

李颖,陈怀亮. 机器学习技术在现代农业气象中的应用. 应用气象学报,2020,31(3):257-266.

DOI: 10.11898/1001-7313.20200301

机器学习技术在现代农业气象中的应用

李颖¹⁾²⁾ 陈怀亮¹⁾³⁾*

¹⁾(中国气象局·河南省农业气象保障与应用技术重点实验室,郑州 450003)

²⁾(河南省气象科学研究所,郑州 450003)

³⁾(河南省气象局,郑州 450003)

摘要

智慧气象和精准农业结合下的现代农业气象工作意味着对包含遥感影像在内的大型农业和气象数据高时效性的分析与处理,机器学习技术是当代自然科学研究和技术发展的主流技术,亦是现代农业气象科研和业务发展的重要工具。该文系统论述了机器学习技术的主要方法及其在现代农业气象中的主要应用方向,比较了不同方法在农业气象不同领域应用的情况,侧重介绍了基于深度学习技术的成果和近年来的最新研究进展。传统浅层机器学习技术中,以支持向量机和人工神经网络应用最为广泛且效果最为理想。近年来,随机森林和梯度提升机等决策树集成方法普遍取得优于核方法的精度,深度学习技术则在某些任务中取得更优于集成学习的精度。未来,有待检验机器学习技术特别是深度学习技术在更多农业气象问题上的适用性和先进性,更好地迎接现代农业气象发展的新挑战与新机遇。

关键词: 机器学习;深度学习;农业气象;农业遥感

引言

农业在全球经济中发挥着基础性且至关重要的作用,国家和地区农业生产力高低直接影响其工业化程度与经济水平,落后的农业技术会带来饥饿、贫穷与社会发展的滞后^[1]。到 2050 年,全球将新增 20 亿左右人口^[2],加之气候变化对农业生产的影响^[3],使得消除饥饿和保障粮食安全成为当今世界可持续发展的重要议题^[4],农业技术的进步可有力应对该挑战^[5]。精准农业被列为农业领域的十大发展之一^[6],其特点是应用密集的数据——以遥感技术^[7]和无线传感器技术^[8]为主要手段采集信息并进行时空处理,提高农业生产效率、作物产量和环境质量。农业气象学是研究农业生产与气象条件之间相互关系及其规律的科学,以促进农业生产为主旨,围绕现代气象与现代农业的智慧化进程也在不断进行

着自身的科学创新,智慧气象和精准农业结合下的现代农业气象工作意味着对纳入遥感可视化数据在内的大型农业和气象数据高时效性的分析与处理^[9],机器学习(machine learning, ML)技术对其发展有很大的助力。

ML 是图像处理和大数据分析不可或缺的技术,广泛应用于医学、药学、经济学、生物学、水文学、农业气象学等诸多科学领域^[10-13]。深度学习(deep learning, DL)和浅层学习中的梯度提升机(gradient boosting machine, GBM)是当前最受瞩目的两项 ML 技术,其中 DL 结构中最著名的卷积神经网络(convolution neural network, CNN)自 2012 年起已成为计算机视觉任务的首选解决方案^[14]。现代农业气象研究不仅涉及大量气象数据、土壤数据、作物观测数据,也涉及到农业遥感中采集自地面、无人机、卫星的海量影像数据,DL 技术的特点及其在机器视觉领域的优势使其在现代农业气象工作中具有

2019-10-22 收到,2019-12-06 收到再改稿。

资助项目:国家自然科学基金项目(41805090),气候变化专项课题(CCSF201934),中国气象局·河南省农业气象保障与应用技术重点开放实验室开放基金项目(AMF201802,AMF201807),河南省气象局气象科学技术研究重点项目(KZ201803)

* 通信作者,邮箱:H.chen@vip.163.com

很大的应用潜力^[15]。

本文对 ML 技术的主要方法及其在现代农业气象中尤其是涉及农业遥感的应用进行系统性介绍。由于所涉及的文献众多,侧重列举代表性文献,对其研究成果进行概要介绍,且重点列举出 DL 技术的应用实例,旨在推动 ML 技术特别是 DL 技术在现代农业气象科研与业务中的深入应用。

1 ML 技术概述

ML 技术蓬勃发展于 20 世纪 90 年代,是人工智能(artificial intelligence, AI)中最受欢迎和最成功的子领域。Arthur Samuel 将 ML 定义为一门不需要通过外部程序指令而让计算机具有自我学习能力的学科。在传统编程中,人类输入规则和需要规则处理的数据,计算机输出答案;而在 ML 技术中,人类输入数据和期望从数据中得到的答案,计算机通过训练找到数据和答案间的统计结构、输出规则,并将这些规则应用于该任务的新数据进而生成答案^[14]。ML 技术与经典统计分析的一项重要区别是 ML 技术倾向于处理大型、复杂的数据集,以及没有已知算法可解决的问题。将 ML 技术用于大数据挖掘可以适应新的数据,发现数据中隐含的模式,减少人工分析工作量,更好地处理并解决传统方法难以应对的复杂问题^[16]。

ML 技术是一个日益庞大的家族,其包含的众多算法与模型可根据不同标准进行归类。其中一种广泛使用的分类方法是根据训练过程中得到的监督的数量和类型,将其分为监督学习、非监督学习、半监督学习和强化学习^[16]。在监督学习中,需要人工为训练数据加标签(即明确的属性标识),其代表性方法包括线性回归(linear regression)、逻辑回归(logistic regression)、朴素贝叶斯(naïve bayes)、高斯判别(gaussian discriminant analysis, GDA)、支持向量机(support vector machine, SVM)、神经网络(neural network, NN)、K 最邻近法(K-Nearest neighbor, K-NN)、决策树(decision trees, DTs)、随机森林(random forest, RF)和梯度提升机等;在非监督学习中,训练数据则不加标签,其主要方法包括以期望最大化算法(expectation maximization)、分层聚类分析(hierarchical cluster analysis, HCA)、K 均值法(K-means)为代表的聚类方法,以主成分分析(principal component analysis, PCA)和局部线性

