。只能用一些特征点的对应关系做射影几何，用一些线条做形状分析。因为图像做不了，所以80年代计算机视觉的研究，很大部分是做几何。

杨：90 年代后，就是数字照相机大量生产了。

朱：在90年代的末期的时候，发生了一个叫做感知器的革命。带动了大数据和机器学习的蓬勃发展。

杨：那机器学习与计算机视觉的关系呢？

朱：计算机视觉是一个domain， 它有很多问题要研究， 就像物理学。 而机器学习基本是一个方法和工具，就像数学和统计学。 这个名词的兴起应该还是最近的事情， 在我看来，是来自于两股人马。 80年代人工智能走入低谷后，迎来了人工神经网络的一个高潮， 所谓的从符号主义到连接主义的过渡。在中国80年代与气功、人体科学一起走红，但这基本是昙花一现。到了90年代初， 退潮之后，就开始搞 NIPS这个会议， 引入统计的方法来做。还有一股就是做模式识别的一些工程人员EECS 背景的。 按道理来说， 这个领域应该叫做 统计学习 （Statistical Learning），因为它的方法都是由概率统计领域拿来的。这些人中的领军人物很有商业头脑， 把统计和物理的数理模型， 改名叫做机器， 比如**模型（model）就叫**机（machine），把一些层次模型（hierarchical model）说成是“网”（net）。这样，搞出了几个“机”和“网”之后， 这个领域就有了地盘。另一方面，我的那些做统计的同事们也都老实、图个清静，不与他们去争论， 也大多无力去争。当然，统计学领域也有不少人参与了机器学习的浪潮。简单说，机器学习中的 “机器”就是统计模型，“学习”就是用数据来拟合模型。 是由做计算机的人抢占了统计人的理论和方法，然后，应用到视觉、语音语言等 domains。 我在计算机和统计两个系当教授， 看得一清二楚。 这个问题我以后可以专

门讨论。

这个机器学习的群体在2000年之后，加上大量数据的到来，很快就成长了， 商业上取得很大的成功。机器学习和计算机视觉大概有百分之六七十是重合的。顺便说一句，2019年我们两个领域会在一起在洛杉矶开CVPR 和 ICML年会， 我是CVPR19的大会主席。因为学习搞来搞去，最丰富的数据是在视觉（图像和视频）。现在这次机器学习的一些大的动作和工程上的推广工作，还是从计算机视觉这边开始的。

杨：谢谢你讲述人工智能,计算机视觉和机器学习的关系。下面我们回到本次访谈的主题。刚才说了这个感知器革命是90年代以后，出了很多的数据要处理了。那么为什么马尔（Marr）在70年代末思考的问题，在面对我们当今处理这个数据的时候，还有意义？就是说马尔用了什么方法？什么思路框架？使它有生命力？

朱：好，就回到1975-1980年这个时间段。我们今天的主题是想初步探讨一下计算机视觉的起源。我们这个领域也没有一个统一的教科书来谈这个事情。我认为视觉的起源，可以追溯到三个人，David Marr， King-Sun Fu 和Ulf Grenander。这三个人代表三个完全不同的方面，为计算机视觉这个领域奠定了基础。

杨：好， 我们逐个来介绍吧。

第三节：视觉的开创者之一：David Marr 的学术思想

朱： David Marr 【1945-1980】，中文音译为马尔， 他奠定了这个领域叫做Computational Vision计算视觉，这包含了两个领域： 一个就是计算机视觉（Computer Vision），一个是计算神经学（Computational Neuroscience）。他的工作对认知科学（CognitiveScience）也产生了很深远的影响。

我们计算机视觉CV，第一届国际会议ICCV 1987年就以David Marr的名字来命名最佳论文奖， 而且一直到2007年之前的20年间， 是CV唯一的奖项和最高的荣誉，两年一次。认知科学年会 （CogSci）也设有一个 Marr Prize给最佳的学生论文。这三个领域在80-90年代走得很近， 最近十多年交叉越来越少了。就是说，原来都是亲戚，表兄弟， 现在很少有人在之间走动了。

