基于无监督特征的时间序列外部回归算法

David Guijo-Rubio* ，Matthew Middlehurst、Guilherme Arcencio、Diego Furtado Silva、Anthony Bagna ll

arXiv:2305.01429v1 [cs.LG] 2 May 2023

摘要—**时间序列外部回归（TSER）是指利用一组训练时间序列，构建一个用于预测连续响应变量的模型，而该变量与回归变量序列没有直接关系。用于比较算法的 TSER 档案库于 2022年 发布，包含 19 个问题。我们将该档案库扩充至 63 个问题，并复现此前对基线算法的比较。随后，我们将比较范围扩展到更多标准回归器，以及此前研究中使用的最新版本 TSER 模型。结果表明，此前评估的回归器均无法胜过对标准分类器旋转森林进行回归改造后得到的模型。我们基于时间序列分类领域的相关研究，提出两种新的 TSER 算法。FreshPRINCE 是一种流水线估计器，先将数据变换为种类丰富的汇总特征，再使用旋转森林回归器。DrCIF 是一种树集成，通过随机区间上的汇总统计量创建特征。研究表明，这两种算法与 InceptionTime 的性能均显著优于测试的另外 18 种回归器。更重要的是，DrCIF 和 FreshPRINCE 这两种方法是仅有的显著优于标准旋转森林回归器的模型。**

索引词——**时间序列、外部回归、区间集成。**

I. 引言

时间序列分析是机器学习和数据挖掘研究中的热门主题。在过去十年中，该领域发表了数千篇研究论文。针对各种应用领域中的不同任务，人们提出了多种算法。这一发展的主要原因是，随着时间推移存储数据的能力不断增强，廉价传感器技术也已普及到大多数科学领域。例如，太阳能电池板依靠传感器来最大限度地发挥其潜力（例如，调整太阳能电池板的倾斜角度，使阳光直射其表面），而医院则会定期记录并存储患者数据，如生命体征。这些海量数据为数据挖掘提供了巨大的潜力。

时间序列机器学习/分析中研究最多的两个任务是分类 [1]–[3]和预测 [4]。

*：通讯作者。

D. Guijo-Rubio 任职于西班牙科尔多瓦 14071 科尔多瓦大学计算机科学与数值分析系，以及英国诺福克诺里奇东安格利亚大学计算机科学学院。电子邮箱：dguijo@uco.es。M. Middlehurst 和 A. Bagnall 任职于英国诺福克诺里奇东安格利亚大学计算机科学学院。电子邮箱：{m.middlehurst,ajb}@uea.ac.uk。G. Arcencio 任职于巴西圣卡洛斯联邦大学计算机系。电子邮箱：garcencio@estudante.ufscar.br。D. Furtado Silva 任职于巴西圣保罗大学数学与计算机科学研究所。电子邮箱：diegof- silva@usp.br。稿件收稿日期：X；修订日期：X。

时间序列分类（TSC）是指基于时间序列（可能是多变量的）为类别型目标变量构建预测模型。时间序列分类与标准分类的区别在于，判别特征通常体现为序列的形状或自相关。预测是根据过去的观测值预测（通常为数值型的）值。通常采用基于模型的算法（例如自回归或指数平滑方法）进行预测，也可以通过滑动窗口将预测问题转化为回归问题，再使用深度学习或 XGBoost 等全局模型。

Tan等人在 [5] 中正式定义了一类相关但不同的时间序列回归问题：时间序列**外部**回归（TSER）。TSER 并非源自预测问题，而是指基于时间序列构建预测模型，以预测一个与训练输入序列不同的实值变量。例如，图 1 展示了可用于估算钾浓度的土壤光谱图。真实值需要通过耗时且昂贵的实验室实验获得。光谱图（有序数据序列，在此视为时间序列）获取成本低，并且可以在任何环境中采集数据。若能构建一个准确的光谱图到浓度的回归器，土地与作物管理将更加高效。

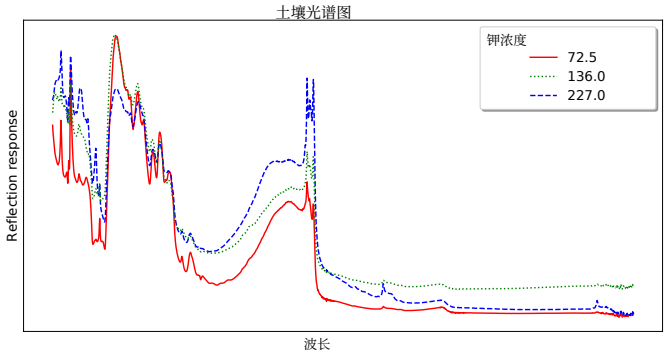


图 1. 钾浓度数据集中用于预测钾浓度的土壤光谱图示例。

TSER 与时间序列分类的关系，类似于传统回归与分类的关系：两者唯一的区别在于，目标变量是实值而非类别值。之所以需要区分外部，是因为预测领域文献中“时间序列回归”一词普遍用于指通过滑动窗口将预测问题转化为回归问题。