嵌入算法(locally-linear embedding, LLE)为代表的降维算法,以及 Apriori, FP-Growth, Eclat 等关联规则学习算法;在半监督学习中,仅需为少量训练数据加标签,或为训练数据加不确定性标签,其重要方法包括拉普拉斯支持向量机(laplacian SVM)、协同过滤算法(collaborative filtering),以及贝叶斯网络(bayesian network)、马尔科夫随机场(markov random field)等概率图模型;在强化学习中,不需要预先给定训练数据,而是通过接收环境对动作的反馈获得学习信息,代表性算法包括策略梯度(policy gradient, PG)、Q 学习(Q-learning)、深度 Q 网络(deep Q network)、Sarsa 算法等^[17-20]。

DL 技术是 ML 技术的一个子领域,代表着一类思想,即以多层结构从数据中学习表示(representation),其结构通常包含数十个乃至上百个连续的表示层。DL 技术起源于神经网络,在神经网络基础上增加了层级,可自动提取复杂特征,近年来,DL 技术以其更高精度和更优性能,在诸多应用领域已经取代了以往支持向量机和集成学习的领先地位^[14]。DL 技术较经典 ML 技术的优势主要体现在以下几方面:①DL 技术完全自动化了浅层 ML 技术的关键步骤——特征工程,对使用者更为简单、友好;②DL 技术具有深层非线性网络结构,采用递增的、逐层的方式开发愈加复杂的特征,具有更强的学习能力,有助于解决浅层 ML 技术难以解决的复杂问题,并可进行迁移学习;③DL 技术可以一次性学习所有特征,并持续在线学习,具有实时运算能力^[21-23]。DL 的基本模型包括卷积神经网络、深度信念网络(deep belief network, DBN)、循环神经网络(recurrent neural network, RNN)、深度自动编码器(deep autoencoder, DA)、递归神经网络(recursive neural network, RNN)等。

2 ML 技术在农业气象工作中的应用

2.1 制图与区划

土地覆盖与作物类型图是农业气象工作的重要基础数据之一。过去的十几年中,随着遥感数据时空分辨率不断提高,以及大量丰富的免费数据源向公众开放,将遥感影像用于土地覆盖与作物类型分类制图方面的研究呈指数增长,ML 技术中多种经典算法、模型已成功应用于该类任务^[24-26],根据 Yu 等^[27]的统计,最大似然分类法使用频率最高,相关

文献中应用比例达 32.34%,最大似然分类法和 K 最邻近法、K 均值法等也是文献中平均精度较低的方法,分类精度较高的方法则是集成分类器、人工神经网络和支持向量机。Khatami 等^[28]进一步统计分析了既往研究,指出传统监督分类方法中支持向量机平均精度最高,紧随其后的是人工神经网络,且当影像空间分辨率和光谱分辨率提高时,支持向量机表现出较人工神经网络更大的优势。此外,ML 技术在作物管理区的划分方面已有成功应用的例子,Pantazi 等^[29]利用 K 均值法和自组织映射结合遥感数据、土壤参数和产量数据进行了作物管理分区,进一步看,农业气候区划(如农作物品质气候区划)、农业气象灾害风险区划、农业保险风险区划^[30-31]等是农业气象工作的重要任务,有待有针对性地将 ML 技术应用于区划工作。

近年来,最受欢迎和最具效率的多源多时相遥感影像土地覆盖与作物制图方法是集成学习和 DL 技术^[32]。针对复杂地区的分类问题,集成学习的经典算法随机森林的分类精度明显优于传统的决策树^[33],DL 结构中的卷积神经网络、深度自动编码器、深度信念网络、循环神经网络等均被用于探索该类任务^[34-38]。Minh 等^[39]利用两种循环神经网络结构结合星载合成孔径雷达(synthetic aperture radar, SAR)影像制作了冬季植被质量分类图,制图精度优于支持向量机和随机森林。Yang 等^[40]研究表明:DL 技术用于土地覆盖分类的精度高于支持向量机等浅层学习模型,且 DL 技术无需人工设计分类特征,并可在分类中使用迁移学习。Kussul 等^[32]将卷积神经网络用于作物制图的精度与随机森林和一种集成的多层感知器(multi-layer perceptrons, MLPs)方法对比,结果表明卷积神经网络精度最高。

2.2 检测与观测

杂草检测是地基农业遥感的一项重要任务,有研究认为杂草是对农作物生产最大的威胁,ML 技术和田间传感器结合可以精确检测田间杂草,进而应用于农业工具和农业机器人的除草作业,最大程度减少除草剂的使用^[41]。Cho 等^[42]使用电荷耦合元件(charge coupled device, CCD)相机和人工神经网络组成的机器视觉系统识别了杂草与萝卜。Karimi 等^[43]将支持向量机用于玉米田杂草和氮素胁迫检测,取得比人工神经网络更高的精度。Binch 等^[44]的对比研究表明:在经典 ML 技术中,支持向量机取得最优的杂草检测效果。近年来的研究表

