Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

《皇家天文学会月报》 **526,** 4148–4156 (2023) https://doi.org/10.1093/mnras/stad2915

在线先行 发表 2023 年 9 月29日

机器学习与修正引力中的结构形成

Jonathan C. Betts, [1★] [1] [2] [2] *Carsten van de Bruck、Christian Arnold* 和李宝玖 谢菲尔德大学数学与统计学院，霍恩斯菲尔德路，*S3 7RH*，谢菲尔德，英国 [2] 计算宇宙学研究所，物理系，杜伦大学，南路，杜伦 *DH1 3LE*，英国

接收于 2023 年 9 月20日。 收到于 2023 年 8 月26日； 原稿 日期为 2023 年5月26日

摘 要

在广义相对论中，基于球形坍缩模型的近似方法，例如 Press–Schechter理论及其扩展，能够预测给定体积内具有特定质量的天体数量。本文利用机器学习算法，检验这类近似方法是否适用于屏蔽修正引力理论。为此，我们使用来自 N 体模拟的数据训练随机森林分类器，研究 Λ冷暗物质（Λ CDM）以及屏蔽修正引力理论中的结构形成，尤其是 f (R) 和 nDGP 引力。训练模型区分最终条件下的结构归属与初始条件下密度场行为所形成的球形聚集体。我们考察从不同引力理论中学习结构形成规律的机器学习模型之间的差异，以及从 Λ CDM 中学习的模型。我们还利用不同强度的 f (R) 和 nDGP 引力理论的数据，检验 Λ CDM 模型的可推广性，从而评估扩展Press–Schechter理论中的球形坍缩方法对这类修正引力理论的可推广性。

关键词： 暗 能 ——暗 物质 ——大尺度 结构 的 宇宙。

1 引 言

宇宙中的结构形成源于引力不稳定性，而这种不稳定性由宇宙极早期产生的微小密度涨落所引发。结构形成自下而上地进行：小于星系的天体首先形成，随后它们再形成星系和星系团等更大尺度的结构。这一理论的一个重要要素是暗物质（DM）的存在，它使结构能够在小尺度上形成，尽管重子物质中存在耗散过程。结构形成图景中的一个关键组成部分是暗物质晕，它们可以被视为宇宙中各种结构的构建基块。这些是受引力束缚的天体，理解其形成与演化是宇宙学的重要目标。由于暗物质晕是高度非线性的天体，要研究暗物质晕的完整演化过程，必须借助 N 体模拟。尽管如此，仍有一些（半）解析近似方法，可用于估算结构进入非线性阶段时的红移，或计算晕质量函数。最早的研究由 Press与Schechter（1974年）完成，之后 Sheth与Tormen（1999年）、Jenkins等人（2001年）和 Tinker等人（2008）等学者对其进行了改进。

这些研究以广义相对论（GR）为基础，即认为结构形成完全由引力造成。然而，自从发现宇宙正在加速膨胀以来（Riess, Filippenko 等，1998；Perlmutter 等，1999年），宇宙学家便开始研究各种能够解释晚期加速膨胀的理论。在最简单的模型中，一种新的能量形式——暗能量（DE）——被认为是宇宙加速膨胀的原因。在广义相对论中引入宇宙学常数，是建立暗能量模型最简洁的方式。在广义相对论中加入宇宙学常数，再加上暗物质，便可得到一个非常成功的模型[Λ冷暗物质（Λ CDM）模型], ，它与许多现有观测数据相符（Peebles & Ratra 2003；Abbott 等，2019；Peebles 2020）。然而，宇宙学常数仍然是个谜，其数值和物理解释至今尚未得到说明。为了解释现有数据并挑战Λ CDM模型，研究者提出了诸如精质模型和修正引力理论等暗能量模型。尤其是在修正引力模型中，新的自由度与物质耦合，从而产生一种额外的力，影响宇宙中的结构形成。在这类理论中，前文提到的 Press–Schechter 等解析模型可能需要进行修正（Pace, Waizmann & Bartelmann 2010；von Braun-Bates & Devriendt 2018）。

本文将采用机器学习方法，深入了解修正引力理论中的结构形成。我们采用 Lucie-Smith 等（2018）提出的方法，将结构形成的演化转化为一个（二元）分类问题。在这一方法中，我们训练一个随机森林分类器，以学习暗物质密度场的初始条件（高红移时）与最终暗物质分布之间的关系。根据初始条件，判断一个粒子最终是否会归入某个暗晕。Lucie-Smith 等（2018）表明，随机森林算法非常有效且准确。正如该论文所示，随机森林分类器的预测结果与扩展Press–Schechter理论高度吻合。本文要探讨的问题是：在修正引力理论中，这种从初始条件到最终条件的映射是否会受到影响。具体而言，我们使用 f(R) 理论和 nDGP（Dvali–Gabadaze–Porrati）引力模拟的数据，并对这些数据进行 Lucie-Smith 等（2018）所述的分析。如果我们的结果表明，随着引力理论逐渐偏离广义相对论，预测效果显著变差，那么就需要将修正引力理论中额外自由度的信息纳入考量。

★ 电子邮件： jbetts3@sheffield.ac.uk

© 2023 作者 由 牛津 大学 出版社 代表 英国 皇家 天文学 学会 出版 。

本文结构安排如下：第 2 节简要概述本文考察的修正引力理论。第 3 节介绍模拟。第 4 节阐述我们的方法，第 5 节呈现研究结果。第 6 节总结研究发现。

2 修正引力理论

在本文中，我们考察两种修正引力理论，即一种 $f(R)$ 模型以及 nDGP 模型。我们将简要介绍这两种模型以铺垫后文。

2.1 $f(R)$ 引力

在这种广义相对论扩展中，Einstein–Hilbert 作用量中加入了一个附加项。该理论由以下作用量描述

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R + f(R)}{16\pi G_\mathrm{N}} + \mathcal{L}_\mathrm{M}\right]. \quad (1)$$

在该方程中，G N 是牛顿引力常数，$\mathcal{L}$ M 是物质拉格朗日量（包括标准模型场和冷暗物质）。函数 f (R) 描述了相对于广义相对论的修正。本文将重点讨论 Hu 和 Sawicki 提出的一种特定形式的 f (R)（Hu & Sawicki 2007）。其形式为

$$f(R) = -m^2\frac{c_1\left(-R/m^2\right)^n}{c_2\left(-R/m^2\right)^n + 1}, \quad (2)$$