Marr 1972年从剑桥大学毕业，博士论文是从理论的角度研究大脑功能，具体来说，是研究的小脑， 主管运动的Cerebellum。1973年受MIT 人工智能实验室主任Minsky的邀请， 开始是做访问学者（博士后）。 1977年转为教职。可是， 1978年冬诊断得了急性白血病。1980年转为正教授不久就去世了， 时年35岁。他在得知来日无多后，就赶紧整理了一本书，就叫 “Vision：A Computational Investigation into the HumanRepresentation and Processing of Visual Information”，《视觉：从计算的视角研究人的视觉信息表达与处理》。他去世后由学生和同事修订，1982年出版。

杨：“Vision”2010年再版了，再版了以后在亚马逊仍然是卖得很好。

朱：它是个经典的东西。我是1989年冬天本科三年级从中科大认知科学实验室的老师那里，读到这本书的中文译本。因为缺乏背景知识，我当时基本读不懂。因为是中文，每句话都明白，但是一段话就不知道是什么意思了。在过去的20多年中， 我每隔1-2年都会再翻一翻这本书。后来我和同事花了大约8年时间，将他的一些思路转化成数理模型，比如primal sketch。

杨：这个人生故事是可以拍电影的。

朱：的确。 很多年前我与他的大弟子 Shimon Ullman饭桌上谈到这段历史， 他说当时大家到处找药，就是救不过来。当年这是一个30多岁正值科学顶峰的、交叉学科的领军人物。顺便说一句， 当年中日友好，1984播放日本电视剧《血疑》， 那是万人空巷， 感人至深。里面的大岛幸子（三口百惠饰）得的就是同样的病。

可惜， 目前计算机视觉这个领域，你如果去问学生的话，他们很多人都没听说过David Marr。“喔，想起来了，好像有个Marr奖吧”。可是你去问认知科学、神经科学的人，他们基本上对Marr非常的清楚。这也是我所担心的， 计算机视觉的发展太工程化、功利化了，逐步脱离了科学的范畴。这是短视和危险的。最近又受到机器学习来的冲击。

我这里顺便说一句， Marr 对我的另外一个间接的影响。他1973年来到MIT， 就租住在JayantShah的房子里， Shah 与 Minsky很熟， 他当时是研究代数几何（Algebraic geometry）的。 而我导师Mumford也是研究代数几何的， 并获得1974年的菲尔兹奖。他们两人很熟，后来在Shah的影响下，Mumford转入计算机视觉， 他们从提取物体边缘开始（boundarydetection），也就是产生了著名的 Mumford-Shah 模型，搞图像处理的应用数学人员基本都是从这个模型开始做。这是后话。关于这段历史，我们以后可以展开谈。

杨：好， 那么 Marr的学术贡献是什么呢？

朱：在我看来，David Marr对我们这个学科最主要的贡献有三条。从而基本上可以说，定义了这个学科的格局。

第一条，就是说在那个时代，60年代开始的时候大家已经很多人研究视觉神经生理学、心理学问题。也有人做一些边缘检测的工作。但是，视觉到底要解决哪些问题？是怎么实现的？大家莫衷一是，谈不清楚，那么David Marr的第一个贡献就是分出了三个层次。他说， 要解决这个问题，可以把它分成计算（其实应该说成是表达）、算法、和实现三层次。首先，在表达的层次，我们问一下这是个什么问题呢？如何把它写成一个数学问题。任务是什么？输出是什么？这是独立于解决问题的方法的。其次，对这个数学问题去求解时，可以选择不同的算法， 可以并行或者串行。再次，一个算法如何在硬件上实现，可以用CPU，DSP， 或者神经网络来实现。 很多观察到的心理学和神经科学的现象都是跟系统硬件有关的东西，比如说人的一些注意机制，记忆力。这些应该从表达层面剔除。这样， 视觉就可以从纯粹的理论、计算的角度来研究了。我们可以参考心理学和神经科学的结论， 但这不是主要的。 打个比方，要造飞机， 可以参考鸟类的结构， 但关键还是建立空气动力学，才能从根本上解释这个现象， 并创造各种飞行器， 走得更远。

杨：他这么一说，今天看来好像很自然的可以理解了，但是在当时，可能没有多少人，是把问题这样分解的。

朱：当时分不开。因为当时站在像神经科学和认知科学角度，是拿一些实验现象来说事，但是不知道这个现象是在哪一层出现的。

比如神经网络和目前的深度神经网络的学习，他们的模型（表达）、算法、和实现的结构三层 是混在一起的。就变成一个特用的计算设备， 算法就是由这个结构来实现的。当它性能不好的时候，到底是因为表达不对，还是算法不对，还是实现不对？ 这个不好分析了，目前的神经网络，或者是机器学习，深度学习，它的本源存在这个问题。

