文章编号:1671-9352(2024)01-0001-10 DOI:10.6040/j.issn.1671-9352.0.2023.512

策略极限理论与策略统计学习

严晓东
(山东大学中泰证券金融研究院, 山东 济南 250100)

摘要:非线性期望是山东大学彭实戈院士开辟的原创性研究方向之一,对各个领域的科学研究越来越重要,而大数据和人工智能的兴起,为非线性期望创新理论与应用研究提供了更强劲的动力。最近,山东大学"非线性期望"团队基于多臂老虎机的策略博弈过程开创了"策略极限理论",是非线性概率理论与强化学习交叉的重大突破性科研成果,变革了传统统计方法研究范式。本文结合徐宗本院士提出的人工智能的 10 个重大数理基础问题,国家自然科学基金委员会发布的 2022 年度重大研究计划项目中关于可解释、可通用的人工智能方法的申报指南,以及科技部发布的数学和应用研究重点专项 2021、2022 年度项目中"数据科学与人工智能的数学基础"理论研究的申报指南,采用"策略"这一概念探寻和揭示人工智能本质和规律,尝试启发、促动人工智能技术变革的激发源和理论依据。不同于传统的大数定律和中心极限定理在独立同分布假设下开展统计学习的研究,策略极限理论打破了数据可交换这一局限,在更大的概率空间中探求最优分布,并提出获得最优分布的最优策略路径,与之对应的统计学习过程被命名为策略统计学习,为复杂机器学习的可解释和可信赖的统计方法研究提供理论支撑。本文介绍策略极限理论的应用包括但不限于:(1) 大规模数据的策略抽样;(2) 数据流的在线学习;(3) 强化学习的中心极限定理;(4) 数据的差分隐私保护;(5) 联邦学习的策略融合;(6) 迁移学习和元学习的信息重构;(7) 知识推理与数据驱动的融合。

关键词:人工智能;策略极限理论;数理基础;大数据分析;强化学习;在线学习;迁移学习;联邦学习;数据隐私保护;知识推理与数据驱动
中图分类号:TP18;TP181;O211.4 **文献标志码**:A
引用格式:严晓东.策略极限理论与策略统计学习[J]. 山东大学学报(理学版),2024,59(1):1-10,45.

Strategic limit theory and strategic statistical learning

YAN Xiaodong
(Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan 250100, Shandong, China)

Abstract: The nonlinear expectation is an original research direction pioneered by Academician Peng Shige of Shandong University, which is becoming increasingly important in various fields of scientific research. The rise of big data and artificial intelligence has provided stronger impetus for innovative theoretical and applied research in nonlinear expectation. Recently, Shandong University's Nonlinear Probability Team has developed the "Strategy Limit Theory" based on the strategic game process of multi-armed bandits, representing a significant breakthrough in the intersection of nonlinear probability theory and reinforcement learning. This has transformed the research paradigm of traditional statistical methods. Based on the proposed 10 basic mathematical problems of artificial intelligence by Academician Xu Zongben, the declaration guide of 2022 major research plan projects issued by the National Natural Science Foundation of China for the research about universal and interpretable artificial intelligence technologies, and the application guide for basic mathematical theory research of artificial intelligence in 2021 and 2022 the key projects of "Mathematics and Applied Research" issued by the Ministry of Science and Technology, this article adopts the concept of "strategy" to reveal the nature of artificial intelligence and explore and the motivation source and theoretical basis for initiating and promoting the innovation of artificial intelligence technology. Different from the applications of the traditional law of large numbers and the central limit theorem in the field of artificial intelligence, we propose novel theory about the strategic law of large numbers and the central limit theorem

收稿日期:2023-12-05; **网络出版时间**:2023-12-22 13:59:37
网络出版地址:https://link.cnki.net/urlid/37.1389.N.20231222.1044.002
基金项目:国家自然科学基金资助项目(12371292); 国家统计局统计科学研究资助项目(2022LY080); 科技部国家重点研发计划资助项目(2023YFA1008701)
作者简介:严晓东(1988—),男,副研究员,博士生导师,博士,研究方向为机器学习、计量经济、金融科技和大数据统计分析.
E-mail:yanxiaodong@ sdu.edu.cn

in the new generation of artificial intelligence. The discussed topics in this work include but not limited to: (1) strategic sampling of massive data; (2) online learning of streaming data; (3) the central limit theorem of reinforcement learning; (4) differential privacy protection of data; (5) strategic integration of federal learning; (6) information reconstruction of transfer learning and meta learning; (7) the fusion of knowledge reasoning and data driving.

Key words: artificial intelligence; strategic limit theory; mathematical foundation; big data analysis; reinforcement learning; online learning; transfer learning; federated learning; data privacy protection; knowledge reasoning and data driving

0 引言

随着新一代科学技术的发展,人工智能(artificial intelligence, AI)技术已经在科学、社会、经济、管理等领域展现出解决各类复杂问题的优势并得到广泛认可,成为创新驱动发展的核心驱动力之一[1-3]。大部分人工智能技术得到很好的落地应用,但是在其他科学问题的挖掘上存在诸多瓶颈,例如基本原理、数学理论等,影响着可解释、可通用的下一代人工智能方法的开发。近年来,国家自然科学基金委员会在重大研究计划项目申报指南以及科技部在数学和应用研究重点专项申报指南中多次以"人工智能技术的数学理论"进行命题,欢迎科研工作者答卷。徐宗本院士提出人工智能的 10 个重大数理基础问题[4],把相关科学问题的理论体系或框架明确细化,为下一步人工智能技术的基础理论研究指明方向。

本文首次采用"策略"这一概念探寻和揭示人工智能本质和规律,尝试启发、促动人工智能技术变革的激发源和理论依据。若干机器学习技术的智能表现源于经验学习,即通过总结历史信息进行下一步决策,例如循环神经网络通过信息在神经网络主题的多次循环,用前一个事件去推理后一个事件,做到持续记忆;强化学习通过引入策略变量记忆历史回报与状态;迁移学习利用以往任务中学出的"知识",比如数据特征、模型参数等,来辅助在新领域中的学习过程,得到更优秀的模型;元学习利用跨任务知识或元知识可以更好地学习每个新任务,提升泛化能力。

