

doi:10.16055/j.issn.1672-058X.2020.0001.010

# 基于最大池化层参数的优化模型 在引力波天文学中的应用

罗华美

(西南交通大学 数学学院数学系, 成都 611756)

**摘要:**针对 Gabbard 等人发表在《Physical Review Letters》上的文章“Matching Matched Filtering with Deep Networks for Gravitational-Wave Astronomy”,提出了一种卷积神经网络优化模型。文章将卷积神经网络应用于引力波信号的识别,研究最大池化层参数对模型分类能力的影响,调整模型中超参数提升引力波信号分类的准确率;将优化后的网络结构与 Gabbard 的卷积神经网络用于相同的模拟数据集,并在测试集上绘制了接受者操作特性曲线(Receiver Operating Characteristic curve,简称 ROC 曲线),计算了 ROC 曲线下的面积;结果证明:相比于未优化的网络,此处的模型在 ROC 下的面积在数值上提高了 0.025 4~0.032 6;同时,改变噪音的振幅,将两种方法应用于新的数据集上,结果同样证明,优化后网络效果更好,鲁棒性强。

**关键词:**引力波天文学;深度学习;卷积神经网络

**中图分类号:** O411

**文献标志码:** A

**文章编号:** 1672-058X(2020)01-0059-06

## 0 引言

在物理学中,引力波是指时空弯曲中的涟漪,这种弯曲是因为质量的存在而导致的,通过波的形式从辐射源向外传播,这种波以引力辐射的形式传输能量。早在 1916 年爱因斯坦就基于广义相对论预言了引力波的存在<sup>[1]</sup>。引力波的存在是广义相对论洛伦兹不变性的结果,因为它引入了相互作用的传播速度有限的概念。

在过去的 60 年里,有许多物理学家和天文学家为证明引力波的存在做出了无数努力。不论是国内还是国外,都在积极地建造或运行各种各样的引力波探测器,例如 Advanced LIGO 就在 2015 年 9 月

开始运行观测。在 2016 年 2 月 11 日,LIGO 科学合作组织和 Virgo 合作团队宣布已经利用高级 LIGO 探测器,首次探测到了来自双黑洞合并的引力波信号<sup>[2]</sup>。在此之前,只能通过它们对双星系统中脉冲星定时的影响间接推断出引力波。根据观测到引力波的日期,将这个引力波信号命名为 GW150914<sup>[2-3]</sup>。

继 LIGO GW150914 首个引力波信号之后,LIGO 合作组织又宣布:位于美国 Hanford 区和 Louisiana 的利文斯顿的两台引力波探测器同时探测到了一个引力波信号,这是人类探测到的第二个引力波信号<sup>[4]</sup>。截至 2018 年 1 月 16 日,类似的事件已被检测到 6 次,有理由相信,在未来几年里,还会陆续观测到更多这样的引力波探测事件,引力波

收稿日期:2019-05-21;修回日期:2019-07-22.

作者简介:罗华美(1994—),女,四川宜宾人,硕士研究生,从事深度学习研究.

天文学即将进入大数据阶段。目前用于检测良好建模的引力波信号的标准方法是匹配滤波,但匹配滤波有其自身的缺点和潜在问题。它的缺点是巨大的计算量导致数据处理很慢,完全不适合后续引力波天文观测预警。它的潜在问题是完备准确的理论波形模板,是其工作的必要条件,而且运用匹配滤波方法很难找到模板以外的引力波信号。

深度学习是机器学习的一个子集,它受到生物神经网络中信息处理和通信模式的模糊启发,但与生物大脑的结构和功能特性存在各种差异<sup>[5-7]</sup>。它将输入数据中的底层特征进行组合,形成稍微抽象和复合的表现形式,这些表现形式与属性类别或者属性特征相关,通过抽象的表现形式,发现数据的分布特征。自 2006 年以来,深度学习算法在众多监督学习问题上取得了卓越的成果,甚至在一些方面的性能远超传统机器学习算法,部分已经超越了人类水平。近年来,深度学习算法在图像处理<sup>[8-10]</sup>、医疗诊断<sup>[11]</sup>、信号识别<sup>[12,13]</sup>方面有广泛应用价值,得到了快速的发展。深度学习在数据处理上体现出来的优势,自然会让人们联想到将深度学习方法应用到引力波天文学中。目前已经有不少文献讨论了深度学习用于引力波数据处理的可行性<sup>[12,14,15]</sup>, Gabbard 等<sup>[16]</sup>也提出,传统的卷积神经网络可以实现匹配滤波的灵敏度。