以前我们审稿的时候，会追问论文贡献是提出了一个新的模型？还是一个新的算法？在哪一个层级上你有贡献，必须说得清清楚楚。2012年，我作为国际计算机视觉和模式识别年会（CVPR）的大会主席， 就发生一个事件。收到神经网络和机器学习学派的一个领军人物 LeCun的抱怨信，他的论文报告了很好的实验结果， 但是审稿的三个人都认为论文说不清楚到底为什么有这个结果， 就拒稿。他一气之下就说再也不给CVPR投稿了，把审稿意见挂在网上以示抗议。2012 年是个转折点。

现在呢？随着深度学习的红火， 这三层就又混在一块去了。 一般论文直接就报告结果， 一堆表格、曲线图。我就是这么做，然后再这么做，我在某些个数据集上提高了两个百分点，那就行了。你审稿人也别问我这个东西里面有什么贡献，哪个节点代表是什么意思，你别问，我也不知道。那算法收敛了吗？是全局收敛还是一个局部收敛？我也不知道，但是我就提高了两个百分点。

杨：或者要用多少数据来训练材料才能够呢？

朱：对，这个也不用管，而且说不清。反正我这个数据集就提高是吧？所以从这个角度来讲，它就很难是一个科学的方法。可以认为它就是一个工程或者是一个经验的，有点像中医。那么要往前再发展的时候，你必须要理清楚这三层的事情。

杨：对。

朱：那么他第二个贡献的话，是理清视觉到底要计算什么。Marr提出了一个系列的表达，从primal sketch（首要简约图）， 到2 ½ D sketch（深度简约图）， 到3D sketch。 这里面还包含了纹理、立体视觉、运动分析、表面形状、等等。比如说我要估计一个物体的深度和形状，我就估计它的光照，和物理材料特性；还有，三维几何形状怎么去表达？ 他试图去建立一个完整的体系。

现在的视觉就基本上被很多人错误地看成一个分类问题，你给我一张图像，我说这个图像里有一只狗或者没有狗，狗在哪儿都不知道。头在哪？脚在哪？不知道。Marr框架是有秩序的，现在的秩序在做深度学习的人眼中还不存在，或者没有忙过来。各人做各人的分类问题，比如说有人算这个动物分类，有的人算这个家具的分类。各种分类以后，他们之间怎么样的关系呢？要对这个图像或者场景要产生一个整体的语义解释。

第三个贡献，Marr提出了一个非常重要的概念，到现在一直还没有一个完整的解答。他说，计算视觉是一个计算的“过程”。这是什么意思？ 我们以前用贝叶斯方法（以及现在的深度网络）认为视觉就是表达成为一个后验概率，寻求一个最优解。这个解就是图像的解释。这个求解过程就会终止。可是Marr说的这个事情，它不是单纯去求一个解，而是一个连续不断的计算过程。我给你一张图像，你越看、越琢磨，你可能看到的东西会越多。

我给你一秒钟，你可能看到某些东西。我给你一分钟，你可能有另外一种理解，这两个理解可能是不一样的。还有一个重要的概念是你的任务决定了你怎么去看这个图像，比如说我在慌忙之中在做饭，那么我对这个场景，只看其中的很小一部分，足够来完成我的任务就行了。里面好多东西改变你根本没注意到。

杨：好像有些魔术就利用了这一点。

朱：就是， 很多心理学实验表明，你眼睛盯着这个图片看的时候，眼睛不眨，我告诉你这个图片在改变。你盯着看，结果它改了你都没看见。在让你看这个图片的时候，把你的注意力引到某个任务所需要计算的关键要素上，其它部分你就视而不见。视觉是受任务驱动的。而任务是时刻在改变之中。 比方说， 视觉求解不是打一个固定的靶子， 而是打一个运动目标。

杨：这听起来是一个耳目一新的概念。

朱：回到人工智能这个问题，视觉，它最后的用途，要给机器人用，机器人目前面临一个什么任务，来决定它要计算什么。这第三个贡献是在算法的层面。就是说我根据我们目前面临的任务，我才决定要计算什么。而且人的任务是在不断变化的，在此时此刻我任务都在变化，那么计算的过程中是没完没了地在改变。这个理念到目前，我们目前在研究这个事情，还没有完全实现。就是说，这将是人工智能和机器人视觉的一个关键。

