评述

机器学习在太阳物理中的应用

刘辉^{1,2}, 季凯帆^{1*}, 金振宇¹

1. 中国科学院云南天文台, 昆明 650216;

2. 昆明理工大学信息工程与自动化学院, 昆明 650500

*联系人, E-mail: jkf@ynao.ac.cn

收稿日期: 2019-01-31; 接受日期: 2019-05-07; 网络出版日期: 2019-08-29

国家自然科学基金(编号: 11873027, 11773072, 11573012, 11833010)资助项目

摘要 太阳物理研究已经进入大数据时代, 而机器学习作为大数据研究的一种良好工具已经获得越来越多的认可. 本文评述了自2007年以来机器学习在太阳物理中的应用. 从结果上看, 最近4年这一领域的研究明显增加. 所利用的数据包括地面和空间的各种仪器、各种类型和波段的太阳观测资料. 研究领域涵盖太阳耀斑、日冕物质抛射、太阳黑子等太阳物理研究的主要方面. 目前虽然获得一些很好的结果, 但尚未有突破性的进展. 使用的机器学习方法涉及分类、回归、聚类、降维以及深度学习等手段, 但经典的算法, 尤其是分类方法依然占据主导地位. 这意味着机器学习在太阳物理的应用还处于起步阶段, 但同样也意味着在这一领域还有很多工作可以深入开展.

关键词 太阳物理, 太阳活动, 机器学习, 深度学习

PACS: 96.60.-j, 96.60.Q-, 96.60.Hv, 96.60.qd

1 引言

天文学是以观测为基础的科学, 作为天文学重要组成部分的太阳物理学, 其观测技术在21世纪有了巨大的发展.

太阳是唯一一颗可以进行高空间、时间、光谱分辨率和高偏振精度观测的恒星. 太阳观测方法的进步, 一方面为研究太阳磁场、流场精细结构等太阳物理基本问题提供了强有力的手段; 另一方面更是极大推动了太阳观测在空间天气监测预报中的应用, 特别是对

各种太阳活动的短时标现象高时空分辨率观测, 以及对太阳活动现象演化过程的长时间监测. 当前, 太阳观测已经发展到从地面到空间, 从高能到射电, 从成像到光谱全方位、多角度、多波段的观测体系, 这为太阳物理学家带来前所未有的海量数据.

太阳物理学以及天文学的很多其他分支, 本质上都是数据驱动的科学. 而如何从海量的数据中挖掘相互关联的物理信息, 研究物理过程的本质, 成为太阳物理学家所面临的前所未有的挑战.

毫无疑问, 太阳物理已经进入大数据时代, 然而,

引用格式: 刘辉, 季凯帆, 金振宇. 机器学习在太阳物理中的应用. 中国科学: 物理学 力学 天文学, 2019, 49: 109601

Liu H, Ji K F, Jin Z Y. The application of machine learning in solar physics (in Chinese). Sci Sin-Phys Mech Astron, 2019, 49: 109601, doi: 10.1360/SSPMA-2019-0031

单纯依赖人类大脑去处理如此巨大的数据,去挖掘海量物理量之间复杂的内在关联事实上已经成了研究方法的瓶颈.现代计算机技术的飞速发展,尤其是日新月异的机器学习技术为我们提供了一种新的思路和方法.

从广义上讲,机器学习是一门人工智能的科学.对于科学研究,我们希望计算机能代替人类自动地从海量的数据中挖掘和总结其背后固有的规律,从而加深我们对研究对象(如太阳)的认识和理解.

机器学习远非一个全新的概念,早在20世纪50年代就开始起步.但正是因为海量数据的出现以及计算能力的显著提升,才最终促成机器学习技术的腾飞.尤其2015年以来,机器学习快速应用到各个领域,其成果已经让越来越多的科学家认识到这种技术革命将为科学研究方法论注入新的活力.但我们也必须看到,虽然机器学习在很多科研领域的应用发展迅猛,但在天文学,尤其是太阳物理学的研究方法上的应用还并不广泛.

本文利用由美国宇航局资助,并由哈佛-史密森天体物理中心管理的SAO/NASA天体物理数据系统(SAO/NASA Astrophysics Data System, ADS)¹⁾,对2007年以来的机器学习在太阳物理研究应用方面的文献进行了检索,使用了太阳物理和机器学习领域多个关键字的组合(如Solar, Solar Activities, Flare, CME, Sunspot, Machine Learning, Deep Learning, Classification, Cluster, Dimension Reduction等)进行多次检索,并最终由人工剔除了其中的重复文献和无关文献,获得了最后选定的68篇文献.本文虽然未检索ADS数据库收录之外的文献,但我们认为这些文献基本上反映了这一领域过去十多年的发展情况.图1显示了随着年份相关文献数量上的变化,从中可以清楚地看到,其总量并不多,但2015年后,机器学习在太阳物理研究方法上的应用明显增加,如图1所示.

本文分析、归纳了这些文献涉及的太阳物理研究问题以及使用的机器学习方法,并对现有技术方法的特点以及未来的发展做了评述和展望.第2节简要介绍机器学习的形式和典型任务;第3节概括了上述文献涉及的太阳物理研究、使用的数据、主要方法以及应用特点;第4节梳理和评述各类机器学习方法在太阳

图1 (网络版彩图)基于机器学习方法的太阳物理研究文献年代分布

Figure 1 (Color online) Chronological distribution of the literature.

物理研究方法上的应用情况;第5节进行总结和展望.

2 机器学习的形式及典型任务

对传统的数据处理而言,一般是给定输入数据,然后用现有的模型(如解析表达式)对数据进行处理,再得到输出结果.在这种情况下,模型是已知的,也就是有了先验知识,只是结果未知.而机器学习努力要解决的问题是在模型未知的情况下,如何通过算法对数据进行归纳和抽象,生成这个模型,然后再对新的输入预测新的输出.

机器学习主要分有监督学习和无监督学习^[1].

