

胡羽聪, 李娜, 江燕, 等. 人工智能算法与生态环境模型耦合研究进展. 应用生态学报, 2023, 34(1): 257-263
Hu Y-C, Li N, Jiang Y, et al. Research progress on coupling artificial intelligence and eco-environmental models. Chinese Journal of Applied Ecology, 2023, 34(1): 257-263

人工智能算法与生态环境模型耦合研究进展

胡羽聪^{1,2} 李娜³ 江燕^{1,2*} 包鑫^{1,2} 李叙勇^{1,2}

(¹中国科学院生态环境研究中心城市与区域生态国家重点实验室, 北京 100085; ²中国科学院大学, 北京 100049; ³中国灌溉排水发展中心, 北京 100054)

摘要 人工智能算法在生态环境领域已有广泛应用,但在揭示自然科学现象规律时存在泛化能力不足、可解释性差等问题。为弥补这些不足,实现优势互补,将人工智能算法与具有物理机制的生态环境模型耦合研究已成为近些年快速发展的一种新型研究方法。本文从应用在生态环境领域的人工智能算法出发,概述了其分类和应用情况,重点梳理了人工智能算法与生态环境模型耦合研究的发展、现状及不足,提出了一个将人工智能与机理模型紧密耦合以重构机理过程的思路,分析了该网络部分参数的理论意义,提高可解释性和泛化能力的可行性,以及模拟机理过程运行的应用前景,并展望了人工智能算法与生态环境模型耦合研究的发展趋势。

关键词 人工智能算法; 生态环境模型; 耦合; 重构机理

Research progress on coupling artificial intelligence and eco-environmental models. HU Yu-cong^{1,2}, LI Na³, JIANG Yan^{1,2*}, BAO Xin^{1,2}, LI Xu-yong^{1,2} (¹State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China; ²University of Chinese Academy of Sciences, Beijing 100049, China; ³China Irrigation and Drainage Development Center, Beijing 100054, China).

Abstract: Artificial intelligence (AI) has been widely used in the eco-environment field, but with shortcomings in revealing the laws of natural science, such as insufficient generalization ability and poor interpretability. In order to overcome these shortages and tap into complementary advantages, coupling AI and eco-environmental models containing physical mechanism has been a new research method with fast development in recent years. We introduced the classifications of AI used in eco-environmental field, outlined its applications, and mainly illustrated the progresses, status and inadequacies for the coupling research. Based on all the summaries, we proposed a new coupling method of physical mechanism and AI for reconstructing mechanism processes, followed by analyses of theoretical significance of partial parameters, feasibility of better generalization and interpretability, as well as prospection of imitating physical mechanism. At the end of the review, we discussed the trend of the coupling method of AI and eco-environment models.

Key words: artificial intelligence; eco-environmental model; coupling; reconstruction of mechanism.

自然科学领域长久以来以总结规律和理论推导为主^[1],建立数学模型,分析自然现象。数学模型主要由关系和变量组成,根据对已知规律描述的精确程度,可以分为黑箱模型、灰箱模型和白箱模型^[2]。生态环境领域常用的灰箱模型介于参数具有明确物理意义的白箱模型与不涉及物理机制的黑箱模型之间,其参数一般利用历史序列数据优化确定,这势必带来峰值捕捉能力差、计算效率低等问

题;黑箱模型的部分参数来源于经验,存在一定误差;而白箱模型的建立需要诸多假设,且忽略了其他机理过程的影响。复杂自然过程具有的参数维度高、非线性程度高等特点,使得数学模型虽然能较好地模拟已知机理过程,但描述的非精确性^[3],以及数据、参数和模型结构的不确定性等难点也给模拟、预测和评价的精度、效率带来极大挑战^[4]。

随着计算机性能的提升,以机器学习(包括深度学习)为代表的人工智能算法(artificial intelligence, AI)为解决上述计算时间成本高、精度受限等问题提供了可能。随机森林(random forest, RF)、支持向量机(support vector machine, SVM)等机器学习