首项时间序列回归（TSER）基准测试工作 [5]介绍了一个包含19个时间序列回归问题的档案库，其中包括4个单变量数据集和15个多变量数据集。他们进行了一项实验

比较了13种算法在这些数据上的表现。从时间序列分类（TSC）文献中改编的两种算法——随机卷积核变换（ROCKET） [6]和深度学习器InceptionTime [7], ——排名最高。然而，表现最好的十种算法在均方根误差（RMSE）方面没有显著差异，这可能是因为数据集数量相对较少，且所采用的多重检验调整方法较为保守。 [5] 的摘要指出“我们表明，该领域仍需开展大量研究，以提高机器学习（*ML*）模型的准确率 *[在时间序列回归（TSER*）中的表现*]*”。尽管这篇论文广受关注，也明确指出了开展创新研究的迫切需求，但在应对这一挑战方面几乎没有进展。我们响应了这一号召，开发并评估了一系列时间序列回归算法。我们提出了新的算法，其表现显著优于 [5]中评估的算法。

我们研究时间序列回归（TSER）的出发点，是将时间序列分类算法改造用于回归。ROCKET 系列分类器都采用随机卷积和池化操作进行变换，随后使用线性分类器或岭分类器。最初的 ROCKET 通过将分类器替换为岭回归器，改造为适用于 TSER 的方法。我们进一步将研究范围扩展到较新的 ROCKET 变体 MultiROCKET [8]。深度学习算法也很容易改造；我们在先前研究的基础上，除了 InceptionTime [7] 的集成回归版本外，还纳入了一种卷积神经网络。

一种用于时间序列分类的替代方法是将大量无监督摘要特征用作变换。一项对多种替代方案的综述 [9] 发现，Fresh Pipeline with RotatIoN forest Classifier（FreshPRINCE）是用于时间序列分类的最佳变换流程。FreshPRINCE 使用基于可扩展假设检验的时间序列特征提取（TSFresh） [10]，随后采用旋转森林（RotF）分类器 [11]。我们将 FreshPRINCE 应用于时间序列回归。

基于区间的分类器也会提取无监督特征，但它们通过集成流水线来实现，其中使用随机选择的区间和快速基分类器。首个用于时间序列分类的基于区间的方法是时间序列森林（TSF） [12]。TSF 生成一组随机区间，并将每个区间的均值、标准差和斜率拼接起来，从而为每个基分类器构建独特的特征空间。aeon 工具包提供了用于回归的 TSF 版本[1]。规范区间森林（CIF） [13],及其后续方法——多样化表示规范区间森林（DrCIF） [14],——采用了与 TSF 类似的模型，但使用了不同的汇总特征和数据表示。CIF 使用规范时间序列特征（Catch22） [15] 特征集。第三节介绍了这些基于变换的算法的细节，以及我们如何将其改造用于时间序列回归。我们的主要贡献总结如下：

1) 我们向 TSER档案库提供了 44 个新数据集，其中包括 24 个单变量数据集和 20 个多变量数据集，使该档案库中的数据集达到 63 个；2) 我们在这些新数据上重复了 [5] 的研究，以检验结论是否适用于更大的数据集集合；3) 我们将最近提出的基于卷积、特征、区间和深度学习的时间序列分类算法应用于时间序列回归；4) 我们使用 21 种回归器开展了广泛的实验研究，并证明，平均而言，基于特征和区间的回归器比任何其他受评估算法都能取得显著更低的均方根误差；5) 我们对提出的两种最佳方法 Fresh-PRINCE 和 DrCIF 进行了全面分析和消融研究。6) 我们提供了与 scikit-learn 兼容的开源实现、清晰的可复现性指南，以及相关代码仓库中的详细结果[2]。

本文其余部分的结构安排如下。第二节介绍背景知识及相关研究。第三节详细描述我们提出的新型时间序列回归算法。第四节概述新的档案库，并介绍我们的实验设置。第五节展示了 21 种方法应用于总计 63 个数据集的实验结果。第六节对这些结果进行更深入的分析。最后，第七节总结我们的研究发现，并指出未来的研究方向。

II. 背景与相关工作

时间序列回归旨在创建时间序列与标量值 [5] 之间的映射函数。时间序列由实值有序观测组成。形式上，长度为 m 的单变量时间序列定义为 $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$。具有 d 个通道的多变量时间序列表示为$\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_d\}$，其中 $\mathbf{s}_k = \{s_{1,k}, s_{2,k}, \ldots, s_{m,k}\}$。我们假设有一个包含 n 个时间序列的列表 $\mathbf{S}$。因此，$s_{i,j,k}$ 表示第 i 个样本的第 j 个观测值在第 k 个通道上的取值。数据集 D 由 n 个时间序列样本及其对应的响应变量组成，即 $D = \{\mathbf{S}, \mathbf{y}\}$，其中$\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$ 为连续输出值，也就是说，输入时间序列 $\mathbf{s}_i$ 与输出 y_i 相关联。

时间序列回归模型是一种映射函数 $\mathcal{T} \to \mathcal{R}$，其中 $\mathcal{T}$是所有时间序列构成的空间， $\mathcal{R}$ 是一个连续值。时间序列回归模型在数据集 D_{TRAIN} 上训练，并在独立测试数据集 D_{TEST}上评估。

时间序列回归与标量对函数回归（SoFR） [16], 有一些相似之处。SoFR是一种函数回归模型，它在回归前将基函数模型应用于序列；也就是说，其目标是拟合一个以标量响应和函数型数据点为预测变量的回归模型 [17]。 [5]在基于 [16]的比较中使用了两种SoFR模型，分别是函数型主成分回归（FPCR）和采用B样条的FPCR（FPCR-Bs）。

时间序列预测通常通过应用滑动窗口，将时间序列构造成$\mathbf{S}$ 和一个回归窗口，并指定预测时域来选择目标$\mathbf{y}$，从而将问题转化为回归。时间序列预测回归（TSFR）中最常用的技术包括深度学习变体和全局模型，其中各通道会被拼接

[1]https://github.com/aeon-toolkit/aeon.