将人工智能技术应用到其他学科领域中碰到的问题,其根本在于算法缺乏解释性、稳健性等,归根结底是概率统计问题,难点是带有策略的机器学习存在分布不确定性,概率分布随着策略改变而变化,无法用传统的线性概率知识体系去解决。基于此,发掘一些新的非线性概率统计理论,构建基于策略的统计分析过程,提供所研究问题的形式化手段、模型化工具和科学化语言是揭示人工智能规律的重要方向。

非常巧合的是,不同于科尔莫戈罗夫(Kolmogorov)的概率论框架体系,在金融风险研究的驱动下,彭实戈院士提出了"非线性期望"框架体系并证明了一系列的非线性极限定理[5],形成了山东大学"非线性期望"研究团队。最近,陈增敬等以机器人博弈学习行为为研究对象,建立了非线性概率与智能机器人之间的联系,得到了一系列的策略极限理论[6-8],包括大数定律、中心极限定理和大偏差定理等极限定理。由于极限理论与策略相关,因此团队得到一系列定理命名为"策略大数定律""策略中心极限定理"和"策略大偏差定理",为带策略的机器学习方法提供了新的数学分析工具,

为了简要、清楚地解释"策略"极限理论,本文以双臂机器人博弈学习模型为例,具体理论结果见附录。相关理论成果已经探索了多臂机器人博弈学习的"策略"极限理论,详细见文献[6]。在概率空间$(\Omega,\mathscr{F},P)$上,假定双臂机器人每一个手臂服从以布朗运动$\{B_1,B_2\}$为驱动的正态分布,用X_t和Y_t代表左、右手臂获得的回报,假定X_t和Y_t满足如下 2 个随机微分方程:

$$\mathrm{d}X_t=\mu_{\mathrm{L}}\mathrm{d}t+\sigma_{\mathrm{L}}\mathrm{d}B_1(t),\quad X_0=0,\quad t\in[0,T], \tag{1}$$

$$\mathrm{d}Y_t=\mu_{\mathrm{R}}\mathrm{d}t+\sigma_{\mathrm{R}}\mathrm{d}B_2(t),\quad Y_0=0,\quad t\in[0,T], \tag{2}$$

其中,μ_{L}、μ_{R}和σ_{L}、σ_{R}是左、右手臂这 2 个随机变量的期望和方差。如果不考虑双臂机器人左、右手臂博弈学习过程,则X_t和Y_t分别代表线性概率空间上的 2 个相互独立的随机变量。若引入"策略"变量

$$\Pi:=\{\pi_t:\Omega\times[0,1]\to\{0,1\}\},$$

其中,$\pi_t=1$表示在t时刻使用左手,$\pi_t=0$表示在t时刻使用右手,便可以通过引入的"策略"$\pi\in\Pi$来描述双臂机器人在时间段$[0,t]$上的策略博弈过程。下面引入X_t和Y_t在策略π下融合成的新的随机变量R_t^{π}[7]:

$$\mathrm{d}R_t^{\pi}=\pi_t\mathrm{d}X_t+(1-\pi_t)\mathrm{d}Y_t,\quad R_0=0,\quad t\in[0,1]。$$

将 R_t^{π} 名命为策略随机变量,因为 R_t^{π} 的分布随着策略 π 变化而变化,构成一个分布族,已经不属于线性概率知识体系。下文将使用 $R_1^{0,\pi}$ 表示基于策略随机变量生成的一个统计量。基于策略随机变量, 严晓东依托陈增敬教授团队的研究成果[9-10]建立了一套新的策略极限理论,并且提出了一系列最优策略的制定方案,以及对应的“最优策略大数定律”与“最优策略中心极限定理”,为最优策略下人工智能技术提供了数学理论支撑。

1　策略极限理论的若干应用

1.1　大规模数据的策略抽样

传统的抽样过程,例如随机抽样、分层抽样、整群抽样、系统抽样等,都基于简单随机抽样,即在一定范围内,每一个样本都会被等概率抽到;大数据对抽样过程提出了新的挑战,传统的 Bootstrap 过程在大规模数据环境下已不再适用[11],因为对大量数据进行等概率有放回抽样,将大大增加计算复杂度。基于小样本等概率采用的 Little Bootstrap[12]受到欢迎,在大大提高运算效率的前提下实现了与原大数据近乎一样的统计分析结果。近两年,另有大量集中于模型驱动的子抽样策略的研究[13],通过非等概率采样出的小批量样本同样可以实现使用大规模数据的估计效果,并且大大降低了计算压力。

目前目标驱动的大数据抽样过程,即在考虑到要解决的统计问题时,如何充分利用问题中的信息呢? 例如关心的统计问题是单边检验, 对应原假设 $\theta>c(c>0)$,亦或是双样本单边检验,对应原假设 $\theta_1-\theta_2>c(c>0)$, 如何在抽样过程中构造检验统计量时使用先验信息 $\theta>0$ 或 $\theta_1>\theta_2$ 呢? 尚缺失基于目标转换的大数据抽样方法,例如将传统的双边检验 $\theta=c$ 这一统计目标转变成 $\theta-c_1=c_2$(c_1 和 c_2 都是事先给定的常数)。这种平移转变虽然与原目标是一样的,但是参照物发生了改变, $\theta=c$ 表示 θ 到原点距离是 c,$\theta-c_1=c_2$ 表示 θ 到点 c_1 的距离是 c_2。基于目标转换的统计假设检验问题,构造统计量抽取的样本也会相应地进行转换,即对应新目标驱动下的抽样过程。同样缺少知识驱动的抽样方法,即在考虑历史信息时,如何评价后续抽样的优劣,例如在记住基于前期抽样的估计结果前提下,当下抽出的样本是否更适合我们的统计分析过程? 是保留还是丢弃? 虽然这个思路与序贯抽样[14]很类似,但是如何去理解在大数据下的基于知识驱动抽样,仍然是一个空白。这一系列影响到抽样过程的问题无疑禁锢着大数据的统计分析。

大数据下抽取到的小批量样本无论是基于模型驱动、目标驱动、目标转换,还是知识驱动,都可以被概括为“策略抽样”。驱动源来自于所研究的问题或历史信息,通过引入策略 π 来表示是否采用这个样本,并且通过驱动源制定最优策略 π^*。最优策略随机变量 $R_t^{\pi^*}$ 同样可以表示基于策略 π^* 采取到的样本集合 $\{R_t^{\pi^*}:t\in[0,T]\}$,并且策略极限理论已经研究了 $R_t^{\pi^*}$ 或者基于样本 $\{R_t^{\pi^*}:t\in[0,T]\}$ 构造的统计量的极限行为[7]。