此处论文在 Gabbard 等<sup>[16]</sup>文章的基础上,对文中的传统卷积神经网络进行优化,并在测试集上绘制 ROC 曲线和计算 ROC 曲线下的面积。结果表明:相比于未优化网络,改进后的模型在 ROC 下的面积在数值上提高了大约 0.025 4 ~ 0.032 6。同时,也改变噪音的振幅,将两种方法应用于新的数据集上,结果同样证明,优化后网络效果更好,鲁棒性强。

## 1 模拟数据

本文使用的模拟数据和 Gabbard 等人在文章<sup>[16]</sup>中模拟的数据相同。区分两种情况,一是双黑洞(Binary Black Hole, BBH)并合信号加上高斯噪

音,另一种情况是只有高斯噪音。其中信号的模拟使用了一个叫 LALSuite 的引力波数据分析库和 Python 下的包 Pycbc。文章中使用的波形是 IMRPhenomD,模拟系统中双黑洞质量  $m_1$  和  $m_2$  在  $5 \sim 95 M_{\odot}$ ,其中  $m_1 > m_2$ ,自旋为零,质量取自天体物理中常用的分布(astrophysically motivated)并且  $m_i \sim \log m_i, i = 1, 2$ ,每个信号中的赤经(right ascension)和赤纬(declination)这两个参数假设是天球的各向同性分布(isotropic on the sky)的先验分布,偏振角和相位的先验是取自  $[0, 2\pi]$  的均匀分布,倾斜角是余弦在  $[-1, 1]$  的均匀分布,输入是以时间序列的形式,波形随机放置在时间序列中,使得每个波形的波峰振幅被随机放置在时间序列的  $[0.75, 0.95]$  部分。为了实现预定义的最优信噪比,波形振幅会进行缩放。最优信噪比的定义如下:

$$\rho_{\text{opt}}^2 = 4 \int_{f_{\text{min}}}^{\infty} \frac{|h(f)|^2}{S_n(f)} df$$

其中  $h(f)$  是引力波应变的频域表示,  $S_n(f)$  是单边检测器噪音功率谱密度(power spectral density, psd),模拟的时间序列在 8 192 Hz 下持续采样 1 s,  $f_{\text{min}}$  是时间序列采样起始时引力波信号的频率。使用 Python 软件,定义了 tukey 窗口、解析命令行参数、利用噪音功率谱密度(psd)生成噪音、获取 psd、获取 snr(信噪比的信号积分)、白化时间序列等一系列的函数,最后调用一系列函数生成训练、验证、测试数据集。在解析命令行参数时,数据集的质量分布选择天体物理中常用的分布,在训练集上每个信号噪音实现数量设置为 25,验证集和测试集设置为 1,对训练集、验证集、测试集设置不同的随机种子。

## 2 优化的模型结构

深度学习是机器学习的子集,它分为监督学习和无监督学习,卷积神经网络是监督学习的一种。监督学习把数据集分为训练集、验证集和测试集。训练集是用于模型拟合的数据样本,验证集用于调整模型的超参数、优化模型和初步评估模型的能力

力,测试集用来评估最终模型的泛化能力。用于目标识别的传统卷积神经网络(Convolutional Neural Network, CNN)大多使用相同原则构建:卷积层和最大池化(Max pooling)层交替出现,后面在跟着少数的全连接层。本篇论文针对 Gabbard 等人使用的网络结构,调整模型的超参数,将优化的结构与 Gabbard 的卷积神经网络模型用于相同的数据集,评估模型的好坏。在训练网络前,先加载模拟的数据,对于模拟的数据集,可以绘制一些时间序列看(图 1)。Class 0 表示只有高斯噪声,此时高斯噪声振幅  $f_{\text{amp}}$  为

$$f_{\text{amp}} = \sqrt{0.25 * T_{\text{obs}} * f_{\text{psd}}}$$

其中  $T_{\text{obs}}$  是观察的持续时间,单位为秒,  $f_{\text{psd}}$  是噪音功率谱密度。Class 1 表示相同的高斯噪声加上 BBH 信号,此时事先给定的信号的 snr 为 6.0。