[2]https://tsml-eval.readthedocs.io/en/latest/publications/2023年/tser_档案扩展/tser档案扩展.html

并应用标准回归器，例如随机森林或 XGBoost。

我们关于时间序列回归 [5] 的主要来源比较了三种标准回归算法（支持向量回归器（SVR） [18],随机森林（RandF） [19], 和极端梯度提升（XGBoost） [20]）；两种 k 近邻模型，分别使用欧几里得距离和动态时间规整距离，并分别采用一个和五个邻居；三种深度学习方法（全卷积网络（FCN） [21], 残差网络（ResNet） [22] 和 InceptionTime [7]）；两种函数型分析方法（函数型主成分回归（FPCR）和基于 B样条的函数型主成分回归 [16]）；以及 ROCKET [6]。标准回归器采用全局预测回归器的方法：将时间序列展平为一个向量，串联所有通道。因此，长度为 m 且具有 d 个通道的多变量时间序列会被转换为长度为 $m \times d$的单个向量。此后，时间序列回归领域几乎没有出现算法上的进展。大多数新颖的发展都具有特定领域针对性，并非面向整个时间序列回归领域。其中，线性状态空间层（LSSL）模型 [23] 已在三个时间序列回归数据集上进行了测试，并取得了较低的误差指标。另一种状态空间模型 Liquid-S4 [24] 也在这三个数据集上进行了评估，并声称取得了更好的结果。一种基于图神经网络、名为 TISER-GCN [25] 的架构已被应用于地震数据，作为一项外部回归任务。在类似背景下， [26] 提出了 Septor，这是一种分层神经网络模型，用于根据波形数据估计地震事件的深度，即一种特定领域的外部回归任务。ROCKET-XGBoost [27]是唯一一种在19个时间序列回归档案库数据集上进行评估的新颖算法，但与 [5]中评估的算法相比，它并未带来显著改进。

A. 时间序列分类（*TSC*）算法

已有大量时间序列分类算法，并在可复现的比较研究中进行了对比 [1]–[3]。总体而言，可根据算法表示时间序列的方式对其进行分组。我们在此简要概述，并重点介绍分类器如何适配或可改编用于时间序列回归。

基于距离的分类器将距离函数与最近邻（NN）分类器等算法结合使用。最常用的两种距离函数是欧氏距离和动态时间规整（DTW）。只需对最近邻的目标变量取平均值，即可轻松将 NN 分类器改编用于回归。对于多变量数据，按照 [28], 中介绍的术语，DTW 可以是独立的（分别计算每个通道上的 DTW 距离，然后将这些值相加），也可以是依赖的（在逐点距离计算中使用所有通道）。

基于特征的算法使用无监督描述性统计将序列转换为特征，然后使用基于新特征集训练的分类器完成整个流程。

基于区间的 分类器是基于特征的流水线分类器的一种扩展。它们并非在整个序列上构建汇总特征，而是将不同区间上提取的特征拼接起来。随后，它们基于不同的随机区间构建集成，而非使用单个估计器。因此，我们将基于特征的方法和基于区间的方法统称为无监督基于特征的分类器。调整将这些算法应用于时间序列回归是我们的主要研究目标，因此我们将在第三节中更详细地介绍这一主题。

基于核函数/卷积的模型从所有可能子序列构成的空间中寻找卷积，并通过某种池化操作将其转化为特征。最流行的方法 ROCKET [6] 会生成随机卷积，并在流水线中与岭分类器配合使用。只需将岭分类器替换为岭回归器 [5]，即可将其改编用于时间序列回归。最近，MultiROCKET [29]作为 ROCKET 的改进版本被提出。ROCKET 使用两种池化操作生成特征：最大池化和正值百分比。MultiROCKET 增加了三种新的池化操作：正值均值、正值索引均值和最长连续正值段。除了原始数据，它还从一阶差分中提取特征。我们以与 ROCKET 完全相同的方式将 MultiROCKET 改编用于时间序列回归。

深度学习仍然在时间序列分类中广受欢迎 [30],，不过据我们所知，InceptionTime [7] 仍然是表现最佳的深度学习器。[5] 中的研究使用了残差网络（ResNet）、全连接网络（FCN）和 InceptionTime。原始的 InceptionTime 论文 [7] 提出采用由五个 InceptionTime 分类器组成的集成来获得最终结果。不过， [5] 在时间序列回归中使用了单个 InceptionTime 模型。我们评估了单个 InceptionTime（Inception）以及忠实于时间序列分类版本的 InceptionTime 集成（InceptionTimeE）在时间序列回归中的表现。我们还评估了基于 [31]中所述分类器的卷积神经网络（CNN）回归器。

基于形状片段的方法 [32], [33] 根据训练数据中选定的相位无关子序列是否出现来进行分类。在分类中，会使用信息增益等监督度量评估形状片段。此外，准确率最高的基于形状片段的方法 [14]采用一对多方法评估形状片段，并在各类别之间平衡搜索过程，以提高多样性。将形状片段方法改用于时间序列回归需要进行重大的内部更改并作出设计决策，这超出了本文的范围。