1.2　数据流的在线学习

我们迎来了大数据时代,日益增长的海量数据呈现“流”特征,亦称“数据流”。此类数据虽然是序列数据;但是与传统时间序列数据不同的是,流数据产生的新数据是基于新个体的样本,因此不存在序列相关性。数据流对存储承载力与计算机的计算性能等方面提出了更高的要求,因此传统的离线优化与学习算法面对极大的挑战。在线学习的出现解决了这一问题。在线学习算法可定义为:在重复决策的过程中,算法基于之前的经验以及当前的数据做出预测,以实现实时决策,并通过不断地对模型进行改进来提高预测的精度。在线学习算法已经被广泛应用于数据流的分析中。在线学习是基于机器学习的方法,对海量数据流进行训练,基于之前的经验(即保存下来的估计)来不断更新得到最佳的预测,而非传统地以批处理的方式运行。在数据流框架下,传统的批量学习方法具有时间和空间成本高、效率低、扩展性差的特点,在线学习模型可从多输入源分布式地输入数据,算法的效率和可扩展性大大提高。

目前为止,大量文献已经提出了各种在线学习方法,包括聚合估计方程[15]、累积更新估计方程[16](cumulatively updated estimating equation, CUEE)、随机梯度下降(stochastic gradient descent, SGD)及其变形[16-18],以及可再生估计器[19]。然而,这些在线更新的方法都要将记忆的历史信息融入到优化过程中,因此大量文献集中研究在线优化的函数形式, 需要小心翼翼地进行数学推导,因为这关乎到优化的速度与统计性质。当在线模型变得复杂时,这项工作将会变得困难,但是,在策略极限理论框架下,可以将历史信息与优

化方法进行分离,即将历史信息放在策略 π_t 中。X_t 和 Y_t 是基于当前数据得到的估计量,分别代表2种不同优化方法或者是2组样本集下得到的2个估计量,策略随机变量 R_t^{π} 根据策略 π_t 在 X_t 和 Y_t 中做出选择,最后基于策略 π 得到的序贯估计量(或策略估计量)集合 $\{R_t^{\pi}: t\in[0,T]\}$ 进行统计分析。因为 X_t 和 Y_t 只是利用当前时刻流数据进行参数优化,所以该方法在保持在线学习效率的前提下,具有很强的可拓展性,不需要纠结复杂模型如何在线优化的难题。

如果流数据来源是多方位的,比如同一批量数据是基于不同的采集时间产生的,或者来自于不同的采集地点,亦或是当前批量数据的采集是使用了不同实验工具等等,从而使多来源数据结构呈现多样化和异质性。假设同一时刻有2个数据源 X_t 和 Y_t,策略随机变量 R_t^{π} 的引入告诉我们每一时刻只采用一个数据源,即 $\pi_t=1$ 使用 X_t,$\pi_t=0$ 使用 Y_t,在时间和空间上进一步优化了传统的在线学习过程[20]。如果出现更多数据源,π_t 可以通过向量化来决定选取哪一个或哪一组数据源[6],并且向量元素也可以取值于0和1之间,表示选取数据源的权重。

1.3 强化学习的中心极限定理

目前强化学习过程是各学科领域热门研究方向,其中,工程问题偏于算法研究,旨在提高运算效率。医学领域使用强化学习模拟医生(智能机器人医生),为病人提供治疗方案;生物制药进行序贯设计实验,检验更有效的药物;经济问题的政策评估与策略方案的制定也离不开强化学习过程。与传统的统计学习算法相比,强化学习过程增加了策略这一变量,导致产生的序贯数据存在分布不确定性,进而导致概率分布框架下的强化学习算法研究的空白。因为传统的线性概率知识体系是无法解释强化学习过程的,所以我们必须使用非线性概率框架。基于以上表述,我们发现强化学习过程缺少统计推断问题以及相应的概率理论。普林斯顿大学范剑青教授也强调了在统计框架下研究强化学习的困难之处以及缺乏相应的统计推断理论,即很难评估一个策略的好坏,以及评估某一个强化学习算法的拟合优度等问题。

传统的强化学习算法(如 ϵ-贪婪算法[21]、梯度 bandit 算法[22]或上置信带算法[23])通过追求平均奖励(即样本平均数)达到最大去估计行动(即策略)。假设考虑2个行动,即 $\pi_t=1$ 获得奖励 X_t,$\pi_t=0$ 获得奖励 Y_t,策略随机变量 R_t^{π} 亦可表示 t 时刻获得真实奖励。强化学习中经典的双臂老虎机算法通过追求最大回报估计最优的策略 π^*,即

$$\pi^* = \arg\max_{\pi} \frac{1}{T+1}\sum_{t=0}^{T} R_t^{\pi}。$$

总之,经典文献中的算法旨在最大化平均回报,是基于大数定律构建的方法,缺乏 R_t^{π} 概率分布的研究,因此强化学习框架下鲜少见到统计推断的研究。

研究 R_t^{π} 概率分布的困难之处是序贯数据的分布随着策略 π 变化而变化,存在分布不确定性,传统的线性概率知识体系是无法解释这一过程的。陈增敬教授团队研究了不同策略 π 下 $R_1^{0,\pi}$ 的概率分布行为,命名为策略中心极限定理,并在此框架下研究了最优策略的制定,以及基于最优策略 π^* 的 R_1^{0,π^*} 的极限分布[7],最后提出了包含2种行为平均回报 μ_L 和 μ_R 的统计量[7],以此作为 μ_L 和 μ_R 的检验统计量。

接下来团队会考虑多行为的以及带有状态的强化学习过程的中心极限定理,使策略极限理论在更一般的强化学习框架下发挥制定最优策略的优势,以及开展若干强化学习的统计推断研究,弥补这一研究方向上的空白。

