**模式识别和机器学习在农业领域应用的国内外研究现状**

**作者**：吕明哲2018Z8020661043；刘成真2018Z8020661066；

杨爽2018Z8020661031；王丽敏2018Z8020661069；芮志清2018Z8020661080

**摘要：模式识别和机器学习在近年来已经成为不少学科研究问题的一种方法，在农业领域中也得到了广泛的应用。本文首先介绍了机器学习目前的发展状况，主要算法和效果。然后在农业生产的多个方面分析其应用情况，其中有土壤墒情预测，农作物空间分布识别，虫害识别以及生长状况预测。同时也分析了农业工控设备的在集成了机器学习后，从机械生产转变为智能生产的过程。**

**关键词：机器学习；农业；设备**

## 引言

模式识别是研究用计算机自动识别事物的科学，其目的是用机器完成类似于人类智能通过视觉、听觉等感官去识别外界环境所进行的工作，机器学习则是模式识别的技术基础之一，通过机器学习算法使得机器像人类处理问题。模式识别与机器学习广泛在其他学科应用中，推动其它学科的发展起着越来越大的作用。

机器学习的研究在提出后经过多年的发展，得到了多种适用于不同环境的算法，将机器学习技术应用中其他领域是近年的研究热点。本文对机器学习算法进行了简述，总结了机器学习在土地土壤墒情预测方面、作物管理、农业工控设备安全中的应用，分析了不同的机器学习算法在农业领域的应用情况。

## 正文

一．机器学习概述

* 1. 机器学习研究背景及研究现状

机器学习于20世纪50年代中叶已被提出, 并其后的30多年时间里, 机器学习的相关研究不断发展。机器学习具有多学科交叉的属性, 其已普遍应用于人工智能领域。机器学习的前沿主题较多, 其涉及多门学科交错的特征较为明显。人工智能流行使机器学习中的随机森林、卷积神经网络等算法发展迅速。“情感分类”“大数据”等热词开始进入机器学习的研究领域, 促使本课题的研究逐步深入。

* 1. 定义

机器学习是一类算法的总称，这些算法试图从大量历史数据中挖掘出其中隐含的规律，并用于预测或者分类，更具体的说，机器学习可以看作是寻找一个函数，输入是样本数据，输出是期望的结果。一个机器学习的程序就是可以从经验数据E中对任务T进行学习的算法，它在任务T上的性能度量P会随着对于经验数据E的学习而变得更好[1]。

* 1. 算法分类

基于训练集是否拥有标记（label），可以分为四类：监督学习、无监督学习、半监督学习、强化学习，当训练样本带有标签时是有监督学习；训练样本部分有标签，部分无标签时是半监督学习；训练样本全部无标签时是无监督学习。强化学习是一个学习最优策略，可以让本体在特定环境中，根据当前状态，做出行动，从而获得最大回报[2]。

* 1. 常见算法

常见的机器学习算法有线性回归算法、支持向量机算法、K-邻近算法、逻辑回归算法、决策树算法、K-平均算法、随机森林算法、朴素贝叶斯算法、降维算法和梯度增强算法等[3]。

常见算法中，决策树是一种类似于树形结构的预测模型，每一个分支是一个分类问题，叶节点表示类标签，分支表示表征这些类标签的连接特征。贝叶斯网络是一种基于概率推理的图形化网络，实质是有向无环图，节点代表随机向量，应用场景包括自然语言理解、故障诊断、计算机视觉、机器人等。支持向量机旨在高维空间中寻找一个最优超平面作为二分类问题的分割，这个超平面要保证最小的分类错误率[4]。