基于字典的算法采用类似词袋的方法，根据近似子序列（模式）出现的次数进行分类。最成功的基于字典的分类器 [34]–[36] 会在一定程度上进行监督式选择，利用准确率进行筛选或加权，或进行特征选择；将其改用于时间序列回归也超出了本文的范围。

III. 基于无监督特征的回归器

通过无监督过程从时间序列中提取特征的方法，已被证明在分类场景中表现良好。ROCKET [6] 和 CIF [13] 的表现不亚于甚至优于形状片段或字典算法等单一表示方法。这些算法的另一优点是复杂度较低，基本上由变换到分类器的流水线或流水线集成构成。近期的一项综述 [3] 还将 ROCKET 和 CIF [13]列为多变量时间序列分类（MTSC）中的顶尖方法。

我们介绍所提取的特征，以及针对另外两种基于无监督变换的算法所做的调整。第一种是 FreshPRINCE [9], ，它是一条使用 TSFresh [10] 特征集的流水线。第二种是 DrCIF [14], ，它是一种基于区间的集成。

A. FreshPRINCE

FreshPRINCE 是一种用于回归的流水线算法，由两个组件构成：TSFresh 特征提取算法将输入的时间序列变换为特征向量，随后由旋转森林（RotF） [11] 估计器构建模型并进行标签预测。TSFresh [37] 是一个特征集合，其中包含从时间序列数据中提取的近 800 个特征。虽然这些特征可以单独使用，但该方法还提供了一种称为 Fresh 的特征选择方法，用于移除无关特征。然而，FreshPRINCE 不使用此特征提取，而是保持变换过程的无监督性，并允许 RotF 决定特征的价值。TSFresh 在数据科学界通常很受欢迎，并且作为 FreshPRINCE [9] 的一部分，已显示出在分类问题上的表现优于其他无监督变换流水线。

RotF是一种树分类器集成，已被证明能够准确地为属性为连续值的问题做出预测 [38]。该分类器已被用作基准，并作为TSC中的其他流水线分类器的一部分 [1], [14],；当与用于TSC的无监督变换结合时，其表现优于岭分类器和 XGBoost [20] [9]。 [11] 和 [38]中提供了RotF算法的完整描述。RotF很容易适用于回归：我们开发的实现移除了类别子采样 [39]„ 将C4.5决策树替换为分类与回归树（CART） [40], ，并对森林中每棵树的标签预测值取平均。完整的TSFresh变换和修改后的RotF共同构成了我们针对TSER的FreshPRINCE改编版本。

B. DrCIF

基于区间的技术会选择具有固定偏移量的依赖阶段的区间，并从中提取汇总特征。这些区间在所有时间序列中的位置都相同，旨在从时间中的特定位置发现判别特征。大多数区间技术采用决策树森林的形式，通过使用不同的区间来提升集成的多样性。尽管有些区间森林确实会使用有监督特征提取， [41], TSF [12] 和 DrCIF [14]在选择区间以及从这些区间提取特征的方法上完全无监督。与分类实现相比，我们为时间序列回归所做的唯一改动是替换所用的决策树算法。TSF 也可以用同样的方式适配回归任务。

对于长度为 m 的序列，考虑所有区间长度和位置时，共有 $m(m-1)/2$ 个可能的区间。即使序列长度较短，从所有可能的区间中提取特征或对其进行评估也不可行。为了解决如何从这个区间池中选择区间的问题，DrCIF 采用了一种基于随机森林的方法。由 CART回归器组成的集成由不同随机区间变换的输出构建而成。算法 1 描述了 DrCIF 的完整构建过程。该变换包含三个步骤。首先，将基础时间序列拆分为三种序列表示：原始时间序列、序列的一阶差分，以及序列的周期图（在算法 1 的第 3 行中进行了表征）。差分和周期图的序列到序列变换已被证明能够为分类方法提供有用信息 [8], [41]–[43]。然后，为每个基础回归器创建不同的变换。首先，从包含 29 个特征的候选池中选择一个由a 特征组成的池（第 6 行）。DrCIF 使用规范时间序列特征（Catch22） [15]。Catch22 是从高度比较时间序列分析（HCTSA）工具箱 [44]中 7000+ 可用的特征中筛选出的 22 个特征组成的多样化集合。Catch22 特征原本是为归一化数据选择的，但我们不作此假设。因此，另外七个汇总统计量也作为候选项：均值、标准差、斜率、中位数、四分位距、最小值和最大值。然后，对于每种数据表示，选择一组 k 随机区间（第 10–13 行），并从随机选择的通道中计算和拼接 a 无监督特征（第 13–15 行）。最后，在每个集成成员所独有的特征集上训练一棵 CART 决策树。图 2 展示了 DrCIF 的变换（左）和集成（右）过程。新样本的预测值通过对基础回归器的预测值取均值得到。

算法 1 DrCIF（含有 n 个长度为 m 、具有 d个通道的样本的列表， $\boldsymbol{T} = (\boldsymbol{X}, \boldsymbol{y})$）

