北京智源人工智能研究院 采访人：卢凯

在人工智能领域，尽管以深度学习为代表的 AI 算法正受业界热捧，但它的“黑盒”式学习模式给整个行业的进一步发展蒙上了隐忧。用中国科学院院士、智源研究院学术委员会主席张钹在某次接受媒体采访时的观点来说，便是现在“AI 奇迹短期难再现，深度学习潜力已近天花板”：算法只有“相关性”，而无“因果性”，将带来人工智能系统根基的脆弱性、可欺骗性等深远隐患。这个观点，已经在中国人工智能业界引起了重大反响和一定程度的共识。

早在 2019 年 4 月，智源研究院便已发布了重大方向“人工智能的数理基础”，并支持诸多中国顶尖数学家，致力于研究、解决人工智能领域最根本的理论挑战，比如可计算性、可解释性、泛化性、稳定性等。那么，让数学家们解决人工智能的问题，他们主要的特点和优势会是什么？他们将给人工智能产业注入哪些令人耳目一新的血液？为此，我们采访了智源研究院“人工智能的数理基础”成员之一、北京大学数学学院教授李铁军。李铁军教授是国内计算数学领域随机算法方面的学术带头人，已经在随机算法和模拟方面的诸多领域取得了突出成果。

1 数据科学，将会和物理科学分庭抗礼

智源：能否谈谈您从事数学研究的主要经历，您为什么会对数学感兴趣？

李铁军: 我小学时便爱上了数学，到了初中时，记得读了一本名叫《数学五千年》的科普书，它讲述的古代希腊、巴比伦、中国还有近代的数学史等，进一步激发了我对数学的浓厚兴趣，由此萌生了将来要做一名数学家的念头。等到高中毕业，我的高考成绩是全县第二，便选择了清华大学数学系，先后读了本科生和研究生。硕士毕业时也曾一度为职业道路而迷惘，因为当时从事数学研究的收入是非常低的，所以很多同学都选择高收入的计算机行业，甚至下海经商等。就在这个人生十字路口，某位师兄对我说了这么一番话：“为什么要为了挣钱去做事？应该去做自己喜欢的事业”。这番话点醒了我。我喜欢的就是数学，它应用面非常广，却有一种简单之美；每当我去学习那些数学算法和定理的时候，总觉得很愉快——数学就是我愿意一辈子去做的事情。于是，我选择了去中国实力最强的北大数学院读博士，毕业后一直留在那里从事数学教研工作。

智源：您在北大的研究方向是什么，是什么契机让您开始关注人工智能领域的？

李铁军：我在北大读博期间，北大应用和计算数学领域的研究方向，正在经历一个比较大的变化。之前受到冷战影响，北大以应隆安、滕振寰老师那一代人为代表，研究方向和国防应用密切相关，比如空气动力学、激波的计算等等，做的主要是双曲型守恒律的计算研究。但是伴随着 1991 年苏联的解体，美国认为最大的威胁已消除，其应用和计算数学领域的研究方向开始变得分散。伴随着计算方法的成熟，新兴领域如计算生物、计算材料、计算化学甚至图像处理等研究方向开始得到长足的发展。当时，北大以张平文为代表的年轻一代老师，他们的研究视野，便代表了应用和计算数学的这一最新潮流，于是滕振寰老师便推荐我在张平文老师的指导下读博士。博士毕业之后，受张平文、鄂维南两位老师的指引，我将自己的的研究方向定位为随机模型及算法，并跟随两

位老师研究“复杂流体”，具体来说就是研究流体中高分子的行为。举个例子来说，鸡蛋清里有很多蛋白大分子，当你去拉蛋清的时候，它会表现出很强的弹性和粘性，如果你想描述这个特性，就需要用随机模型来描述蛋清中蛋白质大分子的微观构型。最近几年，我主要侧重研究生物学里的随机模型。