有监督学习是指在已知输入及其对应输出的情况下,通过对这些数据的自主学习(也称训练)来发现他们之间的对应关系.目前监督学习主要解决两类问题:分类和回归.分类问题是指对于已知的数据,我们能明确它们分别对应什么类别,但对应的数学模型并不明确.比如太阳上的耀斑预测,有耀斑爆发和无耀斑爆发就可以认为是两个类型^[2].在这两种类型下,分别对应很多的观测物理量,也就是已知数据.这些数据和耀斑爆发之间的关系(也就是分类)就是有监督学习的一个典型案例.一旦建立了这些物理量和耀斑爆发与否之间的对应关系,那就可以预测下一次太阳活动是否可能伴随耀斑爆发.分类最终要解决的问题还是预测,但无论是二分类(有无耀斑),还是多分类(不同级别的耀斑),其结果表现为一种离散的数学形式.而回归问

1) <http://adsabs.harvard.edu/>

题则是预测连续状态值的变化. 譬如太阳黑子数的预测, 所关心的不仅是在未来某个时期有无黑子, 更关心有多少黑子, 这是一个连续值^[3]. 目前有监督学习是机器学习的一个普遍应用的方法, 相应的算法也较为成熟, 如经典的分类算法包括支持向量机(Support Vector Machine, SVM)^[4]、决策树(Decision Tree, DT)^[5]、随机森林算法(Random Forest, RF)^[6]、人工神经网络(Artificial Neural Network, ANN)^[7]、多层感知器(Multi-layer Perceptron, MLP)^[8]、极限学习机(Extreme Learning Machine, ELM)^[9]; 经典的回归算法包括逻辑回归(Logistic Regression, LR)^[10]、支持向量回归(Support Vector Regression, SVR)^[11]、广义局部线性模型树(Generalized Locally Linear Model Tree, GLO-LIMOT)等^[3].

与有监督学习不同, 无监督学习关注的情况是只有输入数据, 而没有输出对应体, 需要依赖于对这些已知数据自身特性的统计来寻找其内在的关联. 现在无监督学习主要应用在两个方面: 降维和聚类. 降维是研究多维数据之间内在的关联性和冗余性, 通过组合并提取其中最核心、最有代表性的特征, 试图用最小的维度来刻画这些数据综合表达的特定情况. 降维经常作为预处理方法广泛地应用在数据处理当中. 主成分分析(Principal Component Analysis, PCA)就是一个经典的线性降维方法. 在太阳活动观测数据中, PCA被用于对多个波段数据进行融合分析, 通过提取出各个波段共有的主要特征分量, 同时剔除次要分量(干扰)的降维处理, 将各个波段的观测数据归一化到相同的基准下, 有效抑制了多源观测数据的不同基准对后续分析的不利影响^[12]. 聚类也是用于分析已知的数据本身是否具有某种内在关联性的算法, 但关注的是它们本身是否“成团”, 也就是它们是否有某种共同的属性, 即是否为同一类数据, 或者分属几类. 分类问题是类别已经确定, 需要抽象出分类的规则. 而聚类问题是完全不知道这些数据有没有共同点, 要通过机器学习来知道到底有什么规律. 二维数据, 甚至三维数据我们尚可以通过可视化来判断数据是否成团, 但如果是更高维的数据, 就只能寄希望于计算机来帮我们从中找到规律性的东西. 其中代表性的方法K均值聚类算法(K-means Clustering Algorithm, K-means)被用于对日冕物质抛射(Coronal Mass Ejection, CME)观测数据的聚类分析, 用于判别观测到的CME现象中是否同时

存在多个CME事件^[13]. 另外还有一些经典的聚类方法, 如均值漂移算法(Meanshift)、密度聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)等^[14].

现在流行的“深度学习”其实并不是一种学习类型, 而是一种机器学习的计算模型, 尤指利用人工神经网络对数据进行分层训练的学习模型. 当人工神经网络具有多个隐层的时候, 称为深度神经网络, 而对应的学习方法就称为深度学习. 现代计算机理论已经证明堆叠起来, 含有多个隐层的深度神经网络可以通过海量数据的训练(学习), 逼近复杂的非线性映射关系^[15]. 这就给我们研究带来了新的机遇. 目前深度学习已经成为机器学习中的一个重要发展方向. 过去几年, 深度学习技术转变了人工智能世界, 在非线性复杂关系的映射、图像分析、语音识别、自动驾驶以及围棋上都以惊人的速度发展^[16]. 深度学习不但是一个技术上进步, 更是一个观念上的革新^[17].

3 应用机器学习方法的太阳物理研究

机器学习技术和方法已经被成功应用到计算机视觉、自然语言处理、模式识别、智能控制等诸多领域. 在太阳物理研究方法中的应用目前虽处于起步阶段, 但也广泛涉及全球各大观测设备所获取的各种观测数据, 并用于各类太阳活动及其相关物理特性的研究之中. 本节从应用目标、使用的数据、主要方法以及应用特色, 对检索到的文献进行分析和归纳, 并列举代表性论文及其主要工作成果. 图2给出了所查阅文献按应用目标分布的情况.

3.1 太阳耀斑预报

太阳耀斑预报是机器学习方法应用于太阳活动分析最为活跃的领域. 文献[2,4-6,8,10,11,18-41]针对各类观测设备下的多波段数据, 采用典型的机器学习方法对太阳耀斑活动进行了分析和研究. 其中具有代表性的工作如表1^[2,6,8,10,11,19,21,41]所示.

通过表1代表性文献的分析可以看到, 应用机器学习方法对耀斑的研究, 主要是以磁图数据为核心, 结合其他波段的多源数据, 如光球像、紫外辐射、X射线等, 应用各种分类方法对耀斑事件是否发生进行判别.

在数据处理上, 分为直接使用观测数据和应用特

图 2 (网络版彩图)按应用目标分类的文献分布

Figure 2 (Color online) Literature distribution by application in solar physics.

征工程方法对数据进行特征提取两种策略. 体现了对观测的物理量与耀斑事件关联性的不同处理思想: 对

于基于特征提取的方法, 是将科学家认为与耀斑爆发有关的物理量进行特征变换, 构成分类预测的特征向量, 体现了数据-知识-判别的思想. 而对于直接使用观测数据的方法, 主要采用深度神经网络, 其思想是用数据驱动算法, 自动挖掘各物理量与事件发生之间的内在关系, 体现了直接由数据到判别, 而不经知识处理的认知模式.

在分类算法上, 以基于最佳分割超平面的SVM、基于集成学习的RF和基于深度学习的卷积神经网络算法为代表. 对于不同的数据特点, 应采用具有针对性的分类算法: 对于原始数据的直接处理, 自然会采用具有自主特征提取的神经网络算法; 对于数据样本较少的情况, 则采用小样本训练条件下性能最优的SVM算法; 而对于包含多个影响因素共同作用的多特征数据, 采用基于集成学习的RF算法则能够把多个特征的作用进行有效的融合.

