一种黑子特征自动提取的太阳耀斑模型

李蓉^{1*}, 黄鑫²

1. 北京物资学院信息学院, 北京 101149;

2. 中国科学院国家天文台, 北京 100012

*联系人, E-mail: lirong@bao.ac.cn

收稿日期: 2018-02-22; 接受日期: 2018-03-28; 网络出版日期: 2018-07-31

北京市智能物流系统协同创新中心、“智能物流系统北京市重点实验室”、2017北京人才培养质量建设-专业建设群(编号: PXM2017_014214_000010)和国家自然科学基金(编号: 11303051)资助项目

摘要 太阳耀斑是最剧烈的太阳爆发之一, 太阳耀斑能够在几分钟内对空间环境产生影响, 为了更好地应对太阳耀斑爆发的空间环境效应, 本文提出一种基于自动特征提取方法的太阳耀斑预报模型. 该模型以太阳黑子的连续谱观测图像为输入, 利用卷积神经网络的自动图像特征提取能力, 建立太阳耀斑预报模型. 传统的以太阳黑子特征为输入的耀斑模型需要以黑子的McIntosh分型作为输入, 黑子的McIntosh分型需要人为指定, 效率低并且具有一定的主观性. 通过实验比较, 本文提出的耀斑预报模型比传统的以太阳黑子McIntosh分型为基础的预报模型具有更好的预报准确度. 总之, 本文提出的太阳耀斑预报模型能够自动提取太阳黑子图像的特征, 具有更高的预报效率和准确率(在48 h M级耀斑预报中, 基于深度学习方法的预报模型在耀斑报准率上比传统预报模型高3.7%, 在非耀斑报准率上比传统预报模型高2.8%), 是一种实用性较强的太阳耀斑预报模型.

关键词 太阳耀斑, 太阳黑子, 卷积神经网络, 自动特征提取

PACS: 94.05.Sd, 96.60.qe, 96.60.Q-

1 引言

太阳耀斑是一种剧烈的太阳爆发现象, 与之伴随发生的高能粒子和辐射对空间环境产生剧烈的冲击, 对空间飞行器或宇航员造成潜在的危害. 当耀斑辐射来到地球附近时, 光致电离使得D层的电子密度增加, 破坏电离层并引起无线电通信中断. 因此提前1-2 d的太阳耀斑短期预报具有重要的实用价值.

到目前为止, 已经提出了许多短期太阳耀斑预报

的方法. 考虑到太阳黑子和耀斑发生之间的紧密联系, 一些太阳耀斑预报模型采用黑子参量作为预报因子. 文献[1]提出太阳黑子的McIntosh分型, 并统计了McIntosh分型所对应的太阳耀斑产率. 基于太阳黑子的McIntosh分型, 美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)空间环境实验室和Colorado大学联合开发了一个专家系统Theo来预报太阳耀斑. 文献[2]采用多维线性回归方法统计了每个McIntosh分型对太阳耀斑的贡献. 美国大

引用格式: 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

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, doi: [10.1360/SSPMA2018-00040](https://doi.org/10.1360/SSPMA2018-00040)

熊湖天文台的Gallagher等人^[3]采用泊松统计的方法, 计算McIntosh分类的三个参量对每一个活动区产生X射线耀斑的可能性来进行预报. 文献[4]利用人工神经网络技术建立耀斑预报模型, 达到了与Theo相同的预报精度. 英国Bradford大学的Qahwaji和Colak^[5]应用神经网络和支持向量机算法建立了混合的耀斑短期预报系统, 以McIntosh分型和太阳黑子数为预报因子, 首先利用支持向量机预报是否有耀斑发生, 如果有耀斑发生, 再利用神经网络预报发生耀斑的级别. Li等人^[6]应用支持向量机-k近邻方法建立了一个耀斑短期预报系统, 并得到了良好的预报精度. 文献[7]统计计算了黑子面积、磁分类、McIntosh分型与耀斑的产率关系, 使用决策树算法建立了树状太阳耀斑预报模型. 文献[8]针对X级的大耀斑给出耀斑的峰值、结束时间等参量的预报. 文献[9]结合了黑子McIntosh分型参量和光球磁场参量, 对太阳耀斑进行了更加准确的预报. 文献[10]建立了X级耀斑的径向基神经网络模型. 文献[11]综述了已有的太阳耀斑预报模型. 文献[12]将主成分分析方法和支持向量机算法结合在一起, 给出太阳耀斑预报模型. 文献[13]提出一种自动给定太阳黑子群McIntosh分型的方法. 基于该方法, 文献[14]给出太阳耀斑的自动预报模型.