杨：明白。

朱：我们现在很多人研究这个智能，比如说分类问题。他都是从谷歌的一些应用，比如搜索图片、广告投放，变成分类问题。 从而忽视了更大的本质问题。如果说人工智能往前发展机器人，要从机器人的角度来用视觉的话，那么它就有很多不同的任务。我现在做饭，我在打球，我在欣赏风景，这个时候我看到的东西是完全不一样的。我怎么样通过这千千万万的任务，而不是简单一个分类，来驱动我的计算的过程，来找到我的需求，来支持我目前的任务，这是一个巨大的研究的方向。David Marr的思想，到今天，反而意义非常重大，因为大家现在一窝蜂的去搞深度学习，把这些基本东西给忘掉了。但是这才是人工智能和机器人视觉的长远发展方向。

我前两年给过几个谈话，说研究视觉要从一个agent（执行者）的角度，带着任务进来的这么一个人或机器人，主动地去激发视觉。

目前的计算机视觉的研究还有一大部分是由视频监控的应用来驱动的，比如说我检测一些异常现象，看这个人是男还是女？那这也是一种被动的，就是说它只是在看，没有去做。要去做的话，就涉及到因果关系和更多的不确定性。所以现在的研究生觉得，他整天在做机器学习， 就在调参数，就在跟别人比拼百分之几的性能。 一些公司的研究所就报道， 他们在某某问题（数据集）上国际领先了，排名第一了。他们自己也觉得这个研究没多少意思。那是因为他们没有接触到这些基本的问题上来。

杨：他们可能还没有发现这个问题本身是多么有趣。

朱：因为作为一个科学来发展的话，那它就是要认认真真的来做，把这个理清楚。当前的火热来源于工业界， 工业界没有多少耐心资助他们的研究人员去做科学研究，大家很现实。 那么，David Marr先谈这么多好不好？以后我们可能还会继续深入谈的。

杨：好。那我们第二个人就谈一下傅京孙。

第四节：视觉的开创者之二：傅京孙（King-Sun Fu）的学术思想

朱： David Marr是从这个神经科学和脑科学这个方向来的。傅京孙【1930-1985】，他当时代表的是计算机科学，搞人工智能的人。他是一个有领导才能的人物。他和其他人于1973年组织了第一届国际模式识别会议（ICPR），并担任主席。会议后来演变成国际模式识别学会IAPR，在1976年成立，并被选为其主席。他重组了另外一个IEEE学会下面的模式识别委员会，并于1974年成为其第一任主席，创办了IEEE模式分析和机器智能（PAMI）会刊，并于1978年担任第一任总编。这是目前计算机视觉和相关领域最权威的一本期刊了。很多中国学生现在不知道，这个领域的老大本来是华人。目前， 国际模式识别学会IAPR设立了一个傅京孙奖， 作为终身成就奖， 是模式识别的最高荣誉。

杨：可惜他1985年去世了。听说去世前他每年都在中国举办讲座，并于1978年担任台湾的中央研究院院士。

朱：我正要说的这一点。他去世的时候55岁，在普渡大学，据说他的实验室是一个Chinatown。1978年中国打开国门，中国最早的一批中科院的计算机人员都到他那里进修，在普渡。所以他对中国计算机的发展，可以说是一个贡献非常巨大的人。我也是受到他的恩惠，我大学一二年级就开始跟着科大陈国良老师学习，他之前去普度进修。周末我有时就到陈老师家听他讲外面的一些研究人员和工作。你想想，计算机界那时候华人在美国站住脚的可能没几个人。

杨：对，他对中国计算机发展真的是有历史性的贡献的。我在科学院上研究生的时候，我们那些老师是说他过世太早了，要不然对中国的研究还会更好，他多活10来年就会好很多。

朱：他1985年拿到一个很大的国家项目，好像是开宴会的时候心脏病突发了。 他要是活着，华人在这个领域的话，不止是现在这个样子。不过在他之后， 稍晚一点我们有另外一个杰出华人，黄煦涛（Tom Huang）。他当时也在普渡任教，培养了大量华人研究人员。 我们以后会专门介绍。