表 1 机器学习应用于太阳耀斑研究的代表性工作^{a)}

Table 1 Machine learning application to flare research^{a)}

作者	年份	数据	机器学习方法	应用特色
Huang等人 ^[2]	2018	1996–2015年的SOHO/MDI、SDO/HMI磁图和强度像、GOES软X射线, 样本数>10万个	CNN	用MDI数据训练模型, 并用HMI数据测试, 48 h的X级耀斑预报 TSS=0.791
Florios等人 ^[8]	2018	2012–2016年的23134个SDO/HMI活动区磁图	MLP, SVM, RF	算法比较, 基于集成学习的RF算法对多特征数据效果最佳, 24 h预报结果: 对M1级以上耀斑TSS=0.74, ACC=0.93; 对C1级以上耀斑, TSS=0.60, ACC=0.84
Liu 等人 ^[6]	2017	2010–2016年的845个活动区的SDO/HMI磁图	RF	提取了13个太阳活动区特征, TSS分别为0.67 (B级耀斑), 0.33 (C级耀斑), 0.50 (M级耀斑), 0.29 (X级耀斑)
Nishizuka等人 ^[19]	2017	2010–2015年的SDO/HMI磁图、SDO/AIA, GOES软X射线, 其中包括29个X级和433个M级耀斑	SVM, KNN等	提取60个太阳活动区特征, KNN效果最好, TSS>0.9
Raboonik等人 ^[21]	2017	2010–2016年的422个活动区的SDO/HMI磁图	SVM	利用Zernike矩从图像提取特征进行分类. 48 h预报, TSS=0.856
Boucheron 等人 ^[11]	2015	2000–2010年的2125个活动区的SOHO/MDI磁图	SVR	基于38个磁场特征, 对耀斑尺寸和时间演化进行回归预测
Bobra和Couvidat ^[41]	2015	2010–2014年的2071个SDO/HMI活动区磁图	SVM	从25个参数中选取了对预测影响值最高的13个参数48 h预报, TSS=0.82, ACC=0.943
Yuan等人 ^[10]	2010	1996–2005年的230个SOHO/MDI活动区磁图	SVM与LR结合	采用3个特征和多级SVM分类器进行耀斑预报, 正确率分别为0.86, 0.72, 0.65, 0.84

a) SDO: Solar Dynamics Observatory, 太阳动力学天文台; HMI: Helioseismic and Magnetic Imager, 日震与磁成像仪; AIA: Atmospheric Imaging Assembly, 大气成像组件; SOHO: Solar and Heliospheric Observatory, 太阳和日球层探测器; MDI: Michelson Doppler Imager, 麦克逊多普勒成像仪; GOES: Geostationary Operational Environmental Satellite, 环境应用静地卫星; CNN: Convolutional Neural Networks, 卷积神经网络.

3.2 日冕物质抛射

日冕物质抛射是机器学习方法应用于太阳活动分析的又一重要领域. 文献[9,13,42–47]应用机器学习方法, 对日冕物质抛射进行了检测、分析和研究. 比较具有代表性的工作如表2^[9,42–44,47]所示.

应用机器学习的思想进行CME研究, 主要的研究的对象是CME事件的检测、预报和CME相关事件的关联性分析和判别, 并试图发现和揭示导致CME的关联事件及其背后更为深刻的物理因素. 所采用的策略主要是应用分类算法对经特征变换的数据进行分析和判别. SVM在此类问题中依然采用最多, 主要原因仍然是CME事件的样本规模相对来说还比较小, 在此情况下, 研究人员倾向于采用统计判别风险最优的SVM算法.

3.3 太阳黑子

黑子与磁场和太阳活动密切相关, 目前, 机器学习算法多应用于黑子数和黑子群的检测和预报. 表3^[3,12,45,48]给出了机器学习方法应用于黑子研究的典型案例.

对于太阳黑子研究, 黑子(群)检测和黑子数预报是机器学习方法的主要应用领域. 主要特点是以历史数据的处理分析为基础, 对后续的黑子状态及演化进

行预测. 特征提取主要是进行目标检测、定位和跟踪, 主要的算法是分类、降维和回归.

3.4 其他应用

机器学习方法还涵盖太阳物理研究的多个方面, 如磁场、暗条、冕洞、日冕亮点等^[49–68]. 表4^[49–53]给出了机器学习方法在这些方面的比较有代表性的工作.

此外, 还有应用图像处理技术进行诸如CME关联事件分析^[69]、活动区^[70]、EIT(EUV Imaging Telescope)波检测^[71]、日冕暗化^[72]、谱斑检测等方面的工作^[73].

在上述这些更为广泛的应用中, 机器学习的主要目标是研究太阳活动事件背后的物理机制. Mercer核函数^[51]、Zernike矩^[52]、灰度共生矩阵^[69]、复小波变换^[69]、数学形态学^[73]等图像分析和模式识别技术, 被广泛用于物理特征的检测、分析和表示.

随着机器学习、图像处理技术的应用在太阳物理领域还形成了一些公开的数据库和软件工具, 以方便研究者使用. 如可以检索太阳活动情况的HEK(Heliophysics Events Knowledgebase) <https://www.lmsal.com/hek/index.html>数据库、围绕SDO数据的LSDO(A Large-Scale Solar Dynamics Observatory Image Dataset for Computer Vision Applications)^[74]和SHARP(Space-Weather HMI Active Region Patches)数据集^[75], 以及从

表 2 机器学习应用于CME研究的代表性工作^{a)}

Table 2 Machine learning application to CMEs research^{a)}

作者	年份	应用方向	数据	机器学习方法	应用特色
Liu等人 ^[42]	2018	CME到达时间估计	SOHO LASCO C2之前观测到的182个部分或全晕CME事件	SVM	预测误差约为5.9 h
Inceoglu等人 ^[43]	2018	CME关联事件分析	2010–2018年的DONKI数据	SVM, MLP	基于18个活动区物理参数, 分析CME和SEPs的关联性, TSS约为0.91.
Zhang等人 ^[9]	2017	CME检测	SOHO/LASCO C2图像	ELM	基于图像亮度和纹理特征检测CME, 并结合空-时连续性排除误判区域
Bobra和Ilonidis ^[44]	2016	M级以上耀斑是否引发CME	2010–2015年的3023个活动区的DONKI数据	SVM	通过特征选择算法, 提取了6个最佳预报特征 TSS约为0.68
Byrne等人 ^[47]	2012	CME检测和跟踪	SOHO/LASCO和STEREO/SECCHI图像以及数值模拟的CME图像	多尺度小波变换、特征匹配	多尺度滤波提取复杂的CME结构特征

a) LASCO: Large Angle Spectrometric Coronagraph, 大角度光谱日冕仪; STEREO: Solar Terrestrial Relations Observatory, 日地关系观测台; SECCHI: Sun Earth Connection Coronal and Heliospheric Investigation, 太阳-地球连接日冕和日光层研究; SEPs: Solar Energetic Particles, 太阳高能粒子; DONKI: Space Weather Database Of Notification, Knowledge, Information, 空间天气公告、知识、信息库<http://kauai.ccmc.gsfc.nasa.gov/DONKI/>.