二．土壤墒情预测

土壤墒情是土壤的基本参数，对作物的生长有着非常重要的作用，是确定灌溉制度，实现适时适量灌溉所必须掌握的参数指标。针对土壤墒情的预测早在20世纪的时候便已经开始[5]，大致可以分为经验公式法[6-8]、水量平衡法[9-10]、消退指数法[11]、土壤动力学法[12]、时间序列法[13]、神经网络法[14-15]以及遥感监测法[16]等。其原理不外乎利用各种算法，探究那些简单易得的数据与土壤墒情之间的关系，以此来间接的预测土壤墒情。其中的经验公式法起步较早，模型简单，操作方便，如栗容前[17]以山西汾河灌区为例，建立了经验公式模型，通过时段初墒值W0、气温、空气饱和差3个变量，实现对墒情的预测。类似地，水量平衡法、消退指数法、土壤动力学法和时间序列法等根据相应的水量平衡法则和能量平衡法则等确定明确的模型表达式实现墒情预测。神经网络法包含的范围较大，随着人们对脑科学研究的深入，对脑神经网运行和神经细胞内部处理机制认识的加深，使得以模拟人脑思维活动为基础发展起来的神经网络蓬勃发展，更加智能。神经网络法的一大特点是不需要创建精确的数学模型，代替人工去探索挖掘隐藏在数据背后的联系，并掌握这种数量关系[18-19]。比较常用的神经网络有BP神经网络，Caojun Huang[20]建立BP模型实现对黑龙江省红星农场土壤含水量的预测，黄令淼[21]建立了基于缺省因子的BP土壤墒情预报简化模型，一些其他的神经网络如动态神经网络、RBF神经网络等也逐渐被人们用来预测墒情，随着不同模型在土壤墒情中的应用，其适用性与精确度值得进一步展开评价。随着遥感技术的快速发展，遥感监测法也逐步应用在大尺度的墒情预测中。

国外的研究起步于20世纪70年代，Lewin J等[22]开展了小麦的土壤墒情预报研究工作，通过研究发现，小麦生长季节根系活动区域主要集中在埋深为 0~90 cm，且该层土壤水分变化率与土壤贮水量基本呈现出线性相关的关系，并据此构建了土壤含水量预报的经验模型。随着科技的发展，预测的手段也在不断更新，Magagi R D等[23]研究了应用地球资源卫星( ERS)数据估算土壤含水量的可能性。印度的Pandeya A等[24]应用微波数据结合人工神经算法，对土壤含水量进行了预测，通过X波段微波散射仪发射与接收不同土壤条件下的微波数据，对人工神经网络进行训练，采用训练后的神经网络对土壤墒情进行预测。

三. 作物管理

3.1作物管理主要内容

作物管理属于精准农业的一个重要方面，将图像处理和机器学习技术应用在作物管理是一种可行的办法[25]。农作物管理包括农作物育种培养，播种生长，病虫害防治，产量预测，收获入仓的全过程，机器学习方法在很多环节均有所涉及，主要通过对光学图像或者高光谱图像的智能分析，有效提高作物各环节管理效率。这些新的数据分析手段给农业生产带来了革命性的变化，但是由于其数据量巨大，如何去除干扰数据，提高分析效率是精准农业作物管理的重要问题。  
3.2作物管理相关机器学习算法

3.2.1作物类型识别与空间分布

农作物类型识别及其空间分布信息是预测区域作物产量、优化种植结构的基本依据，更是制定农业政策、确保粮食安全的科学支撑。黄双燕在卫星遥感数据集上使用随机森林分类器对多种干旱区作物进行分类，总体精度达到89%以上[26]。宋茜对不同分类器的性能进行了对比，SVM 对小容量的特征集信息萃取能力优于 RF，而对于高维特征空间来说，RF有更强的泛化能力和抗噪能力；二分类问题中 SVM 分类器的推广能力优于 RF 分类器，而在多分类问题上，随机森林的泛化能力显著优于支持向量机[27]。