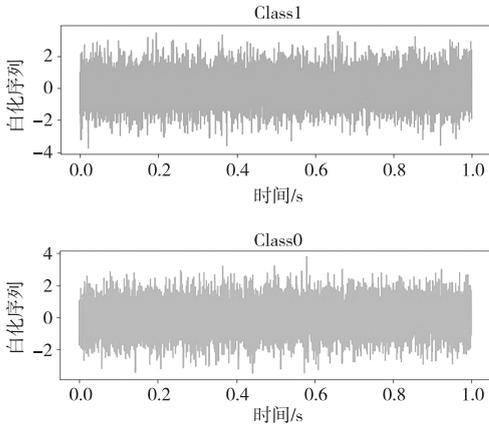


图 1 模拟数据集中一些时间序列例子

Fig. 1 Some time series examples in the simulation data set

加载了数据之后,使用 keras 学习框架训练网络,为了最优化网络,多个超参数必须要调整,将自由选择的参数定义为超参数。例如网络中层的类型和数量、每个层内的神经元数量、权重向量的大小、Max pooling 层内的参数、激活函数的类型、输入数据的预处理、学习速率、特定深度学习技术的应用。通过多次测试和在误差上调整超参数,最优网络结构将会被确定。由于未优化网络存在过拟合问题,故调整最大池化层的参数,表 1 是本文使用的最优网络结构,其中包含了 6 个卷积层,3 个隐藏层。本文采用反向传播方法微调模型中的参数,损

失函数是交叉熵损失函数,自适应学习率优化算法采用 Nadam。损失函数定义如下:

$$f(\theta) = - \sum_{i \in S} \log(\theta_i^S) - \sum_{i \in N} \log(\theta_i^N)$$

Nadam 优化算法参数更新如下:

$$g_t = \nabla_{\theta_{t-1}} J(\theta_{t-1}), \hat{g}_t = \frac{g_t}{1 - \prod_{i=1}^t \mu_i}$$

$$m_t = \mu_t * m_{t-1} + (1 - \mu_t) * g_t, \hat{m}_t = \frac{m_t}{1 - \prod_{i=1}^{t+1} \mu_i}$$

$$n_t = v_t * n_{t-1} + (1 - v_t) * g_t^2, \hat{n}_t = \frac{n_t}{1 - v^t}$$

$$\bar{m}_t = (1 - \mu_t) * \hat{g}_t + \mu_{t+1} * \hat{m}_t, \Delta\theta_{t-1} = -\eta * \frac{\hat{m}_t}{\sqrt{\hat{n}_t} + \varepsilon}$$

第  $t$  次参数  $\theta_t$  更新如下:

$$\theta_t = \theta_{t-1} + \Delta\theta_{t-1}$$

其中,  $J(\theta_{t-1})$  是  $t-1$  次对应的目标损失函数,  $g_t$  是  $J(\theta_{t-1})$  对  $t-1$  次对应的参数  $\theta_{t-1}$  的梯度,  $m_t$ 、 $n_t$  分别是对梯度的一阶矩估计和二阶矩估计,可以认为是期望  $E_{|g_t}$  和  $E_{|g_t^2}$  的估计,  $\hat{m}_t$ 、 $\hat{n}_t$  分别是对  $m_t$ 、 $n_t$  的校正,近似为对期望的无偏估计,  $\mu_t$  和  $v_t$  都是衰减因子,  $\eta$  是学习率。在论文中,取  $\mu = 0.9$ ,  $v = 0.999$ ,  $\varepsilon = 10^{-8}$ ,  $\eta = 0.002$ 。