1.4 数据的差分隐私保护

2021年9月1日正式实施的《中华人民共和国数据安全法》进一步提升了国家数据安全的保障能力和数字经济的治理能力,是对长期以来"重数据搜集,轻安全保护"现象的治理。此外,2021年11月1日实施的《中华人民共和国个人信息保护法》的内容具体到个人信息数据保护,意味着个人信息数据处理者在未经本人允许的条件下不可以擅自应用个人信息数据,进一步加强了对个人数据的保护。对于数据安全保护除了需要严格的法律政策以外,也需要从技术上对数据进行隐私处理。对数据进行隐私化处理是指在任何用户访问数据库的过程中,无法获得任意个体的确切信息,可以看作是对数据的一种加密处理。一般来说,在数据共享机制中,数据持有者简单地删除数据集中较为敏感的信息,然后将数据分享给数据需求者,这种方式通常需要考虑外来的恶意攻击者可能拥有的所有信息(识别哪些数据属于敏感信息),并且难以比较隐私

保护的水平(应该删去多少条数据),因此无法得知隐私保护的可靠性。目前另一种比较主流的隐私保护方式是差分隐私[24]。差分隐私保护模型弥补了传统隐私保护方式的缺陷,既不需要考虑攻击者掌握的信息,又能够提供隐私保护的量化方法,这也是差分隐私方法迅速被业界认可的重要原因。差分隐私是基于这样一种思想:添加差分隐私算法后的目标数据集转化为差异隐私数据集,且不可能检测到数据集中添加或删除一条数据后对该结果的影响。这意味着,如果数据库删除了某条数据,被删除的个体也不会被识别出来,使得数据需求者既可以从数据集中获取所需信息(包括个人识别信息),又能够将敏感信息(例如被删除的个体)进行保护。

目前隐私保护的研究大多集中在计算机与工程领域,研究者只关注如何设计隐私保护的算法,鲜有研究关注隐私保护如何影响数据价值的。如中国工程院院士方滨兴[25]在《释放数据使用权将成为未来技术发展取向》一文中强调了如何在保护数据隐私的前提下,最大限度地挖掘大数据价值是目前不少企业和机构面临的难题。基于隐私保护数据的统计推断的输出结果会不会被明显改变?数据分析研究者仍然能够通过分析差异隐私数据集获得所需要的统计结果么?这是留给统计学家的一份试卷。

虽然学者们从不同的角度对差分隐私机制进行应用,并提出不同的改进方法,但这些方法大多是基于同一数据单元进行的,从而扰动误差服从单一概率分布[24]。如果涉及到多个数据单元,例如以每一个用户的信息作为数据单元,则需要局部差分隐私保护[26-27],即由于每一个用户数据的灵敏度以及用户可接受的隐私保护强度都不一样,因此导致扰动误差的概率分布随着数据源变化而变化,存在分布不确定性,传统的线性概率知识体系采用单一概率分布[24]是不可行的。也就是说,如何实现对多数据源的差分隐私保护同时挖掘数据价值呢?这将是非线性概率统计学家去解答的问题。

假设我们考虑多个数据源$\{D_1,D_2,\cdots,D_K\}$,在每一个源头的数据集下,考虑 2 种分布的扰动获得差异化隐私算法,即

$$\hat{\theta}^{\mathrm{dp}}(D_t)=\hat{\theta}_t+R_t^{\pi},\quad t=1,2,\cdots,K,$$

其中 R_t^{π} 代表具有误差含义的策略随机变量,即 $\mu_{\mathrm{L}}=\mu_{\mathrm{R}}=0$,但是 $\sigma_{\mathrm{L}}\neq\sigma_{\mathrm{R}}$。$R_t^{\pi}$ 与异方差误差的区别是含有策略 π,它的作用就是根据历史信息决定第 t 个原始估计量 $\hat{\theta}_t$ 加扰动 X_t 还是 Y_t,这样不仅能够实现差分隐私保护,而且还可以利用策略极限理论对感兴趣参数 θ 进行统计分析,因此,我们提出的策略误差扰动的差分隐私保护过程不仅包括了传统的差分隐私保护[24]方法,而且还实现了数据价值的挖掘。

1.5　联邦学习的策略融合

近年来,我国逐步加强数据保护,陆续出台相关政策,如国家互联网信息办公室近期起草的《数据安全管理办法(征求意见稿)》表明,数据在安全合规的前提下进行交流,是人工智能技术首要解决的问题。例如 2 个公司甚至公司间的部门都认为彼此的数据具有巨大的潜在价值,即便考虑到利益共享,这些机构也不会提供各自数据进行共享,因为涉及到数据隐私保护,最终导致数据孤岛的形成。为此,联邦学习[28]应运而生,原因主要是在考虑隐私的情况下,解决了绝大多数企业存在的数据量少、数据质量差导致的不足以支撑人工智能技术实现的问题。

由于联邦学习的数据来源不一样,而且不同来源数据也存在个体重叠的现象(如纵向联邦学习),因此导致联邦学习的多源数据分布不再满足传统机器学习假设的独立同分布(independent identically distributed, IID)性质,这是联邦学习核心难点之一。如何保证多源训练模型结果依然可以被有效地全局融合(global aggregation)呢?

基于此,本文提出策略融合的方案。以 2 个客户端为例,X_t 和 Y_t 分别代表 2 个客户端的模型拟合结果,引入策略 π 融合新的估计量 $R_1^{0,\pi}$,其中在联邦学习框架下 π 的制定会稍微复杂,因为需要全局考虑所研究的统计问题及其目标,例如 2 个来源数据的相似程度、样本不平衡程度,特别是需要考虑到联邦学习的反馈过程,即如何将中间结果有效地下发到各个客户端,这是联邦学习的个性化问题(personalization)。最后在既定的统计目标下,制定最优策略 π^*,使得策略融合估计量 R_1^{0,π^*} 达到最好的学习效果。

1.6　迁移学习和元学习的信息重构

迁移学习[29]是机器学习的一个重要研究分支,侧重于将已经学习过的知识迁移应用于新的问题中,以增强解决新问题的能力,提高解决新问题的速度。从目标上看,元学习[30-31]和迁移学习并无本质区别,

都可以看作是利用源域信息在目标域中增加学习器,从而提升在多任务泛化能力,但元学习更偏重于任务和数据的双重采样,而迁移学习强调从一个任务到其他任务的能力迁移,不太强调任务空间的概念。