其中，m 2 设定了暗能量（DE）的能标，其值为 $m^2 \equiv H_0^2\Omega_\mathrm{M}$ 。参数 c 1、c 2 和 n 是自由参数，选取这些参数是为了使该理论的背景演化与 Λ CDM 相匹配。除度规外，f (R) 理论还包含一个额外的标量自由度，即标量子 f R ≡ d f /d R。这个自由度与所有形式的物质普遍耦合，并传递一种新的力。为了使 f (R) 理论与实验相容，可以采用屏蔽机制；采用该机制的必要条件是，在高曲率区域中 | f R| ≪ 1 。该模型中的标量子始终处于有效势的极小值处，因此参数与标量子在当今的取值之间存在一定关系（Arnold、Leo 和 Li 2019b）

$$\frac{c_1}{c_2^2} = -\frac{1}{n}\left[3\left(1 + 4\frac{\Omega_\Lambda}{\Omega_\mathrm{M}}\right)\right] f_{R0}. \quad (3)$$

Hu–Sawicki 最初的理论表述利用太阳系内屏蔽效应的理论论证，给出了 | $_0$| f R $\lesssim 10^{-6}$ 的约束。附近宇宙的距离指标（Jain, Vikram & Sakstein 2012）、矮星系（Vikram et al. 2013）给出的其他约束进一步降低了这一数值。然而，满足这些约束的修正效应十分微弱，因此其对宇宙学尺度上结构形成的影响可以忽略，因为它们不允许影响大尺度结构的聚集（Liu et al. 2021）。尽管存在这些局限，能够使第五种力影响球形坍缩的有效模型，仍可作为检验大尺度引力的试验平台。近年来，对大质量暗晕聚集的约束（Cataneo et al. 2015）在这些有效模型中给出了 | f R$_0$| $< 10^{-4}$.79 的限制。

我们采用 N 体模拟数据，其 n = 1 | $_0$| = 和 f_R, 被选用。下文将 10^{-4} 10^{-5} 10^{-6} 这些模型分别记为 F4、F5 和 F6 。

2.2 nDGP 引力

DGP 模型源于额外维度理论（Dvali, Gabadadze & Porrati 2000 ）。在该模型中，我们的宇宙是一张嵌入于五维时空中的膜，该时空被称为体空间。作用量由下式给出。

$$S = \int_\mathrm{bulk}\mathrm{d}^5x\sqrt{-g^{(5)}}\frac{R^{(5)}}{16\pi G^{(5)}} + \int_\mathrm{brane}\mathrm{d}^4x\sqrt{-g}\frac{R}{16\pi G}. \quad (4)$$

在该方程中，g (5) 表示五维时空度规的行列式，g 表示膜上诱导度规的行列式。G 是牛顿引力常数，G (5) 是五维引力耦合。由于该理论中存在两种引力耦合，它们的比值定义了一个特征尺度，称为交叉尺度，其定义为该比值

$$r_\mathrm{c} = \frac{G^{(5)}}{2G},$$

在该尺度之上，作用量中的第一项占主导，并且预言会出现相对于广义相对论的偏差。该模型包含两个分支。本文考察的第一个分支是“正常”分支（nDGP），第二个分支是自加速分支。我们聚焦于前者，因为该分支不存在鬼不稳定性。在交叉尺度之上，引力变得更强；而在更小的尺度上，由于 Vainshtein 屏蔽，引力的行为类似于广义相对论。

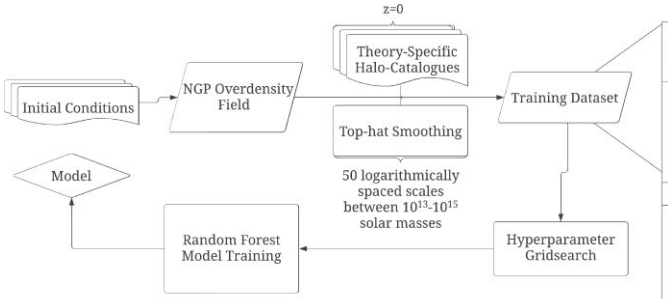
对于本文所采用的模型，我们用乘积 H 0 r c（即 r c 与当前哈勃半径 1/ H 0 的比值）表示相对于广义相对论的偏差。关于膜张力的马尔可夫链蒙特卡罗研究（Lombriser 等，2009 年）表明，交叉尺度远高于哈勃尺度，从而得到约束 H 0 r c > 3。对于本文考察的模型，我们采用 H 0 r c = 1 和 H 0 r c = 5，分别记为 N1 和 N5。

3 模拟

在本研究中，我们使用了 SHYBONE（Simulation HYdrodynamics BeyONd Einstein）模拟套件（Arnold、Leo 和 Li，2019a；Hern´andez-Aguayo 等，2021）。该套件包含一组高分辨率流体动力学星系形成模拟，涵盖广义相对论、f（R）引力和 nDGP 模型。这些模拟使用 AREPO 流体动力学模拟代码（Springel, 2010）执行，并对其进行了修改，以加入能够求解 f（R）和 nDGP 模型中额外标量场的修正引力求解器。在本分析中，仅使用 SHYBONE 的暗物质运行（不含重子）。

这些模拟涵盖六种引力模型——广义相对论、F6、F5、F4、N5 和 N1——并在大小为 62 h^{-1} Mpc 的立方体盒中演化 512 3 个暗物质粒子，时间范围从 $z = 129$ 到 $z = 0$。所有运行均从相同的初始条件开始，因为在 $z = 129$ 时修正引力对物质聚集的影响可以忽略不计。使用 SUBFIND（Springel 等，2001）晕搜寻器识别群目录，并采用 M 200c 作为晕质量定义。该定义包含被半径为 R 200c 的球体包围的所有粒子；在该球体内，平均密度在该天体的势能极小值周围为 $\bar{\rho} = 200 \times \rho_\mathrm{crit}$ ，对应于晕红移。

模拟采用普朗克2015宇宙学（Planck Col- laboration 2016 = 0 9667, = ），其宇 0 6774 $\Omega_\Lambda = 0.6911, \Omega_\mathrm{b} = 0\,0486$ $\Omega_\mathrm{m} = 0\,3089$, $\sigma_8 = 0\,8159$, 宙学参数为 n s . h . , , . 和 ., 其中 $h \equiv H_0/\left(100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}\right)$, n s 和 σ_8 分别为原初密度涨落的谱指数和当今 8 h^{-1} Mpc 范围内物质密度的均方根涨落，而 Ω_b 、 Ω_m 和 Ω_Λ 分别为当今以下各项的密度参数：



ParticleID	ML Feature 0	...	ML Feature 50	Label IN (1) or OUT (0) of halo at z=0
P_0	Value of smoothed overdensity field at first mass scale and position of particle P_0	Value of smoothed overdensity field at intermediate mass scale and position of particle P_0	Value of smoothed overdensity field at last mass scale and position of particle P_0	1
...	...	...	...	...
P_n	Value of smoothed overdensity field at first mass scale and position of particle P_n	Value of smoothed overdensity field at intermediate mass scale and position of particle P_n	Value of smoothed overdensity field at last mass scale and position of particle P_n	0

