arXiv:2301.10358v1 [hep-ex] 25 Jan 2023

将 INFERNO 应用于基于 CMS 开放数据的顶夸克对产生截面测量

Lukas Layer[1,2], Tommaso Dorigo[1,3], 和 GilesStrong[1]

[1] 意大利国家核物理研究所，帕多瓦分部[2]
那不勒斯费德里科二世大学[3] 环球科学研究网络

摘要。近年来，人们通过神经网络构建非线性汇总统计量，并最小化以推断为目标的损失函数，开发出了新型推断技术。其中一种技术是 INFERNO（P. de Castro 和 T. Dorigo，《计算物理通讯》第244卷（2019年）170），研究表明，在含有干扰参数的置信区间估计问题中，它在玩具问题上的表现优于经典汇总统计量。为在实际应用中测试并评估该算法，我们利用 CMS 开放数据重现了 CMS 实验开展的一项完整且由系统误差主导的分析："在质子-质子碰撞中，喷注道的 $t\bar{t}$ 产生截面测量"（CMS 合作组，《欧洲物理学杂志 C》，2013年）。将 INFERNO 驱动的神经网络架构应用于这项分析，展示了它在真实大型强子对撞机分析中降低系统不确定性影响的潜力。这项工作还体现了利用开放数据重现大型强子对撞机分析的可行程度。

1 引言

过去十年间，机器学习驱动的分类和回归模型在高能物理（HEP）领域日益流行，常用于构建强大的汇总统计量以进行推断。尽管这些方法取得了成功，但它们依赖于标准交叉熵损失函数的最小化，以及针对学习任务的标准性能度量；当观测值依赖于干扰参数时，这两者都与推断目标不一致。因此，在推断过程中，干扰参数的存在会降低汇总统计量的统计效能。

近年来，一种名为 INFERNO 的新方法应运而生。INFERNO [1]，是 Inference-Aware Neural Optimization（推断感知神经优化）的缩写，它旨在构建针对特定分析目标进行优化的机器学习汇总统计量。作者提出了一种方法，通过随机梯度下降最小化由推断目标驱动的损失函数，从而构建非线性汇总统计量。该算法可用于直接最小化关注参数期望不确定度的近似值，并充分考虑所纳入干扰参数的影响。

在上述研究中，研究人员通过一个受典型截面测量条件启发的合成示例对该算法进行了测试。结果表明，使用基于 INFERNO 的汇总统计量得到的置信区间比使用二元分类得到的更窄，而且往往更接近使用真实模型似然进行推断时预期得到的置信区间。研究还表明，考虑的干扰参数越多，INFERNO 相较于二元分类的改进就越明显。尽管这些结果令人鼓舞，但使用合成数据集进行性能测试，仍未能明确该算法在面对完整 LHC 分析的复杂性时表现如何。在本研究中，我们填补了这一空白，介绍了一个可使用真实 LHC 数据通过 INFERNO 算法进行训练的框架。为对该框架进行基准测试，我们利用 CMS 开放数据复现了 CMS 实验中一项由系统误差主导的完整分析：“在 pp 碰撞中 $\sqrt{s} = 7$ TeV 下 $\tau+$ 喷注道的 $\mathrm{t\bar{t}}$ 产生截面测量”[2] 。本研究的代码已向公众发布 [3]。

–

2 推断感知神经优化

INFERNO算法 [1] 旨在直接最小化通过基于模拟的非参数合成似然得到的关注参数（POI）的期望方差。神经网络的参数通过随机梯度

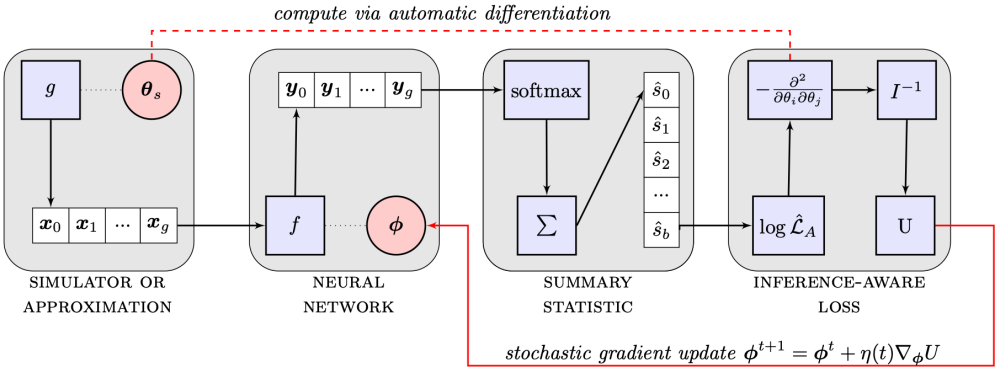


图1：INFERNO算法示意图。模拟器生成的批次数据经由神经网络处理，并构造出一个可微分的汇总统计量，以便计算POI的方差。随后，使用随机梯度下降（SGD）更新神经网络的参数。本图经许可转载自 [1]。

通过自动微分进行下降优化，其中损失函数考虑了统计模型的细节，尤其是干扰参数的影响。原始算法已使用 TensorFlow 1 [4] 实现。图 1 展示了 INFERNO算法的示意图。通过优化神经网络 f 的参数 $\boldsymbol{\phi}$，学习一种能够感知推断的汇总统计量，以降低每个 d 的维度

输入观测值 $\boldsymbol{x}$:

$$f(\boldsymbol{x};\boldsymbol{\phi}):\mathbb{R}^d\to\mathbb{R}^b\ . \tag{1}$$

使用参数为 $\boldsymbol{\theta}_s$ 的模拟器 g 生成模拟样本批次 $G_s=\{\boldsymbol{x}_0,\ldots,\boldsymbol{x}_g\}$，并用其训练神经网络。这里，$G_s$ 表示一个样本批次。神经网络最后一层的节点数决定了汇总统计量的维度 b。由于直方图不可微，原始算法使用 Softmax 函数作为神经网络输出的可微近似 y:

$$\hat{s}_i(G_s;\boldsymbol{\phi})=\sum_{\boldsymbol{x}}\frac{e^{f_i(\boldsymbol{x};\phi)/\tau}}{\sum_{j=0}^{b}e^{f_j(\boldsymbol{x};\boldsymbol{\phi})/\tau}} \tag{2}$$