表 3 机器学习应用于太阳黑子研究的代表性工作^{a)}Table 3 Machine learning application to sunspots research^{a)}

作者	年份	应用方向	数据	机器学习方法	应用特色
Drews和van der Voort ^[12]	2017	黑子半影微喷流	2010年6月SST Ca II 8542 Å 数据	PCA, KNN	通过降维和分类研究PMJs的统计特性
Padinhatteeri等人 ^[48]	2016	δ黑子自动检测和分类	2010–2012年21个活动区的SDO/HMI, NOAA/SWPC的分类数据	SMART-DF	能对δ黑子进行准实时检测, 及时将活动区域标记为潜在的耀斑区域
Henwood等人 ^[45]	2010	长生命周期(>20 d)复发黑子群	1959–1964年格林尼治天文台的光球数据(GPR)	ANN	训练好的神经网络模型, 可以扩展到长时间范围(1874–1976年)的复发黑子群识别
Mirmomeni等人 ^[13]	2007	黑子数预报	用1700–1920年的黑子数样本作为训练集, 用1921–1994年的黑子数样本构造4个测试集	GLOLIMOT	基于局部线性模糊描述子逼近非线性回归模型, 用于黑子数预报

a) SST: Swedish 1-m Solar Telescope, 瑞典1米太阳望远镜; NOAA/SWPC: National Oceanic and Atmospheric Administration/Space Weather Prediction Center, 美国国家海洋和大气管理局/空间天气预报中心; PMJs: Penumbra microjets, 半影微喷流; GPR: Greenwich Photo-heliographic Results, 格林尼治太阳照相观测数据汇编; SMART-DF: Solar Monitor Active Region Tracker-Delta Finder, 太阳活动区监视系统-磁δ探测器.

表 4 机器学习在其他太阳物理应用中的典型案例^{a)}Table 4 Miscellaneous application of machine learning in Heliophysics^{a)}

作者	年份	应用方向	数据	机器学习方法	应用特色
Díaz Baso和Asensio Ramos ^[49]	2018	太阳光球和磁场图像增强	SDO/HMI光球和磁场以及模拟数据	CNN	通过数值模拟数据的训练, 用于SDO/HMI的分辨率提升, 并与Hinode数据进行比较验证
Asensio Ramos等人 ^[50]	2017	太阳光球水平速度场测量	光球数值模拟数据	CNN	用于水平速度场测量, 尤其能探测到非常小的速度变化
Teng ^[51]	2015	光球磁场反演	用2014年的SDO/HMI的偏振数据, 构造了18690组训练样本	ANN, SVR	基于Mercer核特征提取, 用回归算法实现磁场反演
Alipour和Safari ^[52]	2015	日冕亮点的统计特性	2010–2014年的SDO/AIA图像、SDO/HMI磁图	SVM	提取图像Zernike矩作为特征, 用于日冕亮点的识别、跟踪, 并进行相关统计分析
Reiss等人 ^[53]	2015	冕洞和暗条识别	SDO/AIA图像、SDO/HMI磁图	SVM, DT和RF多种算法比较	SVM给出了最佳性能, TSS约为0.9

a) Hinode: 日出太阳望远镜; CPBs: Coronal Bright Points, 日冕亮点.

SOHO/EIT图像序列中自动、实时检测太阳喷射事件的软件包NEMO(Novel EIT Wave Machine Observing)^[76]等. 这些都为机器学习在太阳物理中的应用提供了相当多的便利条件.

从上述几个方面的应用分析和归纳可以看出: (1) 机器学习已应用于各种太阳活动(耀斑、日冕物质抛射和黑子等)及其相关现象和物理机制的分析; (2) 机器学习方法所处理的观测数据涵盖世界上主要观测设备(如SDO/HMI, SDO/AIA, GOES, SOHO, Hinode等)所获得的各类数据; (3) 数据处理由以磁场为核心, 向

多源数据融合拓展, 数据处理策略由特征提取为主向数据挖掘延伸.

4 应用于太阳物理的机器学习方法

在太阳活动分析研究中, 有监督学习中的分类和回归, 无监督学习中的聚类、降维以及深度学习等大类机器学习算法领域都有所涉及. 论文数量的分布如图3所示.

各类太阳活动的观测数据处理中, 对机器学习算

法的应用呈现以下特点.

(1) 以SVM为代表的分类算法占据绝对主导

所查阅68篇文献中直接采用分类算法的文献达33篇, 占据近50%的比例. 事实上, 分类算法在机器学习算法中, 相对更为成熟、应用更为广泛, 其应用效果在众多方法中也较为突出. 在太阳物理研究的应用中, SVM分类器算法的应用最为普遍(19篇), 该算法被广泛地应用于各种太阳活动区域的检测和预报, 如耀斑^[4,8,10,19,21,28,35,38,41]、CME^[42-47]、冕洞和暗条^[53,54]等. 遗憾的是, 各文献都仅给出实验的性能比较, 未深入分析SVM取得成功的原因和理论依据. 我们仔细阅读了文献中SVM的实验方法, 发现这些实验中SVM性能较优的原因主要是基于SVM的最佳分割面特性, 以及它在小样本训练集条件下的分类误差风险最小的统计学习特性. 换句话说, 在SVM取得较好性能的应用中, 所处理的数据主要还是针对较短时段和局部观测的小样本数据. 虽然能得到较高的TSS值, 但TSS值与测试数据集的选择密切相关, 对不同的测试对象, 其性能随之变化, 故其有效应用于新数据的泛化能力还有待考察.

分类算法中另一个采用较多的算法是随机森林算法(8篇). 这类算法被应用于耀斑^[6,8,31]、黑子^[3]、冕洞及暗条^[53,54,64]等的研究. 它们的成功依然离不开其理论基础, 也就是说, 当某种太阳活动的影响因素较为明确, 但因素间的相互关系和影响程度尚未清晰的情况下, 集成学习的方法(如RF)往往能够获得不错的

效果.

(2) 各种算法的应用尚处于初期试验阶段

在基于机器学习的太阳活动分析文献中, 另一个突出的特点是: 很多文献对同一组数据常会应用多种算法进行处理和分析, 再来比较性能优势. 例如, Sadykov等人^[31]比较了3种机器学习分类算法: SVM, KNN和RF对耀斑预测的准确度, 结果表明, 他们采用的基于SVM的预报方法对于M和X级耀斑预报结果优于SWPC/NOAA的预报结果. Florios等人^[8]使用MLP, SVM和RF算法, 比较它们对M1和C1级耀斑爆发预测的性能, 结论是RF对所选择的测试样本性能最佳, 其次是基于熵目标函数的MLP. Nishizuka等人^[19]开发了一种利用机器学习的耀斑预测模型, 也比较了3种机器学习算法: SVM, KNN和极端随机树. 结果表明, KNN算法的性能最好. Reiss等人^[54]将机器学习算法与图像分割技术结合使用, 在图像分割的基础上, 分别应用SVM, DT和RF分类算法对冕洞和暗条进行分类等.