图1. 用于为广义相对论、三种 f（R）理论和两种 nDGP 理论生成模型的完整流程。每次模拟的初始条件用于生成一张 NGP密度对比度场图像。随后对该密度图像应用特征工程中的平滑处理，为每个粒子生成50个描述其密度环境的特征。接着，利用$z = 0$ 处的晕目录为这些特征添加标签，以进行监督训练。最后，通过网格搜索优化超参数，并使用所得数据集为每种理论分别训练一个随机森林分类器。

重子、（全部）物质和宇宙学常数。粒子质量分辨率为 $\simeq 1.5 \times 10^8 h^{-1}$ M ⊙。

4 机器学习方法

本文提出了以下机器学习流程设计，用于研究修正引力中的结构归属及其影响；该设计遵循 Lucie-Smith 等（2018）的方法。第3节中的 N 体模拟用于生成监督训练数据集，以训练能够根据初始密度行为对$z = 0$ 处粒子的结构归属进行分类的模型。随后可考察这些模型的性质，以研究以密度对比度为中心的结构形成描述（Press与Schechter，1974年）对 f（R）和 nDGP 屏蔽引力理论的可推广性。图1展示了该流程的高层次概览。

监督训练数据集由一组带有类别归属标签的对象组成，并为每个对象提供一组特征，这些特征旨在包含与类别归属相关的信息。在本例中，对象是粒子；根据 $z = 0$ 处的结构归属，为它们赋予 IN/OUT 标签。特征则由初始条件下以每个粒子的位置为中心、并在越来越大的非局域质量尺度上平滑后的密度对比度值构成。随后，可以考察基于此类数据训练的模型，以确定哪些特征与分类结果最相关，从而了解密度对比度中包含多少结构归属信息，以及这些信息对应哪个尺度。[1]

4.1 结构归属

目前有多种可用的晕识别算法，能够将 N 体模拟中的质量归并为目录，列出给定红移切片上的暗晕。为定义 IN 和 OUT 类别，我们采用$z = 0$ 处的 SUBFIND晕目录，具体如第 3 节所述。尽管这类晕搜寻器能够识别暗物质晕中的子结构，但由于问题具有二元性质，在本例中相关的只有晕归属，而非子结构归属。

粒子被归类为属于结构内的粒子，如果它们在引力作用下束缚于质量为 $M_{\text{暗晕}} \geq 1.8 \times 10^{12}$ M ⊙，而 OUT 如果它们完全不属于任何结构，或者位于质量低于该阈值的暗晕中。该值选取为一个典型尺度，用于将暗晕目录划分为两个二元类别；它可能取决于所考虑的 N 体模拟的分辨率。[2] 对于多类别而非二元的结构归属问题，也可以考虑更多阈值。引入引力束缚这一附加条件，既保留了球形坍缩形成暗晕的假设，也将处于无关丝状结构中的粒子归入 OUT 类。

需要强调的是，在我们的暗晕识别算法中，引力束缚的定义与广义相对论（GR）中的定义相同（即通过比较给定粒子的动能与其所属结构的牛顿势来判断），而不受修正引力模型的影响。我们发现，这一假设对结果的影响可以忽略不计。更详细的讨论见附录A。

4.2 特征工程

现在，必须将初始条件下粒子周围的局部密度行为转化为特征，以区分最终进入暗晕（IN）和未进入暗晕（OUT）的粒子。基础密度对比场采用最近网格点（NGP）方法生成。[3] 随后，对该场进行球对称平滑处理，以生成如下特征。

对于给定的平滑尺度，平滑后的密度对比度可以定义为 R

$$\delta(\mathbf{x}; R) = \int \delta(\mathbf{x}') W_{\text{top-hat}}(\mathbf{x} - \mathbf{x}'; R) d^3\mathbf{x}', \tag{5}$$

采用实空间顶帽窗口函数

$$W_{\text{top-hat}} = \begin{cases} \frac{3}{4\pi R^3} & \text{for } |\mathbf{x}| \leq R, \\ 0 & \text{for } |\mathbf{x}| > R. \end{cases} \tag{6}$$

给定 R 的卷积 (5) 通过傅里叶变换、在 k 空间中进行滤波，然后进行逆傅里叶变换来计算。给定 R 时，W top-hat (x , R) 对应于特征质量尺度 M smoothing $= \bar{\rho} V$ top-hat (R)，其中 V top-hat (R) = (4//3)πR ^{3} 。对于给定粒子，每个特征随后就成为初始条件中以该粒子为中心、选定平滑质量下的平滑密度对比度值。完整特征集针对一系列对数间隔的...

[1] 本工作使用 Python 编程框架。我们使用 PYNBODY 库（Pontzen 等，2013）和 AREPO 的 Python 集成（Arnold 等，2019a）解析模拟数据。我们使用 PYLIANS 的 Cython 库计算基础密度对比场并进行平滑卷积 5。我们使用适用于 Python 的 SK-LEARN 机器学习库进行模型训练、推理和分析。

[2] 暗晕质量函数截断是由于较小尺度上未解析的暗晕所导致的结构阈值的一个良好指标。

[3] 云中单元方法也被用于生成基础密度对比度场，但发现由于单元之间的边界，在较小尺度上进行平滑时会产生分辨率效应。

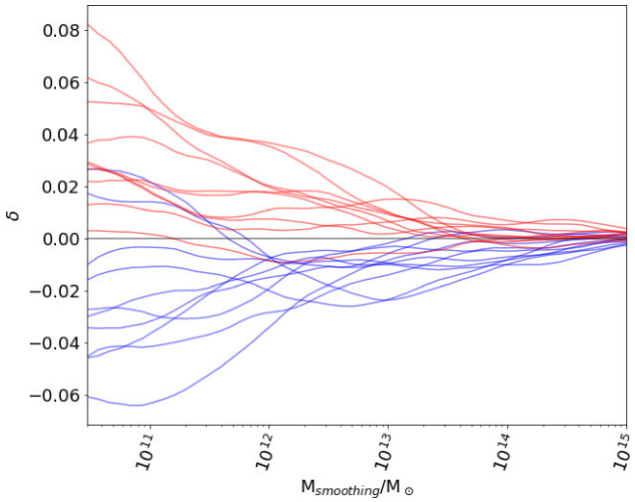


图 2. 广义相对论中 IN （红色）和 OUT （蓝色） 类别 中 示例 粒子 的 密度 轨迹。

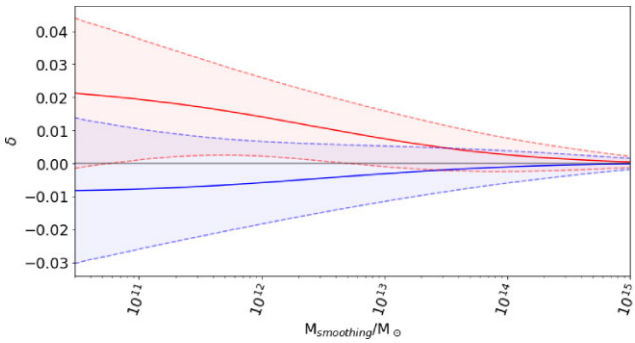


图 3. 在 包含 50 000 个 粒子 的 训练 数据 样本 中， GR 下 IN （红色） 和 OUT （蓝色） 的 平均 轨迹 。 阴影 区域 显示 每个 类别 中的1σ偏差 。