近年来, 在机器学习研究中, 深度学习方法取得了快速的发展^[15,16], 并成功应用在语音识别、自然语言处理、目标识别和分类等领域^[17,18].

与之前方法不同的是, 深度学习方法能够从原始观测数据中自动学习特征参量并建立模型. 该方法替代了经验性的特征提取方法, 大大降低了建立模型的难度. 其中卷积神经网络是在图像识别中应用较为广泛的一种深度神经网络.

在上述耀斑预报模型中, 非常重要的输入是黑子的McIntosh分型参量. 然而, McIntosh分型通常由观测人员指定, 具有一定的主观性, 并且实时性较差. 此外, 建立耀斑预报模型需要统计建立太阳耀斑和太阳黑子参量的关系, 依赖于统计样本和方法. 为了解决上述问题, 本文直接以太阳黑子的连续谱观测图像作为输入, 利用卷积神经网络^[19]自动特征抽取的能力, 自动提取太阳黑子的分布式模式, 并基于此建立太阳耀斑预报模型, 与之前基于McIntosh分型参量的太阳耀斑预报模型进行比较.

2 数据

强耀斑的发生经常和太阳黑子密切相关^[20]. 本文使用太阳黑子观测数据预测太阳耀斑的发生. 太阳黑子主要由McIntosh分型进行描述. McIntosh分型是在太阳黑子Zurich分类的基础上改进而来. McIntosh分型的形式可表达为: Zpc , 其中, Z 表示修改的Zurich分类; p 表示黑子类型; c 表示黑子群分别间隔的紧凑度. 太阳黑子观测图像使用SOHO (Solar and Heliospheric Observatory)卫星MDI (Michelson Doppler Imager)观测的太阳黑子连续谱图像.

太阳耀斑数据来源于国家地球物理数据中心(National Geophysical Data Center, NGDC)整理的数据列表(<ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/>). 太阳耀斑根据其软X射线峰值流量分为B, C, M, X四级. 我们将一段时间内的耀流量总和作为耀斑重要程度的指标

$$I_{\text{tot}} = \sum c + 10 \times \sum m + 100 \times \sum x, \quad (1)$$

其中, c , m , x 分别为C级耀斑、M级耀斑、X级耀斑所对应的系数. 一个耀斑预报模型通常关注在未来一定时间内大于某一级别的耀斑发生情况, 本文考虑在预报时限内发生的所有耀斑的情况, 以耀斑重要程度指标 I_{tot} 的数值作为耀斑是否发生的准则. 如果一个太阳黑子在48 h内的 I_{tot} 大于10, 认为该黑子样本为耀斑样本, 否则认为该黑子样本为非耀斑样本^[21].

本文选取从1996–2008年日面中心30°范围内的太阳黑子数据. 数据集共包含21656张太阳黑子图像, 其中1252张黑子图像在48 h内发生了指定级别的耀斑, 剩余的20404张黑子图像在48 h内没有指定级别的耀斑发生. 我们从20404张没有发生耀斑的黑子图像中随机抽取1252张图像, 与1252张发生耀斑的黑子图像一起, 构成训练预报模型的数据集合.

3 方法

到目前为止, 太阳耀斑爆发的物理机制并不完全清晰, 太阳耀斑预报还是以活动区特征参量为基础的统计或机器学习方法的预报. 传统的基于太阳黑子耀斑预报模型(图1)需要提取活动区的黑子特征参量, 并采用机器学习或统计方法找出太阳活动区参量与太阳

耀斑爆发之间的关系, 即建立预报模型. 太阳活动区的黑子特征提取问题成为整个系统的关键, 并影响预报结果.

而近年来迅速发展的深度学习方法能够实现从原始观测数据中自动提取特征, 并建立预报模型. 其中卷积神经网络常常用来处理图像数据. 本文使用卷积神经网络自动提取太阳黑子图像的特征, 并建立太阳耀斑预报模型. 一个典型的卷积神经网络的结构如图2所示.

由图2可以看出, 一个卷积神经网络主要由如下部分组成:

(1) 卷积层. 卷积层计算输入数据和指定滤波器的卷积计算, 并将卷积的结果作为输出. 卷积层常用来提取对象的低层特征.