此类对方法和性能比较的文献超过10篇, 表明目前在太阳活动观测数据的处理中, 太阳物理学家处在对各种机器学习方法的尝试和验证阶段, 尚未形成较为成熟的解决方案. 不同的数据得到的结果并不一致.

(3) 预处理和特征提取方法需要新的突破

数据预处理和特征提取对于模式识别问题起着至关重要的作用. 在基于机器学习的太阳活动数据分析文献中, 各类问题的分析研究都涉及特征提取与选择, 如Alipour和Safari^[52]用图像的矩特征描述强度、质心、方向等属性, 以及其他形态结构特性, 他们先将时间切片上反映出来的暗化结构投影到Zernike矩空间进行预处理, 然后再由SVM对矩特征集进行分类, 用于CBPs自动检测. Raboonik等人^[21]也采用Zernike矩预处理并结合SVM算法对数据进行分析, 利用SDO/HMI磁图计算了M级和X级耀斑在爆发前18 h的发生概率. Teng^[51]利用Mercer核函数, 将SDO/HMI的偏振数据从原始参数空间投影到高维特征空间, 将非线性回归问题有效地转化为线性回归问题后, 再用于光球磁场反演. Reiss等人^[53]研究了使用Haralick等人在1979年提出的基于共生矩阵的纹理特征表示方法, 来分析SDO图像中冕洞和暗条所包含的纹理信息, 通过结合一阶统计量和形状度量, 测试了几个分类器, 找到了最适合的决策规则对冕洞和暗条进行区分.

上述工作将数据预处理和特征提取与机器学习算

图3 (网络版彩图)按机器学习算法分类的文献比例

Figure 3 (Color online) Literature distribution by machine learning algorithms.

法有效结合,并取得了较好的效果.但大多数工作是基于目前已掌握的物理关系,通过算法从原始数据中计算相关物理量,交由机器学习算法进行判别.对于如何将这此物理量进一步进行特征选择和降维,构成更加适合于最终判别和预测的特征向量的方法和应用,所查阅文献鲜有涉及.由此可见,对于物理机制尚未明确的太阳活动,观测数据中的哪些信息是对物理现象的最佳解释,以及如何对这些信息进行融合应用的方法亟待研究,这也正是太阳物理研究关注的核心问题.由此,我们也可以设想,太阳物理研究需要具有更好抽象能力以及干扰抑制能力的机器学习算法,这也对天文信息处理提出了更高的要求和挑战,为机器学习研究提供了更加广阔的前景和空间.

(4) 深度学习太阳物理研究中受到关注和应用

自2015年以来,随着人工智能技术的发展和应,在太阳活动分析和研究中,深度学习算法也随之受到关注和应用.天文学家们开始探索用新的技术手段,分析和处理太阳活动观测数据,尤其是从海量的观测数据中挖掘和获取尚未明确和得到公认的物理机制.

Hernandez^[38]将深度卷积神经网络(Convolutional Neural Networks, CNN)应用于太阳耀斑预测,并研究了这一技术在自动识别、预测分类和回归等问题上的优点,及其在挖掘磁场能量存储和释放的机制上的可能性.

Jonas等人^[39]在处理2010年5月到2014年5月SDO观测到的3000多个活动区、8.5 TB数据时,构建了一个预先定义了卷积核结构的CNN,并用线性分类器训练各个卷积核参数.用于分析和发现太阳耀斑的物理过程和触发机制.

Huang等人^[2]采用深度CNN构建耀斑预报模型,该深度CNN包括两组卷积、线性校正和池化层(每个卷积层都包含64个11×11像素大小的卷积滤波器),网络从磁图中提取耀斑模式特征,然后再将所提取的特征作为输入送至一个3层的全连接ANN进行分类,从而将活动区磁图分为耀斑爆发和非耀斑爆发两类.

Asensio Ramos等人^[50]则设计了一个名为DeepVel的端到端的深度神经网络,该网络采用全卷积体系结构,由20个级联的“残差块”构成,由于输入和输出间的跳连,残差块结构能够很大程度提升网络训练速度^[77],该网络能够接受任意尺寸的图像,可以从两帧连续图像中得到大气中每一个像素、每一个时间步长

以及3个不同高度的速度估计.

Díaz Baso和Asensio Ramos^[49]基于上述DeepVel网络基本架构,训练了一个用于SDO/HMI图像分辨率增强的深度CNN.为实现分辨率提升,新的网络在原网络的输出层之前加入了一个上采样层,较之另一种在输入层之后加入上采样层的方案,具有更快的训练速度.

深度学习的有效应用,在于深度学习的关键思想是从给定的数据中自动、分层地提取高层次的特征.通过逐层特征变换,将样本在原空间的特征表示变换到一个新特征空间,从而使分类或预测更加容易.与人工规则构造特征的方法相比,从数据直接学习特征表示,更能够刻画数据中丰富的内在联系,有利于从原始数据直接获取有效信息,是对经典机器学习方法的有益补充.另一方面,也避免了物理意义不明确情况下,人工特征工程所带来的局限和困难.对太阳物理研究从观测获得认知的特点而言,是一种极有前景的方法和途径.

(5) 所采用的方法以经典原型算法为主

太阳物理研究目前所采用的机器学习方法主要涉及有监督学习(以分类算法和回归算法为代表)和无监督学习(以聚类算法和降维算法为代表),强化学习算法尚未有文献采用.所采用的各种方法也仅是典型方法当中的一少部分.很多更加灵活和较为复杂的算法尚未被应用.

分类算法中,以经典的SVM, KNN, DT, RF算法为主,贝叶斯方法、基于关联规则的分类的方法还未见报道.

回归算法中,逻辑回归和最小绝对值收缩和选择算法(Least Absolute Shrinkage And Selection Operator, LASSO)得到应用,岭回归、弹性网络、局部散点平滑估计等经典算法尚未应用.

聚类算法中,所查阅的文献仅有K-means算法的应用,层次聚类、期望最大化、DBSCAN等经典算法尚未应用.

降维算法中,仅有PCA方法得到应用,线性判别分析、局部线性嵌入等经典方法尚未应用.

异常检测算法在所查阅的文献中没有应用案例.

深度学习模型中,仅有CNN应用的文献,其他如循环神经网络、变分自编码器、生成对抗网络等较新的模型尚未有应用报道.

可以看出, 机器学习方法在太阳物理中的应用尚处于初期试验阶段, 只是用经典方法做了一些实验, 还未形成比较成熟的针对某一应用特别适合的方法. 虽然有一些文献对同一应用的不同算法进行了比较^[8,19,31,53], 但仍处于比较浅层次的结果和指标(如TSS)之间的对比, 深层次的机理还未深入分析并获得结论. 随着机器学习的发展和在其它领域的成功应用, 也提出了很多更为复杂或者更新颖的算法, 这些算法有可能在以上提到的太阳物理应用领域获得更好的效果; 也有望拓展到其他尚未使用机器学习方法的领域.