平滑质量的等间距取值范围下限受模拟分辨率限制，上限受模拟盒大小限制。在本研究中，3×10^{10} 平滑质量$_\odot < M < 10^{15}$ 平滑质量 ⊙。此前已有研究发现，在这一范围内取 50 个平滑质量值，即使用 50 个特征，足以区分类别（Lucie-Smith 等，2018）。本研究也得出了相同结论：对于广义相对论以及 F4、F5、F6、N1 和 N5，增加特征数量并未改善分类性能。对于给定粒子，其 50 个经过平滑的密度对比度特征可以可视化为随质量尺度增加而变化的密度轨迹。图 2 展示了广义相对论中 IN 和 OUT 类别的一组选定示例粒子轨迹，图 3 展示了各类别的平均轨迹。这说明，提供给分类器的特征可以用于区分类别。

4.3 模型 选择 与 训练

有许多机器学习（ML）算法可以学习分类。对于结构形成而言，各个密度对比度特征彼此相关。研究发现，随机森林分类器不受这一点影响（Lucie-Smith 等 2018 年），因此我们在本研究中继续使用它们，以确保其可审计性。随机森林是一种集成模型，它运用群体智慧原理，通过一组决策树对分类结果进行投票（Breiman 等 1984 年；Breiman 2001年）。决策树方法试图划分特征空间，使得能够根据各类别成员的特征值识别其所属类别。在非简单情形下，构建一棵完美的单一决策树在计算上难以实现（G´eron 2019 年），而由此得到的模型也无法很好地推广到新数据。因此，使用训练数据的随机子集训练的一「森林」决策树会对分类结果进行投票。这能减少预测中的偏差和方差，并提高模型的可推广性。

使用随机森林的另一个次要优势，是能够考察模型的特征重要性属性。模型训练完成后，可以衡量每个特征对分类决策的重要程度。具体而言，通过计算随机森林中使用该特征的各个节点的类别不纯度下降的平均值来实现这一点。本研究采用基尼不纯度（Gini impurity）（Farris 2010）。考虑森林中给定决策树的节点 m。将节点 m 处的数据记为 Q m，其中包含 n m 个样本。节点 m 的目标是将每个对象 y∈ Q m 划分为 k 个类别。在这一二分类情形下，k∈ [0, 1]。节点 m 处属于类别 k 的观测值比例为

$$p_{\mathrm{mk}} = \frac{1}{n_{\mathrm{m}}} \sum_{y \in Q_{\mathrm{m}}} I(y = k), \tag{7}$$

随后将其汇总为该 节点的 基尼不纯度 度量，

$$H_{\mathrm{gini}}(Q_{\mathrm{m}}) = \sum_{k} p_{\mathrm{mk}}(1 - p_{\mathrm{mk}}). \tag{8}$$

对于单棵决策树中的特征 i，N i 是使用特征 i 的节点集合，N t 是树中的节点总数。特征 i 的特征重要性定义为

$$F_{\mathrm{i}} \equiv \frac{\sum_{n \in N_{\mathrm{i}}} H_{\mathrm{gini}}(Q_{\mathrm{n}})}{\sum_{m \in N_{\mathrm{t}}} H_{\mathrm{gini}}(Q_{\mathrm{m}})}. \tag{9}$$

在随机森林中，会针对森林中的每棵树分别计算每个特征的 F i。随后对所有树的这些值取平均，得到每个特征在整个森林中的重要性统计量。最后，将森林的特征重要性归一化到 0 到 1 之间。在本研究中，每个特征表示以某个粒子为中心、给定质量尺度上的密度对比度行为，因此，特征重要性可理解为该质量尺度上的密度场行为对于结构形成过程的重要程度。

除了计算用于描述待分类对象的相关特征之外，训练此类机器学习模型时还有一些值得一提的细节。算法的计算参数包括森林中的树木数量、每棵树的最大深度，或一个节点在最终确定数据集划分之前必须包含的最少训练样本数。这些值称为超参数。超参数的选择会对最终模型的性能产生显著影响，例如可能导致过拟合。过拟合是指模型过度偏向其接收的训练数据，因此在验证时面对新的实例并进行分类，表现会很差。本文采用的投票架构等方法可以在一定程度上缓解这一问题，因为单棵树的偏差会得到平滑（Breiman 1996）。不过，也有一些交叉验证（Stone 1974）技术可在训练过程中使用，以帮助选择能够使模型良好学习并实现泛化的超参数。本文采用的第一种方法是对训练数据结构进行调整，称为 k 折交叉验证（Mosteller & Tukey 1968）。k 折交叉验证通过轮换训练数据集来进行

训练/验证过程。训练数据被划分为 k 个大小相等的子集，模型在其中 k – 1 个子集上进行训练，然后在剩余的一个子集上进行验证。随后以不同的 k – 1 个子集重复这一过程，依此类推，直到每个子集都被用于验证。第二种方法是在一组算法超参数上进行更高层次的迭代，称为网格搜索。网格搜索定义一个超参数空间，从中抽取样本，使用选定的超参数训练模型，然后根据验证数据集测试其性能。该过程迭代进行，并比较所有模型的验证得分，从而选出表现最佳、因而使用最佳超参数组合的模型。超参数空间的抽样既可以穷举进行，也可以随机进行；如果性能趋势显现出来，还可以重复抽样。

在本文中，我们将重点关注四种模型性能度量：受试者工作特征曲线下面积、精确率、召回率和 F1 分数（Bradley 1997；Olson & Delen 2008）。受试者工作特征曲线（有时也称为接收者操作特征，缩写为 ROC 或 ROC 曲线）是分类问题的主要评估方法。它汇总了给定模型在带有已知类别标签的测试数据集上生成的混淆矩阵输出。混淆矩阵根据给定的决策阈值，将模型的预测分为真正例、假正例、真负例和假负例。例如，考虑我们的两个示例类别 y 0 和 y 1。对于二元决策的分类矩阵，令 y 0 为负类，y 1 为正类。对于特征向量 x，模型将输出类别概率 P(y 0) 和 P(y 1)。如果决策阈值设为 0.5，那么当 P(y 1) > 0.5 时，具有属性 x 的对象将被分类为 y 1 类。如果该预测正确，则在决策阈值 0.5 下得到一个真正例；如果错误，则该预测为假正例。负类 y 0 的情况则相反。为生成 ROC 曲线，需要在 0 到 1 之间的一系列决策阈值下，计算正类预测中正确预测所占的比例（真正率），以及假正例占负例的比例（假正率）。随后以假正率为横轴、真正率为纵轴绘图，即得到 ROC 曲线。所得曲线下面积（AUC）是评估模型的有用度量，取值范围为 0 到 1，1 表示完美得分。