要求：决策树的数量： r; 每棵决策树在每种表示中包含的区间数量： k; 以及每棵决策树子采样的属性数量： a（默认值为 $r = 500$、$k = 4 + (\sqrt{d}\sqrt{rm})/3$ 和 $a = 10$） { 其中 rm 是表示序列的长度}1: 令 $\boldsymbol{DT} =$($DT_1 \dots DT_r$) 表示森林中的决策树 2: 令 $\boldsymbol{V}$ 为一个 $3 \times n \times d$ 矩阵，其中包含长度可变的序列，包括基础序列、周期图和一阶差分 3: $\boldsymbol{V} \leftarrow transform(\boldsymbol{X})$ 4: 对于 $i \leftarrow 1$ 到 r，执行 5: $\boldsymbol{X'} \leftarrow []$6: 随机选择 a 个属性索引（$u_1, \dots, u_a$） 7: 对于 $re \leftarrow 1$ 到 3，执行 8: $\boldsymbol{F} \leftarrow$ 初始化一个维度为 $n \times (ak)$ 的矩阵 9: 对于 $j \leftarrow 1$ 到 k，执行 10: $b = rand(1, |\boldsymbol{V}_{re}| - 3)$ { 区间位置 }11: $l = rand(3, |\boldsymbol{V}_{re}|/2)$ { 区间长度 }12: $o = rand(1, d)$ { 区间通道 }13: 对于 $t \leftarrow 1$ 到 n，执行 14: 对于 $c \leftarrow 1$ 到 a，执行 15: $\boldsymbol{F}_{t, a(j-1)+c} \leftarrow summaryStat(u_c, \boldsymbol{V}_{re}, t, o, b, l)$ 16: $\boldsymbol{X'} \leftarrow \boldsymbol{X'} + \boldsymbol{F}$ { 拼接特征向量 }17: $DT_i.buildCART(\boldsymbol{X'}, \boldsymbol{y})$

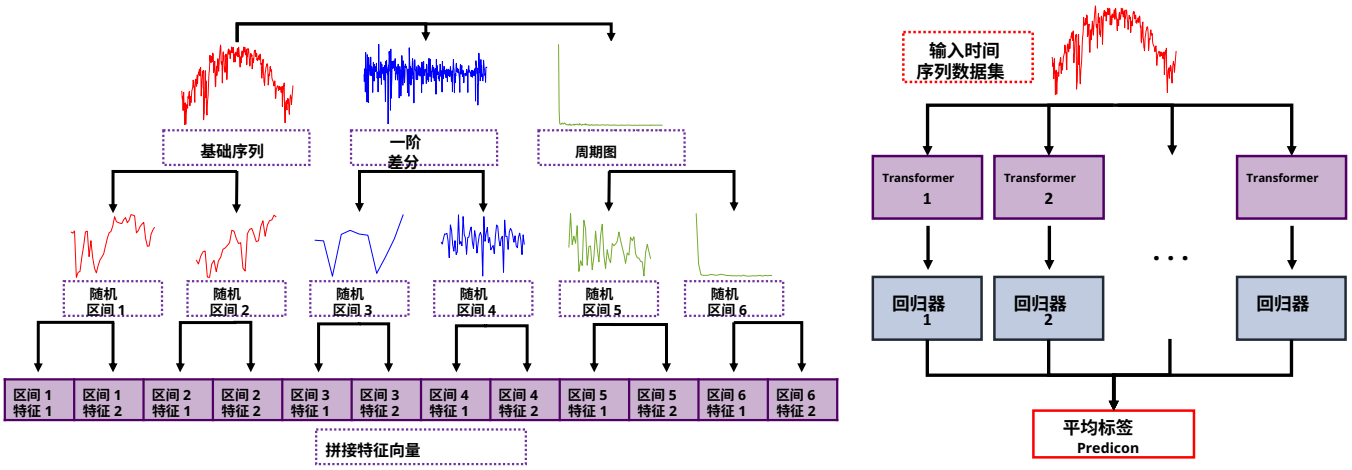


图 2. DrCIF变换（左）和DrCIF集成结构（右）的示意图。

IV. 方法

我们总结了新增的问题、实验中使用的回归器，以及对实验方法的描述。

当前版本的时间序列回归档案库包含19个不同的数据集。我们已将档案库中的数据集数量增加到63个。目前有28个单变量问题和35个多变量问题，通道数从2到24不等。数据集大小从93到16,000多个不等。70%用于训练，30%用于测试。序列长度从14到7500不等。其中9个问题包含缺失值，2个问题包含不等长序列。

新数据集取自 Kaggle 竞赛，以及与应用研究相关的其他档案库和代码仓库/网站。表 I 总结了收集的数据。有关数据集的更多详细信息可在相关代码仓库中获取。所有数据集均未进行归一化。其中一个新问题包含不等长序列。在实验中，按照 [5]，的惯例，序列中的缺失值采用线性插值法处理，不等长序列则截断至最小值序列的长度。完整描述，以及不等长序列和包含缺失值的序列，可在相关网站上获取。

A. 回归算法

第五节中评估的21个回归器（及其对应缩写）完整列表如下：

- 标准机器学习回归器：
  - 岭回归（Ridge）。- 网格搜索支持向量回归（Grid-SVR）[18]。- 随机森林（RandF）[19]。- RotationForest（RotF）[11]。- 极端梯度提升（XGBoost）[20]。
- 函数线性模型：
  - 函数型主成分分析（函数型主成分回归）[16]。- 基于B样条的函数型主成分回归（FPCR-Bs）[16]。
- 基于卷积的：
  - ROCKET [6].- MultiROCKET [8].
- 基于距离：
  - k 采用欧几里得距离的最近邻（kNN-ed），其中 $k = \{1, 5\}$。- k 采用动态时间规整距离的最近邻（kNN-dtw），其中 $k = \{1, 5\}$。
- 基于特征的：
  - FreshPRINCE [9]。
- 基于区间：
  - 时间序列森林（TSF）[12]。- DrCIF [14]。
- 深度学习方法：
  - 卷积神经网络（CNN）[30]。- 全连接神经网络（FCN）[30]。- 残差神经网络（ResNet）[30]。- 单一 Inception 神经网络（Inception）[7]。- Inception 神经网络集成（InceptionE）[7]。