3.3.2 农作物病虫冻害识别

农作物病虫冻害是严重影响作物产量的不利因素，通过卫星遥感数据实时分析监测片区农作物生长状态，具有重要的现实意义。魏传文采用卫星影像对油菜冻害指标进行了反演预测，相比于线性函数，二次多项式函数等曲线拟合算法和KNN算法，基于EVI，MNLI，PVI三个最优变量的RFR模型预测效果最好[28]。谭文学设计了病害图像预处理方法：包括整形算法，方位多样性仿真，亮度多样性仿真等，结合支持向量机构造病变模式的识别算法，提出随机反馈深度受限玻尔兹曼机半监督学习的病害图像特征提取，提出了弹性动量深度卷积病变图像特征提取识别网络一体化方法，解决了过程和目标失配问题[29]。

3.2.3生长发育情况预测

农作物生长发育情况直接关系到产量，如果发生大面积歉收将引起粮食市场的巨大波动，因此准确预测产量和发育状况有助于决策部门制订合理的粮食生产政策。乌玲瑛采用支持向量机方法，建立以适应区域尺度生产指导为目的的水稻发育预测模型[30]。通过整合水稻发育期数据和气象数据，构建训练集与测试集，并应用支持向量机建立针对5个不同发育阶段，模型经过实际测试，最终预测精度大于80%，取得了很好的效果。

四 农业工控设备安全

4.1 工控设备安全

工业控制系统（Industrial Control System，ICS）[31]是一种表示用于工业控制的几种类型的控制仪器和相关仪器的通用总成。工业控制系统（ICS）通过整合信息技术（IT）来提高效率和可控性，但是也因此引入了对ICS的网络威胁[32]。ICS开始面临像IT同样的漏洞环境，经受IT一样的安全考验。而且越来越多的攻击开始针对性的选择ICS进行攻击。如Stuxnet和乌克兰电网的网络攻击，导致大规模的电力瘫痪，对社会造成巨大影响[33]。但是由于这些系统的重要性和独特性，很难找到通用的方式去防御攻击。目前重要的是设计新颖的防御机制，包括对底层物理结构的防御，并能在早期阶段检测到攻击[32]。

4.2 农业工控设备安全

随着生产力的提升，智慧农业在农业生产中的应用越来越广泛。大量的传感器和无线通信技术应用于农业生产，广泛用于环境监测，精确农业。通讯协议有基于ZigBee的WSN和无源，半无源和有源RFID[34]。还有基于M2M的机器和过程控制，建筑和设施自动化以及基于RFID的可追溯系统[35]。农业工控设备的安全性至关重要，由于农业设备的脆弱性，设备被劫持后被应用于DDOS攻击的影响十分巨大。

4.3 传统的工控设备防御方法

由于工控设备的的脆弱性和重要性，研究者对工控设备安全性的防御进行了很多研究。包括如下方面：（1）提升工控设备软硬件的稳健性，在生产阶段对软件的源代码进行静态分析[36-38]，对设备进行动态分析[39,40]；（2）在生产阶段对网络的流量，控制反馈进行实时的监控。

4.4 基于机器学习的工控设备防御方法

传统的工控设备防御方法难以应对大规模的工控设备的安全问题。随着机器学习技术的广泛应用，研究者开始使用训练机器学习模型以解决工控设备的安全问题。基于机器学习进行流量分析是其中一种实现方式。Li等[33]通过基于遥测分析的入侵检测，优化系统模型，捕获系统中不同节点之间的通信包，然后使用机器学习算法对数据集进行训练，实现恶意流量检测，进一步区分攻击类型。其优势在于遥测意味着它不需要进入工业控制系统网络，而是通过远程捕获数据包来实现入侵检测。Keliris等[32]针对底层集成电路进行了攻击模拟，重点研究可编程逻辑控制器（PLC）的控制系统实现。攻击分为三类，传感器攻击、执行器攻击和控制器攻击，并将攻击将用于训练基于机器学习的攻击检测模块。为了检测和对抗ICS上的恶意活动，Keliris利用基于模型和经验的动态过程中的典型时间信号模式及其动态特性的知识来确定攻击的状态，并训练了一个主SVM来检测正在进行的攻击的存在，以及一组辅助的、独立的SVM来检测特定类型和类型的攻击。SVM除了提供简单的基于范围的分类器外，还具有显著的鲁棒性。通过对多个信息流之间运行的相关性，支持向量机支持高检测精度。相比之下，简单的基于范围的攻击检测器在设置足够窄的范围以可靠地检测攻击时会遇到大量的误报，或者在扩大范围以减少误报时会遇到大量的误报。