本文由国家重点研发计划项目(2019YFB2102901, 2019YFB2102902)资助。

2022-01-28 收稿, 2022-09-08 接受。

* 通讯作者 E-mail: yanjiang@rcees.ac.cn

方法通过提取数据特征,使模拟和预测的精度和速度有了显著提高^[5]。为去除机器学习筛选特征时的人为影响,深度学习技术从输入数据开始,建立端到端的联系,进一步拓展了神经网络的模拟和预测能力。然而,虽然由数据驱动的人工智能算法在自然科学领域逐渐展露优势,但数据的时空相关性、数据序列缺省等因素,也会引起数据数量、质量不满足训练需求^[6]、结果可解释性差^[7]、参数缺乏实际意义^[6]等问题。由数据驱动的人工智能算法与基于物理机制的数学模型对实际自然过程的描述各有优劣。如何利用两者的优势、弥补各自短板是当前研究者所关注的一个主要问题。

本文从生态环境模型和人工智能算法的特点和面临的问题出发,梳理人工智能算法在生态环境领域应用现状,总结两者现有的耦合方式,分析当前研究的不足,提出了一种人工智能算法与机理模型紧密耦合的思路,并对耦合研究的发展趋势进行展望。

1 人工智能算法的基础应用

人工智能算法在自然科学领域得到了广泛应用,在水资源管理^[8]、生态安全^[9]、天气预报^[10]等生态环境领域^[11]已取得较好的应用成果。在满足一维序列数据和二维图像数据的模拟和预测需求方面有大量研究。

1.1 数据处理

自然科学领域的研究一般需要高质量的时间数据^[12]。然而,由于设备故障、网络中断、自然条件干扰等不利因素的影响,实测数据经常缺失。过去常用的两种补充缺省值方法(时间序列自回归和邻近站点插补)逐渐难以满足精度需求^[13]。随着 AI 挖掘数据潜在规律的能力日益凸显,Mital 等^[13]通过考虑邻近站点与缺省位置的相关性,使用随机森林算法按时序插补,得到可靠的插值结果。Nguyen 等^[14]采用长短期记忆神经网络(long-short term memory, LSTM)显著改善了 3 h 内的平均面积降水量预测值。通过数据增强(data augmentation),在不额外增加数据量的前提下,数量有限或样本不平衡的二维数据能够发挥更大作用^[15-16]。经过预处理后的数据,能够明显提升算法的精度。

1.2 模拟和预测

模拟和预测自然过程的变化规律,探究和调控人类活动对生态环境的影响,是使用模型和 AI 这两种方法的共同目的。模型能够提供较为全面的结果,而不同算法关注的角度和相应的精度却有所不

同, AI 算法在生态环境领域的应用大致可以分为以下 4 类^[17]。

1.2.1 基于概率统计类 概率统计的研究对象是自然过程中的随机事件、随机变量等,在假设的前提下通过概率判断事物发展的内在规律,常用算法有贝叶斯定理(Bayes' theorem)、马尔可夫链(Markov chain)等。贝叶斯定理可以估计某种情况发生的可能性,如模拟树木死亡率^[18]、评价水质状况^[19]等。而马尔可夫链可根据当前状态随机决定下一状态,在预测害虫的发生程度^[20]、分析生态安全的空间变化效应^[21]和土地利用景观格局变化^[22]等方面有广泛应用。

1.2.2 基于空间几何类 空间几何主要指算法在维度上的变化,如一维线条划分平面、二维平面划分空间以及向量运算提取特征等^[17],常用的算法有支持向量机等。Chen 等^[23]采用 3 种方法[多层感知机(multi-layer perceptron, MLP)、支持向量机(support vector machine, SVM)和径向基函数神经网络(radial basis function network, RBF)]分别模拟黑河流域中游的地下水位和流量,并与数值模拟方法的结果进行对比,结果显示, SVM 和 RBF 的模拟精度最高,基于描述物理过程的数值模型泛化能力最高,并得出了机器学习适用于执行大量计算而无须考虑机理过程的观点。由于 SVM 较高的精度,其各种衍生方法也广泛应用于微生物群落演替^[24]、生态系统健康评价^[25]等领域。

1.2.3 基于图论类 基于有向图等传统图结构、欧拉图、树结构等特殊图的人工智能算法包括决策树(decision tree, DT)、随机森林等,且在数据量有限时,后者所代表的集成学习能够得到更准确的误差。Wei 等^[26]改进了梯度提升回归树(gradient boosting regressor, GBDT),用于在高光谱遥感影像中估计土壤钾含量。Guio Blanco 等^[27]根据高程数据、遥感影像和几类植被指数,采用随机森林和 10 折交叉验证,较精准地预测了高山稀疏草地生态系统内研究区的土壤持水量。张文强等^[28]使用随机森林大大提高了归一化植被指数(NDVI)的量化精度。类似地,随机森林在悬浮泥沙估计^[29]、气象和植物生长周期预报^[30]等分类和回归任务中均有良好的表现。