(2) 池化层. 池化层用于计算某一指定区域的统计特性, 包括计算输入数据在指定范围的统计参量, 并将统计结果作为输出. 池化层通常用于降低数据的维度, 防止模型过拟合. 本文使用最大值池化方法.

(3) 归一化层. 归一化层实现了从输入到输出的非线性变换. 该层用于在模型中增加非线性特征. 常用的归一化函数有S形激活函数($f(x) = \frac{1}{1+e^{-x}}$)和ReLU

(Rectified Linear Units, $f(x) = \max(0, x)$)激活函数. 由于ReLU函数具有更快的模型收敛能力, 本文选自ReLU函数作为归一化函数.

(4) 全连接层. 全连接层中的每个神经元与前一层的所有神经元相连. 通常全连接层是卷积神经网络最后几层的结构, 前面几层的输出在全连接层中进行高阶的推理和计算, 并给出模型最终的输出结果.

在卷积神经网络的设计过程中, 首先需要设计模型的结构. 模型的输入使用SOHO卫星MDI观测的太阳黑子连续谱图像. 网络中的权值由观测数据确定, 确定网络权值的过程被称为学习或训练. 模型结构确定后, 使用随机梯度下降算法训练模型, 训练的最终目的是最小化训练误差, 从数据中学习网络权值^[19], 使模型在新数据上具有良好的预测能力. 一旦网络的权值被确定, 训练过程就结束了, 模型的性能可使用测试数据来评估.

4 结果与分析

本文使用伯克利视觉和学习中心(Berkeley Vision and Learning Center)开发的深度学习框架Caffe^[22]建立

图 1 (网络版彩图)传统预报模型的结构图

Figure 1 (Color online) Example of a traditional neural network.

图 2 (网络版彩图)卷积神经网络的结构图

Figure 2 (Color online) Example of a typical convolutional neural network.

卷积神经网络进行太阳耀斑预报. 模型的输入是太阳黑子的连续谱观测图像, 第二层是一个卷积层(50个5×5的滤波器), 第三层是一个最大值池化层(3×3), 第4–6层是一个3层的全连接层(100×10×2个节点), 模型的两个输出分别代表未来48 h是否产生太阳耀斑.

上述模型有两类输出(耀斑或非耀斑), 我们将耀斑样本认为是正类样本, 非耀斑样本是负类样本, 因此模型的测试结果会出现如表1所示的4种情况. 其中TP定义为对正类样本预报为正类的数目; TN为对负类样本预报为负类的数目; FP为对负类样本预报为正类的数目; FN为对正类样本预报为负类的数目.

根据表1的4种输出, 我们定义如下4个指标反映模型的测试性能.

TP rate为正确的正类样本数和真实的正类样本数的比

$$\text{TP rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (2)$$

TN rate为正确的负类样本数与真实的负类样本数的比

$$\text{TN rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (3)$$

FP rate为虚报率, 定义为错误的否样本数和预报的正类样本数的比

$$\text{FP rate} = \frac{\text{FP}}{\text{TP} + \text{FP}}. \quad (4)$$

FN rate为漏报率, 定义为错误的肯定样本和真实的负类样本的比

$$\text{FN rate} = \frac{\text{FN}}{\text{FP} + \text{TN}}. \quad (5)$$

对于整个数据集, 我们采用十折交叉验证法测试太阳耀斑预报模型的性能. 在该方法中, 将数据集分成10份, 将其中9份作为训练数据集, 剩下的1份作为测试数据集, 测试轮流进行10次, 保证整个数据集都进行测试.

为了比较本文预报模型与传统基于McIntosh分型进行太阳耀斑预报模型的性能, 将本文的实验结果(表2)与文献[7]的实验结果(表3)进行了对比(表4).

在太阳耀斑预报中, 只有少部分太阳黑子产生大的太阳耀斑, 其余黑子没有产生大的耀斑. 在机器学习中, 此类问题被称为“类不平衡问题”^[23]. 在处理类

表1 两类输出预报模型的4种模型测试情况

Table 1 2×2 contingency table for binary solar flare forecasting

	观测耀斑	观测非耀斑
预报耀斑	TP	FP
预报非耀斑	FN	TN

表2 卷积神经网络模型的测试结果

Table 2 Testing results of convolution neural network model

	观测耀斑	观测非耀斑
预报耀斑	1008	231
预报非耀斑	291	978

表3 C4.5决策树预报模型的测试结果^[7]

Table 3 Testing results of C4.5 decision tree model [7]