其中，温度超参数 τ 调节算子的平滑程度。温度较低时 $\tau\to 0^+$，最大分量的概率趋近于 1，而其他分量的概率趋近于 0。借助这种近似，可以通过计算 Asimov 泊松计数似然[1] $\hat{\mathcal{L}}_A$，为每个批次构造汇总统计量:

$$\hat{\mathcal{L}}_A(\boldsymbol{\theta};\boldsymbol{\phi})=\prod_{i=0}^{b}\mathrm{Pois}\left(\hat{s}_i\left(G_s;\boldsymbol{\phi}\right)|\hat{s}_i\left(G_s;\boldsymbol{\phi}\right)\right)\ . \tag{3}$$

Asimov 似然的极大似然估计（MLE）就是用于生成模拟数据集 G_s 的参数向量 $\boldsymbol{\theta}_s$，即 $\mathrm{argmax}_{\boldsymbol{\theta}}\left(\hat{\mathcal{L}}_A(\boldsymbol{\theta};\boldsymbol{\phi})\right)=\boldsymbol{\theta}_s$。可通过改变混合模型的混合系数、对部分特征进行平移，或采用条件密度比重加权，来纳入关注参数和主要干扰参数的影响。下一节将介绍该算法的应用，并对此进行示例说明。随后，根据 Asimov 似然，通过自动微分计算 Fisher 信息矩阵:

$$\boldsymbol{I}(\boldsymbol{\theta})_{ij}=\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(-\log\hat{\mathcal{L}}_A(\boldsymbol{\theta};\boldsymbol{\phi})\right)\ . \tag{4}$$

如果 $\boldsymbol{\theta}$ 是 $\boldsymbol{\theta}$ 值的无偏估计量，则可以通过 Fisher 信息矩阵的逆估计协方差矩阵 ^i:

$$\mathrm{cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\geq I(\boldsymbol{\theta})^{-1}\ . \tag{5}$$

也可以纳入用于约束干扰参数的辅助测量，其特征由似然函数 $\left\{\mathcal{L}_C^0(\boldsymbol{\theta})\ldots,\mathcal{L}_C^c(\boldsymbol{\theta})\right\}$ 表示，并通过考虑增广似然函数 $\mathcal{L}'_A$:

$$\hat{\mathcal{L}}'_A(\boldsymbol{\theta};\boldsymbol{\phi})=\hat{\mathcal{L}}_A(\boldsymbol{\theta};\boldsymbol{\phi})\prod_{i=0}^{c}\mathcal{L}_C^i(\boldsymbol{\theta})\ . \tag{6}$$

用于优化神经网络参数 $\boldsymbol{\phi}$ 的损失函数可以是 $\boldsymbol{\theta}_s$处协方差矩阵的任意函数，具体取决于所研究的推断问题。对角元素 $I_{ii}^{-1}\left(\boldsymbol{\theta}_s\right)$ 对应于以下参数的期望方差:

[1]在统计学中，阿西莫夫数据的概念也称为“饱和模型”。

参数 θ_i。因此，如果目标是对某个参数 $\omega_0 = \theta_k$进行最优推断，一种可能的损失函数是：

$$U = I_{kk}^{-1}\left(\boldsymbol{\theta}_s\right) \tag{7}$$

这对应于 ω_0 置信区间宽度的近似预期值 .

3 在 τ+喷注道中复现顶夸克对 产生截面的测量

为了让由数千名成员组成的大型合作组相信新型分析方法的实用性和正确性，最好在尽可能贴近实际的数据集上测试并评估这些算法。因此，我们开发了一个能够复现完整 CMS 分析且向公众开放的数据集，以便对比各种新颖的推断方法。本研究复现了一项已发表的 CMS 分析，该分析基于 CMS 实验 Run 1 的真实遗留数据，这些数据可在欧洲核子研究中心开放数据门户 [5] 上获取。CMS 实验于 2013年 首次发表了这项复现的分析。该分析测量了质子-质子碰撞中 τ+喷注道的 $t\bar{t}$ 产生截面，碰撞能量为 $\sqrt{s} = 7$ TeV [2]。在.....的背景下 –

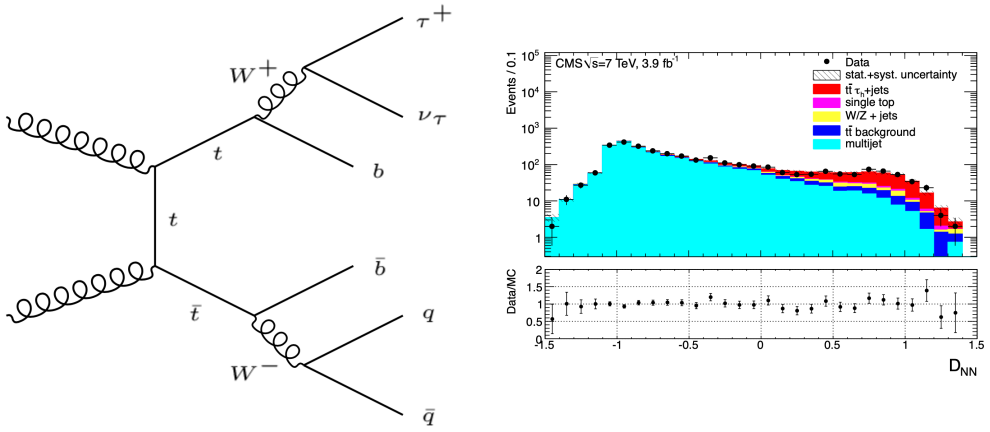


图 2：左图：顶夸克对衰变为 τ+喷注的费曼图。右图：原始分析中所用神经网络分类器的数据与模拟的一致性 [2]。

在对 INFERNO 算法进行测试和基准评估时，这项分析值得关注，因为其结果受系统不确定性的主导，而且它使用神经网络分类器构建汇总统计量，作为推断的输入。因此，这是一个 INFERNO 技术有望提高测量精度的应用案例。

数据集	运行范围	触发器	$\mathcal{L}\,(\mathrm{pb}^{-1})$
Run2011A [6]	$160431-165969$	QuadJet40_IsoPFTau40	357.5
Run2011A [6]	$165970-166782$	QuadJet45_IsoPFTau45	363.5
Run2011A [6]	$166783-171049$	四喷注40_隔离粒子流τ40	514.7
Run2011A [6] Run2011B [7]	$171050-178420$	QuadJet45 _IsoPFTau45	2930.2
总积分亮度			4165.9