杨：傅京孙的故事也可以拍电影。

朱：这是我们这个领域的不幸，两个奠基人很快就走了。他们刚刚把这个地基打起来，人就没了。

杨：那傅的主要贡献是什么呢？

朱：傅京孙的贡献， 我也谈三点。第一个贡献应该就是对这个学科和学会的建设，以及工程师的培养上面，他起到了开创性的作用。一般公认他是模式识别的开山鼻祖，模式识别与计算机视觉分不开的。第二个作用，就是关于他的这个句法结构性的表达与计算，就是句法模式识别，Syntactic Pattern Recognition这个词，这个词其实非常深刻。他在走之前，他那个时候也没有多少数据，那么他只是画一些图，图表性的东西，来表达他的概念，他从计算机这边来的，你想很自然就会用到形式语言，因为计算机里面的几个基础之一是形式语言。逻辑、形式语言，对吧？

杨：这好像是在编译原理里面学到过，因为编译的基础是形式语言。

朱：我们这个世界的模式， 一个最基本的组织原则是composition。一张图像就像语言、句子符合语法结构， 视频中的一个事件也有语法结构。寻找一个层次化、结构化的解释是计算视觉的核心问题。从傅京孙1985年丢下来这个摊子后，基本很少有人去碰。差不多18年以后，我和我第一个博士生继续做图像解译Image Parsing这个方向，于2003年得了Marr马尔奖。然后我和我导师专门于2006年写了一本小书，总结了图像的随机语法。我刚才谈到了，在做识别，做分类的时候，只是单独在分类某一个东西，怎么去把各个识别器和分类器给它整合在一起，变成一个统一的表达？就必须产生一个结构上的表达。现在机器学习界把它换了另外名字，叫做结构化的输出，其实是一个东西。他们提出一个新的名词，把原创的图像解译名称覆盖住，这事现在经常发生。所以我说机器学习领域经常到别人那里偷概念，改头换面。数学界不允许这样做的。我还是坚持把它叫做解译、语法。

因为语法，它就是一些规则，其实语法并不见得是一个确定性的，它可以跟统计连在一块，它也可以跟目前的一些神经网络结合，这个都没问题。它表达了一个骨架或者支柱，形成一个统一表达。

第三点，从算法的角度来讲，有一个层次化的表达以后，意义就不一样了，比如自底向上或自顶向下的计算的过程就可以在上面体现出来，就是马尔说的计算的过程，就可以在这里面体现出来。视觉的计算过程应该是由大量的自底向上（bottom-up）和 自顶向下（top-down）过程交互和同时进行的。顺便再说一句，当前的深度神经网络就是一个feedforward的自底向上的计算， 缺乏自顶向下的过程。而在人脑计算中，自顶向下的计算占据很大一部分。

杨：那就是说， 这个语法结构对计算过程有了规范和表达的途路。

朱：对，你的搜索的过程，这个计算的过程是什么？马尔他提出了第二个概念，说视觉是个计算的过程，那么这个计算过程你什么时候算哪个，这是个调度的问题，就像操作系统。那么David Marr计算的过程，没完没了的，随着你的任务不断改变，那么它就有一个调度的问题。所以说我现在要去做饭，或者我要欣赏风景，或者说我要去走路，开车，那么它的不同的任务产生了不同的进程。这个进程，要在层次化的表达里面的统一起来调度。从这个意义看，感知是计算一个解译图（parse graph）， 认知是对这个parse graph进一步推理扩大， 而机器人的任务规划（task planning）也是一个同样结构的parse graph， 那就更别说语言是用parse graph来表达的。所以，人工智能的一个核心表达就是随机的语法和解译图。

杨：对。

朱：这个是绕不掉的，不管谁来做，都要做这个事情。当然，现在有人千方百计想绕过去，重新发明一套名词， 让新来的学生忘记历史， 这样他们就可以变成社会公认的大师。有些教授、研究人员在学术上没什么原创贡献， 却在网上、社会上成了当红明星，学科代言人。用社会上的知名度再给学术界施压。

总结一下，傅京孙三点主要贡献：一是学科的人才和组织基础，二是他提出这么一个的语法表达方法， 三是这个表达支撑了自底向上或自顶向下的计算的过程。他去世后， 这个方向一直处于一种休眠状态，我的研究有一条线是跟着这个方向做。2011年马里兰大学周少华他的导师有一个演讲，题目叫：语法模式识别--从傅到朱 （From Fu to Zhu）。我们在继承他的框架往前走。