1.2.4 基于交叉学科的群智能类 生命现象和计算机科学的结合是近些年主要的交叉领域之一。这种结合形成了遗传算法、粒子群算法等仿生优化算法以及深度学习两类发展方向。

仿生优化算法主要应用于模型参数率定,可大

幅提高寻找最优参数的速度。灰箱模型中,能通过物理方法测定的参数较少,大部分需要依靠观测数据率定,即在给定范围内迭代寻找参数使计算结果与实际值的偏差尽可能小。然而,在复杂的非线性计算中往往会出现“局部最优解”、“维数灾难”等问题,虽然粒子群算法、遗传算法等方法在寻找“全局最优解”上有明显优势,但由于率定过程在理论上完全依赖输入数据、约束条件和目标函数,因此,会出现“异参同效”等问题^[31]。

深度学习在生态环境领域的发展迅速。在水文学领域, Sit 等^[32]按使用频率总结了常用的深度学习技术,其中,图像预测最常使用卷积神经网络(convolution neural network, CNN),其采用卷积核将邻近点集成,以处理数据量较大的二维数据。Kang^[33]将改进的 CNN 模型用于识别照片中的波浪。然而,卷积同样可以使用在一维序列, Duan 等^[34]开发了一种时间卷积神经网络(temporal convolution neural network, TCNN)模型,用于气候变化条件下的日径流量预测,在均值和最值的预测上表现出色。Laloy 等^[35]研究表明, CNN 还能学习到更高维度的数据。而序列预测常用 LSTM、人工神经网络(artificial neural network, ANN)^[32]等方法。单独使用 LSTM 可以改进对长时间序列的应用效果,与 ANN 相比,其精度更高^[36],且与小波分析等手段结合,精度还能进一步提高^[37]。Kratzert 等^[38]证明了 LSTM 在未知流域的模拟精度比物理模型在已标定流域的模拟结果更好,传统的物理模型对数据的利用程度还有很大不足。

虽然大量研究证实了人工智能算法在生态环境领域应用的可行性,且效果可能比基于物理机制的模型更好,但由于其对自然过程的解释能力较弱,有时会产生与实测值偏差很大的模拟结果^[39-40]。考虑到机理模型具有更好的泛化能力和可解释性,研究人员尝试耦合 AI 算法与生态环境模型,在解决环境模型不足的同时,补足人工智能算法在应用时暴露的短板,如,异常事件精度差、泛化能力弱、可解释性差等。

2 耦合研究

环境本质上具有很强的异质性,且目前还存在未知的机理过程。因此,根据机理建立精确的生态环境模型难度较大,模型结构的不确定性、参数率定的数据需求以及模型运行的时间成本等问题也同样突出。人工智能算法本身不考虑自然过程机理,对

已有数据的利用程度很高,大大降低了求解难度,但同时也存在泛化能力差、异常事件预测精度差、对数据质量要求较高等不足。因此,探究生态环境模型与人工智能算法的耦合方式,使模拟结果具有更好的精度、泛化能力和更低的求解成本,是当前研究中一个极具挑战的方向。目前,人工智能算法与生态环境模型的耦合应用主要体现在以下两方面。

2.1 松散耦合

为了提高人工智能算法的模拟和预测能力,往往将模型模拟的结果加入算法输入数据中,通过提高输入数据质量的办法来提高精度。Li 等^[41]通过分别基于人工神经网络和 K 近邻算法(K-nearest neighbors, KNN)估计径流量和径流量误差,不断用模型输出作为下一时段的输入,解决了传统人工神经网络模型只能提前一个时段预报径流量的问题。Yang 等^[42]将全球水文模型(global hydrological models, GHMs)的模拟结果作为 LSTM 的输入,显著提升了预测效果。Noori 等^[43]将 SWAT 模型与人工神经网络结合,将 SWAT 模型模拟的日流量与降水量作为人工神经网络的输入数据,采用 SWAT-CUP 验证输出结果的精度。Rahmani 等^[44]反向操作将算法结果输入模型中成功减小了水温预测值的误差。松散耦合在一定程度上弥补了 AI 算法的缺陷,而更深层次的研究也在迅速展开。