下面仅以迁移学习为例介绍策略极限理论的应用。迁移学习指的是给定一个基于目标域的学习任务 A,可以从学习任务 B 的源域得到帮助,通过学习任务 A 发现并迁移源域和学习任务 B 潜在的可迁移知识,提高学习函数的性能,因此迁移学习的核心是找到源领域和目标领域之间的相似性。

迁移学习的主要目的是解决训练数据不足,与一般的深度学习方法相比,迁移学习放松了训练数据和测试数据必须是独立同分布的假设,将知识从源域迁移到目标域,策略极限理论[7]恰好回答了在这种"差异"的情况下,如何引入策略对数据进行重构,并强调了策略极限理论在"数据重构"上的优势,因为:(1)有差异数据(即不同分布)重组会变成不独立,使用基于正态分布的传统方法都是错误的,即源域的数据分布 X_t 和目标域的数据分布 Y_t 之间只要存在差异,基于策略融合而成的统计量 $R_1^{0,\pi}$ 就不是正态分布,而且在既定的目标制定的最优策略 π^* 下,R_1^{0,π^*} 一定比正态分布好,例如文献[7]已经证明 R_1^{0,π^*} 的方差更小,即通过策略迁移的"数据重构"使信息汇聚得更集中,而且文献[7]建立的策略极限定理制定了最优策略,即告诉我们如何迁移;(2)独立+独立=相关,也是策略极限理论建立的重要结论,意思是即便源域的数据分布和目标域的数据分布是独立的,即信息不可压缩,我们也可以通过引入策略 π 使得它们相关,使其信息可以进行压缩,这种现象是非线性概率所独有的,完全违背了传统的线性概率知识体系的结论。

1.7 知识推理与数据驱动的融合

徐宗本院士梳理了人工智能研究发展的 3 次浪潮[4],从以符号推理/知识库运用为特征且需要人工设定的知识表示,到基于数据驱动的机器学习的知识自动表示,最后一个阶段是前 2 次浪潮学习能力的叠加,实现自主学习和环境自适应并具备持续自主学习能力。特别是,后深度学习时代[32]必然追求数据驱动的机器学习与知识驱动的符号计算相融合的新型人工智能理论和方法,研究知识表示与推理框架、知识数据双驱动的决策推理,使得深度学习在保持强大的数据学习能力基础上,具有更明确的可解释性和更强的泛化性。

徐宗本院士强调,数据驱动与知识驱动的融合模型应该能够同时处理 2 类不同的变量——数据变量(连续或离散的实数形式)和逻辑变量(符号形式),以及处理 2 类不同的运算——实数运算和逻辑运算,因此如何在数据驱动的机器学习算法中联通数据蕴含的知识或语言表达的知识,是融合系统首要解决的问题,这需要新型数学符号的设计与数学理论支撑。

接下来,本文尝试使用策略随机变量的构建与策略极限理论结果来分别解释知识的数据化(数字化)(从抽象到具体/示例)和数据的知识化(从具体/示例到抽象)这一融合系统。我们可以把"策略"看作是在数据空间和语义空间之间建立一个中间空间,来联通数据与知识[33]。策略随机变量与传统随机变量的区别是,策略随机变量里面包含了知识,例如 X_t 代表不同压强下水的沸点,Y_t 代表不同压强下酒精的沸点,如果用三元组表示知识图谱,即"水的平均沸点大于酒精的平均沸点"可以简化地表示为"μ_{L} 大于 μ_{R}",通过附录中策略分布的定义,我们发现,这个知识已经被包含在最优策略 π^* 的设计中,因此策略的引入将知识进行了数据化。另外,X_t、Y_t 和策略 π 的融合可以看成是数据与知识的融合,融合系统就是生成的策略统计量 $R_1^{0,\pi}$,那么,关于 $R_1^{0,\pi}$ 策略极限理论的探索就可以看作是数据到知识的转化过程,即基于数据的知识化。

2 结论

本文首次采用"策略"这一概念探寻和揭示若干机器学习技术智能表现优异的原因,将"策略"看作是促动人工智能技术更新和变革的核心动力,人工智能技术表现的优劣,关键看历史信息(知识)学习得好不好,通过总结经验进行下一步决策。山东大学"非线性期望"团队最近发展了带策略的机器学习的概率理论,是非线性概率知识体系在人工智能领域交叉研究方面的非常有价值的探索,提出了"策略大数定律"和"策略中心极限定理",为后续智能机器学习技术的统计问题的研究,例如统计推断,提供了基础数学理论。更重要的是,策略极限理论的研究是面向以各类机器学习,包括深度学习、强化学习、迁移

学习等为代表的人工智能方法鲁棒性差、可解释性劣、可拓展性差等基础科学问题,挖掘智能机器学习的基本原理,发展可解释、可通用的下一代人工智能方法,并推动人工智能方法在若干科学领域的创新应用。

本文虽然强调了策略极限理论在研究人工智能技术上的优势;但是在交叉研究方面仍需要仔细推敲,因为人工智能技术的优化过程还是很复杂的,优化手段以及估计函数形式都会影响着最优策略的制定以及策略极限分布的形式。

如同在线性概率框架下的正态分布是研究简单或复杂统计模型的理论基础,以正态分布为基础的卡方分布、t 分布、F 分布等也是研究复杂估计量渐近行为的概率基础。本文旨在强调以非线性概率发展的策略分布(附录 A)将会是替代传统正态分布的研究人工智能技术的重要工具,属于基础理论工具。

在传统的数据分析中,我们刻画数据的随机性或估计量的渐近行为时采用了正态分布,正态分布是基础的但又是理想的。自 1733 年 De Moivre 的理论发现,正态分布已统治概率和统计学学科近 400 年,正态分布在概率论中起着核心作用,并促使数据驱动的 AI 方法的数学理论研究取得了丰富的理论成果,然而,带策略的机器学习将会打破完美的正态分布。基于此,我们通过开发新的概率理论体系来揭开若干复杂 AI 技术数学理论的神秘面纱,例如强化学习、迁移学习、元学习、深度学习等。我们预计,此项研究将成为更多应用研究的起点, 如量子纠缠、混沌分析或物理和生物组学中的布朗棘轮,包括基因组学、蛋白质组学和代谢组学,都将与此类研究相关。