我关注人工智能，是深受鄂维南老师的影响。我觉得几乎可以说，鄂老师是中国数学领域最早对“AI 的数学基础”感兴趣的学者；他作为应用数学领域的大师，对于应用数学领域新兴动向的眼光、见识绝对是超人一筹的，比如早在 2000 年左右，他就开始做“稀有事件”、“多尺度”等数学理论了。记得还在 2012 年前后，鄂老师就几乎每次在我们应用数学圈子的会议里，不停地宣传一个观点：“数据科学将会和物理科学分庭抗礼”。当应用数学领域还在津津乐道于物理科学范畴：计算物理、计算化学、计算生物和计算材料等，并普遍认为数据科学属于计算机科学的范畴，一个数学家却跳出来说了这样的判断，这给人感觉的份量是不一样的。当然，我相信当时很多人是怀疑这个判断的，但事实正越来越证明他是对的。比如现在有些做计算物理的人，开始用深度学习（Deep learning）来建模，或者用深度学习来求解密度泛函理论，求解偏微分方程等，可以说深度学习正在科学的各个领域里迅速壮大，这样的发展趋势正在印证鄂老师的那句话。总而言之，在鄂老师的影响下，我也萌生了对数据科学、对 AI 的兴趣，想看看从数学的角度，基于 AI 能否挖掘出一些值得去研究的问题，而了解的结果是，它还真的存在一系列问题亟待去解释清楚。

2 数学看人工智能：“逼近论”是一个根本问题

智源：从数学家的角度，您觉得当今 AI 算法里有哪些问题亟待解释清楚？

李铁军：一个很基本的问题就是所谓“逼近论”问题，它包括：能否逼近？如何逼近？逼近之后如何论证它的泛化性？

在传统数学领域，对于一个函数空间，我们会用多项式、分片多项式去逼近。

但这些方法会遇到“维数诅咒”（Curse of Dimensionality），即遇到高维就失效了——它的自由度随着维数的增加而指数增加，目前来说到了“4 维”这些算法已经基本上没有办法了。更高维的现在一般用“蒙特卡洛算法”，但它不在通常的研究范畴之内。转机发生在 2016 年 AlphaGo 出现以后，应用数学领域的人发现原来在计算机科学领域，大家正一直在尝试着另外一种路径：Deep learning。

对于高维逼近，Deep learning 整个运算非常简单：它通过一层层的输入（input），每层与一个权重矩阵（weight）相乘，然后再加上偏置量（bias），再通过激活函数做个非线性变换，然后进行一层层的函数复合。它借鉴了人脑神经元的工作方式，而且表达出来的是一种相对比较简单的模式，却非常有效地解决了高维逼近的问题。这深深地刺激了应用数学圈子，因为之前几乎没有人做这方面的研究。但是这个思路，萌发了“数学理解 AI”的最根本问题之一：能否对神经网络逼近任一高维函数的有效性进行论证？最关键的是，能否对它的逼近做精细的误差估计？

当构建了一个神经网络去逼近一个已知函数后，接下去的问题就是如何去训练它？就是说，如何设置有效的算法去自动调配那一大堆权重和偏置量更好地逼近待求的函数。现在这一领域，最基本的方法是随机梯度下降法，就是 SGD (Stochastic Gradient Descent)。但问题是我们能不能将 SGD 理解得很清楚，从而改进和优化它？尽管目前已经有很多尝试了，比如关于 SGD 的各种变形，但它还是需要被进一步地理解和深化。鄂维南老师已经做了一个研究，他把 SGD 这一离散时间迭代法，用动力系统的语言，即随机常微分方程进行了描述，并证明了 SGD 是这一方程的弱一阶逼近。这类研究的意义在于，将机器学习里不太容易理解的离散算法，建立其与连续型数学问题之间的联系，从而可以借助相关成熟的数学工具进一步去分析和处理它。