这种度量存在一些局限性，例如对类别不平衡缺乏稳健性。如果训练样本中 90% 的对象为 y 1，那么将所有对象都分类为 y 1 会使真正率飙升，即使对负类的分类效果很差也是如此。这会导致 AUC 分数虚高。幸运的是，将上文列出的其他度量与 AUC 分数结合起来，可以更全面地评估分类器的性能。精确率和召回率分别定义，但通常结合使用，以便从类别相对的角度了解模型。二者也都根据给定决策阈值下的混淆矩阵计算得出。精确率 P 定义为

$$P \equiv \frac{T_{\rm p}}{T_{\rm p} + F_{\rm p}}, \tag{10}$$

其中 $T_{\rm p}$ 是 真正例 的 数量 ，而 $F_{\rm p}$ 是 假正例 的 数量 。 召回率 R 定义 为

$$R \equiv \frac{T_{\rm p}}{T_{\rm p} + F_{\rm n}}, \tag{11}$$

其中，F n 是假阴性的数量。从语义上讲，精确率表示被判定为正例的结果中真实正例所占的比例，而召回率表示真实标签中的正例被模型正确识别的比例。这一区别微妙却十分重要。精确率高但召回率低的模型可能无法识别出所有实例。某个类别的成员，但它确实识别出的该类别成员很可能被正确分类。相反，召回率高但精确率低的模型能够捕获某个类别的大多数实例，但对这些预测是否正确则不太确定。根据问题的不同，这两种情况都可能是理想的。本研究使用的最终性能指标是 F 1 分数，其定义为精确率和召回率的调和平均数。

$$F_1 \equiv \frac{2PR}{P + R}. \tag{12}$$

F 1 分数的取值范围也是 0 到 1，通常作为更全面评估模型的一部分，用于解决类别不平衡问题。鉴于本研究要解决的问题是对 N 体尺度结构内部和外部的粒子进行分类，因此结构内部的粒子类别数量会少于结构外部的粒子类别数量。将结合 AUC 和 F 1 分数对超参数进行网格搜索评估。

针对广义相对论、F4、F5、F6、N1 和 N5，各生成一系列训练数据集，每个数据集均包含 50 000 个粒子 [4] 。第一个数据集将用于上述超参数的 k 折网格搜索。其余数据集将用于训练一系列模型，所用超参数则针对相应的引力机制选定。我们将汇总这些模型的特征重要性，以展示密度对比度在多大程度上决定粒子的结构归属，进而反映它对相应修正引力下解析结构形成的预测能力。

5 结 果

5.1 单个 模型 属性

图 4 展示了针对每种宇宙学构建的模型的特征重要性属性及其相应的平均 AUC 分数表现。对于我们的广义相对论模拟，密度对比度不仅能完整描述结构形成（正如预期），而且大部分信息都集中在 5.9×10^{11} M ⊙。随着 f (R) 修正偏离广义相对论的程度增强（即结构受修正影响的屏蔽效应减弱），结构分类最相关的尺度向内移至3.8×10^{11} M ⊙。对于 nDGP 引力也有类似情况：随着引力增强效应变强，相关信息也从范围 3.8×10^{11} $\mathrm{M}_{\odot} < M$ （平滑尺度为/ $\mathrm{M}_{\odot} < 5.9 \times 10^{11}$ M ⊙）向内移至 3.8×10^{11} M ⊙。这些与结构形成相关的信息变化似乎也有其物理依据，体现在各模拟的结构分布中。广义相对论模拟在 $z = 0$ 处包含∼ 115 000 个结构，而 F4 中为 ∼ 101 000 个，N1 中则为 ∼ 109 000 个。已知，在 f (R) 和 nDGP 模型中，增强第五种力会增加大质量暗晕的丰度，同时牺牲较小结构，并使结构之间的空洞更加空旷（Schmidt 等人 2008；Mitchell 等人 2021）。作者推测，这种行为导致了图 4 中特征重要性峰值的移动，因为密度信号的这种聚集使信息局部化。

图 5 分别 展示了 这种 效应 对 IN 类别 成员资格 在 广义相对论 和 F4 模拟中的影响。 由于 这些 初始 条件 经过 平滑 后用于生成 机器 学习 特征，要捕捉 最终 进入结构中的所有粒子，就需要更大的径向 （因此