附录 A　策略极限理论

本文并不对策略极限理论作详细论述,相关原创性研究成果可以参考文献[6-7]。下面只展示策略极限理论的最核心理论结果,即策略大数定律和策略中心极限定理。

下面先介绍策略大数定律,我们得到了策略强大数定律和策略弱大数定律。

定理 1(策略大数定律)　令 $S_T^{\pi}=\sum_{t=1}^{T}R_t^{\pi}$,

(1)(策略强大数定律)采用 $h\in[\underline{\mu},\bar{\mu}]$,且 $\underline{\mu}=\min(\mu_{\mathrm{L}},\ \mu_{\mathrm{R}})$, $\bar{\mu}=\max(\mu_{\mathrm{L}},\ \mu_{\mathrm{R}})$,其中 h 对应的形式是

$$h=\gamma\bar{\mu}+(1-\gamma)\underline{\mu},\quad \gamma\in[0,1],$$

实验者可以采用策略 π^{γ} 使得

$$\lim_{T\to\infty}\frac{S_T^{\pi^{\gamma}}}{T}=h,\ P\text{-a.s.}。$$

其中,策略强大数定律构造的策略 $\pi^{\gamma}=(\pi_1^{\gamma},\pi_2^{\gamma},\cdots,\pi_t^{\gamma},\cdots)$ 如下:

第 1 步:选择左手臂,即 $\pi_1^{\gamma}=1$。

第 2 步:选择右手臂,即 $\pi_2^{\gamma}=2$。

第 $k(\geqslant 3)$ 步:分为 3 种情况进行讨论。

情况 A　对于 $i>1$,当 $k=2^i-1$ 时, 选择左手臂 L, 即 $\pi_k^{\gamma}=1$。

情况 B　对于 $i>1$,当 $k=2^i$ 时, 选择右手臂 R, 即 $\pi_k^{\gamma}=2$。

情况 C　对于 $i>1$,当 $2^i<k<2^i+1-1$ 时,令 m_k^{L}(相应地, m_k^{R})代表在前 k 步使用左手臂 L(相应地, R) 的次数。μ_k^{L} 和 μ_k^{R} 分别对表示截至第 k 步时左、右手臂的平均数,即

$$\mu_k^{\mathrm{L}}:=\frac{\sum_{t\leqslant k,\pi_j^{\gamma}=1}X_t^{\mathrm{L}}}{m_k^{\mathrm{L}}},\quad \mu_k^{\mathrm{R}}:=\frac{\sum_{t\leqslant k,\pi_j^{\gamma}=2}Y_t^{\mathrm{R}}}{m_k^{\mathrm{R}}}。$$

(2)(策略弱大数定律)对于任意 $\varepsilon>0$,

$$\lim_{T\to\infty}\inf_{\pi\in\Pi}P\left(\underline{\mu}-\varepsilon<\frac{S_T^{\pi}}{T}<\bar{\mu}+\varepsilon\right)=1,$$

并且对于 $\varepsilon>0$, $h\in[\underline{\mu},\bar{\mu}]$,

$$\lim_{T\to\infty}\sup_{\pi\in\Pi}P\left(\left|\frac{S_T^{\pi}}{T}-h\right|<\varepsilon\right)=1。$$

定理 1 中的强大数定律可以估计左、右手臂的上期望,弱大数定律可以用于检测 Parrondo 悖论不成立的条件。下面介绍策略中心极限定理。

定理 2(策略中心极限定理) 令 $\bar{\mu}=\max(\mu_{\mathrm{L}},\mu_{\mathrm{R}})$, $\underline{\mu}=\min(\mu_{\mathrm{L}},\mu_{\mathrm{R}})$, $\sigma_{\mathrm{L}}=\sigma_{\mathrm{R}}=1$,对时间 T 进行归一化处理,即 $T=1$。

(1)(基于最大概率的策略中心极限定理) 对于任意 $0\leqslant a<b<\infty$,存在最优策略

$$\pi_t^{*}=\begin{cases}\dfrac{\underline{\mu}-\mu_{\mathrm{R}}}{\mu_{\mathrm{L}}-\mu_{\mathrm{R}}}, & R_1^{0,\pi^{*}}>\dfrac{a+b}{2}-\dfrac{\bar{\mu}+\underline{\mu}}{2},\\ \dfrac{\bar{\mu}-\mu_{\mathrm{R}}}{\mu_{\mathrm{L}}-\mu_{\mathrm{R}}}, & R_1^{0,\pi^{*}}\leqslant\dfrac{a+b}{2}-\dfrac{\bar{\mu}+\underline{\mu}}{2},\end{cases}\quad t\in[0,1], \tag{3}$$

使得

$$P(a\leqslant R_1^{0,\pi^{*}}\leqslant b)=\sup_{\pi\in\Pi}P(a\leqslant R_1^{0,\pi}\leqslant b)。$$

$R_1^{0,\pi^{*}}$ 落在 $[a,b]$ 的概率是

$$P(a\leqslant R_1^{0,\pi^{*}}\leqslant b)=\Phi\left(-\left|\frac{a+b}{2}-\frac{\bar{\mu}+\underline{\mu}}{2}\right|+\frac{\bar{\mu}-\underline{\mu}}{2}+\frac{b-a}{2}\right)-\exp\left(-\frac{\bar{\mu}-\underline{\mu}}{2}(b-a)\right)\Phi\left(-\left|\frac{a+b}{2}-\frac{\bar{\mu}+\underline{\mu}}{2}\right|+\frac{\bar{\mu}-\underline{\mu}}{2}-\frac{b-a}{2}\right), \tag{4}$$

其中 $\Phi(\cdot)$ 是标准正态的分布函数。$R_1^{0,\pi^{*}}$ 的概率密度函数是

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(|z-h|+k)^2}{2}\right)+k\exp(-2k|z-h|)\Phi(-|z-h|+k), \tag{5}$$

$f(z)$ 被称作是尖峰分布,其中 $h=\dfrac{\bar{\mu}+\underline{\mu}}{2}$, $k=\dfrac{\bar{\mu}-\underline{\mu}}{2}$。