明:DL 技术可有效从图像中自动提取特征,在目标识别中取得优于支持向量机的精度^[45-47]。王璨等^[48]应用卷积神经网络准确识别了幼苗期玉米与杂草。Dyrmann 等^[49]应用卷积神经网络在农田影像中识别了 22 种杂草与作物物种,并在后续研究中将全卷积神经网络应用于存在严重叶片遮挡的情况下,从谷物田中成功识别单株杂草^[50]。除杂草检测,张雪芬等^[51]利用支持向量机结合 CCD 影像实现了作物发育期的图像自动识别,余卫东等^[52]在对中国农业气象自动化观测的展望中提及计算机视觉技术的应用,ML 技术在今后的农业气象自动化观测中有待发挥更重要的作用。

DL 技术可以从高维海量数据中强有力提取复杂的结构信息^[21],近年来在植株表型观测、病虫害检测、农田障碍检测、果实检测等任务中得到成功应用^[53-55],可极大提升农业气象自动化观测水平。Christiansen 等^[56]对比了 DL 技术和经典 ML 技术在农田障碍和异常检测中的应用效果,结果显示 DL 技术具有最高精度和最快运算速度。Yalcin 等^[57]利用卷积神经网络自动提取图像特征,识别农业植被的物候期,精度优于基于手工设计特征的经典 ML 技术。Jin 等^[58]应用更快速的区域卷积神经网络和区域生长法从 Lidar 3D 点云中分割单株玉米,可准确测量植株高度。Ubbens 等^[59]开发了可用于叶片计数等植物表型任务的 DL 平台。Xiong 等^[60]开发了基于卷积神经网络的图像分割软件用于水稻穗部分割,可实现水稻表型自动化测量,段凌凤等^[61]和张领先等^[62]开展了类似研究。Baweja 等^[63]使用 CCD 相机和卷积神经网络组成的机器视觉系统自动计算茎秆数并测量茎宽。黄双萍等^[64]利用卷积神经网络和穗株高光谱图像提取不同尺度穗瘟病斑分布式特征,实现水稻穗瘟病害的精准检测。Mohanty 等^[65]通过迁移学习和重新训练两种方式训练卷积神经网络,有效识别 14 种作物物种和 26 种作物病害,孙俊等^[66]开展了类似工作。Rahnemoonfar 等^[67]提出一种 DL 结构用于果实计数,即使水果处于阴影下,或被树叶、树枝遮挡,或水果之间存在一定程度的重叠,也能有效计数。薛月菊等^[68]利用 DL 技术中的 YOLOv2 网络检测未成熟芒果,表明该方法在复杂场景下的检测精度优于更快速的区域卷积神经网络。

2.3 产量预测

在健康的生态系统下以最低成本取得最大作物

产量是农业生产最重要的目标之一^[69]。作物产量预测是农业气象工作中的一项重要任务,关系到粮食安全、种植结构调整、作物管理、农业保险等,提高预测的准确性有助于增加作物产量和商业利润。经典统计方法难以准确预测作物产量的非线性时空变化,当前更有效的产量预测方法包括作物生长模拟和 ML 技术^[70]。作物生长模型将作物生长阶段的动态机制以数学模型抽象表达^[71],建模过程耗时且昂贵,其运行所需参数集在发展中国家尤难获取。另一方面的研究中,经典 ML 技术与遥感数据、气象数据、土壤数据等结合实现了不同尺度下不同作物产量的准确预测^[72-73],并有学者对比不同方法的预测能力。Fortin 等^[74]研究表明:在马铃薯产量预测中,多层感知器预测效果优于多元线性回归。Rup^[75]对比了多层感知器、回归树、径向基核函数网络和支持向量回归用于冬小麦产量预测的精度,表明支持向量回归预测结果最准确。González 等^[70]对比了多元线性回归、M5-Prime 回归树、人工神经网络、K 最邻近法和支持向量回归对大规模种植的多种作物产量的预测能力,结果显示:M5-Prime 回归树表现最优,作者同时指出变量和属性的选取直接影响不同算法、模型的预测精度,这是其研究与前人研究结论存在差异的主要原因。与产量预测紧密相关的农业气象灾害风险评估是农业气象学领域中研究的热点^[76-80],较之传统技术手段,ML 技术对此有很大的应用潜力,目前已有研究将 ML 技术用于农业气象灾害遥感监测与产量影响评估,如 Park 等^[81]利用随机森林等 ML 技术与多源遥感数据结合,对农业干旱进行监测评估。

近几年,将 DL 技术应用于作物产量预测的研究陆续出现^[82]。Kuwata 等^[83]对比使用 DL 技术和支持向量回归预测玉米产量,显示 DL 技术预测精度更高。Kim 等^[84]将支持向量机、随机森林、极限随机树和 DL 技术用于玉米产量预测,显示 DL 技术精度最高,且可克服一般 ML 技术应用中的过拟合问题,更稳定。You 等^[85]利用公众可获得的遥感数据预测大豆产量,使用了卷积神经网络和循环神经网络中的长短期记忆(long-short term memory, LSTM)网络与 3 种对比方法,显示卷积神经网络和长短期记忆网络可自动学习原始数据中的有效特征并取得最高的预测精度。Wang 等^[86]将深度迁移学习和回归模型分别与遥感数据结合用于作物产量预测,表明使用迁移学习的长短期记忆网络预测精度

最高。

2.4 参数估算

农业气象研究相关的水文、土壤、作物参数通过站点观测无法取得其连续准确的空间分布情况,且某些参数测量难度大、费用昂贵。ML 技术与气象数据、遥感数据等相结合,可简单、高效地实现参数估算,且使其时空连续性监测与预报成为可能。农业气象工作关注的热点参数包括蒸散、土壤湿度、土壤温度、氮素含量、叶面积指数、生物量等。