杨：真好！那么咱们下面就谈第三个人Ulf Grenander。

朱：这个人的话，知道的人非常少。

杨：我翻看了网上资料，他是这个领域里头真正的是大神了，但绝对是个小众人物。

第五节：视觉的开创者之三：Ulf Grenander的学术思想

朱：Ulf Grenander 【1923-2016】是很少有人知道的。感觉有点像金庸小说《天龙八部》里的在藏经阁扫地的灰衣老僧。武功和思想都出神入化，但是，他基本是世外高人，不参与江湖争斗， 金庸也没有交代他的名字。所以江湖上的人大多没听说过他。 这样也好， 他自自在在活了93岁， 今年刚刚去世的。国际应用数学季刊邀请我和其他人写纪念文章，正准备出版专刊呢。

杨：对，我读他的生平，他这个人简直就是把欧洲美洲的，还有俄国的所有的精华的人物都接触过。

朱：那是，他出身在瑞典，他的导师叫Harald Cramér。概率论里面的一个重要的定理，还有数论里的一个猜想是用他命名的。然后，他也跟 Bohr（波尔），Kolmogorov（科尔莫戈罗夫）他们走得比较近。他的起点就是做概率统计， 时间序列， 随机过程，因为你现在想概率论和统计学的一些重要应用，就是那个时候发力了。

杨：从保险业开始了，北欧那边因为航海，保险业非常发达，所以这也有点道理。

朱：关于概率和统计学对于科学、视觉、以及人工智能的重要意义， Mumford 1999年写了一篇论文，是在一个大会的发言，叫做《随机性时代的曙光》（Dawning of the Age of Stochasticity）。

杨：对，那是你们老师写的， 网上能找到。

朱：他总结说，过去两千多年的西方科学的发展是建立在亚里士多德以来的数理逻辑基础之上的。但是，后面一千年包括人工智能、人的思维这些东西是随机性过程。人的思维应该是建立在概率推理基础之上。其实， 我们看到现在的机器学习， 人工智能完全就是从这个方向走了。

杨：你的导师说，整个世界的数学可以用概率的这套思想重新写一遍，就像罗素和怀特海的写这个数学原理似的，可以把数学重新建立起来，用概率的这种思想。

朱：这个工作已经有人做了。E. T. Jaynes就是发明最大熵原理的那个人，他写了一本很厚的书，《Probability Theory: The Logic of Science》， 他就是用这个原理去写。这也是一篇遗作。他没写完就过世了。这也是以后可以谈的话题。

朱： Ulf Grenander就诞生在这么一个概率发源的中心的地带，跟几个大师学习，博士毕业后出来游历，做概率论随机过程的这些东西。到六、七十年代的时候，他就开始提出来，想用数学来把这个模式识别与智能的现象的问题定义清楚。我们前面谈到的David Marr 是从神经科学、认知科学来的。傅京孙是一个计算机科学与工程的人。这两者基本没有多少严格的数学定义，提出的框架是漂浮的。Ulf是从数学的角度，奠定基础。他提出来一个应用数学的分支，叫做 Pattern Theory。他的出发点完全不同， 就是要给世界上的各种模式、现象， 建立一个数学的框架来研究。格局就很宏伟。而不是急于去解决某种实际问题， 后者叫做模式识别 （pattern recognition）。 他在90岁高龄出版了最后一本书， 想用数学来研究人的思想是从哪里来的。 你看我们脑袋里的念头、主意也往往是随机产生，像冒泡一样， 所谓思如泉涌。到底怎么来的？

杨：那太了不起了。这个事说起来，我想到当时我的老师是让我读Geman and Geman 1984年的吉布斯采样算法，那就已经了不起了。

朱：Grenander最后落脚在布朗大学应用数学系，Geman是他当年（70年代末80年代初）招到组里的年轻教员之一。这个吉布斯采样（Gibbs Sampler）的算法是一个里程碑的东西，在80年代初引起轰动。但那只是这个学派的诸多贡献的一个片段。