	观测耀斑	观测非耀斑
预报耀斑	961	274
预报非耀斑	291	978

表4 卷积神经网络模型和C4.5决策树模型性能比较

Table 4 Performance evaluation measures of convolution neural network model and C4.5 decision tree model

测试参数	卷积神经网络模型 (%)	决策树模型 ^[7] (%)
TP rate	80.5	76.8
TN rate	80.9	78.1
FP rate	18.6	22.2
FN rate	20.2	23.2

不平衡问题时, 通常将两类样本处理平衡后再进行建模分析. 本文的实验就是在产生耀斑的黑子图像和没有产生耀斑的黑子图像数量相等的条件下进行的. 为了讨论太阳耀斑预报中的类不平衡问题, 逐步增加没有产生耀斑的太阳黑子图像的数据, 并建立预报模型. 不同样本比例下, 太阳耀斑预报模型性能如图3所示. 图3的测试结果表明, 随着没有产生耀斑的太阳黑子图像数量的增加, 预报模型对非耀斑的报准率不断升高, 但是模型对耀斑的报准率却显著下降. 该实验表明, 在太阳耀斑预报中, 将两类样本处理平衡后再进行预报建模的策略是合理的.

文献[24]中对黑子特征的刻画仍然使用McIntosh分型等经典特征参量. 本文的工作不再使用黑子的McIntosh分型等经典特征参量, 而是从大量黑子数据

图 3 (网络版彩图) 太阳耀斑样本和非耀斑样本数量变化对预报结果的影响

Figure 3 (Color online) Influence of prediction result from number changes in flare and non flare samples.

中自动提取刻画黑子性质的特征参量. 通过对比实验, 我们发现直接以太阳黑子连续谱观测图像为输入的太阳耀斑卷积神经网络预报模型比以太阳黑子 McIntosh 分型为输入的传统太阳耀斑预报模型性能略好. 本

文提出的太阳耀斑预报模型能够实现太阳黑子特征的自动提取, 避免了人工指定太阳黑子分型的主观误差, 模型运行效率和准确率都有所提高, 因此该耀斑预报模型具有更好的实用性.

5 结论

本文以太阳黑子的连续谱观测图像作为输入, 利用卷积神经网络的图像特征自动提取能力, 建立太阳耀斑预报模型. 传统的太阳耀斑预报模型需要人工区分太阳黑子的 McIntosh 分型, 然后以太阳黑子的 McIntosh 分型作为模型的输入, 进行太阳耀斑预报. 与传统太阳耀斑预报模型相比, 本文提出的方法避免了人工特征提取的过程, 具有更高的效率, 避免了人为因素引入的误差. M 级耀斑 48 h 预报测试结果显示, 基于深度学习方法的预报模型在耀斑报准率上比传统预报模型高 3.7%, 在非耀斑报准率上比传统预报模型高 2.8%. 综上所述, 本文提出的太阳耀斑预报模型具有更强的实用性, 值得在业务运行中进行推广.

参考文献

- McIntosh P S. The classification of sunspot groups. *Sol Phys*, 1990, 125: 251–267
- Bornmann P L, Shaw D. Flare rates and the McIntosh active-region classifications. *Sol Phys*, 1994, 150: 127–146
- Gallagher P T, Moon Y J, Wang H. Active-region monitoring and flare forecasting I. Data processing and first results. *Sol Phys*, 2002, 209: 171–183
- Bradshaw G, Fozzard R, Ceci L. A connectionist expert system that actually works. *Adv Neural Inf Process Syst*, 1989, 1: 248–255
- Qahwaji R, Colak T. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Sol Phys*, 2007, 241: 195–211
- Li R, Wang H N, He H, et al. Support vector machine combined with K-nearest neighbors for solar flare forecasting. *Chin J Astron Astrophys*, 2007, 7: 441–447
- Li R, Cui Y M. Short-term solar flare forecasting model using data mining technique (in Chinese). *Sci Sin-Phys Mech Astron*, 2011, 41: 1342–1350 [李蓉, 崔延美. 应用数据挖掘技术的短期太阳耀斑预报模型. 中国科学: 物理学 力学 天文学, 2011, 41: 1342–1350]
- Guo C, Xue B S, Lin Z X. Study for prediction method of X-ray flare characteristic parameters (in Chinese). *Chin J Space Sci*, 2012, 32: 771–777 [郭策, 薛炳森, 林兆祥. 太阳射线耀斑特征参数预报方法研究. 空间科学学报, 2012, 32: 771–777]
- Li R, Zhu J, Cui Y M. Solar flare forecasting model with active region photospheric magnetic field properties and sunspot factors. *Chin Sci Bull*, 2013, 58: 1845–1850 [李蓉, 朱杰, 崔延美. 结合活动区光球磁场参量和黑子参量的太阳耀斑预报模型. 科学通报, 2013, 58: 1845–1850]
- Pei S X, Li Z Y, Cui F P, et al. A study of the prediction for X-class and above flares based on RBF artificial neural network (in Chinese). *J Northwest Normal Univ (Nat Sci)*, 2014, 50: 26–32 [裴世鑫, 李仲怡, 崔芬萍, 等. 基于 RBF 人工神经网络的 X 级以上太阳耀斑预报研究. 西北师范大学学报(自然科学版), 2014, 50: 26–32]
- Li R, Zhu J, Huang X. Progress in solar flare forecasting (in Chinese). *Chin Sci Bull*, 2014, 59: 2452–2463 [李蓉, 朱杰, 黄鑫. 太阳耀斑预报研究进展. 科学通报, 2014, 59: 2452–2463]
- Yuan F, Lin J B, Deng Y Y, et al. Solar flare forecasting model of combining principal component analysis with support vector machine (in Chinese). *Sci Bull*, 2016, 61: 2316–2321 [袁飞, 林佳本, 邓元勇, 等. 结合主成分分析和支持向量机的太阳耀斑预报模型. 科学通报, 2016,