[4] 更 大的 样本 规模 也 可以 使用 ，但 对 模型 性能 的 提升 可能 会 逐渐 减少。

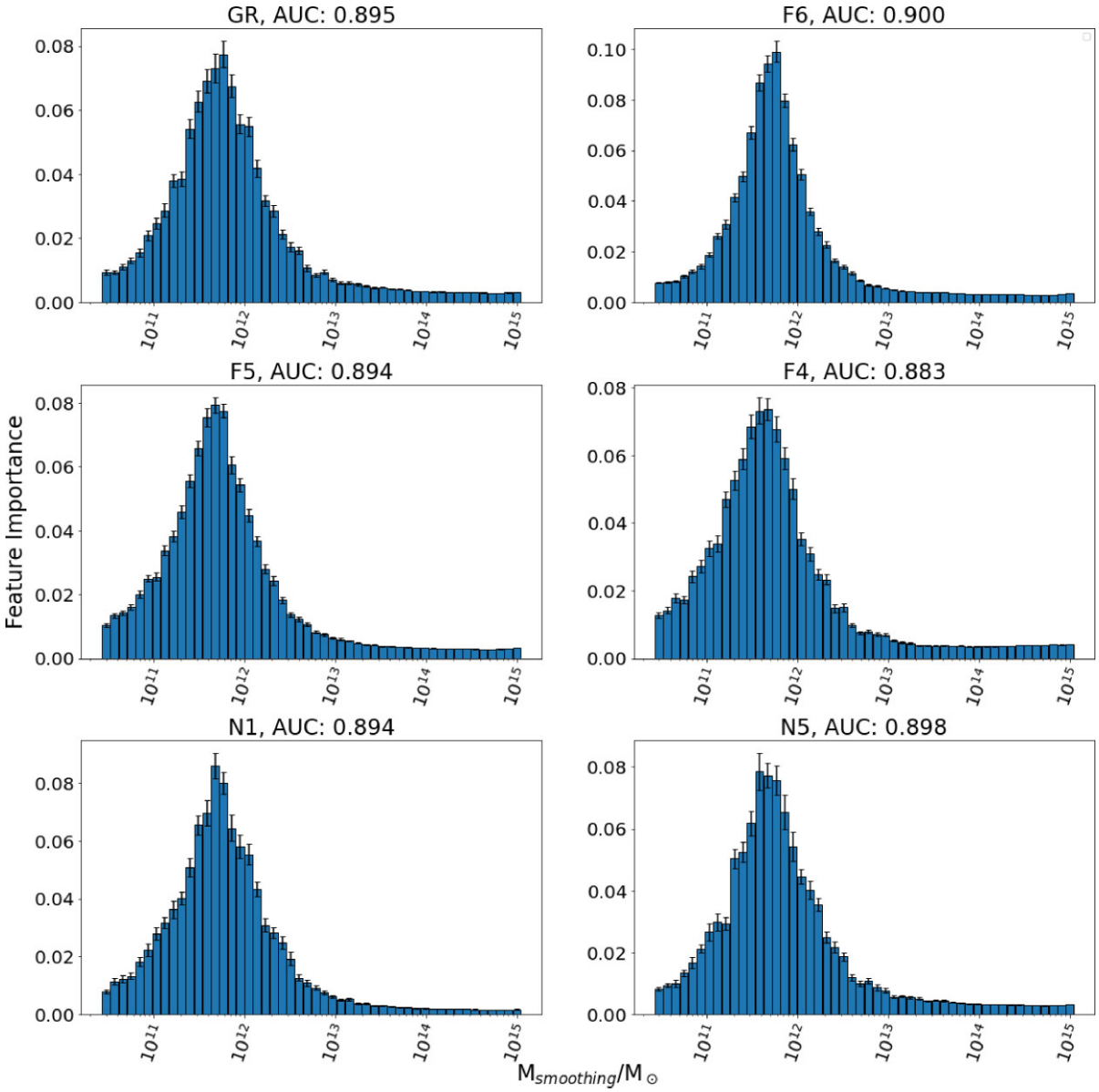


图 4. 在 GR、F6、F5、F4、N1 和 N5 上分别优化超参数并进行训练的一组模型的特征重要性。针对 GR、F6、F5、F4、N1 和 N5，各生成 50 个训练数据集。每个数据集均用于训练一个分类器，从而为每种宇宙学模型得到一组包含 50 个模型的样本。还给出了每组模型的平均 AUC 分数，作为其性能指标。

质量）尺度。在广义相对论中，必须采用更大的质量尺度，才能捕捉到最终进入结构中的所有粒子。然而，在 F4 中，由于上文所述的第五种力的影响，IN 类粒子分布得更为局部。较小的质量尺度就足以捕捉最终进入特定结构的所有粒子，这或许可以解释随着修正引力理论的强度增加，图 4 中特征重要性峰值的移动。

正如 AUC 分数所示，尽管理论之间的差异体现在它们判定结构归属时采用的不同质量尺度上，但对于各自在相应引力理论下训练的模型而言，密度对比度对结构形成的描述，与广义相对论模型一样完整。需要指出的是，对于由相同初始条件生成的不同宇宙学模型，特征重要性峰值的相对移动反映了力学变化，但其在每组模拟中的位置会因分辨率而异。

5.2 （cid:2）CDM 模型 的 可推广性

除了在单个模拟上训练的模型所表现出的特性外，广义相对论模型本身似乎也能很好地泛化到其他引力理论。图 6 展示了广义相对论模型应用于非广义相对论数据时，性能如何变化。针对每种修正引力，均生成了 50 组数据集，每组包含 50 000 个粒子。随后将每组数据集输入广义相对论模型，并与真实标签进行比较，从而构建广义相对论模型在面对其他引力理论时性能的概率分布函数（PDF）。

对于 f (R)，在广义相对论数据上训练的球形坍缩模型在三组模拟中都表现良好。然而，图 6 确实显示，该模型在 F4 数据上的表现不及其他情况。这表明，对于 F4引力而言，从初始密度场中球形聚合得到的信息，在确定最终状态下的结构归属方面作用较小。对于 F5 和 F6，该模型的表现

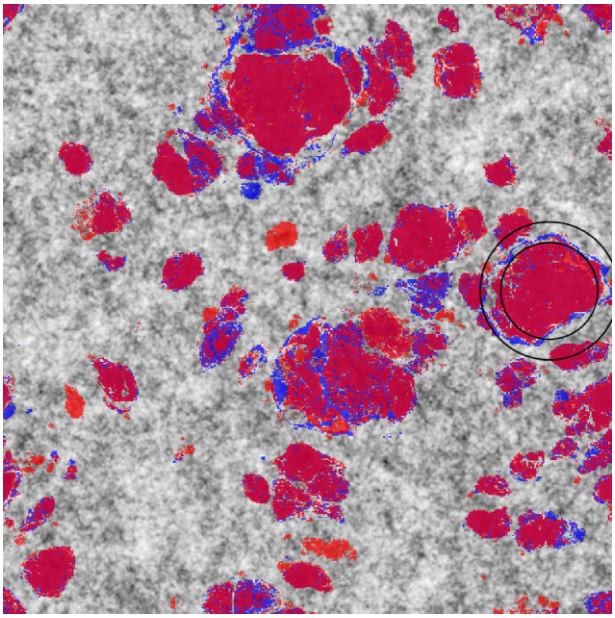


图 5. 初始密度场的 一个 切片 ， 其中 广义相对论 中的 IN 类 粒子 以 蓝色 表示 ， 而 F4 中的 IN 类 粒子 以 红色 表示 。黑色 圆圈 展示了 特征工程 过程中 一 组 平滑 半径 的 示例 。

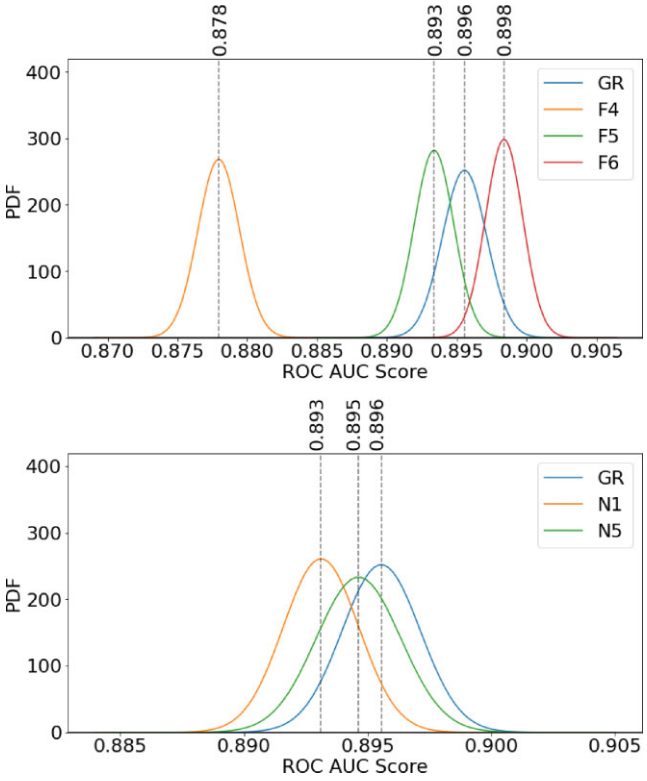


图 6. 将广义相对论分类器应用于由其他各类修正引力生成的数据集后，得到的 ROC AUC 性能指标概率密度函数。针对广义相对论、F6、F5、F4、N1 和 N5，各生成了 50 个测试数据集。随后，将这些批次的测试数据交由使用广义相对论数据训练的模型进行推断。广义相对论模型在各批次上的表现用于生成每种引力对应的概率密度函数。除 F4 外，广义相对论模型在所有引力下的表现似乎大致相同。