2.2 紧密耦合

随着自然科学领域研究人员对人工智能算法的认知加深,人们逐渐认识到算法需要与物理机制结合才能更好地应对复杂多样的自然过程。Roweis 等^[45]首先验证了因子分析、混合高斯模型与机器学习耦合的可行性。Nearing 等^[3]将这种思路应用在降雨径流模型;Pelissier 等^[46]将混合高斯模型与地表模型耦合核算土壤水分,研究表明,这种耦合思路能够减小训练数据覆盖不全导致的误差。在总结前人耦合研究的基础上, Karpatne 等^[47]探讨了在材料、化学和水文学等领域实施耦合研究的可行性,提出了以机理为导向的耦合思路(theory guided data science, TGDS)。de Bézenac 等^[1]以海面温度预报模型(sea surface temperature, SST)为例,提出了将一般形式的偏微分方程引入深度学习,并说明了该方法在遵循辅助扩散原则基础上进行动态研究的应用前景。

已有的耦合研究思路主要分为两种:一种是在算法中添加含有约束条件的隐藏层,如 Ma 等^[48]在 LSTM 模型中新增一层考虑污染物浓度时空变化机

制的隐藏层,从而预测时间维度上 $PM_{2.5}$ 浓度变化,并对观测范围内 $PM_{2.5}$ 浓度进行空间插值,拓展了时空维度的泛化能力;Daw 等^[49]通过在神经网络层的连接中添加物理约束、在隐藏层增加具有物理意义的神经元,构建湖泊温度模拟模型以量化不确定性,并在有限的数据集获得了更好的泛化能力和更贴近观测数据的温度分布模拟结果,因此认为该结构可对机理模型的不确定性分析提供校正;Kraztzer 等^[38]用 LSTM 学习到了不同流域的相似性和差异性,以此作为物理约束提高了精度;Hoedt 等^[50]对于现有的 LSTM 算法模拟结果存在的质量不守恒问题,提出了在神经元连接间添加质量守恒约束的办法,较好地改善了降雨径流模拟过程中峰值精度差的情况,并提高了算法的可解释性。另一种则是用人工智能算法替换物理过程,如 Zhao 等^[51]采用神经网络模拟 Penman-Monteith 模型中的经验参数——地表阻抗的形成过程,将模拟结果作为 PM 方程的一部分,并以此判断了输入变量的影响程度。经检测,混合模型的泛化能力和对异常事件的适应能力比纯机器学习模型更好,因此,认为加入物理机制能够提高机器学习的精度。Reichstein 等^[7]对此提出了人工智能在自然科学领域中与机理模型耦合的 5 种方式,并列出了面临的 5 种挑战。

除以上两种基于已知机理的耦合方式外,还有更多的耦合角度正在被发掘。如,基于物理信息指引的神经网络 (physics-informed neural networks, PINNs) 从求解模型的角度,试图通过神经网络降低复杂的偏微分方程的求解难度^[52]。Mo 等^[53]以全连接的卷积神经网络 (DenseNet) 重现了偏微分方程在三维坐标上的变化。以上研究整体可归到添加物理机制、助力 AI 算法的研究角度中。Iten 等^[54]提出神经网络能够根据观测数据学习到有物理意义的参数,反过来证明 AI 有助于促进机理研究。基于自编码器 (auto-encoder) 的 SciNet 成功地仅从地火、地日间距时序观测数据中模拟出了地球与火星和太阳所呈角度的时间序列。

3 耦合研究的不足及构建探索

当前研究已经认识到,应用在自然科学领域的人工智能需要与机理过程更紧密的结合,才能进一步探索自然界的科学问题,但上述耦合方式仍有许多可以完善的方面。

3.1 不足之处

1) 耦合程度不高、方向发展不均衡。模型由数

据、参数和结构 3 部分构成。前文所总结的几种耦合方式主要聚焦于数据的改善和参数的调整。除上述应用外,人工智能算法的优势主要在于挖掘数据的潜在规律。然而,目前大多数耦合研究对已知机理的依赖更高,耦合方向聚焦于在算法中添加物理约束,对物理机制的利用还有很大的发展空间。

基于优势互补的理念,理论上除了 PINNs 这类有助于方程求解的耦合方式,AI 对机理模型也应当发挥一些积极作用,但相关研究,特别是在生态环境领域的研究还很少。沿着这个方向,还存在一些问题需要探索,如:AI 挖掘到的数据之间的潜在规律能否作为物理机制的一种获取和评价手段?或者能否作为参考,去探寻已获得机理的不足之处以及真正的未知机理过程?随着耦合程度的加深,这些问题还需要一一解答。