表1：采用所选触发器的数据集、对应的运行编号和积分亮度。数据集的版本为12Oct2013-v1。

从物理学角度来看，$\mathrm{t\bar{t}}$ 产生截面的测量是对标准模型（SM）的一项重要检验，因为顶夸克质量很高，在许多标准模型扩展中都发挥着关键作用。在 $\tau+$ 喷注末态中直接测量 $\mathrm{t\bar{t}}$ 截面，有助于探究顶夸克可能存在的质量依赖或味依赖耦合。此外，带电希格斯玻色子可能导致质子—质子碰撞中的顶夸克对截面增强。如图 2 左图所示的费曼图所示，这一衰变的末态包含四个强子喷注，其中两个源自底夸克，此外还有一个强子衰变τ子和一个τ中微子。原始 CMS 分析测得的 $\sqrt{s} = 7$ TeV 处 $\mathrm{t\bar{t}}$ 产生截面为：$\sigma_{tt} = 152 \pm 12$（统计。）± 32（系统。）± 3（亮度。）pb，与标准模型预测一致。神经网络分类器的数据与模拟的一致性见图 2 右图。

3.1 数据集与触发器

数据与模拟样本的选择遵循原始 CMS 分析中的方案。采用多喷注触发器记录 $\mathrm{pp} \to \mathrm{t\bar{t}} \to \tau_{\mathrm{h}}$ +喷注事例。该触发器要求存在四个量能器喷注，其中一个被识别为 τ轻子。由于数据采集期间瞬时亮度上升导致事例率增加，因此开发了两个版本的触发器：QuadJet40_IsoPFTau40 和 QuadJet45_IsoPFTau45，后者提高了喷注和 τ轻子的 p_{T} 阈值。所用的多喷注触发器属于 CMS 在 Run 1 期间采集的 $\mathit{MultiJet}$ 主数据集。2011 RunA 和 RunB 的 $\mathit{MultiJet}$ 数据集已向公众开放，可在 CMS 开放数据数据库中获取。表1总结了各数据采集时段所选用的触发器及相应的积分亮度。使用 CMS 开放数据分析的数据集总积分亮度为 4.16 fb^{-1}。使用 CMS 开放数据集重新计算了触发效率，分别测量单个喷注和单个强子衰变τ轻子满足触发器要求的效率。按照原始

分析，在模拟中，通过将能量最高的三个中心喷流的效率与 τ候选粒子的触发效率相乘，对触发效率进行建模。

采用旧版CMS开放数据的蒙特卡罗（MC）模拟（称为“Summer11 模拟”）来估算信号效率以及电弱背景过程的贡献 [8–16]。$t\bar{t}$ 信号和背景事例以及 W/Z + 喷注事例由 madgraph 生成器 [17] 使用 CTEQ66 [18] 部分子分布函数集进行模拟。短寿命粒子的部分子簇射、碎裂、强子化和衰变（τ 轻子除外）使用 pythia [19]进行模拟。τ 轻子使用 tauola [20]进行衰变。单顶夸克事例由 powheg [21]与 pythia 和 tauola 接口进行模拟。采用的顶夸克质量值为 172.5 GeV，并假定 $t\bar{t}$ 截面的次次领头对数（NNLL）值为 164 ± 10 pb [22]。

3.2 事例筛选

事例筛选与原始分析 [2] 基本一致，要求至少存在四个粒子流喷注和一个粒子流 τ 轻子候选体。喷注使用 anti-k_T 聚类算法 [23] 重建，其距离参数为 $R = 0.5$。筛选后的事例必须至少包含四个赝快度为 $|\eta| < 2.4$的粒子流喷注。排除与轻子在 ΔR（喷注，轻子）范围内重叠的喷注 > 0.4 。为与触发器保持一致，要求三个喷注的横向动量达到 $p_T > 45$ GeV，第四个喷注达到 $p_T > 20$ GeV。喷流候选体必须与触发器中使用的喷注对象在 $\Delta R < 0.4$范围内匹配。筛选后的事例必须至少包含一个 b 标记喷流，该喷流使用推荐的组合次级顶点算法（CSV）中等工作点进行识别 [24]。横向缺失能量（MET）使用粒子流算法计算。通过施加横向缺失能量选择条件 MET > 20 GeV，来排除量子色动力学背景。

强子衰变 τ 轻子候选体使用强子加条带（HPS）算法进行重建 [25]。所选 τ 候选粒子必须满足隔离标准：要求在 τ 候选粒子周围半径为 $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.5$ 的隔离锥内重建的额外带电强子和光子的横向能量总和~~小于 1 GeV。~~ 此外，τ 候选粒子必须通过针对 μ 子和电子的鉴别器。若 τ 候选粒子的领头径迹被识别为 μ 子，则将其排除，以抑制 μ 子污染。另外，带电 τ 候选粒子不得被识别为最小电离粒子，因此要求与该 τ 候选粒子相关联的电磁量能器和强子量能器中的能量沉积总和与领头径迹动量之比大于 0.2。为符合触发条件，τ 候选粒子的横向动量必须满足 $p_T > 45$ GeV，且 τ 候选粒子必须与触发器中使用的 τ 对象在 $\Delta R < 0.4$ 范围内匹配。

为抑制将电子和 μ 子误识别为 τ 候选粒子，我们对是否存在隔离要求较宽松的电子和 μ 子实施否决。隔离要

求定义为 $I/p_{\mathrm{T}} < 0.15$，其中 I 是电磁量能器和强子量能器中的横向能量沉积之和，p_{T} 是半径为 $\Delta R = 0.3$ 的圆锥内径迹动量的标量值。

本分析中占主导地位的背景来自量子色动力学（QCD）过程产生的高喷注多重数（多喷注）事例，其中一个喷注被误识别为强子衰变τ轻子。多喷注本底采用数据驱动方法，通过反转 b 标记要求来估算，即 否决存在使用 CSV算法选定的 b 标记喷流。电弱过程产生的较小贡献则根据模拟事例估算，并按理论截面和总积分亮度进行归一化。

3.3 多变量分类器

训练一个神经网络分类器，以区分模拟的 $\mathrm{t\bar{t}}$ 信号事件与模拟的量子色动力学背景事例，并得到用作统计分析输入的最终判别变量。基于选定喷注、τ轻子和缺失横向能量的运动学变量进行特征工程，以构造能够更好地区分信号与背景的高层特征。计算了以下变量，并将其作为多变量分类器的输入：