3.结论

机器学习与模式识别方法论是很多领域突破研究进展的一种有效途径。随着本文的介绍，可以看出模式识别与机器学习技术在农业方面取得了丰硕的成果。并且领域内还在不断地进行深化研究，完善整个智能农业生产系统。但目前的研究力度还不够深厚，未取得实质性的突破。应当加强在作物管理方面的理论研究，深入研究问题本身，这样才能更好的将模式识别技术融合进去。**参考文献**：

[1]兰欣,卫荣,蔡宏伟,郭佑民,侯梦薇,邢磊,那天,陆亮.机器学习算法在医疗领域中的应用[J].医疗卫生装备,2019,40(03):93-97.

[2]叶雷. 机器学习算法在医疗数据分析中的应用[D].华中师范大学,2017.

[3] BASTANLAR Y, OZUUSAL M. Introduction to machine learning[J]. Methods Mol Biol, 2014, 1 107:105-128.

[4]敬思,许建飞. 于知识图谱的机器学习领域研究进展分析[J]. 江苏科技信息2018,35(13):18-21.

[5]彭世彰,胡玲,张利昕.国内外灌溉预报研究现状与动态分析[J].灌溉排水学报,2004,24(1):6-10.

[6]尚松浩,雷志栋,杨诗秀.冬小麦田间墒情预报的经验模型[J].农业工程学报,2000,16(5):31-33.

[7]孙倩倩,刘晶淼,梁宏.东北地区土壤湿度的区域性预报模型研究[J].自然资源学报,2014,29(6):1065-1075.

[8]Wei H，Bing C S. Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition[J].Geoderma,2012,193-194(2013),180-188.

[9]李明生,刘震.土层水量平衡模型在土壤墒情预报中的应用[J].东北水利水电,2005,23(1):49-56.

[10]张胜平,苏传宝.土壤墒情变化规律及预报模型研究[J] .东北水利水电,2005,23(1):49-56

[11]马孝义,王君勤,李志军.基于土壤消退指数的田间土壤水分预报方法的研究[J]水土保持研究,2002,9(2):93-96.

[12]康绍忠,张富仓,梁银丽.玉米生长条件下农田土壤水分动态预报方法的研究[J].生态学报,1997,17(3):245-251.

[13]白冬妹,郭满才,郭忠升.时间序列自回归模型在土壤水分预测中的应用研究[J].中国水土保持,2014(2):42-45.

[14]陈坤,雷晓云,李芳松,文静.基于BP网络的膜下滴灌加工番茄墒情预报研究[J].中国农村水利水电,2012(9):1-4.

[15]冀荣华,张舒蕾,郑丽华.基于多值神经元复数神经网络的土壤墒情预测[J].农业工程学报,2017,33:126-131.

[16]刘永红,叶彩华,王克武，高燕虎.RS和GIS技术支持下的北京地区土壤墒情预报技术[J].农业工程学报,2008,24(9):155-160.

[17] 栗荣前,康绍忠,贾云茂，张宝忠，韦怡冰.汾河灌区土壤墒情预报方法研究[J].中国农村水利水电,2005(10):92-95.

[18] 梁烨妮.人工神经网络的发展及应用[J].高新技术产业发展,2014,(12):2-3 .

[19]龙伟,张金,黄杰.人工神经网络发展前景[J].机械,1998,25(1):47-50.