所有算法的参数设置见补充材料。

B. 实验设计

每个数据集都提供了默认的训练/测试划分。我们将每项实验重复 30 次，以获得更显著的结果。

[3]https://perma.cc/DAP4-JC2A [4]
https://perma.cc/PA63-7GVU [5]
https://perma.cc/CH23-UVRJ [6]
https://perma.cc/8X2R-PRUD [7]
https://perma.cc/9V27-BF3F [8]
https://perma.cc/TP5Y-KS6M [9]
https://perma.cc/3 2 10 KPY-YHW
https://perma.cc/7865-ZAAD [11]
https://perma.cc/ZUS5-E26E [12]
https://perma.cc/B6FV-SLCG [13]
https://perma.cc/3 43 14 2APP-L
https://perma.cc/A7FG-KFLT [15]
https://perma.cc/BAG4-W8SL [16]
https://perma.cc/RQW7-QH7L [17]
https://perma.cc/J6LK-99Q5

表 I 新的时间序列回归数据集

Name	预测问题（响应变量）
经济分析	
每日油气价格	每日天然气价格及石油价格时间序列[3]。
能源监测	
建筑能源预测因子	估算建筑物中不同类型的能源消耗[4]。
OccupancyDetectionLight	根据传感器测量数据得出的办公室每小时平均占用情况 [45]。
SolarRadiationAndalusia	根据大气测量得出的每小时平均太阳辐射[5]。
TetuanEnergyConsumption	根据大气测量值 [46]得土安三个区域的每日平均电力消耗。
WindTurbinePower	基于扭矩测量时间序列[6]的风力涡轮机每日功率输出。
环境监测	
马德里噪声污染	LAeq[7] 声压级的第 1 百分位数。
非洲土壤化学	从非洲土壤信息服务（AfSIS）土壤化学 [8]衍生出的 12 个问题集合。
BeijingIntAirportPM25Quality	北京机场大气数据中颗粒物的日均值。 [47]。
每日温度纬度	根据每日气温年度时间序列推断城市的纬度[9]。
达卡每小时空气质量	使用本地化颗粒物时间序列计算达卡的空气质量指数[10]。
MadridPM10Quality	根据气体测量值计算西班牙马德里市的颗粒物周平均值[11]。
地铁州际交通流量	美国某条道路基于大气变量的日均交通量[12]。
伯明翰停车	每小时停放车辆总数得出的日占用率 [48]。
PrecipitationAndalusia	西班牙安达卢西亚的降雨量年平均值，基于气象测量值[13]。
内华达山脉积雪	基于温度时间序列的积雪量 [49]。
设备监测	
电动机温度	基于扭矩读数时间序列的电动机温度[14]。
液化石油气/甲烷监测家庭活动	来自气体传感器的液化石油气和甲烷浓度 [50]。
气体传感器阵列乙醇/丙酮	基于 16 个金属氧化物传感器测得的两种分析物（丙酮）的浓度 [51]。
波浪张力数据	基于波面高度时间序列的弦张力[15]。
健康监测	
BarCrawl6min	使用加速度计测得的经皮酒精含量 [52]。
COVID-19安达卢西亚	西班牙安达卢西亚的死亡人数与感染人数之比 [53]。
呼吸机压力	根据吸气电磁阀的控制输入和输出测得的该阀气压[16]。
情感分析	
加密货币情绪	基于相同日期每小时价格的四种加密货币情绪[17]。
天然气价格情绪	基于每日天然气价格的天然气价格情绪评分 [54] 。

分析。第一个实验使用默认数据。后续实验通过合并训练集和测试集的数据进行重采样，并按照与原始数据相同的训练集/测试集比例随机划分数据。为符合 [5]，使用均方根误差衡量性能。为了比较回归器，我们首先计算所有重采样的均方根误差平均值。所有统计检验均使用秩。对于多个数据集上的多个回归器，我们采用临界差异图 [55]，的改编版本：将事后 Nemenyi 检验替换为使用成对 Wilcoxon 符号秩检验比较所有分类器，并使用 Holm 校正 [56]，[57]形成团。

V. 结果

我们的实验结构如下。在第五节-A 中，我们在原始的 19 个数据集上重现 [5] 中呈现的结果。随后，我们将分析扩展到更大的数据集集合，以测试 [5] 中得出的结论是否能推广到包含 63 个问题的新档案库。在第五节-B 中，我们纳入了前一项研究中所用回归器的两项改进，以及开源工具包中提供的其他回归器。在第五节-C 中，我们将前述实验中表现最佳的回归器，与我们提出的新算法 FreshPRINCE 和 DrCIF 进行比较。