(2)(基于最小概率的策略中心极限定理)弱采用策略

$$\pi_t^{*}=\begin{cases}\dfrac{\bar{\mu}-\mu_{\mathrm{R}}}{\mu_{\mathrm{L}}-\mu_{\mathrm{R}}}, & R_1^{0,\pi^{*}}>\dfrac{a+b}{2}-\dfrac{\bar{\mu}+\underline{\mu}}{2},\\ \dfrac{\underline{\mu}-\mu_{\mathrm{R}}}{\mu_{\mathrm{L}}-\mu_{\mathrm{R}}}, & R_1^{0,\pi^{*}}\leqslant\dfrac{a+b}{2}-\dfrac{\bar{\mu}+\underline{\mu}}{2},\end{cases}\quad t\in[0,1], \tag{6}$$

使得

$$P(a\leqslant R_1^{0,\pi^{*}}\leqslant b)=\inf_{\pi\in\Pi}P(a\leqslant R_1^{0,\pi}\leqslant b)。$$

$R_1^{0,\pi^{*}}$ 落在 $[a,b]$ 的概率是

$$P(a\leqslant R_1^{0,\pi^{*}}\leqslant b)=\Phi\left(-\left|\frac{a+b}{2}-\frac{\bar{\mu}+\underline{\mu}}{2}\right|+\frac{\underline{\mu}-\bar{\mu}}{2}+\frac{b-a}{2}\right)-\exp\left(-\frac{\underline{\mu}-\bar{\mu}}{2}(b-a)\right)\Phi\left(-\left|\frac{a+b}{2}-\frac{\bar{\mu}+\underline{\mu}}{2}\right|+\frac{\underline{\mu}-\bar{\mu}}{2}-\frac{b-a}{2}\right), \tag{7}$$

其中 $\Phi(\cdot)$ 是标准正态的分布函数。$R_1^{0,\pi^{*}}$ 的概率密度函数是

$$g(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(|z-h|-k)^2}{2}\right)-k\exp(2k|z-h|)\Phi(-|z-h|-k), \tag{8}$$

$g(z)$ 被称作是双正态分布。

下面仅基于最大概率的策略中心极限定理对应的尖峰分布作出一些解释。以定理 2 表明,当且仅当 $\bar{\mu}=\underline{\mu}$ 时,$R_1^{0,\pi^{*}}$ 退化为传统的标准正态分布 $N(\bar{\mu},1)$。如图 1 所示,2 个分布距离越远,即 k 越大,策略分布的概率密度函数图象越尖峰,表明 2 个数据分布差异越大,在最优策略 π^{*} 的作用下,数据重构使信息汇聚得更集中。同时,定理 2 也告诉我们,如果统计目标是一个区间,总可以找到一个最优策略 π^{*} 以最大概率覆盖这个区间,而且这个概率总比通过传统的正态分布计算的概率大。如图 2 所示,无论解决的问题目标是什么,即考虑 3 个区间 $[0,1]$、$[0.5,1.5]$、$[1,2]$,总可以找到一个目标驱动的最优策略 π^{*}(式(3)),使 $R_1^{0,\pi^{*}}$ 覆盖任何区间的概率总是最大的,这是数据重构使信息汇聚得更集中的原因。

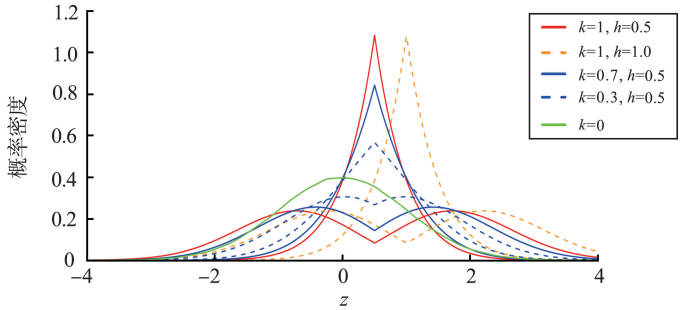
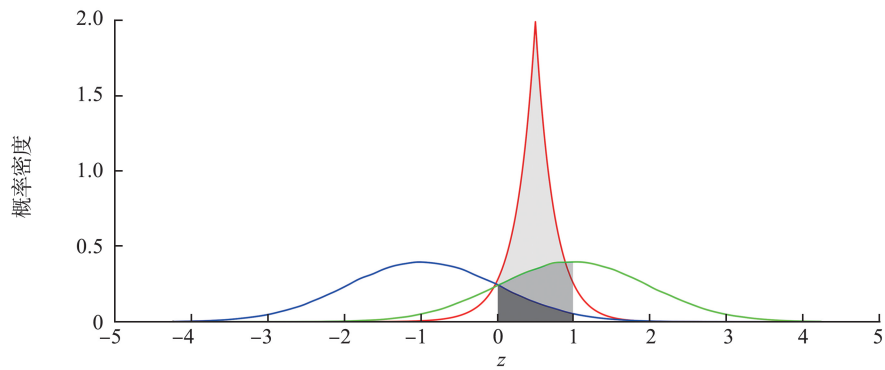
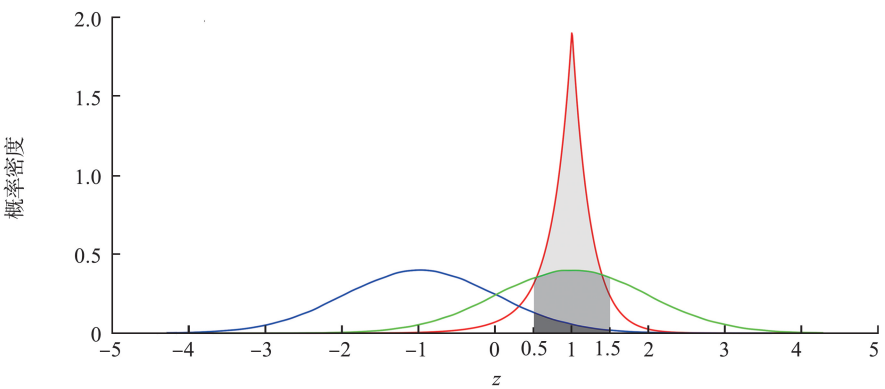