Grenander的理论解释起来的确有点费劲，既然谈历史，我先从我个人的经历谈一下。

他1994年出了一部总结性的书，900多页，叫做《General Pattern Theory》，广义模式理论。有点爱因斯坦做广义相对论的意思。但这本书很抽象， 没多少人读。我1995年在哈佛研究纹理模型（texture models），因为我用的学习算法就是吉布斯采样，在训练的时候，跑一遍要等两个星期才收敛，机器被占了，我就有时间，也是耐着性子把这本书读完了。我估计世界上不超过20人，能有耐心完整地读他的书。然后，我1996年1月答辩论文，我导师和我每周开车去布朗大学参加讨论。波士顿的冬天很冷， 哈佛到布朗1个小时左右，漫天大雪， 我们有时在高速上车被陷住，下来铲雪。到了6月， 我导师从哈佛提前退休，带着我一起加入布朗的应用数学系。那在当时是一个学术思想的中心。组会里有Grenander，Mumford， Geman 还有其他20来人， 一坐就是2个多小时。这些人都明察秋毫， 做报告的人无法含混过去的， 一步一步都必须理清楚，说不清楚你就下去想， 下次再来。

我一直认为计算机视觉和模式识别领域亏欠Grenander, 因为统计建模和随机计算逐渐成为我们领域的核心理论基础，而大家并不知道，很多思想、算法都源于这个人或者他的学派。所以，2012年， 我主持CVPR（国际计算机视觉和模式识别）大会， 特意放到布朗大学附近召开，我和另外两个主席一说，大家立即就同意了。并特制了一个银质的大奖章， 在大会上颁给他，表达我们的敬意。这里发生很多故事，我们以后再谈吧。

杨：那你能简短总结一下Grenander对计算机视觉、甚至人工智能的主要贡献吗。

朱：还是谈三点主要的吧。 首先，他提出了一个思想， 叫做 analysis-by-synthesis， 这是所谓 产生式建模的核心理念。当你要去识别、分析一个模式，比如一个动物，人脸， 一个事件， 你首先要建立一个数理模型， 这个模型通过数据来拟合， 也就是当前的机器学习。 那么， 判断这个模型好坏， 或者模型是否充分，的一个依据是什么呢？产生式建模的方法就是对这个模型随机抽样，也就是，合成（synthesis）。 我把这个过程直观叫做**“计算机之梦”**。计算机模型一开始初始化为空（完全随机）， 那它做的梦就是白噪声， 或者一张白纸。通俗来说， 这个模型就是一个“白痴”。人脑有这个功能，我们把眼睛一闭，没有外界输入了，就能做梦， 白日梦就是想象力的体现。一个好的模型采样产生的图片（模式）， 与真实观察的图片（模式）， 就应该是真假难辨。如果你能分辨，那说明这个模型不到位。 现在很多机器学习的方法是没法去随机合成图片的。 举个例子来说，我要检验你是不是真的听懂和理解中文，就看你能不能说流利的中文。如果你说话语法有错，词汇量不够，或者有口音，那就揭示你在哪方面还需要提高。

杨：这个要求好像比光是听懂 要更严格。

朱：的确。我们当年考英语， 多半是读，说和写都不行。我们考TOEFL， GRE Verbal的时候， 就算没搞懂， 也能蒙个60%-70%。 新东方的题海战术也很奏效。当你做了大量考题， 就算不懂， 也能考好。当前大数据、机器学习就用题海战术。 这个方法强调在实战中检验，考什么就拼命复习什么，不考的东西就不学，这也很有道理，很直接， 来得快。 但是， 因为你的模型没有真正理解， 没有“真懂”，考试大纲外面的东西更不懂， 那么后遗症就是， 遇到新考题， 缺乏泛化能力，遇到新问题，缺乏创造力。

想一想， 如果我的学生一步步考试都是靠题海战术这么学过来的， 那多可怕，要让他们去搞研究、创新，那就基本不可能。很遗憾的是，现在中国学生从幼儿园开始，就是在题海中泡大的。机器人、人工智能，靠题海战术是可以演示不少功能的， 但是， 那还离真正的智能比较遥远。

杨：好， 我明白这个analysis-by-synthesis 的意义了。他的第二贡献呢？

朱：他提出了一整套建模的理论和方法。把代数、几何、概率整合起来。 代数指的是一些结构，比如群论， 记得在科大本科我学过 群、环、域这些概念吧？也就是说我有一些基本元素，叫 generator，连接成为图graph，然后是群group，在上面进行操作，产生了各种各样的变化。还有很多几何， 变换， 在连续情况就产生形变。通过组合，语法、产生丰富的图模式。然后，再在这个图模式的空间上定义距离（测度）和概率。