[20]Huang C J, Li L, Ren S H, Zhi S Z. Research of soil moisture content forecast model based on genetic algorithm bp neural network [J].Springer Berlin Heidelberg,2010,345:309-316.

[21]黄令淼,任树梅,杨培岭,税朋勃，曹建武，周嵘.基于缺省因子的BP-ANN土壤墒情预报简化模型[J]. 中国农业大学学报,2013, 18(5):166-172.

[22]Lewin J. A simple soil water simulation model for assessing the irrigation requirements of wheat[J] . Israel Journal of Agricultrical Research,1972, 22(4):201- 213.

[23]Magagi R D, Kerr Y H.Retrieval of soil moisture vegetation characteristics by use of ERS-1Wind Scatterometer over arid and semi arid areas [J].Journal of Hydrology, 1997, 188 - 189:361 - 384

[24]Pandeya A, Jha S K,Srivastava J K , Prasad R. Artificial neural network for the estimation of soil moisture and surface roughness [J]. Russian Agricultural Sciences, 2010, 36:428-432

[24] 李阿伦,杨卫中,卢娟.区域墒情的空间分析方法与应用[J].中国农学通报，2012，28(21):311-316

[25]阎庆. 基于局部线性嵌入的降维算法研究及其在精准农业中的应用[D]. 安徽大学, 2014.

[26] 黄双燕,杨辽,陈曦,姚远.机器学习法的干旱区典型农作物分类[J].光谱学与光谱分析,2018,38(10):3169-3176.

[27]宋茜. 基于GF-1/WFV和面向对象的农作物种植结构提取方法研究[D].2016.

[28]魏传文. 基于多源数据的油菜冻害遥感机理与方法研究[D]. 2018.

[29]谭文学. 基于机器学习的作物病害图像处理及病变识别方法研究[D].

[30]乌玲瑛,徐奂,蔡喨喨,严力蛟.基于机器学习的水稻发育期预测模型构建[J].扬州大学学报(农业与生命科学版),2012,33(03):44-50.

[31]Wikipedia. Industrial control system[EB/OL]. [20190328]. https://en.wikipedia.org/wiki/Industrial\_control\_system.

[32] Keliris A, Salehghaffari H, Cairl B, et al. Machine learning-based defense against process-Aware attacks on Industrial Control Systems[C]. 47th IEEE International Test Conference, ITC 2016, November 15, 2016 - November 17, 2016, 2017: IEEE Philadelphia Section; Institute of Electrical and Electronic Engineers, Inc. (IEEE).

[33] Li H, Qin S. Optimization and implementation of industrial control system network intrusion detection by telemetry analysis[C]. 3rd IEEE International Conference on Computer and Communications, ICCC 2017, December 13, 2017 - December 16, 2017, 2018: 1251-1254.

[34] Ruiz-Garcia L, Lunadei L, Barreiro P, et al. A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends[J]. Sensors, 2009, (9): 4728-4750.

[35] Wang N, Zhang N, Wang M. Wireless sensors in agriculture and food industry—Recent development and future perspective[J]. Computers and Electronics in Agriculture, 2006, 50(1): 1-14.

[36] 张健, 张超, 玄跻峰, et al. 程序分析研究进展[J]. 软件学报, 2019, 30(1): 1-31.

[37] Feng Q, Zhou R, Xu C, et al. Scalable Graph-based Bug Search for Firmware Images[C]. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016: 480-491.

[38] Dewey D, Giffin J T. Static detection of C++ vtable escape vulnerabilities in binary code[C]. NDSS, 2012.

[39] Chen J, Diao W, Zhao Q, et al. IOTFUZZER: Discovering Memory Corruptions in IoT Through App-based Fuzzing[C]. Proceedings of the 2018 Network and Distributed System Security Symposium, 2018.

[40] 邹权臣, 张涛, 吴润浦, et al. 从自动化到智能化:软件漏洞挖掘技术进展[J]. 清华大学学报(自然科学版), 2018:1-16.