5 总结与展望

从对2007年以来的相关工作的总结、分析可以看出: 机器学习方法在太阳物理研究方法中的应用逐渐开展, 自2015年开始有明显增长, 但总体仍不够活跃. 所处理的数据覆盖了世界主要观测机构和观测设备采集的各类型数据. 应用中最多的目标分别是: 耀斑、日冕物质抛射、黑子、磁场, 并由单一现象的研究向关联性分析发展. 采用的数据以磁场为主, 并向多种数据融合方向发展. 符合太阳活动以磁场为核心, 以耀斑和日冕物质抛射为主要对象的研究特点^[78].

机器学习方法主要集中在基于分类和聚类的预报方面, 在回归(尤其是高维度非线性回归)、维度约简、异常数据检测和处理等重要的应用方面较少. 使用的方法多为机器学习中的代表性原型方法, 针对物理背景变形优化的方法较少; 传统模式识别中的数据预处理与特征提取对于太阳物理研究方法而言需要新的突破; 基于神经网络的人工智能、深度学习算法研究得以应用, 但基本采用深度CNN, 新的模型方法尚无应用; 强化学习尚无文献涉及. 可以说, 机器学习在太阳物理研究上的应用还处于起步阶段, 开始有一些太阳物理学家关注这一领域, 并尝试将这些方法和自己的研究联系起来. 研究对象也基本覆盖了的主要的太阳物理数据处理问题.

太阳物理学的研究方法包含了观测、数据处理与分析、数值模拟和耀斑、CME等特殊事件的预报等很多方面. 目前机器学习的应用主要集中在太阳活动的预报方面, 本文作者在光球磁场定标和数据增强方面也有所尝试, 但基本都是围绕现有的机器算法尝试

解决太阳物理研究中遇到的具体困难. 这些具体问题虽然是机器学习的重要应用领域, 但是这些领域还不能发挥机器的最大潜能. 机器学习更大的优势是在大数据中通过学习, 寻找传统研究手段尚未发现的统计规律或关联性. 现在太阳物理研究的特点恰恰是数据非常巨大, 观测数据从高能、极紫外、光学、红外到射电, 数值模拟也产生海量的数据. 利用传统的研究方法, 很难充分利用如此多的不同波段、不同设备、不同方法所获得的数据. 随着深度学习、人工智能技术的发展和进步, 面向大数据分析和挖掘的各类机器学习技术在太阳物理学研究和应用中将会有更为广阔的空间和前景. 人工智能和大数据时代正在改变我们研究天文的方式, 也许下一个十年中, 机器学习就将逐渐取代一些传统的分析技术.

在机器学习方法逐渐应用于太阳物理的趋势下, 我们也必须冷静思考和解决目前所面临的问题: 一方面, 虽然研究对象覆盖的范围较大, 但尚未有突破性的成果出现, 甚至公认的被能广泛接受的成果也不多见. 在收集到的文献中, 大约有1/3以上来自会议报告. 这说明两个问题: (1) 我们对机器学习到底能解决什么样的太阳物理研究问题还不确定, 尚处于摸索阶段; (2) 目前研究方法和结果要让天文学家广泛接受还存在认知上的障碍. 究其原因, 这是一种和传统方法完全不同的数据处理方式. 但我们相信, 并且切身感受到, 随着更多的研究开始使用机器学习技术, 太阳物理学界对这一问题的认识一定会逐渐加深; 另一方面, 从技术角度看, 目前采用的机器学习算法还比较经典, 并没有能将最新的算法引入到太阳物理研究方法当中, 这说明太阳物理学家和计算机专家的合作还不够深入. 如何打破这两个学科之间的知识壁垒, 将双方的优势密切结合到一起是交叉学科发展的一个关键因素.

机器学习在工程方面的应用已经非常成功, 在日常生活中几乎随处可见. 但在科学研究中, 尤其像天文学这种基础科学研究, 机器学习到底能承担什么样的角色, 这是一个需要深入思考的问题. 长期以来我们更依赖于解析表达式来建立物理模型. 解析表达式中的每个参数, 都有物理意义支撑. 而机器学习是通过建立大规模数值模型来表征客观世界的映射关系, 数值模型(比如深度CNN)中的参数可以成千上万, 目前尚未明确这些参数的物理含义的. 虽然哈佛大学的

Lin和麻省理工学院的Tegmark给出了机器学习的物理层面解释, 认为机器学习不仅仅是一个数学或者计算机的问题, 更是具有可解释的物理思想^[17]. 但具体到一个个真实的案例, 如何去理解机器学习得到的模型

和结果就成为一种新的挑战. 如果有一天能从机器学习的结果中归纳产生新的物理理论, 那将对科学研究方法带来翻天覆地的变化, 我们也非常期待那一天的到来.