朱：比如一个概率模型， 是定义在一个什么样的结构上，它是个什么样的解空间？这个数理上你必须交代清楚，否则你的论文写不下去了。现在它的一个很大的应用在医疗图像上面，比如说一个病人，他的肝变形了，那么他的肝的形状和正常人的肝的形状之间怎么定义一个合理的距离？两张人脸，怎么定义这个距离的呢？这个距离定义在一个流型上，数学的流型（manifold）。

杨：这些东西真用上了吗？

朱：他有个Postdoc，名叫Michael Miller， 现在是Johns Hopkins 大学图像中心主任， 就用这一套方法来做医疗图像、脑科学（Brain Mapping）等方面的应用。

杨：他的第三方面的贡献呢？

朱：第三个方面主要是算法上面。当我们去做求解的时候，在一个解空间，这个求解空间肯定是一个非凸的，他有千千万万的局部最优解local minimum 在里面。

杨：对。这是当时八十年代的时候提出来一个很尖锐的问题，好像有什么模拟熄火方法。

朱：很多蒙特卡洛算法都是他和这个学派的人提出来的。这个解空间是一个异构空间，空间里面非常复杂的，包含有很多子空间，子空间里面又包含又子空间，每个子空间维度又不一样，他们之间，从一个解跳到另外一个解的时候，这跳转必须是可逆的。在计算机里面就叫可以回溯。从这个学派走出来的人，他们设计算法每一个步骤都是有章法的，要做到合规合矩。包括上面提到的吉布斯采样算法、可逆蒙特卡洛跳转法，还有变分法（variational methods）和偏微分方程式， 还有一些随机下降法（stochastic gradient）， 这后者是目前训练深度学习模型的主要办法。他也开创了非参数模型的学习方法。这里面东西太多，先谈到这里吧。

正因为很多人没有接触过Grenander的理论， 缺乏这方面的理论素养， 造成我们学科发展的一个巨大的问题：很多教授、博士、研究生就是用别人的模型（机），拿来调试，基本缺乏自己发明新模型、新算法的能力。我们这个领域，很多美国名牌大学助理教授、副教授、教授， 他们的论文中的公式错误百出。现在干脆大家在论文中都不写公式了，直接报告最后的实验结果，提高了几个百分点。这就“一俊掩百丑”了。 英文有个类似的说法叫做 “sweep the dirt under the carpet把污垢扫到地毯下”。 这些人在大量培养博士、他们出来的人评审论文。 这样一来，学科的发展堪忧！

第六节：结束语

杨：听了你番谈话，我明白很多。记得我当时念研究生，包括念博士生的时候，实际上是很糊涂的。就是对这个领域到底做多少东西，没有信心。觉得很多研究像画鬼一样，原理不清楚。我觉得那样的话，与其那样做事情，那不如干脆到工业界那更快乐。

朱：正因为我们这个领域很多历史、框架性的东西，没有搞清楚，培养出来的博士，缺乏分析能力。大家被一些工程的任务和数据驱动，被一些性能的指标牵制，对科学的发展比较迷茫。

杨：好， 谈了很多， 我们做个总结吧。

朱：那我就说两点。

首先， 我在开场白中提到 “一个民族如果忘记了历史,她也注定将失去未来。”一个学科要健康发展，需要研究人员、研究生们理解自己领域的历史和大的发展方向，建立文化的认同。否则，自己家的东西，被别人偷取，浑然不知。就像日本打入中国，想把我们的地名改掉，大家开始说日语，把名字都改做山本太郎之类，感觉很酷吗？ 或者是韩国人把中国的文化拿去申报世界文化遗产，这都是要制止的。否则，过了一代人，还真说不清楚了。我记得刚来美国的时候，美国同事把汉字叫做“Kang-ji”，说是日本字。 我们领域很多人对保护这个领域的文化和传统缺乏清醒认识。皮之不存，毛将焉附？

其次，一个学科内部，大家互相不够了解，各自为政。特别现在会议审稿人很多是研究生，以自己的狭窄的眼光和标准去评判别人的方法，造成很多混乱。搞工程的看不到理论的重要性，反之亦然。大家又都疏远心理学和认知科学的研究。我提倡我们的研究人员、学生要提高理论修养、培养长远眼光，向相关学科取经，取长补短。

我希望这个微信公众号，能够帮助大家正视问题，让计算机视觉这个领域健康、稳健、可持续地发展。