2) 参数意义不明。神经网络内每层参数的数量以及层与层之间的连接方式,基本由研究者先主观确定方案,再结合搜索结果和需求反复调整,导致无论输出精度多高,一旦环境条件发生改变又需要重新调参,使得参数缺乏实际意义,调整参数的过程也缺乏合理性和可解释性^[7]。

3) 变量重复参与计算。目前的松散耦合主要将 AI 和机理模型顺序连接,进行二次模拟以提高精度,但有些数据既作为机理模型的输入,同时也是 AI 算法的输入,造成部分变量重复参与计算的问题,这种耦合方法虽然提升了输出精度,但也造成过度干预 (over-interpretation)^[55]。

4) 数据不足。无论是机理模型,还是 AI 智能算法,都需要数据作为支撑。与工业领域的不同,自然科学领域的数据完整度有待提升。一方面是采集周期较长,易受其他因素影响 (天气、设备、疫情等) 导致部分缺失;另一方面是得到的影像数据难以给定标签,不便于后续的图像分类和预测。

以 TGDS 为核心的耦合方式对数据的需求暂且能够通过 AI 满足,但 PINNs 系列的研究需要大量的初始条件和边界条件以求解方程,因此,训练数据的制备难度较大^[56]。

5) 时空尺度不匹配。人工智能算法所使用数据的时间尺度与机理模型的时间尺度需要匹配,一般可以通过求和、平均值等方法升尺度,也可以采用插值的方法降尺度,但对于突发事件,降尺度插值存在较大误差。其次,机理模型既可以进行时间序列预测,也可以模拟研究区域内任意位置的参数,虽然 CNN 系列算法能够很好地模拟一维、二维数据,但

三维卷积要结合地形地貌特征,需要大量的数据和很高的时间成本^[56]。如何以较高的精度和速度在空间尺度上匹配,完成目标点的模拟和预测,也是耦合研究需要考虑的问题。

3.2 构建探索

3.2.1 思路分析 为解决 AI 算法对机理模型的补充作用研究不足、耦合程度不紧密以及参数缺乏意义这 3 个问题,在神经网络具有万能近似定理(universal approximation theorem)^[57]这一性质的基础上,借鉴 Reichstein 等^[7]和 Zhao 等^[51]用机器学习代替物理模型中半经验公式的想法,提出一个神经网络块与机理模型紧密耦合,重构机理计算过程的思路。

以 3 层神经网络和生态环境领域的水环境模型为例,从模型结构的角度提出如图 1 所示的耦合方式,使形成的神经网络能够隐式地表达数学规律和公式。

在图 1 的输入层部分,对于现有的 7 种实测数据(A~K),首先进行相关性分析,这里认为相关性强的数据,实际联系也更紧密,因此,将彼此间相关性强的数据分为一组。假设最终分成 3 组,把每组紧密相连的数据投入到一个 3 层神经网络进行学习,输出的 F_1 、 F_2 、 F_3 等价于神经网络学到了这组数据的特征集合,也就相当于分别完成了水环境机理模型的部分公式计算(如: F_1 集成了输入数据中与污染物浓度有关的特征、 F_2 集成了输入数据中与降

水有关的特征、 F_3 集成了输入数据中与空气流动有关的特征)。但由于神经网络仅仅是 $f(ax+b)$ 的层层连接形式,因此,与经验公式和复杂物理机制构成的水环境模型相比,在特征提取和计算上更具优势,且 3 个隐藏层参数被赋予了实际意义。

将输入数据按相关性分组后再训练的神经网络,可仅以小型神经网络的规模快速运行;但在利用神经网络计算优势的同时,分组后的输出参数却囊括了该组数据的主要实际意义。几个神经网络的组合近似地按照白箱模型的形成方式,将公式融入到神经网络的框架中,完成了按照机理模型输出计算结果的过程,赋予了部分参数实际意义,便于后续从模型结构上调参。

3.2.2 研究意义 上述耦合方式仿照已知的机理过程,将数据分组再集合,能够部分继承机理模型的优势,进一步探索对算法可解释性、异常值捕捉能力等不足之处的改善程度。该方法与 SciNet 存在很大不同^[54],学习的特征更复杂,网络构成更符合机理的形成过程,耦合程度更加深入。虽然理论上较有明显的应用前景,但实际应用时也还有许多问题需要考虑,如:网络层数的确定、用哪种方法能更好地避免梯度问题、输入变量的具体分组方式等。因此,应先从简单的经验公式做起,再逐步尝试在复杂模型中应用,进一步探究 AI 在验证已知机理和探究未知机理所能发挥的作用。