A. 在 *19* 个时间序列回归数据集上复现结果

我们在档案库当前的 19 个数据集上运行了 [5] 中报告的 13 个回归器，并使用五次重采样，以遵循原始工作的实践。图 3 展示了我们的结果与 [5] 中呈现的结果的临界差异图。总体而言，算法的排序相同，各组也相似。不过，排序仍存在一些差异：在我们的实验中，ROCKET 和全卷积网络的排名较低，而 Inception 和残差网络的排名高于原始结果。我们的分组也更加多样化。这是因为我们针对多重测试所采用的校正方法，比 [5], 中使用的方法更为宽松；后者采用了完整的 Bonferonni 校正，而不是 Holm 校正。

在图 4 中，我们比较了用于图 3 的 13 个回归器在包含 63 个数据集的更大档案库上的表现。从现在开始，我们还将重采样次数从五次扩展到 30 次。我们注意到，随机森林现在表现最佳，显著优于 ROCKET、Inception 和残差网络。因此，此前为时间序列回归提出的特定于时间序列的方法，并不比使用带有拼接特征的现成回归器更好。

B. 新 *TSER* 档案库的基准测试

在接下来的实验中，我们选取图 4 中排名前五的算法，并将它们与一些针对时间序列特定算法的替代性改编方法进行比较。良好的

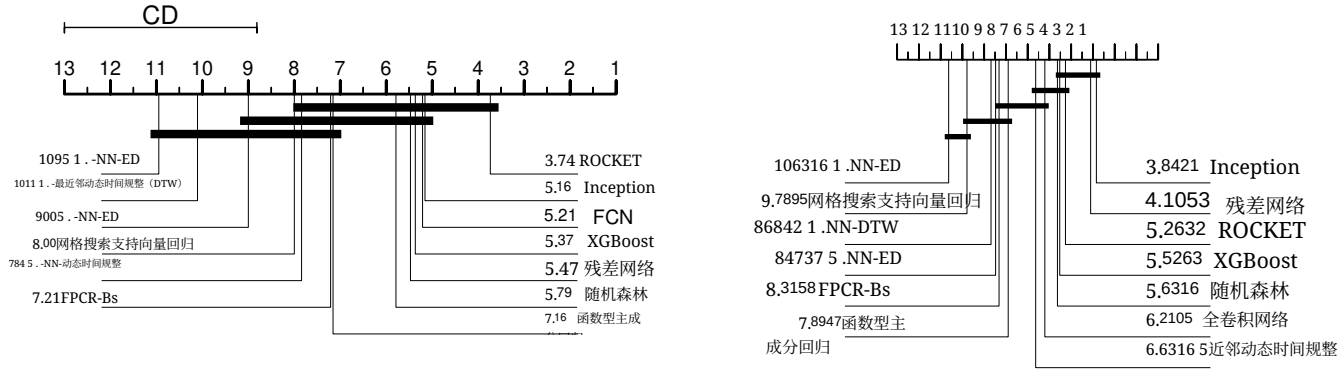


图 3. 均方根误差排名的复现　　在原始档案库（19 个数据集和 5 次重采样）上。左侧是来自 [5]的原始图像，右侧是我们的复现版本。　　.

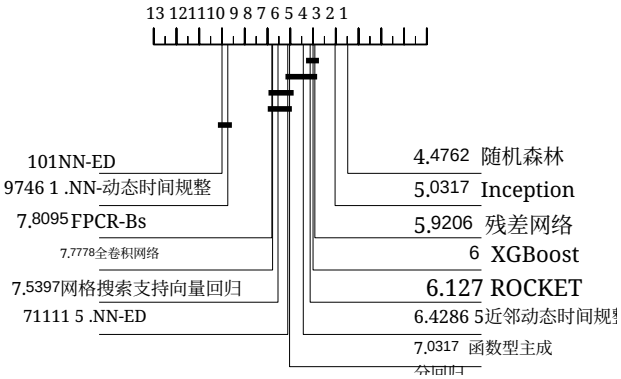


图 4. [5] 在 63 个时间序列回归数据集上使用的 13 种回归器的均方根误差排名。

XGBoost 和随机森林的表现表明，我们不应忽视标准分类器。旋转森林（RotF） [11] 是一种分类器，只需通过简单的平均操作 [39]即可轻松将其调整为回归器。研究表明，对于所有属性均为实值属性的问题，包括时间序列问题 [1], [38]，该方法尤其有效。因此，我们在本轮实验中加入了相应的回归版本。此外，为了使评估更加完整，我们还加入了标准岭回归器。另外，开源工具包 aeon[18] 包含两种此前尚未在时间序列回归（TSER）背景下进行评估的回归实现。时间序列森林回归器（TSF）是对时间序列森林分类器 [12]的改编，而卷积神经网络回归器（CNN）则是基于 [31]中所述版本的卷积神经网络。在进一步调查后，我们发现， [5]中 InceptionTime 的结果是通过单个 InceptionTime 模型（Inception）得出的。然而，在原始研究 [7],中，支持 InceptionTime 作为分类器的结果实际上是通过由五个 InceptionTime 模型组成的集成得出的。因此，我们加入了用于回归的 InceptionTime 集成模型（InceptionE）。此外，ROCKET算法的一个改进版本最近已发表，称为 MultiROCKET [29]。我们据此将其调整为适用于 TSER 范式。我们提供了 MultiROCKET、RotF 和 InceptionE 的实现，位于】【。