参考文献

- Murphy K P. Machine Learning: A Probabilistic Perspective. Cambridge: The MIT Press, 2012
- Huang X, Wang H, Xu L, et al. Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms. *Astrophys J*, 2018, 856: 7
- Mirmomeni M, Lucas C, Araabi B N, et al. Forecasting sunspot numbers with the aid of fuzzy descriptor models. *Space Weather*, 2007, 5: S08004
- Yang X, Lin G, Zhang H, et al. Magnetic nonpotentiality in photospheric active regions as a predictor of solar flares. *Astrophys J Lett*, 2013, 774: 1201–1205
- Yu D, Huang X, Wang H, et al. Short-term solar flare prediction using a sequential supervised learning method. *Sol Phys*, 2009, 255: 91–105
- Liu C, Deng N, Wang J T L, et al. Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm. *Astrophys J*, 2017, 843: 104, arXiv: [1706.02422](#)
- Qahwaji R, Colak T, Al-Omari M, et al. Automated prediction of CMEs using machine learning of CME-flare associations. *Sol Phys*, 2008, 248: 471–483
- Florios K, Kontogiannis I, Park S H, et al. Forecasting solar flares using magnetogram-based predictors and machine learning. *Sol Phys*, 2018, 293: 28, arXiv: [1801.05744](#)
- Zhang L, Yin J, Lin J, et al. Detection of coronal mass ejections using multiple features and space-time continuity. *Sol Phys*, 2017, 292: 91
- Yuan Y, Shih F Y, Jing J, et al. Automated flare forecasting using a statistical learning technique. *Res Astron Astrophys*, 2010, 10: 785–796
- Boucheron L E, Al-Ghraibah A, McAteer R T J. Prediction of solar flare size and time-to-flare using support vector machine regression. *Astrophys J*, 2015, 812: 51, arXiv: [1511.01941](#)
- Drews A, van der Voort L R. Microjets in the penumbra of a sunspot. *Astron Astrophys*, 2017, 602: A80, arXiv: [1702.06078](#)
- Byrne J P. Investigating the kinematics of coronal mass ejections with the automated CORIMP catalog. *J Space Weather Space Clim*, 2015, 5: A19, arXiv: [1506.04046](#)
- Rodriguez A, Laio A. Clustering by fast search and find of density peaks. *Science*, 2014, 344: 1492–1496
- Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313: 504–507
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521: 436–444
- Lin H W, Tegmark M, Rolnick D. Why does deep and cheap learning work so well? *J Stat Phys*, 2017, 168: 1223–1247, arXiv: [1608.08225](#)
- Nishizuka N, Sugiura K, Kubo Y, et al. Solar flare prediction using machine learning with multiwavelength observations. *Proc IAU*, 2017, 13: 310–313
- Nishizuka N, Sugiura K, Kubo Y, et al. Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms. *Astrophys J*, 2017, 835: 156, arXiv: [1611.01791](#)
- Benvenuto F, Piana M, Campi C, et al. A hybrid supervised/unsupervised machine learning approach to solar flare prediction. *Astrophys J*, 2018, 853: 90, arXiv: [1706.07103](#)
- Raboonik A, Safari H, Alipour N, et al. Prediction of solar flares using unique signatures of magnetic field images. *Astrophys J*, 2016, 834: 11, arXiv: [1610.03222](#)
- Raboonik A, Safari H, Dadashi N, et al. Prediction of M and X solar flares by using machine learning algorithm. In: 41st COSPAR Scientific Assembly. The Istanbul Congress Center (ICC), Turkey, 2016
- Tremblay B, Strugarek A, Charbonneau P. Sandpile model and machine learning for the prediction of solar flares. In: Solar Heliospheric and Interplanetary Environment (SHINE 2018). Cocoa Beach, FL, 2018
- Winter L M. X-ray and gamma-ray observations of solar flares and SEP events. In: American Astronomical Society, SPD Meeting #47, 2016

- 25 Muranushi T, Shibayama T, Muranushi Y H, et al. UFCORIN: A fully automated predictor of solar flares in GOES X-ray flux. *Space Weather*, 2015, 13: 778–796, arXiv: [1507.08011](#)
- 26 Ilonidis S, Bobra M, Couvidat S. Solar flare predictions using time series of SDO/HMI observations and machine learning methods. In: IAU General Assembly, Meeting #29, 2015
- 27 Winter L M, Balasubramaniam K. Using the maximum X-ray flux ratio and X-ray background to predict solar flare class. *Space Weather*, 2015, 13: 286–297, arXiv: [1504.00294](#)
- 28 Ilonidis S, Bobra M. Solar flare forecasting using time series of SDO/HMI vector magnetic field data and machine learning methods. In: Joint American Astronomical Society/American Geophysical Union Triennial Earth-Sun Summit, Meeting #1, 2015
- 29 Winter L M, Balasubramaniam K S. Towards predicting solar flares. In: Joint American Astronomical Society/American Geophysical Union Triennial Earth-Sun Summit, Meeting #1, 2015
- 30 Bobra M, Couvidat S P. Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. In: American Geophysical Union, Fall Meeting, 2014
- 31 Sadykov V, Kosovichev A, Kitiashvili I N. Using machine-learning methods and expert prediction probabilities to forecast solar flares. In: Solar Heliospheric and Interplanetary Environment Conference (SHINE 2018). Cocoa Beach, FL, 2018
- 32 Ahmed O W, Qahwaji R, Colak T, et al. Solar flare prediction using advanced feature extraction, machine learning, and feature selection. *Sol Phys*, 2013, 283: 157–175
- 33 Qahwaji R, Ahmed O, Colak T. Automated feature detection and solar flare prediction using SDO data. In: 38th COSPAR Scientific Assembly. Bremen, 2010. 4
- 34 Qahwaji R, Colak T. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Sol Phys*, 2007, 241: 195–211
- 35 Bobra M. Predicting solar activity using machine-learning methods. In: American Geophysical Union, Fall Meeting, 2017
- 36 Colak T, Qahwaji R. Automated solar activity prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather*, 2009, 7: S06001
- 37 Fu G. Solar Activity Detection and Prediction Using Image Processing and Machine Learning Techniques. Dissertation for Doctoral Degree. Newark: New Jersey Institute of Technology, 2007. 100
- 38 Hernandez T. Forecasting flare activity using deep convolutional neural networks. In: American Geophysical Union, Fall Meeting, 2017
- 39 Jonas E, Bobra M, Shankar V, et al. Flare prediction using photospheric and coronal image data. *Sol Phys*, 2017, 293: 48, arXiv: [1708.01323](#)
- 40 Li R, Huang X. Solar flare forecasting model based on automatic feature extraction of sunspots (in Chinese). *Sci Sin-Phys Mech Astron*, 2018, 48: 119601 [李蓉, 黄鑫. 一种黑子特征自动提取的太阳耀斑模型. 中国科学: 物理学 力学 天文学, 2018, 48: 119601]
- 41 Bobra M G, Couvidat S. Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys J*, 2015, 798: 135, arXiv: [1411.1405](#)
- 42 Liu J, Ye Y, Shen C, et al. A new tool for CME arrival time prediction using machine learning algorithms: CAT-PUMA. *Astrophys J*, 2018, 855: 109, arXiv: [1802.02803](#)
- 43 Inceoglu F, Jeppesen J H, Kongstad P, et al. Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs. *Astrophys J*, 2018, 861: 128, arXiv: [1806.07117](#)
- 44 Bobra M G, Ilonidis S. Predicting coronal mass ejections using machine learning methods. *Astrophys J*, 2016, 821: 127, arXiv: [1603.03775](#)
- 45 Henwood R, Chapman S C, Willis D M. Increasing lifetime of recurrent sunspot groups within the Greenwich photoheliographic results. *Sol Phys*, 2010, 262: 299–313, arXiv: [0907.4274](#)
- 46 Bobra M, Ilonidis S. CME prediction using SDO, SOHO, and STEREO data with a machine learning algorithm. In: American Geophysical Union, Fall Meeting, 2015
- 47 Byrne J P, Morgan H, Habbal S R, et al. Automatic detection and tracking of coronal mass ejections. II. Multiscale filtering of coronagraph images. *Astrophys J*, 2012, 752: 145, arXiv: [1207.6125](#)
- 48 Padinhatteeri S, Higgins P A, Shaun Bloomfield D, et al. Automatic detection of magnetic $\delta\Delta$ in sunspot groups. *Sol Phys*, 2016, 291: 41–53, arXiv: [1510.06413](#)
- 49 Díaz Baso C J, Asensio Ramos A. Enhancing SDO/HMI images using deep learning. *Astron Astrophys*, 2018, 614: A5, arXiv: [1706.02933](#)
- 50 Asensio Ramos A, Requerey I S, Vitas N. DeepVel: Deep learning for the estimation of horizontal velocities at the solar surface. *Astron Astrophys*, 2017, 604: A11, arXiv: [1703.05128](#)