模型中的非线性特性是由非线性激活层来体现的,在前面的卷积层和前两个隐藏层使用的都是 elu 激活函数,最后的输出层使用 softmax 函数,其中 elu 函数表示为

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0, \alpha > 0 \end{cases}$$

本文使用 elu 激活函数,是因为相比于 sigmoid 和 ReLU,它融合了这两者的特性,左侧具有软饱和性,能够让 elu 对输入变化或噪声鲁棒性更高;右侧线性特性使得 elu 能够缓解梯度消失,并且,它的收敛速度更快,因为输出均值接近于 0。输出层使用 softmax 函数,它可以用于多分类问题,也可以用于二分类,本文中它将两个神经元的输出,映射到 (0, 1) 区间上的一个向量。该向量的分量数值代表了输入样本属于该位置对应类别的概率。找到数值大于 0.5 的分量代表的类别,将其认为是输入样本

应属于的类别。表 1 为文中使用的网络框架。

表 1 网络框架

Table 1 Network framework used in this paper

参 数 (选项)	层								
	1	2	3	4	5	6	7	8	9
类型	C	C	C	C	C	C	H	H	H
神经元数量	8	8	16	16	32	32	64	64	2
卷积核	64	32	32	16	16	16	na	na	na
最大池化层	na	10	na	8	na	6	na	na	na
Drop out	0	0	0	0	0	0	0.5	0.5	0
激活函数	elu	SMax							

表 1 中,na 表示不应用参数(not applicable)。

### 3 结 果

为了评估文中所提的网络性能,将本文的模型和 Gabbard 卷积神经网络用于模拟数据集的比较。在微调网络的参数和训练网络之后,展示了两种方法在数据集上的结果,并在测试集上绘制了接受者操作特性曲线(Receiver Operating Characteristic curve,简称 ROC 曲线),又称为感受性曲线(sensitivity curve),计算了 ROC 曲线下的面积。ROC 曲线与坐标轴围成的面积定义为  $f_{AUC}$  (Area Under Curve, AUC),分类器的  $f_{AUC}$  值等价于将随机选择的正样本排序在随机选择的负样本之前的概率<sup>[17]</sup>。 $f_{AUC}$  越大,说明该分类器分类效果越好。结果证明,优化后模型的  $f_{AUC}$  比 Gabbard 的卷积神经网络模型提高了 0.025 4 ~ 0.032 6。图 2 是  $snr = 6.0$  和  $snr = 8.0$  时,未改变振幅情况下,两种方法应用于相同数据集绘制的 ROC 曲线。其中曲线①②③④含义如下:① 优化模型,  $snr = 8.0$ ; ② 未优化模型,  $snr = 8.0$ ; ③ 优化模型,  $snr = 6.0$ ; ④ 未优化模型,  $snr = 6.0$ 。

$$f_{FAP} = \frac{f_{FP}}{f_{FP} + f_{TN}}$$

$$f_{TAP} = \frac{f_{TP}}{f_{TP} + f_{FN}}$$

其中:  $f_{TP}$   $f_{FN}$   $f_{FP}$   $f_{TN}$  分别是真阳性、假阴性、假阳性、真阴性的样本数量。

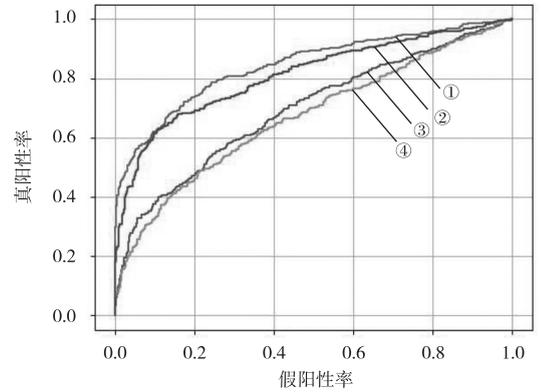


图 2 ROC 曲线( $f_{amp} = \sqrt{0.25 * T_{obs} * f_{psd}}$ )