第三步的问题也很重要，就是训练完的神经网络，如何评估它的泛化性（generalization）？现在神经网络进行训练的方法，基本上是 SGD 及其变种，

它们伴随着神经网络能量下降的过程——即“损失函数”。如果用简单梯度下降法进行训练，训练对象会从一个初始状态开始，最终会停止于被称为“局部最小”的地方；如果用 SGD 训练，它在过程中会伴随着噪声，能量呈现一定程度的起伏，但最终也会停留在“局部最小”的地方。此时，我们往往就认为一个神经网络已经训练好了。但问题的关键就在这里，损失函数是个“高度非凸函数”，得到的“局部最小”一定不是“全局最小”，你凭什么说基于“局部最小”而训练出来的一堆参数（如权重，偏置量等）就是好的呢？比如一个图片分类应用，基于猫，狗，猴等带有标签的图片进行训练，仅是基于已有数据的局部最优（并非全体数据，也非全局最优）训练出了一个神经网络系统，当我们输入一张不是训练集的图片，而结果却是系统准确地给出了结果，这是为什么呢？它为什么有比较好的泛化能力？这个问题至今也没有人解释得清楚。

以上三方面，我觉得是人工智能领域中涉及的几个基本问题，单凭计算机科学家不一定能回答它，因为对其理解一定需要大量的数学。

3 数学家 VS 计算机科学家：“为什么”和“怎么做”的分野

智源：请归纳一下，数学家和计算机科学家，在面对人工智能时，研究视角的主要不同点在哪里？数学家的主要优势是什么？

李铁军：我的理解，计算机科学的人去研究人工智能，往往是设计某个聪明的算法去解决一个具体的问题，而做数学的人总是要去理解：为什么是这样，并建立某种更深层的理论框架去解释它。我从事应用数学十多年，期间分别和做生物、物理以及做计算机工程的人合作过，我发现他们思考问题的方式是这样的：现在有一个问题，我要找到一个很巧妙的方法去解决它。而这个方法为什么好，他们似乎不太关心。但我想强调的是“理解很重要”，为什么呢？你只有理解了这个方法为什么好，你才可能设计出更好的方法去改进。在人工智能领域，我觉得数学家如果要能比计算机科学家做得更好，就不能仅仅停留在实验阶段，而是要深入研究其内在的机理。比如现在的神经网络，

要么解决一个分类问题，要么应对一个回归问题，至于搭建的网络到底应该设多深多宽，到底需要什么特殊网络结构，都不知道，反正就是做“试验”，最简单的就是用“多层感知机”搭建起来，随后则利用 Tensorflow 这样的软件，让你连程序都几乎不用怎么写，直接把数据输入进去，软件自动帮你做一个 SGD 确定参数，等结果出来后再输入新的数据试验，随后再看看效果、看看误差等……但这种方法为什么好，好到一个什么程度，把这个方法修改后能否达到一个更好的逼近？没有任何一个定理可以告诉我们。我觉得数学家应该往下更进一步，去解释它为什么。

智源：那相比于计算机科学家，数学家研究人工智能有哪些劣势？

李铁军：数学家的劣势在于，他发表成果论文一般没有计算机科学家那么多、那么快；在解决具体问题的方法论设计上，计算机科学家也往往有更多的经验，更加熟练。计算机科学家有点像经验丰富的老中医，看到病人的症状，立刻知道他（她）大概需要吃什么药；而数学家则更像西医，力求归根结底地将病因机理解释清楚。

4 能量景观分析：从蛋白质中汲取的 AI 数学灵感

智源：您在北京智源人工智能研究院承担的课题是“深度神经网络的能量景观分析”，请谈谈什么是“能量景观分析”？

李铁军: 这个课题和我之前做的“稀有事件”研究有一定关系。什么是稀有事件？它假定有一个能描述物理对象的能量函数。我们能看到的物体，往往是它能量状态比较低的状态——能量状态越低，越稳定，被看到的概率越大。比如山峰上放一个小球，就不容易被看到，因为它如果被风一吹，受到扰动，就往势能更低、也更稳定的地方去了。假如用能量函数描述这个特征，它一定是非凸函数，因为它不止一个“井”（即：局部极小值），但如果遇到噪声，就会在井底被往上推，时间足够长之后，它可能就会翻滚到另一个局部极小值的井底去