与其在广义相对论数据上的表现一致。位于 F4 和 F5 之间的一些中间模拟数据，将有助于进一步探究这一行为。然而，这些结果似乎确实表明，对于满足 $|f R_0| < 10^{-4}$ 的模型，在广义相对论中的 Press–Schechter 球形坍缩数据上训练的机器学习模型能够很好地泛化到 f (R) 引力。

对于 nDGP 引力，球形坍缩的 GR 模型对 N1 和 N5 都有良好的可推广性，尽管 N1 违反了 $H_0 r_c > 3$ 约束（Lombriser 等人，2009）。随着 $H_0 r_c$ 的强度从 N5 增加到 N1，开始出现一些分歧，但程度并不像 f (R) 中那么显著。违反 nDGP 引力的约束似乎对球形坍缩的可推广性影响较弱。然而，仍有一些用于约束 nDGP 引力的框架，其结果尚待正在进行和即将开展的天空巡天观测公布（Mitchell 等人，2021）。这些项目的结果或许能进一步揭示 nDGP 约束的敏感性以及这一结果，但从 GR 中的球形坍缩学习结构形成的机器学习模型，尽管受到当前约束的影响，显然仍具有良好的可推广性。

6 结 论

在本研究中，我们利用两种修正引力理论的 N 体模拟训练了一个随机森林分类器，以寻找高红移（$z = 99$）时密度场的初始条件与当前状态之间的关系。研究发现，这种方法在 Λ CDM 模拟中颇有价值（Lucie-Smith 等 2018）。具体而言，研究表明，在 Λ CDM 中，初始密度场包含足够的信息，可用于预测暗物质晕的形成。其准确度被证明与基于球形坍缩模型的（半）解析框架相当。本文的主要结果是，这些框架对于本文考察的屏蔽修正引力理论——即 f (R) 理论和 nDGP 模型——仍然具有很好的描述能力。

本研究构建的 机器学习 模型 还 捕捉 到了 其他 修正 引 力 行为 。首先 ， 它们 似乎 展现 了 其 影响 广义相对论与两种屏蔽引力理论之间聚集行为的变化。随着第五种力增强，小型星团受到抑制、空洞和较大结构得到增强，以及密度信号局域化等现象，都被机器学习模型捕捉到。这体现在模型对结构形成所作判断的特征重要性上。其次，广义相对论模型的表现与关于第五种力强度的既有约束相吻合。当屏蔽机制的参数被调节到超出已知会影响聚集行为的约束范围时，在 f (R) 情形下，广义相对论模型的表现出现了统计上显著的下降。在 nDGP 情形下，性能下降不那么明显，但仍然存在。机器学习模型揭示了这些行为，这也表明，球形坍缩对屏蔽修正引力的可推广性这一关键结果，远非统计上的偶然。特征重要性结果和广义相对论模型的性能结果都为后续研究提供了自然方向，例如检验标量场信息对结构归属的影响。本文所述方法也可能推广至其他具有可缩放额外作用力的修正引力理论，从而使结构形成成为检验新引力理论的天然试验平台。最后，我们希望本研究的结果能进一步证明，机器学习可作为开展有意义的宇宙学研究的工具。

致谢

JB 感谢 Luisa Lucie-Smith 就其流程 的转换 事宜 进行了 有益 的 交流 ， 并 将其 用于 本 工 作 。JB 还 想 要 特别 感谢

感谢 Adam Seed 和 James Coyle 就数据流程的开发提供有益的意见和反馈。我们也感谢 Adam Moss 对测试流程提出的深刻意见，并感谢 Hans Winther 提供测试模拟数据。CvdB 的部分经费来自 STFC 资助项目 ST/T001038/1 下的兰开斯特—曼彻斯特—谢菲尔德基础物理联盟。CA 和 BL 获得 ERC Starting Grant（ERC-StG-PUNCA-716532）资助。BL 还获得 STFC 综合资助 [ST/P000541/1, ST/T000244/1]。本研究使用了由宇宙计算研究所代表 STFC DiRAC HPC Facility 管理的 DiRAC@Durham 设施（www.dirac.ac.uk）。设备经费来自 BEIS 资本拨款，通过 STFC 资本资助 ST/K00042X/1、ST/P002293/1、ST/R002371/1 和 ST/S002502/1、杜伦大学以及 STFC 运营资助 ST/R000832/1 提供。DiRAC 是国家电子基础设施的一部分。

数据 可用性

本文 所 用 数据 由 CA 和 BL 在 杜伦 DiRAC 设施 生成， 相关 详情 见 上 文。 本 研究 中 生成 的 衍生 数据 可 在 合理 请求 下 向 通讯 作者 获取。