准确估算蒸散对农业灌溉水资源时空优化配置至关重要,同时该参数测量难度较大。Yang 等^[87]利用支持向量机结合通量观测数据与 MODIS 遥感数据实现了大尺度蒸散的时空变化预测。Jung 等^[88]使用一种模型树集成的 ML 技术集成站点观测蒸散与遥感数据和气象数据,估算全球尺度的多年蒸散。Patil 等^[89]将 ML 技术与气象数据结合估算参考蒸散,表明单层前馈神经网络中的极限学习机(extreme learning machine, ELM)和最小二乘支持向量机估算精度高于经验模型。Mehdizadeh 等^[90]利用支持向量机、基因表达式编程、多元自适应回归样条与气象观测资料结合估算干旱与半干旱地区的月平均参考蒸散,显示支持向量机和多元自适应回归样条效果最好。

土壤温湿度和养分含量直接影响作物生长发育和产量形成,其信息获取对农业生产中科学高效的水肥管理具有重要意义。Baghdadi 等^[91]利用多层感知器和 SAR 数据结合估算了农业区的土壤表面粗糙度和土壤湿度。Srivastava 等^[92]对比了支持向量机、关联向量机、人工神经网络和广义线性模型在土壤湿度降尺度中的应用效果,显示基于人工神经网络的降尺度方法提高遥感反演土壤湿度空间分辨率效果最好。Nahvi 等^[93]利用极限学习机与气象数据结合估算了不同深度的日平均土壤温度,效果理想。Morellos 等^[94]利用偏最小二乘回归、主成分回归,以及两种 ML 技术(最小二乘支持向量机和 Cubist 算法)与地面光谱数据结合估算土壤全氮、有机碳和含水量,结果显示:最小二乘支持向量机对含水量和有机碳估算效果最好,而 Cubist 算法对全氮估算效果最好。

叶面积指数和生物量是反映植被生长状况的重要参数,与植被生产力和作物产量密切相关。Ali 等^[95]的综述显示,人工神经网络、支持向量回归和随机森林回归是生物量和叶面积指数遥感反演中使

用频率较高的算法。Prasad 等^[96]利用人工神经网络实现菠菜不同生育期叶面积指数和生物量等参数的反演。Jia 等^[97]利用人工神经网络与地面散射计数据和 RADARSAT-2 影像结合反演水稻生物量,取得理想精度。Wang 等^[98]将随机森林回归和支持向量回归、人工神经网络两种对比方法用于小麦多个生育期生物量的遥感反演,结果显示:随机森林回归估算精度最高,且其稳健性与支持向量回归相当,优于人工神经网络。Mao 等^[99]对比了人工神经网络、支持向量回归、随机森林回归、高斯过程回归和梯度升压回归树(gradient boosting regression tree, GBRT)等 ML 技术在棉花叶面积指数遥感反演中的应用效果,表明梯度升压回归树估算精度最高且稳健性最好,支持向量回归则计算效率最高。ML 技术与遥感数据结合,还可反演作物的其他生物物理和生物化学参数,并可监测作物生长过程的相关活动。Liu 等^[100]利用一种神经网络模型与微波亮温数据结合,反演了小麦植株含水量。Yang 等^[101]分别利用支持向量回归和逐步多元回归与高光谱反射率结合反演水稻叶面积指数和叶绿素含量,表明支持向量回归在水稻生理生化参数估算方面优于逐步多元回归。Abdel-Rahman 等^[102]利用随机森林回归和逐步多元回归分别与 Hyperion 高光谱数据结合估算甘蔗叶片氮素含量,表明随机森林回归估算精度高于逐步多元回归。Van Wittenberghe 等^[103]利用高斯过程与光谱数据结合成功估算了叶片含水量、叶绿素含量、氮素含量和比叶面积等生化与结构参数。Maimaitijiang 等^[104]分别利用偏最小二乘回归、支持向量回归和极限学习机回归与多源遥感数据结合反演大豆的生理参数(叶面积指数和生物量)与生化参数(氮素含量和叶绿素含量),显示极限学习机回归效果最理想。

近些年,将 DL 用于上述参数估算的研究陆续见诸报道。Song 等^[105]提出一种联合深度信念网络与宏观细胞自动机(macroscopic cellular automata, MCA)的 DBN-MCA 模型,结合环境变量预测土壤湿度,显示其预测精度高于经典 ML 技术,为预测高度非线性特点的土壤湿度提供了一个强有力的工具。王璨等^[106]利用卷积神经网络和近红外光谱预测土壤含水率,结果显示:卷积神经网络预测精度优于人工神经网络、偏最小二乘回归和最小二乘支持向量机等对比方法。Ma 等^[107]利用田间数字影像和卷积神经网络估算冬小麦早期生育阶段的地上生

物量,结果显示该方法具有良好的稳健性。马浚诚等^[108]利用可见光图像和卷积神经网络估算冬小麦冠层叶面积指数和地上生物量,表明卷积神经网络估算精度优于支持向量机和随机森林两种对比方法。此外,DL 技术还被用于农业气象条件预测、畜牧业等相关研究,如 Sehgal 等^[109]在作物规划中利用长短期记忆网络预测天气和土壤属性。

3 小 结

本文系统概述了 ML 技术的主要方法及其在现代农业气象中的主要应用方向,有针对性且全面涵盖了现代农业气象中特别是涉及农业遥感的 ML 技术的研究及应用情况,并纳入近年来最新的研究进展。