变量	描述
H_T	所有选定喷注和 强子τ轻子候选体
非平面性	$A = \frac{3}{2}\lambda_1$ 其中 λ_1 是动量的最小特征值 张量 $M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i \lvert\vec{p}_i\rvert^2$，其中 i 遍历 喷注和 τ候选粒子的数量， α， $\beta = 1$， 2，3 表示三个 动量的空间分量。
球形度	$A = \frac{3}{2}(\lambda_1 + \lambda_2)$，其中 λ_1， λ_2 是该 动量张量 $M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i \lvert\vec{p}_i\rvert^2$ 的最小特征值，其中 i 遍历 喷注和 τ 候选粒子的数量。
$q \times \lvert\eta(\tau_{\mathrm{h}})\rvert$	τ轻子候选粒子的电荷乘以绝对值 赝快度
MET	横向缺失能量
$\Delta\phi(\tau_{\mathrm{h}},$ MET)	强子τ轻子候选体与 横向缺失能量方向之间的方位角
$M(\tau_{\mathrm{h}},$ 喷注)	选定喷注与强子τ轻子的 不变质量
$M_T(\tau_{\mathrm{h}},$ MET)	强子τ轻子候选体与横向 缺失能量的横向质量

非平面性和球形度描述了顶夸克衰变产物的球形拓扑结构，而 $q \times |\eta(\tau_{\mathrm{h}})|$ 变量则利用了 $\mathrm{t\bar{t}}$ 事例的电荷对称性，以区别于 W + 喷注事例。

顶夸克对信号事例总数为 43,570，量子色动力学背景事例总数为 11,176。数据样本分为包含 20,000 个顶夸克对信号事例和 5000 个量子色动力学背景事例的训练集；其余事例用于验证。训练神经网络分类器时使用 PyTorch 软件包。我们选择了具有两个隐藏层、采用 ReLU 激活函数并在最后一层使用 Sigmoid函数的前馈神经网络架构。损失函数采用标准二元交叉熵（bce），批量大小为 256。训练输入为上文所述的八个高层特征。这些特征经过重新缩放，使其均值为零、标准差为一。对于最终的 bce 模型，20

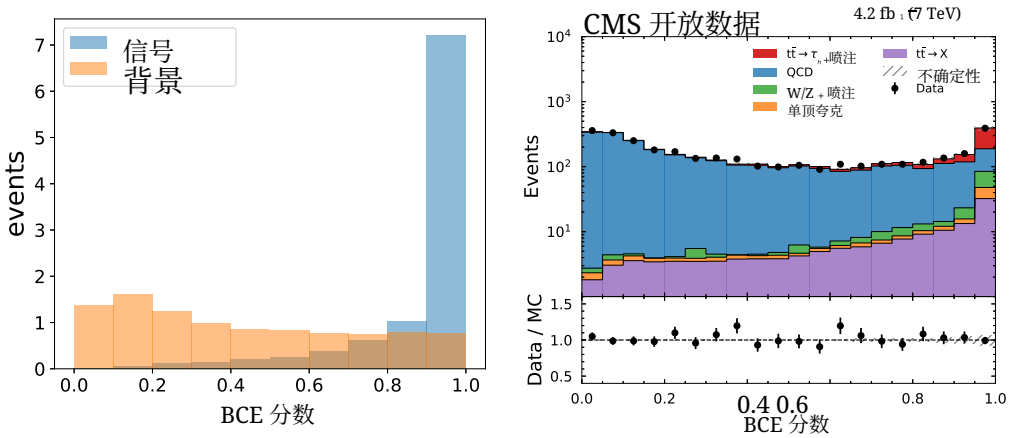


图 3：左图：使用二元交叉熵训练的神经网络的分类器分数。右图：神经网络分类器分数的数据与模拟的一致性。

每层的神经元数量和 0.001 的学习率是经过选择的。模型使用 ADAM 优化器训练 100 轮；将验证集上二元交叉熵损失最低时的权重保存下来。图 3 左图展示了验证集上神经网络模型的分箱分类器分数。将神经网络应用于 CMS 开放数据后，得到右图所示的数据与模拟的一致性。数据驱动的量子色动力学样本的归一化设为进行简单对数似然拟合后得到的最佳拟合值。可以观察到数据与模拟吻合良好。

3.4 系统不确定性

系统不确定性的主要来源是 Jet Energ 不确定性所导致的影响 y
标度（喷注能量标度）、PDF变化以及τ能量标度。系统不确定性的计算 c

遵循原始 CMS 分析中的描述，但采用旧版 CMS 开放数据推荐的详细方案。喷注能量标度（JES）对应的不确定性通过将喷注能量按一个标准差的不确定范围向上和向下移动来估算。对于喷注能量分辨率不确定性的系统效应，将喷注能量分布按一个标准差进行涂抹。按照 CMS 开放数据说明中推荐的方案，将这些校正传播至缺失横向能量测量。按照原始分析，将τ能量校正的不确定性通过将τ能量值上下移动± 3%来评估。强子衰变τ鉴别效率的不确定性估计为 6% [2]。对于在模拟事例中应用于 b、c 和轻味喷注的 b标记尺度因子，其不确定性通过将所用尺度因子的值按一个标准差的不确定范围移动来估算。对于多喷注数据样本采用的 b误标记重加权方法，按照原始分析的选择，不确定性估计为 5%。粒子流喷注和粒子流τ的触发效率对应的统计不确定性，通过对所有信号和背景过程中的喷注及τ效率施加 $\pm 1\sigma$ 统计变化，重新计算触发权重来确定。按照原始分析，为τ腿触发效率测量计入 ± 5%的不确定性，以考虑到用于估算τ触发效率的参考样本主要由误识别为τ轻子候选粒子的喷注组成。矩阵元与部分子簇射关联带来的不确定性估计为 3% [2]。PDF选择对信号接受度造成的不确定性，通过将与 CTEQ6 相关的 2×22参考部分子分布函数的变化以平方和开方的方式合并来估算。不同模拟背景过程的截面不确定性取自原始分析所用的理论计算。顶夸克质量带来的不确定性按测得截面的 3% 相对不确定性评估；对于 $\mathrm{t\bar{t}}$ 过程，重整化标度和因子化标度依赖性估计为 2% [2]。光度测量的不确定性估计为 2.2%。

3.5 截面测量

原始 CMS 分析通过对分类器输出分布进行负最大似然拟合，估算 QCD 多喷注背景和 $\mathrm{t\bar{t}}$ 信号占比。拟合前会从数据中扣除次要背景。系统不确定性通过改变模板来计算：针对相应系统来源，将模板变化 $\pm 1\sigma$ ，然后重复拟合。然而，为与现代大型强子对撞机分析保持一致，CMS 开放数据分析采用基于 cabinetry 软件包 [27] 和 CMS combine 工具 [28], 的剖面似然拟合进行推断，同时考虑所有系统变化。为检验信号强度 μ 的某个假设值，使用以下剖面似然比：

$$t_\mu = -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{8}$$

^^其中， $\hat{\boldsymbol{\theta}}(\mu)$表示给定强度参数$\mu$时， $\hat{\boldsymbol{\theta}}$ 的条件最大似然估计量；这意味着在给定 μ 的值时，它使 L 最大化。分母是最大化后的（无条件）似然函数，即，$\hat{\mu}$，而 $\hat{\boldsymbol{\theta}}$ 是它们的最大似然估计量。通过使用形变技术，将模板的形状变化纳入考虑。