- 51 Teng F. Application of kernel based machine learning to the inversion problem of photospheric magnetic fields. *Sol Phys*, 2015, 290: 2693–2708
- 52 Alipour N, Safari H. Statistical properties of solar coronal bright points. *Astrophys J*, 2015, 807: 175
- 53 Reiss M A, Hofmeister S J, De Visscher R, et al. Improvements on coronal hole detection in SDO/AIA images using supervised classification. *J Space Weather Space Clim*, 2015, 5: A23, arXiv: [1506.06623](#)
- 54 Reiss M, Hofmeister S, DeVisscher R, et al. Differentiation between coronal holes and filament channels from SDO image data using machine learning algorithms. In: EGU General Assembly Conference Abstracts, 2015
- 55 Deo R C, Downs N, Parisi A V, et al. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environ Res*, 2017, 155: 141–166
- 56 Hao Q, Chen P F, Fang C. Automated detection methods for solar activities and an application for statistic analysis of solar filament. *Proc IAU*, 2018, 13: 101–107, arXiv: [1804.03320](#)
- 57 Camporeale E, Carè A, Borovsky J E. Classification of solar wind with machine learning. *J Geophys Res Space Phys*, 2017, 122: 10910–10920, arXiv: [1710.02313](#)
- 58 Muranushi T, Hattori T, Jin Q, et al. Comparative studies of prediction strategies for solar X-ray time series. In: AGU Fall Meeting, 2016
- 59 Guennou C, Pariat E, Vilmer N. Investigating the reliability of solar photospheric eruptivity proxies. In: American Geophysical Union, Fall Meeting, 2016
- 60 Susanna K. Predicting Major Solar Eruptions. AAS Nova Highlights, 2016, 6: 958
- 61 Shahamatnia E, Dorotović I, Fonseca J M, et al. An evolutionary computation based algorithm for calculating solar differential rotation by automatic tracking of coronal bright points. *J Space Weather Space Clim*, 2016, 6: A16
- 62 Schuh M A, Angryk R A, Martens P C. Solar image parameter data from the SDO: Long-term curation and data mining. *Astron Computing*, 2015, 13: 86–98
- 63 Angryk R. Mining of solar big data: Challenges and opportunities. In: AGU Fall Meeting Abstracts, 2014
- 64 Schuh M A, Banda J M, Bernasconi P N, et al. A comparative evaluation of automated solar filament detection. *Sol Phys*, 2014, 289: 2503–2524
- 65 Colak T, Qahwaji R. Prediction of extreme ultraviolet variability experiment (EVE)/extreme ultraviolet spectro-photometer (ESP) irradiance from solar dynamics observatory (SDO)/atmospheric imaging assembly (AIA) images using fuzzy image processing and machine learning. *Sol Phys*, 2013, 283: 143–156
- 66 Cheung M. From emergence to eruption: The physics and diagnostics of solar active regions. In: American Astronomical Society, SPD meeting #48, 2017
- 67 Yuan Y. Development of Advanced Algorithms to Detect, Characterize and Forecast solar Activities. Dissertation for Doctoral Degree. Newark: New Jersey Institute of Technology, 2011. 136
- 68 He H, Wang H, Du Z, et al. Solar activity prediction studies and services in NAOC. *Adv Space Res*, 2008, 42: 1450–1456
- 69 Yang Y H, Tian H M, Peng B, et al. Multi-label learning for detection of CME-associated phenomena. *Sol Phys*, 2017, 292: 131
- 70 Verbeeck C, Higgins P A, Colak T, et al. A multi-wavelength analysis of active regions and sunspots by comparison of automatic detection algorithms. *Sol Phys*, 2013, 283: 67–95
- 71 Long D M, Bloomfield D S, Gallagher P T, et al. CorPITA: An automated algorithm for the identification and analysis of coronal “EIT waves”. *Sol Phys*, 2014, 289: 3279–3295, arXiv: [1403.6722](#)
- 72 Attrill G D R, Wills-Davey M J. Automatic detection and extraction of coronal Dimmings from SDO/AIA data. *Sol Phys*, 2010, 262: 461–480
- 73 Barata T, Carvalho S, Dorotović I, et al. Software tool for automatic detection of solar plagues in the Coimbra Observatory spectroheliograms. *Astron Comput*, 2018, 24: 70–83, arXiv: [1811.08389](#)
- 74 Kucuk A, Banda J M, Angryk R A. A large-scale solar dynamics observatory image dataset for computer vision applications. *Sci Data*, 2017, 4: 170096
- 75 Bobra M G, Sun X, Hoeksema J T, et al. The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: SHARPs—Space-weather HMI active region patches. *Sol Phys*, 2014, 289: 3549–3578, arXiv: [1404.1879](#)
- 76 Podladchikova O, Vuets A, Leontiev P, et al. Recent developments of NEMO: Detection of EUV wave characteristics. *Sol Phys*, 2012, 276: 479–490, arXiv: [1103.4637](#)
- 77 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, 2016

- 78 Gan W, Deng Y, Li H, et al. ASO-S: Advanced space-based solar observatory. In: Proceedings of the SPIE Solar Physics and Space Weather Instrumentation VI, vol. 9604. San Diego, 2015. 96040T

The application of machine learning in solar physics

LIU Hui^{1,2}, JI KaiFan^{1*} & JIN ZhenYu¹

¹ Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, China;

² Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Solar physics has entered the era of big data, and machine learning has gained more and more recognition as a good tool for big data research. This paper reviews the application results of machine learning in solar physics since 2007. Our studies have shown that research in this field has increased significantly during the last four years. Massive solar observation data obtained from various instruments on the ground and in space have been applied, and the topics have covered major aspects of solar physics, such as solar flares, coronal mass ejections, sunspots. Although some good results have emerged and proved that machine learning is suitable for data analysis of solar physics, there has not been a breakthrough yet. The machines learning methods that used in this field involve classification, regression, clustering, dimensionality reduction, and deep learning. However, classical algorithms, especially classical classification method is more popular. This means that the application of machine learning in solar physics is still in its infancy, but it also means that there is still a lot of work in this field that can be studied in the future.

solar physics, solar activity, machine learning, deep learning

PACS: 96.60.-j, 96.60.Q-, 96.60.Hv, 96.60.qd

doi: [10.1360/SSPMA-2019-0031](https://doi.org/10.1360/SSPMA-2019-0031)