本文将 ML 技术在农业气象工作中的主要应用归纳为 4 个方面:制图与区划、检测与观测、产量预测和参数估算。在制图与区划方面,ML 技术与遥感影像结合实现了不同尺度的土地覆盖与作物类型制图,亦已结合遥感数据、土壤数据、统计数据用于作物长势、植被质量等专题图的制作与作物管理区划分;在检测与观测方面,ML 技术成功用于田间影像中的杂草检测,DL 技术在植株表型观测、病虫害检测、农田障碍检测、果实检测等方面,取得了理想精度,可极大提升农业气象自动化观测水平;在产量预测方面,ML 技术与遥感时间序列数据、气象数据、土壤数据结合在不同尺度成功预测了不同作物的产量,与之相关,ML 技术在农业气象灾害评估中也有很大的应用潜力;在参数估算方面,农业气象研究关注的以蒸散、叶面积指数、土壤湿度、氮素含量等为代表的水文、土壤、作物参数均可利用 ML 技术与气象数据、遥感数据等的结合实现精确反演或预测。

综合看,传统浅层 ML 技术中以支持向量机和人工神经网络在本文涉及的诸多任务中应用最为广泛且效果最为理想。近年来的方法对比类研究中,随机森林和梯度提升机等集成学习方法普遍取得优于支持向量机和人工神经网络的精度,DL 技术则在某些任务中取得更优于集成学习的精度,且可解决浅层 ML 技术较难解决的一些问题,如在农业气象观测中可精准实现植株表型的自动化观测。从应用时间上看,人工神经网络和支持向量机等浅层 ML 技术自 20 世纪 90 年代开始在农业气象和农业

遥感中应用并逐渐繁荣,其中人工神经网络的应用略早于支持向量机,而以支持向量机为代表的核方法则较人工神经网络更具优势;决策树自21世纪开始受到学界的关注,2010年后随机森林和梯度提升机等决策树集成方法在很多方面被认为是较核方法更好的选择;DL技术自2012年前后重回主流学界的视野,伴随卷积神经网络和长短期记忆网络的成功,自2015年后被广泛认为在诸多应用中的表现超越了支持向量机和集成学习^[14]。尽管将DL技术应用于农业问题的先驱探索始于2010年^[110],但大部分研究成果发表于2015年以后^[15],且有逐年增多的趋势,国内相关研究相对滞后,多数发表于2017年以后,且涉及到的现代农业气象中的问题尚十分有限。

目前有待验证ML技术特别是DL技术在更多农业气象问题上的适用性和先进性,如农业气象灾害遥感监测与损失评估、农业气象灾害风险评估与区划、农业气候区划、气候变化对作物生长的影响评估等任务,同时,伴随ML技术的发展,特别是类似长短期记忆网络的DL时间维算法和新结构的发展,有望更好地结合以风云系列气象卫星数据为代表的遥感时间序列数据,在制图、估产、预测等诸多已开展研究的任务中取得更高的精度和准确性。特别地,将ML技术与智能手机等移动终端结合,可为农业管理者和生产者提供功能强大且智慧化的农业气象信息服务;将DL技术与地基观测、无人机遥感等结合,开发相关业务系统,可以显著提升农业气象自动化观测水平。同时,需要认识到ML技术中没有一种方法可以取代其他所有方法,在现代农业气象工作中需要根据具体任务和数据情况选择最适用的ML技术,如梯度提升机在当今被普遍认为是处理非感知数据的最好算法之一,而当训练数据有限时,浅层ML技术往往比DL技术更适用。

未来ML技术的各种算法和模型必将进一步发展,程序库和普适性的训练数据集也将更加丰富,继承并突破现代DL技术核心思想的新方法也将出现。需要熟悉ML技术在现代农业气象所涉及的各种问题上的适用情况,并及时追踪掌握科技前沿技术,使农业气象科研与业务工作可以最大程度地受益于每一次信息技术的革命性突破,以ML技术特别是DL技术带动农业气象服务模式的创新,更好地迎接现代农业气象发展的新挑战与新机遇。

参考文献

- [1] Gollin D, Parente S, Rogerson R. The role of agriculture in development. *American Economic Review*, 2002, 92(2): 160-164.
- [2] Dodds F, Bartram J. The Water, Food, Energy and Climate Nexus: Challenges and an Agenda for Action. New York: Routledge, 2016.
- [3] Howden S M, Soussana J F, Tubiello F N, et al. Adapting Agriculture to Climate Change//Proceedings of the National Academy of Sciences, 2007, 104(50): 19691-19696.
- [4] United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. General Assembly 70 Session, 2015.
- [5] Piao S, Ciais P, Huang Y, et al. The impacts of climate change on water resources and agriculture in China. *Nature*, 2010, 467(7311): 43-51.
- [6] Crookston R K. A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. *Crop Science*, 2006, 46(5): 2253-2262.
- [7] Mulla D J. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 2013, 114(4): 358-371.
- [8] Wang N, Zhang N, Wang M. Wireless sensors in agriculture and food industry-Recent development and future perspective. *Computers and Electronics in Agriculture*, 2006, 50(1): 1-14.
- [9] Kamilaris A, Kartakoullis A, Prenafeta-Boldú F X. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 2017, 143: 23-37.
- [10] 韩丰, 龙明盛, 李月安, 等. 循环神经网络在雷达临近预报中的应用. *应用气象学报*, 2019, 30(1): 61-69.
- [11] 陆虹, 翟盘茂, 覃卫坚, 等. 低温雨雪过程的粒子群-神经网络预报模型. *应用气象学报*, 2015, 26(5): 513-524.
- [12] Cramer S, Kampouridis M, Freitas A A, et al. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, 2017, 85: 169-181.
- [13] 王在文, 郑祚芳, 陈敏, 等. 支持向量机非线性回归方法的气象要素预报. *应用气象学报*, 2012, 23(5): 562-570.
- [14] Ketkar N. Deep Learning with Python. Shelter Island: Manning, 2017.
- [15] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 2018, 147: 70-90.
- [16] Géron A. Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media, 2017.
- [17] Bishop C M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
- [18] Alpaydin E. Introduction to Machine Learning. London: MIT