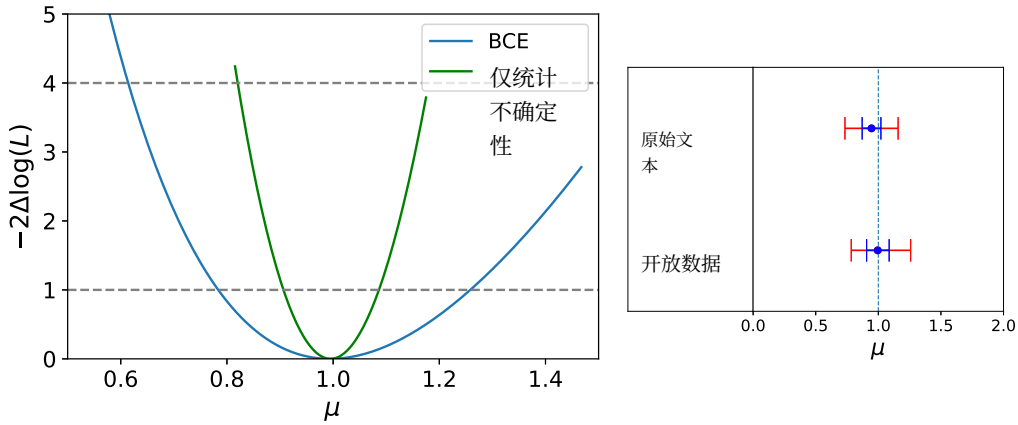


图4：左图：使用通过二元交叉熵训练的神经网络分类器得到的汇总统计量，对信号强度 μ 进行剖面似然扫描。右图：比较原始分析与使用 CMS 开放数据复现的分析所得的信号强度 μ 。蓝色条表示统计不确定性的大小。

对于阿西莫夫数据，计入系统不确定性后的信号强度为：

$$\mu = 1.00^{+0.24}_{-0.19}\ (\text{syst.}) \pm 0.08\ (\text{stat.}) \tag{9}$$

而在观测数据上，信号强度的计算结果为：

$$\mu = 0.99^{+0.25}_{-0.19}\ (\text{syst.}) \pm 0.09\ (\text{stat.}) \tag{10}$$

图 4 展示了对剖面似然的显式扫描。原始分析给出的数值为：

$$\mu = 0.94 \pm 0.08\ (\text{stat.}) \pm 0.19\ (\text{syst.}) \tag{11}$$

使用 CMS 开放数据进行的分析结果与原始结果非常吻合。CMS 开放数据分析中的系统不确定性略大，这可以用计算不确定性时采用了不同方案来解释。总之，基于 CMS 开放数据复现的分析是一项符合实际的测量，因此可用于在真实案例中研究 INFERNO算法。

4 INFERNO 的应用

下文将介绍如何将 INFERNO 算法应用于一个典型的高能物理问题，并根据对 t¯t

在 $\tau+$ 喷注道中测量的产生截面来评估其性能。该测量基于上一节介绍的 CMS 开放数据，在 $\sqrt{s} = 7$ TeV 的质子-质子碰撞中进行。

4.1 将 INFERNO 扩展至高能物理数据

参考文献 [1] 中介绍的 INFERNO 算法原始代码，是针对一个合成问题的结构开发的。然而，高能物理中通常遇到的系统不确定性具有特殊结构，因此需要对该算法进行调整，以适应类似高能物理的系统误差。值得注意的是，高能物理中的系统误差通常以 $\pm 1\sigma$ 变化的形式给出，这会导致变量 [29]的整体形状发生畸变。喷注能标不确定性就是一个例子，它会使事例中所有喷注能量朝同一方向偏移。可以通过改变蒙特卡罗模拟中的参数（例如能标），并重新计算偏移后的分布，对这类畸变进行建模。例如，将能标提高和降低一个标准差，并重新计算分布，便可分别得到能标位于负一倍标准差、中心值和正一倍标准差时，分布形状及归一化的三种度量。另一种常见情况是，系统不确定性以事例权重的形式给出，对应于 $+1\sigma$ 和 -1σ 变化。在这种情况下，可以对标称分布中的每个事例加权，以获得分布形状（及归一化）的三种度量。通过引入一个“形变”参数，可以将谱形的这三种度量转换为随能标因子变化的各个区间内的连续估计。这一方法首次在 PyTorch 版的 INFERNO [30], 中实现，该版本也复现了合成问题的结果。此外，我们还加入了另一种插值算法，该算法用于 CMS 标准拟合工具 combine [28], 中；还加入了处理非对称归一化不确定性的函数，以及一种可微的替代汇总统计量。通过采用插值技术并对系统变化进行适当的预处理，INFERNO 算法得以扩展，可处理任意数量的类似高能物理的系统误差。该算法的代码已发表于 [3]。

4.2 使用一个干扰参数进行训练和推断

为了量化将 INFERNO算法应用于第3节所述分析时的表现，我们开展了若干项研究，通过人为施加系统偏移来影响分类器的形状及归一化。研究得出结论：INFERNO有望减轻影响分类器形状的干扰参数所带来的影响。下文以使用最重要的干扰参数之一——喷注能标——进行训练为例对此展开研究。我们训练了一个由两层神经网络组成的INFERNO模型，每层包含60个神经元，学习率为0.001，温度 τ 为0.1，训练100轮。通过在 $\pm 1\sigma$ 种变化之间进行插值，将喷注能标变化的影响纳入 INFERNO训练。因此，INFERNO模型包含两个参数：预期信号事件数 s ，以及对应于喷注能标的干扰参数 θ^{JES}。将 s 的方差选作损失值。随后，将该 INFERNO模型与一个在相同数据集上训练的优化 BCE模型进行比较。近似协方差矩阵可以

在 BCE模型的训练过程中，通过对模型预测结果分箱，并使用 INFERNO算法计算 Fisher 信息矩阵的逆矩阵，即可得到该矩阵。图5展示了近似 2×2 协方差矩阵随训练的变化情况；该矩阵经过监测