61: 2316–2321]

- 13 Colak T, Qahwaji R. Automated McIntosh-based classification of sunspot groups using MDI images. *Sol Phys*, 2008, 248: 277–296
- 14 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
- 15 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313: 504–507
- 16 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521: 436–444
- 17 Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. *J Mach Learn Res*, 2011, 12: 2493–2537
- 18 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, 2012
- 19 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE*, 1998, 86: 2278–2324
- 20 Tian Z D, Li S J, Wang Y H, et al. A hybrid prediction model of smoothed monthly mean sunspot number (in Chinese). *Sci Sin-Phys Mech Astron*, 2016, 46: 119601 [田中大, 李树江, 王艳红, 等. 太阳黑子数平滑月均值的混合预测模型. *中国科学: 物理学 力学 天文学*, 2016, 46: 119601]
- 21 Cui Y, Li R, Zhang L, et al. Correlation between solar flare productivity and photospheric magnetic field properties. *Sol Phys*, 2006, 237: 45–59
- 22 Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*. New York: ACM, 2014
- 23 Qing M M, Qiu J L, Lu P C, et al. AdaBoost-based class imbalance learning algorithm (in Chinese). *Appl Res Comput*, 2017, 34: 3229–3232 [秦孟梅, 邱建林, 陆鹏程, 等. 基于AdaBoost的类不平衡学习算法. *计算机应用研究*, 2017, 34: 3229–3232]
- 24 Li R, Zhu J. Solar flare forecasting based on sequential sunspot data. *Res Astron Astrophys*, 2013, 13: 1118–1126

Solar flare forecasting model based on automatic feature extraction of sunspots

LI Rong^{1*} & HUANG Xin²

¹ *The Institution of Information, Beijing WuZi University, Beijing 101149, China;*

² *National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*

The solar flare is one of the most violent solar eruptions. It can affect the space environment in a few minutes. In order to better deal with the space environment, which is influenced by solar flares, a solar flare forecasting model based on the automatic feature extraction method is proposed. The input of this forecasting model is established by taking the continuum image of a sunspot. The convolution neural network with the ability of automatic image feature extraction is used to build the forecasting model. The McIntosh classifications of the sunspots is considered as the inputs of the forecasting model in traditional methods. The McIntosh classifications of the sunspot are specified by an expert with a certain subjectivity. Compared with the traditional forecasting model based on the McIntosh classifications of the sunspot, the proposed flare forecasting model has better forecasting accuracy (In the forecast of M-class flares with a 48-h lead time, the success rate of the deep learning method is increased by 3.7% compared to that of the traditional flare forecast model, whereas the success rate for the non-flare prediction is increased by 2.8%). In conclusion, the proposed solar flare forecasting model, which can automatically extract the features of a sunspot, is a practical solar flare forecasting model.

solar flare, sunspots, convolutional neural networks, automatic feature extraction

PACS: 94.05.Sd, 96.60.qe, 96.60.Q-

doi: [10.1360/SSPMA2018-00040](https://doi.org/10.1360/SSPMA2018-00040)