- Press, 2009.
- [19] Murphy K P. Machine Learning: A Probabilistic Perspective. London: MIT Press, 2012.
- [20] Witten I H, Frank E, Hall M A, et al. Data Mining: Practical Machine Learning Tools and Techniques. Cambridge: Morgan Kaufmann, 2016.
- [21] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436-444.
- [22] Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*, 2015, 61: 85-117.
- [23] Goodfellow I, Bengio Y, Courville A. Deep Learning. London: MIT Press, 2016.
- [24] Duro D C, Franklin S E, Dubé M G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens Environ*, 2012, 118: 259-272.
- [25] 李颖, 李耀辉, 王金鑫, 等. SVM 和 ANN 在多光谱遥感影像分类中的比较研究. *海洋测绘*, 2016, 36(5): 19-22.
- [26] 戴建国, 张国顺, 郭鹏, 等. 基于无人机遥感可见光影像的北疆主要农作物分类方法. *农业工程学报*, 2018, 34(18): 122-129.
- [27] Yu L, Liang L, Wang J, et al. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int J Remote Sens*, 2014, 35(13): 4573-4588.
- [28] Khatami R, Mountrakis G, Stehman S V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens Environ*, 2016, 177: 89-100.
- [29] Pantazi X E, Moshou D, Mouazen A M, et al. Data Fusion of Proximal Soil Sensing and Remote Crop Sensing for the Delineation of Management Zones in Arable Crop Precision Farming // HAICTA, 2015: 765-776.
- [30] 任义方, 赵艳霞, 王春乙. 河南省冬小麦干旱保险风险评估与区划. *应用气象学报*, 2011, 22(5): 537-548.
- [31] 张蕾, 霍治国, 黄大鹏, 等. 10—11月海南省瓜菜苗期湿涝风险评估与区划. *应用气象学报*, 2015, 26(4): 432-441.
- [32] Kussul N, Lavreniuk M, Skakun S, et al. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(5): 778-782.
- [33] Rodriguez-Galiano V F, Ghimire B, Rogan J, et al. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 67: 93-104.
- [34] Chen Y, Lin Z, Zhao X, et al. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2094-2107.
- [35] Geng J, Fan J, Wang H, et al. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(11): 2351-2355.
- [36] Chen Y, Zhao X, Jia X. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2381-2392.
- [37] Rußwurm M, Korner M. Temporal Vegetation Modelling Using Long Short-term Memory Networks for Crop Identification from Medium-resolution Multi-spectral Satellite Images // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 11-19.
- [38] Ienco D, Gaetano R, Dupaquier C, et al. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1685-1689.
- [39] Minh D H T, Ienco D, Gaetano R, et al. Deep Recurrent Neural Networks for Mapping Winter Vegetation Quality Coverage via Multi-temporal SAR Sentinel-1 // arXiv Preprint arXiv: 1708. 03694, 2017.
- [40] Yang J, Zhao Y Q, Chan J C W. Learning and transferring deep joint spectral-spatial features for hyperspectral classification. *IEEE Trans Geosci Remote Sens*, 2017, 55(8): 4729-4742.
- [41] Liakos K, Busato P, Moshou D, et al. Machine learning in agriculture: A review. *Sensors*, 2018, 18(8): 2674.
- [42] Cho S I, Lee D S, Jeong J Y. AE-automation and emerging technologies: Weed-plant discrimination by machine vision and artificial neural network. *Biosystems Engineering*, 2002, 83(3): 275-280.
- [43] Karimi Y, Prasher S O, Patel R M, et al. Application of support vector machine technology for weed and nitrogen stress detection in corn. *Computers and Electronics in Agriculture*, 2006, 51(1/2): 99-109.
- [44] Binch A, Fox C W. Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. *Computers and Electronics in Agriculture*, 2017, 140: 123-138.
- [45] Huang F J, LeCun Y. Large-scale Learning with SVM and Convolutional Nets for Generic Object Categorization // Proc Computer Vision and Pattern Recognition Conference (CVPR'06), 2006.
- [46] Sharif R A, Azizpour H, Sullivan J, et al. CNN Features Off-the-shelf: An Astounding Baseline for Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014: 806-813.
- [47] Ishii T, Nakamura R, Nakada H, et al. Surface Object Recognition with CNN and SVM in Landsat 8 Images // 2015 14th IAPR International Conference on Machine Vision Applications (MVA). IEEE, 2015: 341-344.
- [48] 王璨, 武新慧, 李志伟. 基于卷积神经网络提取多尺度分层特征识别玉米杂草. *农业工程学报*, 2018, 34(5): 144-151.
- [49] Dyrmann M, Karstoft H, Midtiby H S. Plant species classification using deep convolutional neural network. *Biosystems Engineering*, 2016, 151: 72-80.