图 1　尖峰分布 $f(z)$ 和双正态分布 $g(z)$ 随着参数 k 和 h 变化的概率密度函数图,其中绿色线是标准正态分布概率密度函数图

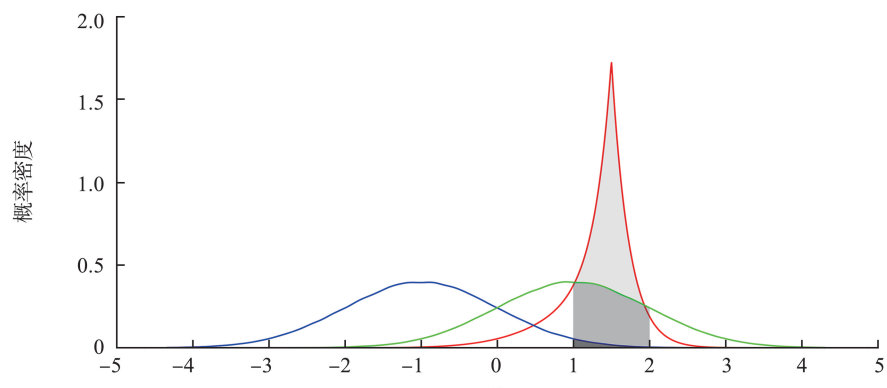
Fig.1　Different probability density functions of the peak distribution $f(z)$ and the binormal distribution $g(z)$ as the parameters k, h change, where the green line is the probability density function plot of the standard normal distribution



(a) 区间 [0,1], a=0, b=1



(b) 区间 [0.5,1.5], a=0.5, b=1.5



(c) 区间 [1,2], a=1, b=2

图 2　3 个尖峰分布概率密度函数 X_t(蓝色), Y_t(绿色)和 R_1^{0,π^*}(红色)在区间[a,b]的覆盖概率,其中 X_t 和 Y_t 的期望分别设置为 $\mu_L=-1$, $\mu_R=1$ 以及同方差 $\sigma_L=\sigma_R=1$,并且考虑 3 个区间[0,1],[0.5,1.5],[1,2]

Fig.2　Coverage probability of the 3 peak distribution probability density functions X_t(blue), Y_t(green), and R^{0,Π^*}(red) in the interval [a,b], where the expectations of X_t and X_t are set to $\mu_L=-1$, $\mu_R=1$ and the variances are set to $\sigma_L=\sigma_R=1$, respectively, and consider the three intervals[0,1], [0.5,1.5], [1,2]

参考文献:

[1] 徐宗本. 用好大数据须有大智慧:准确把握、科学应对大数据带来的机遇和挑战[N]. 人民日报, 2016-03-15 (07).
XU Zongben. To make good use of big data requires great wisdom: accurately grasp and scientifically respond to the opportunities and challenges brought by big data[N]. People's Daily, 2016-03-15 (07).

[2] 徐宗本. 把握新一代信息技术的聚焦点:数字化、网络化、智能化[N]. 人民日报, 2019-03-01 (09).
XU Zongben. Grasp the focus of the new generation of information technology: digitalization, networking, and intelligence [N]. People's Daily, 2019-03-01 (09).

[3] 徐宗本,唐年胜,程学旗. 数据科学:它的内涵、方法、意义与发展[M]. 北京: 科学出版社, 2021.
XU Zongben, TANG Niansheng, CHENG Xueqi. Data science: its connotation, method, significance and development[M]. Beijing: Science Publishing, 2021.

[4] 徐宗本. 人工智能的 10 个重大数理基础问题[J]. 中国科学: 信息科学, 2021, 51:1967-1978.
XU Zongben. Ten fundamental problems for artificial intelligence: mathematical and physical aspects[J]. Science China: Information Sciences, 2021, 51:1967-1978.

[5] PENG S. Nonlinear expectations and stochastic calculus under uncertainty: with robust CLT and G-Brownian motion[M]. Berlin: Springer Nature, 2019.

[6] CHEN Z, EPSTEIN L G. A central limit theorem for sets of probability measures[J]. Stochastic Processes and Their Applications, 2022, 152:424-451.

[7] CHEN Z, FENG S, ZHANG G. Strategy-driven limit theorems associated bandit problems[EB/OL]. 2022-04-09[2023-12-05]. https://arxiv.org/abs/2204.04442.

[8] CHEN Z, EPSTEIN L, ZHANG G. A central limit theorem, loss aversion and multi-armed bandits[J]. Journal of Economic Theory, 2023, 209:105645.

[9] CHEN Z, YAN X, ZHANG G. Strategic two-sample test via two-armed bandit process[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2023, 85:1271-1298.

[10] CHEN Z, FENG X, LIU S, et al. Optimal distributions of rewards for a two-armed slot machine[J]. Neurocomputing, 2023, 518:401-407.

[11] ZHAO T, CHENG G, LIU H. A partially linear framework for massive heterogeneous data[J]. Annals of Statistics, 2016, 44(4):1400-1437.

[12] KLEINER A, TALWALKAR A, SARKAR P, et al. A scalable bootstrap for massive data[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2014, 76(4):795-816.

[13] AI Mingyao, YU Jun, ZHANG Huiming, et al. Optimal subsampling algorithms for big data regressions[J]. Statistica Sinica, Forthcoming, 2021, 31:749-772.

[14] ETIKAN I, ALKASSIM R, ABUBAKAR S. Comparision of snowball sampling and sequential sampling technique[J]. Biometrics and Biostatistics International Journal, 2016, 3(1):55.

[15] LIN N, XI R. Aggregated estimating equation estimation[J]. Statistics and its Interface, 2011, 4:73-83.

[16] SCHIFANO E D, WU J, WANG C, et al. Online updating of statistical inference in the big data setting[J]. Technometrics, 2016, 58(3):393-403.

[17] CHEN X, LEE J D, TONG X T, et al. Statistical inference for model parameters in stochastic gradient descent[J]. The Annals of Statistics, 2020, 48:251-273.

[18] ZHU W, CHEN X, WU B. Online covariance matrix estimation in stochastic gradient descent[J]. Journal of the American Statistical Association, 2021, 118(154):393-404.

[19] LUO L, SONG P X K. Renewable estimation and incremental inference in generalized linear models with streaming data sets [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2020, 82:69-97.

[20] CUI W, JI X, KONG L, et al. Opposite online learning via sequentially integrated stochastic gradient descent estimators [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington: AAAI Press, 2023, 37(6):7270-7278.

[21] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M].[S.l.]: MIT Press, 2018.

[22] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992, 8(3):229-256.

(下转第 45 页)