Fig. 2 The ROC curve for  $f_{amp} = \sqrt{0.25 * T_{obs} * f_{psd}}$

另外,为了验证模型对于噪音改变的鲁棒性,改变噪音振幅,分别在  $snr = 6.0$  和  $snr = 8.0$  情况下,将两种方法应用于相同数据集绘制了 ROC 曲线,结果展示在图 3 和图 4 中。

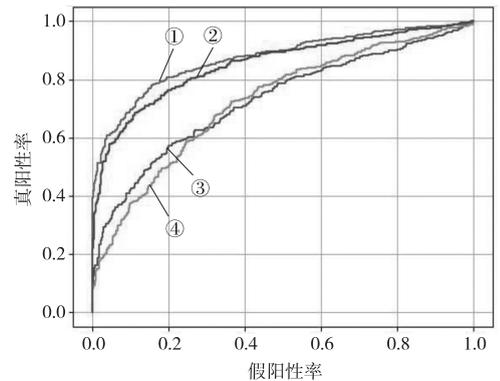


图 3 ROC 曲线( $f_{amp} = \sqrt{0.2 * T_{obs} * f_{psd}}$ )

Fig. 3 The ROC curve for  $f_{amp} = \sqrt{0.2 * T_{obs} * f_{psd}}$

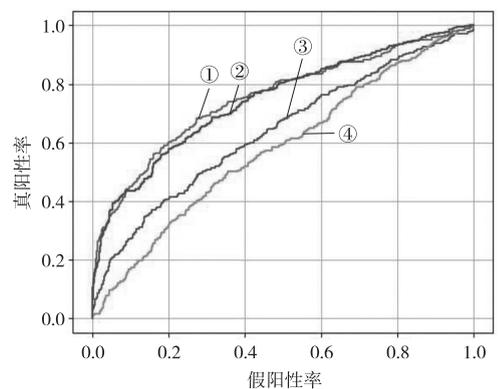


图 4 ROC 曲线( $f_{amp} = \sqrt{0.3 * T_{obs} * f_{psd}}$ )

Fig. 4 The ROC curve for  $f_{amp} = \sqrt{0.3 * T_{obs} * f_{psd}}$

另外,也计算了不同情况下 ROC 曲线的面积数值,结果展示在表 2 中。

表 2 不同情况下 ROC 曲线的面积数值

Table 2 Area values of ROC curves under different conditions

条 件	准确率/%		
snr 和振幅	未优化 网络	优化后 的网络	提升 数值
snr=6.0,未改变振幅	66.81	69.35	2.54
snr=8.0,未改变振幅	81.03	84.29	3.26
snr=6.0,第一次改变振幅	72.44	72.71	0.27
snr=8.0,第一次改变振幅	85.11	87.22	2.11
snr=6.0,第二次改变振幅	57.92	63.66	5.74
snr=8.0,第二次改变振幅	74.58	75.14	0.56

## 4 小结与展望

引力波的成功观测打开了引力波天文学观测宇宙的新窗口,目前在引力波的数据处理中,匹配滤波仍占据着主要位置。但是它需要准确完备的波形模板,使得无法检测到理论模板外的引力波,这对利用引力波探测发现未知事物的可能进行了限制。近年来,深度学习发展迅速,如果能够将深度学习应用到引力波的观测中去,对识别出模板以外的引力波信号是非常有帮助的。本篇文章基于模拟的数据集,将优化后的模型与 Gabbard 等人模型进行了比较,结果表明:前者在 ROC 曲线下的面积提高了 0.025 4 至 0.032 6。同时,也展示了改变噪音振幅后的效果。但是本文只讨论了最简单的情况,数据中有信号存在,或者数据中只有噪音。在论文里,也只展示了优化后的模型应用于 BBH 并合的结果,然而,这个方法也可以应用于其他的并合类型,比如双中子星信号。该方法也能扩展到其他良好建模的引力波目标,例如快速旋转的非轴对称中子星的连续发射。当然,我们也会寻找适合引力波信号的更有效的深度学习方法。在接下来的研究中,我们将致力于找到既能提升准确率,又能提升速度的适用于引力波信号的模型。