了。由于这个事件发生的概率极低，等待的时间很长，便叫做“稀有事件”。

稀有事件，在蛋白质折叠领域是一个非常根本的问题。我们都知道蛋白质是一个氨基酸序列，但它只有在三维空间里呈现特定的构象，才会真正发挥作用。这种结构，便是所谓基态，它正好对应了蛋白质能量最低的状态。这样，当我们知道了蛋白质的氨基酸序列，知道了原子的空间位置，就可以利用数学公式计算出它的势函数，而蛋白质真正稳定的状态，便是势函数的值最低的状态。势函数的全局极小化，这也是计算生物学领域一直想要解决的基本问题，也叫“蛋白质折叠”。根据大家积累的研究经验，我们已经知道蛋白质的能量势函数，就是一个高维的高度非凸函数，因为假定一个蛋白质有 1000 个原子，那就有了 3000 个维度。这样一来，它的问题情境和神经网络就非常类似了，两者都是一个高维非凸函数，对应于蛋白质折叠过程中的原子位置是一个高维空间，神经网络中的权重、偏置量等也是一个高维空间，它们都要经历一个进行“极小化”的运算过程。鉴于在蛋白质领域，经过几十年的研究，已经积累了一些可行方法，来构建势函数从初始状态，到局部极小，到全局极小的路径图谱——我们将它称作为“能量景观”；所以我萌生了这个想法，通过借鉴蛋白质折叠问题的能量景观理论，寻找不同神经网络结构的“能量景观”构型，为解释神经网络的根本问题，以及算法设计提供一些洞见（insight）：比如解释神经网络为什么具有好的泛化能力，力图提出一些优化或者改进 SGD 的算法等。

智源：“能量景观分析”应用于人工智能，目前国内外主要有哪些研究进展？

李铁军：国际上有一些研究，但主要基于计算的某个“局部极小”来做局部的能量景观分析，目前他们得出的主要经验结论是：在能量景观图谱里，邻域比较“平坦”的局部极小，泛化能力比较好，相反“尖锐”、“陡峭”的局部极小，泛化能力比较差，但是也有文献对此持不同的观点。在这一问题的研究中有一个非常有趣的结果：研究者针对残差神经网络 (ResNet) 的能量景观在一些局部极小态处进行过一些二维切片式的探索。他们发现，对于残

差神经网络，如果不加入所谓的跳联 (skip connection)，此时得到的能量景观二维切片是一个非常粗糙、高度非凸的能量面；而如果加入了跳联之后，能量景观的二维切片就变得非常光滑而且接近于一个简单的凸函数。研究者基于这样的结果来说明由微软的何恺明等人所提出的 ResNet 为什么在加入跳联之后的训练变得有效且易于实现。这一探索对于神经网络损失函数的构造提供了一些洞见。但是总的说来，他们做的还只是局部、低维的能量景观分析，而我们希望做的是神经网络的全局能量景观。

智源：您觉得在研究中将面临的主要挑战是什么？预期什么时候能产生研究成果？

李铁军：这需要一步一步来，构建能量景观并不容易。我希望先在半年之内有个初步的结果：先构建出简单的神经网络的能量景观图，然后再根据它们作进一步的细化分析。我们将根据四种典型例子来构建神经网络，前两种都是 toy model（玩具模型），先构建一个回归网络、一个分类网络，它们的数据相对简单，比如三、四个混合高斯，然后将它们进行一番神经网络训练，根据训练得到的数据集再进一步构建我们的能量景观；另外的，我们将采用 MNIST 和 CIFAR 两类典型数据集来构建神经网络的能量景观。