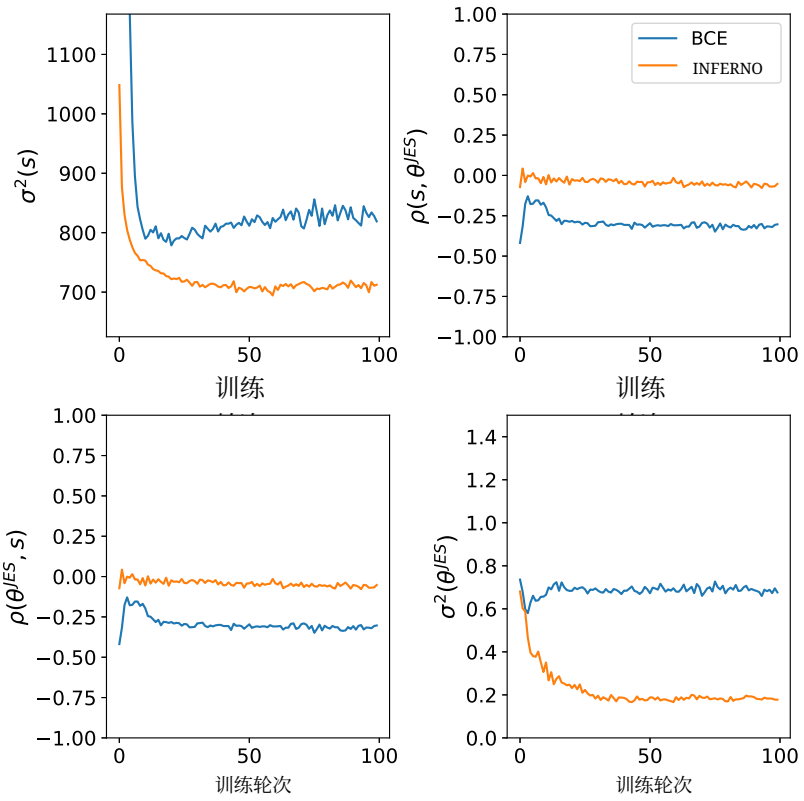


图5：训练过程中在验证集上计算的协方差矩阵的变化。对角线元素（左上和右下图）显示了预期信号事件数的方差 s 以及JES干扰参数的方差θ^{JES}。非对角线元素显示了相关系数 $\rho(s,\ \theta^{\mathrm{JES}})$。

在验证集上每个训练轮次之后。图中的对角线元素显示了预期信号事件数的方差 $\sigma^2(s)$ 以及干扰参数的方差 θ^{JES}，分别记为 $\sigma^2(\theta^{\mathrm{JES}})$。非对角线元素显示了 s 与 θ之间的相关系数 ρ ，其定义为两个变量的协方差除以它们的标准差之积：

$$\rho(\mu, \theta^{\mathrm{JES}}) = \frac{\mathrm{Cov}(\mu, \theta^{\mathrm{JES}})}{\sigma(\mu)\sigma(\theta^{\mathrm{JES}})} \ . \tag{12}$$

评估协方差矩阵的演变情况可见，与使用二元交叉熵训练的分类器相比，INFERNO模型中的关注参数 s 的方差收敛到更低的值。 ρ 与 s 之间的相关系数收敛

到更接近零的值。与二元交叉熵分类器相比，INFERNO模型中的 θ^{JES} 方差收敛到更低的值。这表明，INFERNO算法能够充分利用数据，使关注参数 s与干扰参数 θ^{JES}去相关，因此，与使用二元交叉熵训练的模型相比， s的方差更低。图6左图展示了INFERNO训练输出在验证集上的类别预测。

使用 inferno 训练得到的汇总统计量，随后用于 cabinetry [27] 软件包中实现的剖面似然拟合。拟合中仅纳入与喷注能标变化相对应的干扰参数θ ，以确保条件与 inferno 训练时相同。使用 MINOS 算法 [31]

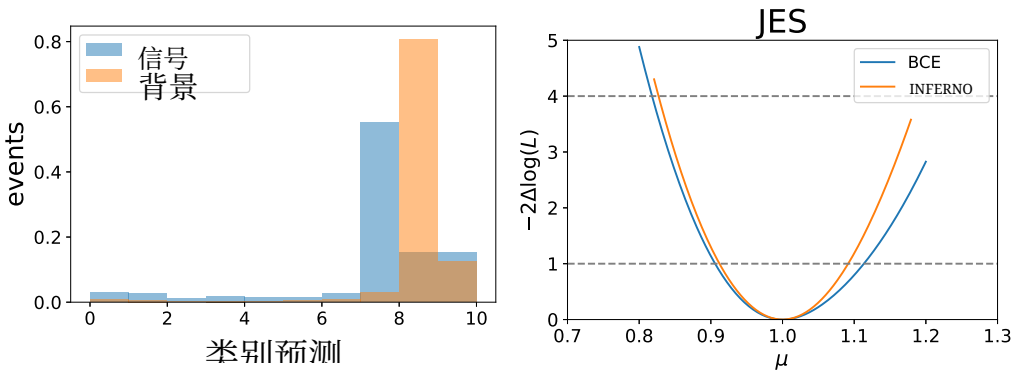


图 6：左图：使用喷注能量标度变化训练的 INFERNO 模型的类别预测。这里的类别并非例如信号和背景，而是神经网络学习得到的汇总统计量的各个分箱：不应期待这些分箱中信号和背景所占比例呈现任何特定顺序。右图：使用喷注能量标度变化训练的 INFERNO 模型的剖面似然扫描。

用于计算 68% 置信区间，而干扰参数的不确定性及相关系数则根据 HESSE 估计计算。基于 INFERNO 和二元交叉熵汇总统计量，在阿西莫夫数据上评估得到的信号强度 μ 的置信区间为：

$$\begin{aligned} \mu^{A}_{\mathrm{BCE}} &= 1.000^{+0.113}_{-0.094} \\ \mu^{A}_{\mathrm{INF}} &= 1.000^{+0.091}_{-0.087} \end{aligned} \quad , \tag{13}$$

根据 HESSE 估计得到的干扰参数 θ 的拟合后取值为：

$$\begin{aligned} \theta^{A}_{\mathrm{BCE}} &= 0.000 \pm 0.905 \\ \theta^{A}_{\mathrm{INF}} &= 0.000 \pm 0.602 \end{aligned} \tag{14}$$

，并已计算 ρ 与 μ 之间的相关系数，其结果为：

$$\begin{aligned} \rho^{A}_{\mathrm{BCE}} &= -0.42 \\ \rho^{A}_{\mathrm{INF}} &= -0.10 \ . \end{aligned} \tag{15}$$