- [50] Dyrmann M, Jørgensen R N, Midtby H S. RoboWeedSupport-Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences*, 2017, 8(2): 842-847.
- [51] 张雪芬, 薛红喜, 孙涵, 等. 自动农业气象观测系统功能与设计. *应用气象学报*, 2012, 23(1): 105-112.
- [52] 余卫东, 杨光仙, 张志红. 我国农业气象自动化观测现状与展望. *气象与环境科学*, 2013, 36(2): 66-71.
- [53] Steen K, Christiansen P, Karstoft H, et al. Using deep learning to challenge safety standard for highly autonomous machines in agriculture. *Journal of Imaging*, 2016, 2(1): 6.
- [54] Pound M P, Atkinson J A, Townsend A J, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience*, 2017, 6(10): gix083.
- [55] Amara J, Bouaziz B, Algergawy A. A Deep Learning-based Approach for Banana Leaf Diseases Classification//BTW(Workshops), 2017: 79-88.
- [56] Christiansen P, Nielsen L, Steen K, et al. DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors*, 2016, 16(11): 1904.
- [57] Yalcin H. Plant Phenology Recognition Using Deep Learning: Deep-Pheno//2017 6th International Conference on Agro-Geoinformatics. IEEE, 2017: 1-5.
- [58] Jin S, Su Y, Gao S, et al. Deep learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *Frontiers in Plant Science*, 2018, 9: 866.
- [59] Ubbens J R, Stavness I. Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science*, 2017, 8: 1190.
- [60] Xiong X, Duan L, Liu L, et al. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods*, 2017, 13(1): 104.
- [61] 段凌凤, 熊雄, 刘谦, 等. 基于深度全卷积神经网络的大田稻穗分割. *农业工程学报*, 2018, 34(12): 202-209.
- [62] 张领先, 陈运强, 李云霞, 等. 基于卷积神经网络的冬小麦麦穗检测计数系统. *农业机械学报*, 2019, 50(3): 151-157.
- [63] Baweja H S, Parhar T, Mirbod O, et al. Stalknet: A Deep Learning Pipeline for High-throughput Measurement of Plant Stalk Count and Stalk Width//Field and Service Robotics, 2018: 271-284.
- [64] 黄双萍. 基于深度卷积神经网络的水稻穗瘟病检测方法. *农业工程学报*, 2017, 33(20): 169-176.
- [65] Mohanty S P, Hughes D P, Salathé M. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 2016, 7: 1419.
- [66] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别. *农业工程学报*, 2017, 33(19): 209-215.
- [67] Rahneemofar M, Sheppard C. Deep count: Fruit counting based on deep simulated learning. *Sensors*, 2017, 17(4): 905.
- [68] 薛月菊, 黄宁, 涂淑琴, 等. 未成熟芒果的改进 YOLOv2 识别方法. *农业工程学报*, 2018, 34(7): 173-179.
- [69] Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 2018, 151: 61-69.
- [70] González S A, Frausto S J, Ojeda B W. Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 2014, 12(2): 313-328.
- [71] Safa B, Khalili A, Teshnehlab M, et al. Artificial Neural Networks Application to Predict Wheat Yield Using Climatic Data//Proceedings of 20th International Conference on IIPS. Iranian Meteorological Organization, 2004: 1-39.
- [72] Irmak A, Jones J W, Batchelor W D, et al. Artificial neural network model as a data analysis tool in precision farming. *Transactions of the ASABE*, 2006, 49(6): 2027-2037.
- [73] Jaikla R, Auephanwiriyakul S, Jintrawet A. Rice Yield Prediction Using a Support Vector Regression Method//2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. IEEE, 2008, 1: 29-32.
- [74] Fortin J G, Anctil F, Parent L É, et al. Site-specific early season potato yield forecast by neural network in Eastern Canada. *Precision Agriculture*, 2011, 12(6): 905-923.
- [75] Ruß G. Data Mining of Agricultural Yield Data: A Comparison of Regression Models//Industrial Conference on Data Mining, 2009: 24-37.
- [76] 郭建平. 农业气象灾害监测预测技术研究进展. *应用气象学报*, 2016, 27(5): 620-630.
- [77] 王春乙, 张继权, 霍治国, 等. 农业气象灾害风险评估研究进展与展望. *气象学报*, 2015, 73(1): 1-19.
- [78] 陈怀亮, 邓伟, 张雪芬, 等. 河南小麦生产农业气象灾害风险分析及区划. *自然灾害学报*, 2006, 15(1): 135-143.
- [79] 侯英雨, 张蕾, 吴门新, 等. 国家级现代农业气象业务技术进展. *应用气象学报*, 2018, 29(6): 641-656.
- [80] 王馥棠. 中国气象科学研究院农业气象研究 50 年进展. *应用气象学报*, 2006, 17(6): 778-785.
- [81] Park S, Im J, Jang E, et al. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agric For Meteorol*, 2016, 216: 157-169.
- [82] Jiang Z, Liu C, Hendricks N P, et al. Predicting County Level Corn Yields Using Deep Long Short Term Memory Models//arXiv Preprint arXiv:1805.12044, 2018.
- [83] Kuwata K, Shibasaki R. Estimating Crop Yields with Deep Learning and Remotely Sensed Data//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2015: 858-861.
- [84] Kim N, Lee Y W. Machine learning approaches to corn yield