主要挑战在于它的计算量大和维数高。生成能量景观图不是一件容易的事情。尽管我们实验的是小规模神经网络，那也至少得有成千上万个参数。在这一个高维的空间里，我们需要用 SGD 产生大量的轨道，每次需要很长的时间进行模拟，得到一个高维空间的一堆点云，然后再基于这些点重构马尔科夫链并进行分析等，所以它的计算量是非常大的。现在我们也是摸着石头过河，一步一步来。科学研究，总归是会有一些不确定的因素。

5 寄语青年学子：在开放和交流中研究

智源：能否结合您自己在研究方法等方面的成功经验，给青年学者、学生们提几个学术成长方面的建议？

李铁军：对于从事应用数学这类交叉学科的人，一定要有开放的心态，重视跟不同人的学术交流。作为“交叉学科”，我们不能“闭门造车”，因为研究的最终目的是要在实际中产生影响。比如我们所在智源研究院的“人工智能的数理基础”课题组，每两个星期就会举办一次交流会，介绍课题领域相关问题的重要文章，包括我们智源学者、智源青年科学家以及各自带的学生们在内，一般有 50~60 人参加。虽然大家做的都是“人工智能的数理基础”，但是从事的具体领域都不一样。这样，你便能有机会了解对他们来说很重要的问题，寻找到有望合作的共同点等等。此外，在这样的学术交流中，我发现即使是学生们的分享也能带来很大收获，他们思维敏锐、开放，有很强的接触新鲜事物的能力。

同时，广泛阅读 paper（论文）也很重要。我常跟学生们说，读论文就是寻找新的刺激，避免大脑走到死胡同里。当你读完一些论文，受到一些新问题、新思想的刺激，已经很明确了你需要深入思考的问题后，此时你不妨暂时“关上门”，创造一个不受干扰的研究环境，专注地去思考它。等解决之后，再“打开门”分享出去。

此外，对于计算和应用数学的学生们，如果条件允许，我建议采用一种“双导师”模式，就是两个导师分别来自数学专业和计算机专业。对于数学专业的导师，可以帮助学生打好数学方面的基础，培养深入思考的能力；对于计算机专业的导师，可以帮助学生们开拓视野，及时了解领域的最新动态。我觉得，双导师模式对于学生们的学术成长可能比较有利。

后记

对李铁军老师的采访，有三点令笔者印象深刻：首先，在李铁军老师不足 20 平米的办公室内，摆放着一辆老款式的自行车；其次，办公室墙上贴着一张普通 A4 纸，上面写着：

“人生就是一次没有终点的长跑，每个人尽他最大的力量和意志到达最远，

有人擅长短跑，但是不一定能坚持到最后；有人会暂时领先，但是并不意味着后来者就应放弃竞争，在这场无休止的长跑中，每个人最大的竞争对手乃是他自身的消极，气馁和惰性，因为惟有战胜自身才是最大的成功！”

李铁军老师解释说是有感于很多比他更优秀的同学们，因为各种原因没有坚持下去，最终放弃了对学术的追求。“在我看来，做学术是一件更有意义的事情”，李铁军老师说。

第三点，便是李铁军老师在访谈中，描述数学如此吸引他的原因——美。这让人初觉有点抽象，随着笔者整理完他对人工智能算法根本性问题的剖析，令笔者一度觉得高不可攀的深度学习圣殿，居然就在李老师聊聊数语的描述中、如“庖丁解牛”般被清晰地拆解了。笔者觉得获得的启发，甚至远超以往阅读过的所有厚厚人工智能书籍，由此深深感受到了一个数学家的深邃和简洁……数学真美！

北京智源人工智能研究院 2020 年 2 月 11 日网络版。感谢采访人卢凯对本文转载的支持。