图6右侧面板展示了针对参数 μ 的剖面似然扫描结果，分别对应 INFERNO 和 BCE模型。

上述结果表明，当存在*ShapeNorm*干扰参数时，与BCE模型相比，INFERNO汇总统计量可使信号强度 μ 的置信区间更窄。结果还表明，INFERNO训练中对协方差矩阵的估计能够很好地逼近通过拟合阿西莫夫数据得到的数值。正如训练过程中观察到的，INFERNO算法降低了 μ 与 θ之间的相关性。这表明，INFERNO带来的主要改进源于感兴趣参数与干扰参数之间相关性的降低。

4.3 使用所有干扰参数进行训练与推断

上一节所述的喷注能标变化研究已针对所有会影响分类器形状和归一化的相关系统不确定性重复进行。图7展示了 μ的MINOS不确定度、相应干扰参数 θ 的不确定度以及相关系数 ρ 。比较

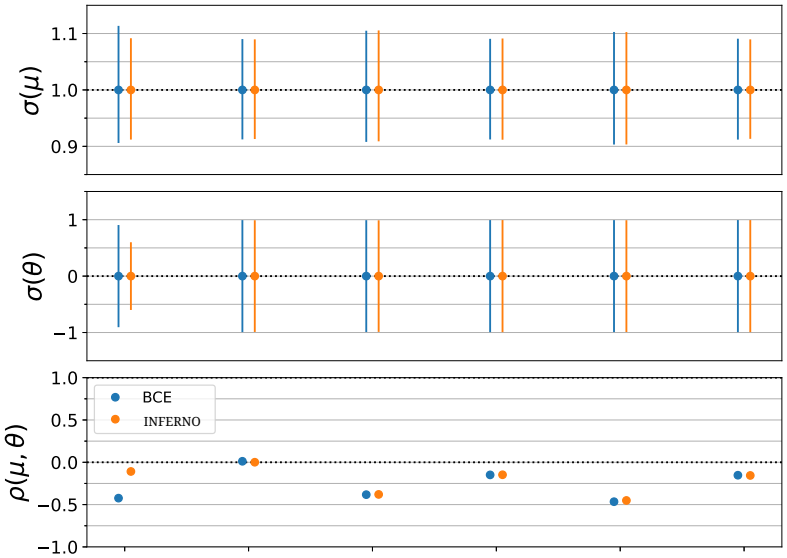


图7：对每个纳入考察的*ShapeNorm* 干扰参数，通过剖面似然拟合得到的 μ 置信区间（上图）、相应*ShapeNorm* 干扰参数 θ 的不确定度（中图）以及相关系数ρ 的比较。

tINFERNO与BCE模型的结果表明，改进主要体现在 y
对于 JES干扰参数，INFERNO 能够将其与

信号强度 μ 去相关。对于其他干扰参数，得到的置信区间和相关性非常相似。这些干扰参数对分类器形状的影响很小。因此，INFERNO 无法将这些参数与感兴趣参数去相关，相较于 BCE，使用 INFERNO 几乎没有带来任何改进。

对于完整的截面测量，我们训练了一个模型，将所有相关的形状归一化 干扰参数纳入考虑。超参数扫描验证了这样一种选择是合适的：由包含 60 个神经元的两层神经网络构成的 INFERNO 模型，学习率为 0.001，温度为 τ 0.1，训练 100 轮。

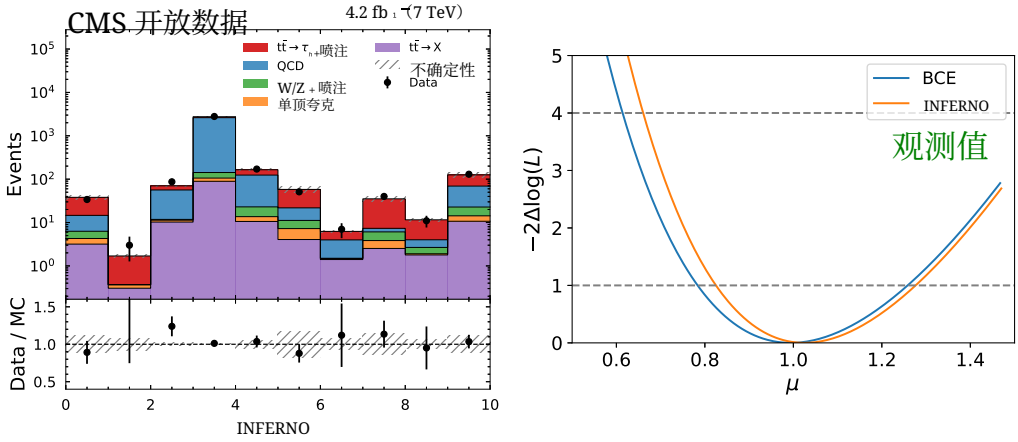


图 8：左图：使用所有相关形状归一化 干扰参数训练的 INFERNO 模型的数据与模拟的一致性。右图：比较使用包含所有相关干扰参数的 INFERNO 模型和 BCE 模型对观测数据进行拟合所得的剖面似然扫描。

INFERNO输出的数据与模拟的一致性见图8左侧面板。观测到数据与模拟之间具有良好的一致性。为与第3节中得到的信号强度 μ 结果进行比较，使用cabinetry进行了完整的剖面似然拟合，其中包含所有相关过程的所有相关干扰参数。此处假设次要背景的影响在INFERNO训练过程中可以忽略，因此得到的汇总统计量仍然是最优的。根据在阿西莫夫数据上评估的INFERNO汇总统计量，得到的 μ 置信区间为：

$$\mu^{A}_{\text{INF}} = 1.00^{+0.22}_{-0.17}\ (\text{syst.}) \pm 0.09\ (\text{stat.}) \tag{16}$$

根据在观测数据上评估的 μ ，得到的置信区间为：

$$\mu_{\text{INF}} = 1.02^{+0.23}_{-0.18}\ (\text{syst.}) \pm 0.09\ (\text{stat.})\ . \tag{17}$$

剖面似然扫描如图 8 右图所示。为进行比较，图中还包含了第 3 节中使用 BCE 训练的分类器所得到的似然扫描。

测得的信号强度中心值 μ 相当一致。与 BCE 汇总统计量相比，使用 INFERNO 汇总统计量可适度提高置信区间的精度。这一改进源于对喷注能标干扰参数的缓解。该结果表明，如果干扰参数会影响分类器的形状，那么 INFERNO 有望通过降低感兴趣参数与相关干扰参数之间的相关性，提高置信区间的精度。