4 耦合应用发展趋势

4.1 合与分

随着算法在自然科学领域的应用加深,机理模型与算法的耦合思路有所拓宽,且逐渐趋向紧密耦合。虽然紧密耦合是重点发展对象,但在急需切实提高精度的实际应用中,松散耦合也仍然有且将长期有广泛的使用优势。在加紧探究紧密耦合方式的同时,算法和生态环境机理也需要各自有更进一步的发展。因为耦合以两种方法的缺点为起点,以集合优势为目标,更好的耦合效果还需要充足的试验数据作为保障,以及清晰的机理过程互相印证。

4.2 数据与算法

在当前算法与机理模型耦合应用的发展趋势下,相关研究对数据和算法的依赖越来越强,生态环境领域也同样如此。然而,当前生态环境数据主要通过人工采集,再辅以设备测量,在时空连续性、准确性等方面还有较大的提升空间;算法的种类大幅增长,各种改进方法层出不穷,但没有算法能够通用地解决各类

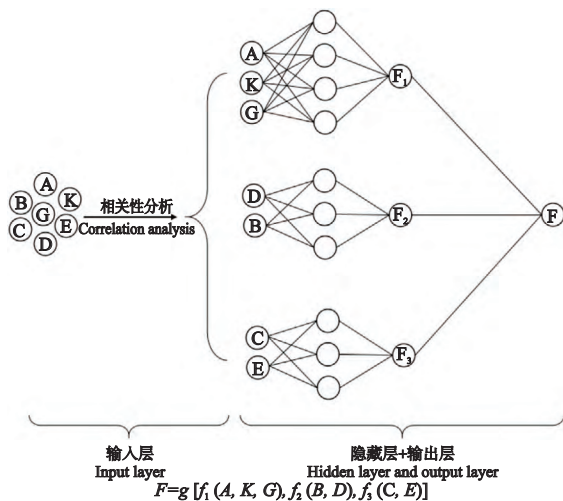


图 1 紧密耦合思路的示意图

Fig.1 Sketch map of the proposed coupling method of physical mechanism and AI.

A~K 为 7 种变量的实测数据 Letter A~K denoted measured data of seven variables; F_1 、 F_2 、 F_3 为分组后小型神经网络的输出,具有相应的实际意义 F_1 、 F_2 、 F_3 indicated outputs of three small neural networks, which had physical significance corresponding to grouped input data; F 为输出 F was the ultimate output.

问题,因此,如何引入并改进算法,以适用于生态环境领域的研究也是需要仔细考虑的问题。

4.3 尺度推绎

数据的潜在关联能够一定程度上反映时空尺度变化的物理机制。从时间线上看,事件尺度(如洪水、雾霾等)、日尺度、月尺度变化密切相关;考虑空间尺度,气、水、土等要素在各响应单元、区域间彼此依赖。但至今也没有较好的理论能清晰地解释不同尺度之间复杂的递推过程。将机理模型与 AI 适度耦合,有望通过挖掘数据的潜在联系,助力尺度推绎机理的研究。

5 小 结

本文概述了人工智能算法的引入对生态环境模型研究的影响,重点分析其与模型耦合的研究进展。在指出现有耦合方式不足之处的同时,从提高紧密耦合程度、改善参数意义和拓宽研究思路等角度,以水环境模型为例,提出一种将模型拆解融入网络结构、重构机理过程运行方式的新耦合角度,并分析了该网络在改善参数缺乏实际意义,模拟机理过程并有助于机理研究的可能性。期望该思路的提出能为人工智能算法与机理模型的紧密耦合提供更多可能。