## 参考文献 (References):

- [1] EINSTEIN A. Nöherungsweise Integration der Feldgleichungen der Gravitation [M]. Berlin: Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 1916
- [2] ABBOTT B P, ABBOTT R, ABBOTT T D, et al. Observation of Gravitational Waves from a Binary Black Hole Merger [J]. Physical Review Letters, 2016, 116(6):061102
- [3] NAEYE R. Gravitational Wave Detection Heralds New Era of Science[N]. Sky and Telescope, 2016(2)
- [4] ABBOTT B P, ABBOTT R, ABBOTT T D, et al. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence [J]. Physical Review Letters, 2016, 116(24): 241103
- [5] MARBLESTONE A H, GREG W, KORDING K P. Toward an Integration of Deep Learning and Neuroscience [J]. Frontiers in Computational Neuroscience, 2016(10):00094
- [6] OLSHAUSEN B A. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images[J]. Nature, 1996, 381(6583): 607—609
- [7] BENGIO Y, LEE D H, BORNSCHEIN J, et al. Towards Biologically Plausible Deep Learning[J]. CoRR,2015(2): 04156
- [8] PEREIRA F, BURGESS C J C, BOTTOU L, et al. Regression Models and Life-Tables[J]. Royal Statistical Society, 1972, 34(2): 187—220
- [9] ZHANG R, ISOLA P, EFROS A A. Colorful Image Colorization[J]. CoRR,2016(3): 08511
- [10] KARPATY A, LI F F. Deep Visual-Semantic Alignments for Generating Image Descriptions [J]. CoRR, 2014(12): 2306
- [11] KONONENKO I. Machine Learning for Medical Diagnosis: History, State of the Art and Perspective[J]. Artificial Intelligence in Medicine, 2001, 23(1): 89—109
- [12] GEORGE D, HUERTA E. Deep Learning for Real-Time Gravitational Wave Detection and Parameters Estimation: Results with Advanced LIGO Data[J]. Physics Letters B, 2018, 778: 64—70

- [13] GEBHARD T, KILBERTUS N, PARASCANDOLO G, et al. ConvWave: Searching for Gravitational Waves with Fully Convolutional Neural Nets[C]//Workshop on Deep Learning for Physical Sciences (DLPS) at the 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017
- [14] GEORGE D, HUERTA E A. Deep Neural Networks to Enable Real-Time Multimessenger Astrophysics[J]. Phys Rev, 2018, D97: 044039
- [15] GEORGE D, SHEN H, HUERTA E A. Deep Transfer Learning: A New Deep Learning Glitch Classification Method for Advanced LIGO [J]. Phys Rev, 2017, D97: 101501
- [16] GABBARD H, WILLIAMS M, HAYES F, et al. Matching Matched Filtering with Deep Networks for Gravitational-Wave Astronomy [J]. Physical Review Letters, 2018, 120(14): 141103
- [17] FAWCETT T. An Introduction to ROC Analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861—874

## Application of Optimization Model Based on Maximum Pooling Layer Parameters to Gravitational Wave Astronomy

LUO Hua-mei

(Department of Mathematics, College of Mathematics, Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** In this paper, the convolutional neural network is applied to the identification of gravitational wave signals, the influence of the maximum pooling layer parameters on the classification ability of the model is studied, and the accuracy of the superparameters in the model to improve the classification of gravitational wave signals is adjusted. The optimized network structure and Gabbard's convolutional neural network were used for the same simulation data set, and the Receiver Operating Characteristic curve (ROC curve) was plotted on the test set and the ROC curve area was calculated. The results show that our model has increased the area under ROC by 0.0254 to 0.0326 compared with the unoptimized network. At the same time, we also changed the amplitude of the noise and applied the two methods to the new data set. The results also prove that the network is better and more robust after optimization.

**Key words:** gravitational wave astronomy; deep learning method; convolutional neural network

责任编辑:李翠薇

引用本文/Cite this paper:

罗华美. 基于最大池化层参数的优化模型在引力波天文学中的应用[J]. 重庆工商大学学报(自然科学版), 2020, 37(1): 59—64

LUO H M. Application of Optimization Model Based on Maximum Pooling Layer Parameters to Gravitational Wave Astronomy[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(1): 59—64