- estimation using satellite images and climate data: A case of Iowa State. *Korean Soc Surv, Geodesy, Photogramm Cartogr*, 2016, 34(4): 383-390.
- [85] You J, Li X, Low M, et al. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data // Thirty-First AAAI Conference on Artificial Intelligence, 2017: 4559-4565.
- [86] Wang A X, Tran C, Desai N, et al. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data // Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. ACM, 2018: 50.
- [87] Yang F, White M A, Michaelis A R, et al. Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. *IEEE Trans Geosci Remote Sens*, 2006, 44(11): 3452-3461.
- [88] Jung M, Reichstein M, Ciais P, et al. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 2010, 467(7318): 951-954.
- [89] Patil A P, Deka P C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Computers and Electronics in Agriculture*, 2016, 121: 385-392.
- [90] Mehdizadeh S, Behmanesh J, Khalili K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture*, 2017, 139: 103-114.
- [91] Baghdadi N, Cresson R, El Hajj M, et al. Estimation of soil parameters over bare agriculture areas from C-band polarimetric SAR data using neural networks. *Hydrology and Earth System Sciences*, 2012, 16: 1607-1621.
- [92] Srivastava P K, Han D, Ramirez M R, et al. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resources Management*, 2013, 27(8): 3127-3144.
- [93] Nahvi B, Habibi J, Mohammadi K, et al. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. *Computers and Electronics in Agriculture*, 2016, 124: 150-160.
- [94] Morellos A, Pantazi X E, Moshou D, et al. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, 2016, 152: 104-116.
- [95] Ali I, Greifeneder F, Stamenkovic J, et al. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 2015, 7(12): 16398-16421.
- [96] Prasad R, Pandey A, Singh K P, et al. Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions. *Advances in Space Research*, 2012, 50(3): 363-370.
- [97] Jia M, Tong L, Chen Y, et al. Rice biomass retrieval from multitemporal ground-based scatterometer data and RADARSAT-2 images using neural networks. *Journal of Applied Remote Sensing*, 2013, 7(1): 073509.
- [98] Wang L, Zhou X, Zhu X, et al. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, 2016, 4(3): 212-219.
- [99] Mao H, Meng J, Ji F, et al. Comparison of machine learning regression algorithms for cotton leaf area index retrieval using Sentinel-2 spectral bands. *Applied Sciences*, 2019, 9: 1459.
- [100] Liu S F, Liou Y A, Wang W J, et al. Retrieval of crop biomass and soil moisture from measured 1.4 and 10.65 GHz brightness temperatures. *IEEE Trans Geosci Remote Sens*, 2002, 40(6): 1260-1268.
- [101] Yang X H, Huang J F, Wu Y P, et al. Estimating biophysical parameters of rice with remote sensing data using support vector machines. *Science China Life Sciences*, 2011, 54(3): 272-281.
- [102] Abdel-Rahman E M, Ahmed F B, Ismail R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *Int J Remote Sens*, 2013, 34(2): 712-728.
- [103] Van Wittenberghe S, Verrelst J, Rivera J P, et al. Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. *Journal of Photochemistry and Photobiology B(Biology)*, 2014, 134: 37-48.
- [104] Maimaitijiang M, Ghulam A, Sidike P, et al. Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 134: 43-58.
- [105] Song X, Zhang G, Liu F, et al. Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. *Journal of Arid Land*, 2016, 8(5): 734-748.
- [106] 王臻, 武新慧, 李恋卿, 等. 卷积神经网络用于近红外光谱预测土壤含水率. *光谱学与光谱分析*, 2018, 38(1): 36-41.
- [107] Ma J, Li Y, Chen Y, et al. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European Journal of Agronomy*, 2019, 103: 117-129.
- [108] 马浚诚, 刘红杰, 郑飞翔, 等. 基于可见光图像和卷积神经网络的冬小麦苗期长势参数估算. *农业工程学报*, 2019, 35(5): 183-189.
- [109] Sehgal G, Gupta B, Paneri K, et al. Crop Planning Using Stochastic Visual Optimization // 2017 IEEE Visualization in Data Science (VDS). IEEE, 2017: 47-51.
- [110] Demmers T G M, Cao Y, Gauss S, et al. Neural predictive control of broiler chicken growth. *IFAC Proceedings Volumes*, 2010, 43(6): 311-316.

Review of Machine Learning Approaches for Modern Agrometeorology

Li Ying¹⁾²⁾ Chen Huailiang¹⁾³⁾

¹⁾ (CMA • Henan Key Laboratory of Agrometeorological Support and Applied Technique, Zhengzhou 450003)

²⁾ (Henan Institute of Meteorological Sciences, Zhengzhou 450003)

³⁾ (Henan Meteorological Service, Zhengzhou 450003)

Abstract

With the development of smart meteorology and precision agriculture, modern agrometeorology tasks demand for efficient analyzing and processing of massive agricultural and meteorological data, including multi-source remote sensing images. Machine learning technology can powerfully contribute to the development of agrometeorology and the innovation of agrometeorological service mode. A targeted overview on the related work of machine learning in modern agrometeorology domains is given, including mapping and zoning, detection and observation, yield prediction, and parameter prediction, with specially focuses on deep learning approaches for agrometeorology and the latest research progress in recent years. From the aspect of mapping and zoning, machine learning technology can be combined with remote sensing images to map land cover and crop types in different scales, and can also be combined with remote sensing data, soil data and statistical data to make thematic maps of crop growth and vegetation quality and to zone crop management areas. From the aspect of detection and observation, machine learning technology is successfully used to detect weeds in field images. Deep learning technology is used in plant phenotype observation, disease and pest detection, obstacles and anomaly detection, fruit counting and so on with high accuracy, which could greatly improve the level of agrometeorology automatic observation. From the aspect of yield prediction, machine leaning technology combined with remote sensing time series data, meteorological data and soil data is successfully used to predict the yield of different crops in different scales. Machine learning technology also has great application potential in loss assessment for agrometeorological disasters. From the aspect of parameter prediction, the hydrological, soil and crop parameters concerned by agrometeorology tasks such as evapotranspiration, leaf area index, soil moisture and nitrogen can be accurately inverted and predicted by the combination of machine learning technology, meteorological data and remote sensing data. Overall, among the traditional machine learning approaches, support vector machine and artificial neural network are the most widely used and the most ideal methods. In recent years, ensemble-based methods such as random forest and gradient boosting machine have generally achieved higher accuracy than kernel methods, while deep learning approaches have achieved higher accuracy than ensemble-based methods in some tasks. In the future, it is necessary to verify the applicability and advancement of more different machine learning approaches, especially deep learning approaches in more different agrometeorological tasks, and choose the most suitable machine learning technology for each specific task in modern agrometeorological services according to the data using, which will help to meet new challenges and opportunities of the modern agrometeorology development.

Key words: machine learning; deep learning; agrometeorology; agricultural remote sensing