参考文献

- [1] de Bézenac E, Pajot A, Gallinari P. Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, 2019, **2019**: 124009
- [2] 刘思峰, 杨英杰, 吴利丰. 灰色系统理论及其应用. 第7版. 北京: 科学出版社, 2014: 3
- [3] Nearing GS, Gupta HV. The quantity and quality of information in hydrologic models. *Water Resources Research*, 2015, **51**: 524–538
- [4] Blöschl G, Bierkens MFP, Chambell A, et al. Twenty-three unsolved problems in hydrology (UPH): A community perspective. *Hydrological Sciences Journal*, 2019, **64**: 1141–1158
- [5] 周信哲, 陈永娟, 蒲文鹏. 机器学习在水质预测中的应用. 2020 中国环境科学学会科学技术年会, 南京, 2020: 3948–3953
- [6] Nearing GS, Kratzert F, Sampson AK, et al. What role does hydrological science play in the age of machine learning? *Water Resources Research*, 2021, **57**: e2020WR028091
- [7] Reichstein M, Camps-Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science. *Nature*, 2019, **566**: 195–204
- [8] Sun AY, Scanlon BR. How can Big Data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letters*, 2019, **14**: 073001
- [9] Liu ZL, Peng CH, Xiang WH, et al. Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese Science Bulletin*, 2010, **55**: 3853–3863
- [10] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS One*, 2018, **13**(3): e0194889
- [11] 方精云. 生态学学科体系的再构建. *大学与学科*, 2021, **2**(4): 61–73
- [12] Hubbard SS, Varadharajan C, Wu YX, et al. Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry. *Hydrological Processes*, 2020, **34**: 3175–3182
- [13] Mital U, Dipankar D, Brown JB, et al. Sequential imputation of missing spatio-temporal precipitation data using random forests. *Frontiers in Water*, 2020, **2**: 20
- [14] Nguyen DH, Bae DH. Correcting mean areal precipitation forecasts to improve urban flooding predictions by using long short-term memory network. *Journal of Hydrology*, 2020, **584**: 124710
- [15] Cubuk ED, Zoph B, Mane D, et al. Autoaugment: Learning augmentation strategies from data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 113–123
- [16] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, **6**: 60
- [17] 陶阳明. 经典人工智能算法综述. *软件导刊*, 2020, **19**(3): 276–280
- [18] Yahya Y, Ismail R. Using uncertainty of Bayesian theorem to predict mortality of tree in forest growth simulation system. Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, New York, USA, 2018: 1–5
- [19] 廖杰, 王文圣, 丁晶. 贝叶斯公式在河流水质综合评价中的应用. *四川师范大学学报: 自然科学版*, 2007, **30**(4): 519–522
- [20] 闫香慧, 赵志模, 刘怀, 等. 应用马尔可夫链理论对褐飞虱和白背飞虱发生程度的预测. *生态学报*, 2009, **29**(11): 5799–5806
- [21] 薛亮, 任志远. 基于空间马尔可夫链的关中地区生态安全时空演变分析. *生态环境学报*, 2011, **20**(1): 114–118
- [22] 何丹, 金凤君, 周璟. 基于 Logistic-CA-Markov 的土地利用景观格局变化——以京津冀都市圈为例. *地理科学*, 2011, **31**(8): 903–910
- [23] Chen C, He W, Zhou H, et al. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. *Scientific Reports*, 2020, **10**: 3904
- [24] Liu R, Gu Y, Shen M, et al. Predicting postmortem interval based on microbial community sequences and machine learning algorithms. *Environmental Microbiology*, 2020, **22**: 2273–2291
- [25] 毕温凯, 袁兴中, 唐清华, 等. 基于支持向量机的湖泊生态系统健康评价研究. *环境科学学报*, 2012, **32**(8): 1984–1990
- [26] Wei L, Yuan Z, Zhong Y, et al. An improved gradient boosting regression tree estimation model for soil heavy