参考文献

Abbott T. M. C. 等，2019，《物理评论快报》，122，171301 Arnold C.，Leo M.，Li B.，2019a，《自然·天文学》，3，945 Arnold C.，Leo M.，Li B.，2019b，《物理评论D》，100，6，064044 Bradley A. P.，1997，《模式识别》，30，1145 Breiman L.，1996，《机器学习》，24，123 Breiman L.，2001，《机器学习》，45，5 Breiman L.，Friedman J.，Olshen R.，Stone C.，1984，《分类与回归树》（CART），Routledge，英国牛津郡，第40页 Cataneo M. 等，2015，《物理评论D》，92，044009 Dvali G.，Gabadadze G.，Porrati M.，2000，《物理快报B》，485，208 Farris F. A.，2010，《美国数学月刊》，117，851 G´eron A.，2019，《动手学机器学习：使用 Scikit-Learn、Keras 和 TensorFlow》，O'Reilly Media, Inc，英国萨里 Hern´andez-Aguayo C.，Arnold C.，Li B.，Baugh C. M.，2021，《皇家天文学会月报》，503，3867 Hu W.，Sawicki I.，2007，《物理评论D》，76，064004 Jain B.，Vikram V.，Sakstein J.，2012，《天体物理学杂志》，779，39 Jenkins A.，Frenk C. S.，White S. D. M.，Colberg J. M.，Cole S.，Evrard A. E.，Couchman H. M. P.，Yoshida N.，2001，《皇家天文学会月报》，321，372 Liu R.，Valogiannis G.，Battaglia N.，Bean R.，2021，《物理评论D》，104，103519 Lombriser L.，Hu W.，Fang W.，Seljak U.，2009，《物理评论D》，80，063536 Lucie-Smith L.，Peiris H. V.，Pontzen A.，Lochner M.，2018，《皇家天文学会月报》，479，3405 Mitchell M. A.，Hern´andez-Aguayo C.，Arnold C.，Li B.，2021，《皇家天文学会月报》，508，4140 Mosteller F.，Tukey J. W.，1968，载于 Lindzey G.、Aronson E. 编，《社会心理学手册》，第2卷。Addison-Wesley，美国波士顿 Olson D.，Delen D.，2008，《高级数据挖掘技术》，Springer，德国柏林、海德堡 Pace F.，Waizmann J.-C.，Bartelmann M.，2010，《皇家天文学会月报》，406，1865 Peebles P. J. E.，2020，《宇宙学的世纪：我们对宇宙的现代认知内幕》。普林斯顿大学出版社，普林斯顿 Peebles P. J. E.，Ratra B.，2003，《现代物理学评论》，75，559 Perlmutter S. 等，1999，《天体物理学杂志》，517，565 Planck Collaboration XIII，2016，《天文学与天体物理学》，594，A13 Pontzen A.，Ro ˇskar R.，Stinson G. S.，Woods R.，Reed D. M.，Coles J.，Quinn T. R.，2013，天体物理学源代码库，ascl:1305.002，pynbody：用于 Python 的天体物理模拟分析 Press W. H.，Schechter P.，1974，《天体物理学杂志》，187，425 Riess A. G. 等，1998，《天文学杂志》，116，1009 Schmidt F.，Lima M.，Oyaizu H.，Hu W.，2008，《物理评论D》，79，083518 Sheth R. K.，Tormen G.，1999，《皇家天文学会月报》，308，119 Springel V.，2010，《皇家天文学会月报》，401，791 Springel V.，White S. D. M.，Tormen G.，Kauffmann G.，2001，《皇家天文学会月报》，328，726 Stone M.，1974，《皇家统计学会杂志B辑：统计方法论》，36，111 Tinker J. L.，Kravtsov A. V.，Klypin A.，Abazajian K.，Warren M. S.，Yepes G.，Gottlober S.，Holz D. E.，2008，《天体物理学杂志》，688，709 Vikram V.，Cabre A.，Jain B.，VanderPlas J.，2013，《天体物理学杂志》，2013，8，20 von Braun-Bates F.，Devriendt J.，2018，《宇宙学与天体粒子物理学杂志》，2018，028

附录 A：修改引力对引力束缚的影响

关于第 4.1 节中结构定义所采用的引力束缚条件，合理的做法是质疑这一假设在修正引力理论背景下，尤其是在 f(R) 引力中的适用性。对于修正引力势，一个更合适的模型可能是 $U = (1 + \beta^2)\,U_G$，其中 U_G 是标准牛顿势，而 $\beta = \sqrt{1/6}$，它考虑了 f(R) 理论中出现的第五种力的影响。在最终条件下定义暗晕时，所有粒子均基于相关暗晕查找器的朋友-朋友方法被判定为束缚状态。随后会进行一个解束缚过程，在此过程中，将粒子的动能与相关结构的引力势进行比较。在判定引力解束缚时，使用 U 而非 U_G 可能会导致 f(R) 中有更多粒子被纳入结构。这种效应在 F4引力中最为显著，因为其中所有结构

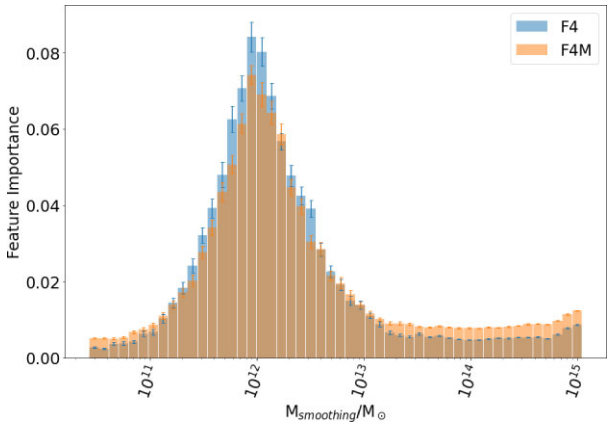


图 A1. F4 （蓝色） 和 F4M （橙色） 数据训练的 模型 的特征 重要性。

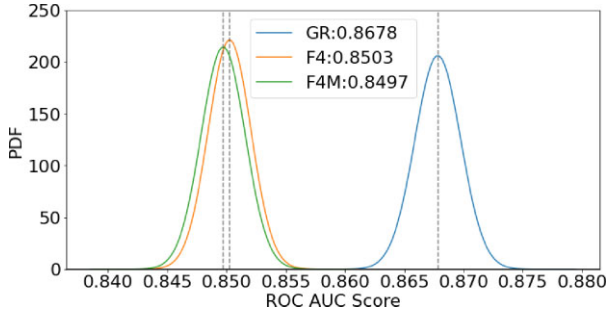


图 A2. 对由广义相对论 分类器 应用于由广义相对论 、F4 和 F4M 生成的数据集 所得到的ROC AUC 性能 度规 的 概率密度函数。

未经过筛选。在确保第 5.1 节和第 5.2 节的结果不受引力束缚中第五种力影响的前提下，我们使用原始 F4 数据以及一个采用 U 进行束缚计算的新 F4 暗晕目录（此处记为 F4M），重复了这些测试。下文中的图 A1 和图 A2 展示了 F4 和 F4M 模型的特征重要性，以及与第 5.2 节中类似的广义相对论（GR）模型性能测试结果。

图 A1 显示， F4M 特征 重要性 在 整个 质量 范围 内 分布 得更 广， 但 峰值 位于 相同 位置， 其 行为 与 F4 相同 。图 A2 显示， 该 广义相对论模型在 F4M 数据上的表现几乎与在 F4 数据上完全相同（差异为 0.06%）。这些结果表明，在用于引力解束缚的晕搜寻器中使用广义相对论势，不会影响本研究结果的可靠性；而这项机器学习研究观察到的行为，也并非源于对引力束缚的这一定义。

本文 由 作者 准备 的 一 个 T E X/L A T E X 文件 排版 而 成 。

© 2023 作者 由 牛津大学出版社 代表 英国皇家天文学会 出版