5 结论

本研究使用 CMS 开放数据，完整复现了一项已发表的 CMS 分析，可为未来类似研究提供基准。借此，我们得以将 INFERNO算法 应用于一个真实的高能物理问题。基于复现分析，我们开展了一项研究，将使用 INFERNO 得到的汇总统计量进行推断，与通过使用二元交叉熵训练模型所得到的汇总统计量进行比较。研究发现，INFERNO 有望通过降低信号强度与干扰参数之间的相关性，减轻影响分类器形状的干扰参数所带来的影响。对于所研究的 CMS 开放数据分析，喷注能量标度干扰参数的影响有所降低；但由于分析中的大多数不确定性仅影响数据归一化，因此改进幅度有限。研究结果表明，对于受系统误差影响、且系统误差会扭曲分布形状的新型高能物理分析，使用 INFERNO 训练汇总统计量可带来显著收益。INFERNO 后续开发的一个潜在方向，是将其应用于一项新的 CMS 物理分析。该算法还可以扩展，以纳入多个背景过程和多个道。本论文所述研究的代码已向公众发布 [3]。



致谢

作者感谢 CMS 实验和 CERN 开放数据门户团队公开提供本分析所用的数据集。本研究由欧盟资助的欧洲创新培训网络 INSIGHTS 提供支持，该项目属于地平线 2020 研究与创新计划，征集编号为 H2020-MSCA-ITN-2017，资助协议编号为 765710。作者感谢 CloudVeneto 提供计算和存储设施。

参考文献

[1] Pablo de Castro 和 Tommaso Dorigo。Inferno：面向推断的神经网络优化。《计算机物理通讯》，244:170–179，2019年。[2] CMS 合作组。在质子-质子碰撞中测量 τ+喷注道的 $t\bar{t}$ 产生截面，碰撞能量为 $\sqrt{s} = 7$ TeV。《欧洲物理学杂志 *C*》，73(4)，2013年。[3] Lukas Layer、Tommaso Dorigo 和 Giles C. Strong。llayer/cmsopen：面向 CMS 开放数据的 Inferno，2022年2月。[4] Martín Abadi 等。TensorFlow：用于异构系统的大规模机器学习，2015年。软件可从 tensorflow.org 获取。[5] Kati Lassila-Perini、Clemens Lange、Edgar Carrera Jarrin 和 Matthew Bellis。在研究中使用 CMS 开放数据——挑战与方向。《*EPJ* 会议网络》，251:01004，2021年。[6] CMS 合作组（2016年）。来自 2011 年 RunA 的 AOD 格式 MultiJet 主数据集（/multijet/run2011a-12oct2013-v1/aod）。欧洲核子研究中心开放数据门户。

[7] CMS 合作组（2020年）。来自 2011 年 RunB 的 AOD 格式 MultiJet 主数据集（/multijet/run2011b-12oct2013-v1/aod）。欧洲核子研究中心开放数据门户。[8] CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 ttjets_tunez2_7tev-madgraph-TAUOLA。欧洲核子研究中心开放数据门户。[9] CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 wjetstolnu_tunez2_7tev-madgraph- TAUOLA。欧洲核子研究中心开放数据门户。[10]CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 dyjetstoll_tunez2_m-50_7tev- madgraph-TAUOLA。欧洲核子研究中心开放数据门户。[11] CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 t_tunez2_t-channel_7tev-powheg-TAUOLA。欧洲核子研究中心开放数据门户。[12]CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 tbar_tunez2_t-channel_7tev- powheg-TAUOLA。欧洲核子研究中心开放数据门户。[13]CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 t_tunez2_tw-channel-dr_7tev- powheg-TAUOLA。欧洲核子研究中心开放数据门户。[14] CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 tbar_tunez2_tw-channel-dr_7tev- powheg-TAUOLA。欧洲核子研究中心开放数据门户。[15] CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 t_tunez2_s-channel_7tev-powheg- TAUOLA。欧洲核子研究中心开放数据门户。[16]CMS 合作组（2016年）。用于 2011 年碰撞数据（标准模型全包）的 AODSIM 格式模拟数据集 tbar_tunez2_s-channel_7tev-powheg-TAUOLA。欧洲核子研究中心开放数据门户。[17] Johan Alwall、Michel Herquet、Fabio Maltoni、Olivier Mattelaer 和 Tim Stelzer。MadGraph 5：更进一步。《高能物理学杂志》，2011(6)，2011年。[18] Hung-Liang Lai 等。用于对撞机物理的新部分子分布。Physical Review D，82(7)，2010年。[19] Torbjörn Sjöstrand 等。PYTHIA 8.2 简介。《计算机物理通讯》，191:159–177，2015年。[20] Nadia Davidson 等。TAUOLA 的通用接口：技术与物理学文档。《计算机物理通讯》，183(3):821–843，2012年。[21] Simone Alioli、Paolo Nason、Carlo Oleari 和 Emanuele Re。在簇射蒙特卡罗程序中实现 NLO 计算的通用框架：POWHEG BOX。《高能物理学杂志》，2010(6)，2010年。

[22] Nikolaos Kidonakis. 顶夸克截面和横向动量分布的次次领头阶软胶子修正。*Physical Review D*，82(11)，2010。[23] Matteo Cacciari、Gavin P. Salam 和 Gregory Soyez。反k_t 喷注聚类算法。*JHEP*，04:063，2008。[24]CMS 合作组。使用 CMS 实验识别底夸克喷注。8(04):P04013–P04013，2013年4月。[25] CMS 合作组。重建并识别衰变为强子的 τ 轻子，以及在 CMS 上的 ν_τ 。11(01):P01019–P01019，2016年1月。[26] Adam Paszke 等。PyTorch：一种命令式风格的高性能深度学习库。载于 H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alché-Buc、E. Fox 和 R. Garnett（编），神经信息处理系统进展第*32*卷，第 8024–8035 页。Curran Associates, Inc.，2019。[27] Kyle Cranmer 和 Alexander Held。使用 cabinetry 构建和调控分箱模板拟合。*EPJ Web Conf.*，251:03067，2021。[28] 2011 年夏季大型强子对撞机希格斯玻色子搜索组合流程。技术报告，欧洲核子研究中心，日内瓦，2011 年。[29] John S. Conway。将干扰参数纳入多源谱的似然函数，2011 年。[30] Giles C. Strong。pytorch_inferno，2021年。[31] Bruce A. Murtagh 和 Michael A. Saunders。大规模线性约束优化。*Mathematical Programming*，14(1):41–72，1978。