- metal (Arsenic) pollution monitoring using hyperspectral remote sensing. *Applied Sciences*, 2019, **9**: 1943
- [27] Guio Blanco CM, Brito Gometz VM, Crespo P, *et al.* Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. *Geoderma*, 2018, **316**: 100–114
- [28] 张文强, 罗格平, 郑宏伟, 等. 基于随机森林模型的内陆干旱区植被指数变化与驱动力分析: 以北天山北坡中段为例. *植物生态学报*, 2020, **44**(11): 1113–1126
- [29] 方馨蕊, 温兆飞, 陈吉龙, 等. 随机森林回归模型的悬浮泥沙浓度遥感估算. *遥感学报*, 2019, **23**(4): 756–772
- [30] 李德, 陈文涛, 乐章燕, 等. 基于随机森林算法和气象因子的砀山酥梨始花期预报. *农业工程学报*, 2020, **36**(12): 143–151
- [31] 芮孝芳, 蒋成煜, 张金存. 流域水文模型的发展. *水文*, 2006, **26**(3): 22–26
- [32] Sit M, Demiray BZ, Xiang ZR, *et al.* A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 2020, **82**: 2635–2670
- [33] Kang L. Wave monitoring based on improved convolution neural network. *Journal of Coastal Research*, 2019, **94**: 186–190
- [34] Duan S, Ullrich P, Shu L. Using convolutional neural networks for streamflow projection in California. *Frontiers in Water*, 2020, **2**: 28
- [35] Laloy E, Héroult R, Jacques D, *et al.* Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 2018, **54**: 381–406
- [36] Fan H, Jiang M, Xu L, *et al.* Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water*, 2020, **12**: 175
- [37] Mouatadid S, Adamowski JF, Tiwari MK, *et al.* Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. *Agricultural Water Management*, 2019, **219**: 72–85
- [38] Kratzert F, Klotz D, Herrnegger M, *et al.* Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 2019, **55**: 11344–11354
- [39] Eisenstein M. Big data: The power of petabytes. *Nature*, 2015, **527**: S2–S4
- [40] Lazer D, Kennedy R, King G, *et al.* The parable of Google Flu: Traps in big data analysis. *Science*, 2014, **343**: 1203–1205
- [41] Li ZJ, Kan GY, Yao C, *et al.* Improved neural network model and its application in hydrological simulation. *Journal of Hydrological Engineering*, 2014, **19**: 04014019
- [42] Yang T, Sun FB, Gentine P, *et al.* Evaluation and machine learning improvement of global hydrological model-based flood simulations. *Environmental Research Letters*, 2019, **14**: 114027
- [43] Noori N, Kalin L. Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, 2016, **533**: 141–151
- [44] Rahmani F, Lawson K, Appling A, *et al.* Process learning of stream temperature modelling using deep learning and big data. AGU 2021 Fall Meeting, New Orleans, LA, USA, 2021; 1
- [45] Roweis S, Ghahramani Z. A unifying review of linear Gaussian models. *Neural Computation*, 1999, **11**: 305–345
- [46] Pelissier C, Frame J, Nearing G. Combining Parametric Land Surface Models with Machine Learning [EB/OL]. (2020-12-21) [2022-01-01]. <https://arxiv.org/ftp/arxiv/papers/2002/2002.06141.pdf>
- [47] Karpatne A, Atluri G, Faghmous JH, *et al.* Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 2017, **29**: 2318–2331
- [48] Ma J, Ding YX, Chen JCP, *et al.* A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for PM_{2.5}. *Journal of Cleaner Production*, 2019, **237**: 117729
- [49] Daw A, Thomas RQ, Carley CC, *et al.* Physics-guided architecture (PGA) of neural networks for quantifying uncertainty in lake temperature modeling. Proceedings of the 2020 SIAM International Conference on Data Mining, Cincinnati, OH, USA, 2020; 532–540
- [50] Hoedt PJ, Kratzert F, Klotz D, *et al.* MC-LSTM: Mass-conserving LSTM. International Conference on Machine Learning, PMLR, Chengdu, China, 2021; 4275–4286
- [51] Zhao WL, Gentine P, Reichstein M, *et al.* Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, 2019, **46**: 14496–14507
- [52] Raissi M, Perdikaris P, Karniadakis GE. Physics Informed Deep Learning (part I): Data-driven Solutions of Nonlinear Partial Differential Equations [EB/OL]. (2017-12-12) [2021-12-12]. <https://doi.org/10.48550/arXiv.1711.10561>
- [53] Mo W, Luo X, Zhong Y, *et al.* Image recognition using convolutional neural network combined with ensemble learning algorithm. *Journal of Physics: Conference Series*, 2019, **1237**: 022026
- [54] Iten R, Metger T, Wilming H, *et al.* Discovering physical concepts with neural networks. *Physical Review Letters*, 2020, **124**: 010508
- [55] Chauvet JM. The 30-year Cycle in the AI Debate [EB/OL]. (2018-12-12) [2021-12-12]. <https://doi.org/10.48550/arXiv.1810.04053>
- [56] Camps-Valls G, Tuia D, Zhu XX, *et al.* Deep Learning for the Earth Sciences [EB/OL]. (2021-12-12) [2022-01-01]. <https://doi.org/10.1002/9781119646181.ch19>
- [57] Hornik K, Stinchcombe M, White H. Multilayer feed forward networks are universal approximators. *Neural Networks*, 1989, **2**: 359–366

作者简介 胡羽聪, 女, 1999年生, 硕士研究生。主要从事人工智能在流域模型上的应用研究。E-mail: ychu2020_st@rcees.ac.cn

责任编辑 杨弘