[18]https://github.com/aeon-toolkit/aeon

相关代码仓库。图 5 展示了图 4 所述实验中表现最好的三种算法以及六种新回归器的结果。由于 MultiROCKET 的原因，我们不得不排除澳大利亚降雨数据集，将研究中的数据集数量减至 62 个。MultiROCKET 方法在该数据集上需要超过 600 GB 的内存，并且需要 15 天以上才能完成。卷积神经网络和岭回归器

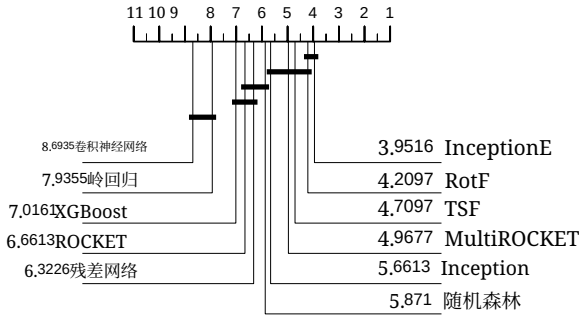


图 5. 11 种回归器在 62 个时间序列回归数据集上的均方根误差排名。

在另外九种算法面前，卷积神经网络和岭回归器缺乏竞争力。InceptionE 显著优于单个 Inception 网络。正如预期，MultiROCKET 显著优于 ROCKET。不过，最突出的结果是 RotF 与 InceptionE 并列表现最佳。RotF 是时间序列分类中表现最佳的标准算法 [1], ，因此这或许并不令人意外。尽管如此，一种用于标准回归的算法胜过除 InceptionE 之外的所有深度学习和时间序列专用方法，这一事实表明时间序列回归领域仍有很大的改进空间。

C. 新时间序列回归算法的评估

现在，我们将上一项实验中表现最佳的六种算法，与基于无监督特征提取的两个回归器 DrCIF 和 FreshPRINCE 进行比较。请注意，未纳入单个 Inception 网络，因为 InceptionE 始终优于它。此外，我们还决定纳入 ROCKET，因为它是 [5] 原始研究中表现最佳的方法。图6显示，这两个新回归器

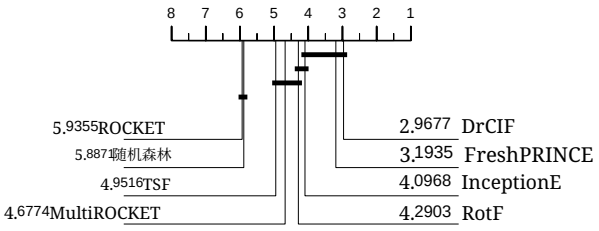


图 6. 两个基于特征的回归器 DrCIF 和 FreshPRINCE，以及在 62 个时间序列回归数据集上表现最佳的六个回归器的均方根误差排名。

都位于顶级组中，并且就均方根误差的平均排名而言，显著优于 RotF；在我们尝试过的其他方法中，RotF 是表现最佳的标准方法。InceptionE 是排名第三的算法。InceptionE 通常表现出色：在 62 个问题中，它有 18 个排名第一。然而，它在许多问题上的表现也会一败涂地。图7展示了所有问题上均方根误差相对偏差的箱线图。数值低于 0.5 表示该算法在相应问题上的表现优于均方根误差中位数。总体来看，只有 DrCIF 和 FreshPRINCE 的表现始终优于中位数性能，且分布较为集中。InceptionE 的分布范围最广。图8

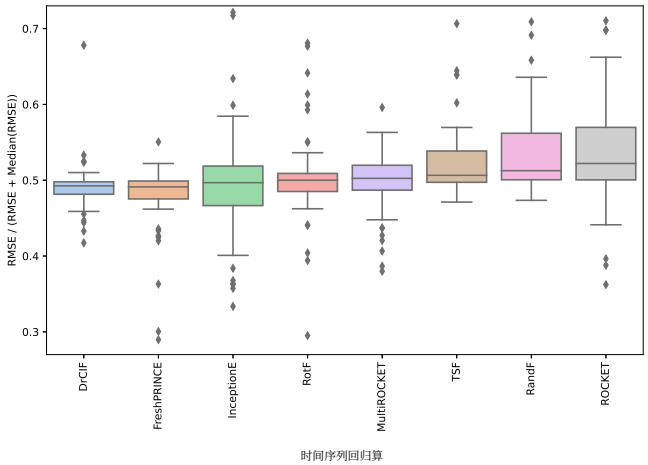


图 7. 八个回归器的相对均方根误差分布（数值越低越好）。

详细分析了结果。在这种情况下，使用新颖工具[19]绘制了显示平均均方根误差结果的热图。可以看出，DrCIF 和 FreshPRINCE 得出的平均结果明显优于其余方法。最后，正如之前所观察到的，InceptionE 的平均均方根误差结果最差，这表明它不是一种非常稳定的方法，因为其排名为第 3 位，但按平均均方根误差计算却位列第 8 位。

VI. 分析

我们进一步详细探讨结果，以更全面地了解这些回归器，并深入分析影响其性能的因素。

[19]https://github.com/MSD-IRIMAS/Multi Comparison Ma_ trix

A. 运行时间

图9展示了图7中八个回归器的均方根误差平均排名与运行时间之间的关系（运行时间采用对数刻度）。可以看到，运行时间与性能之间存在直接的权衡。除 InceptionE 在图形处理器上运行外，所有算法都在单线程 CPU 上运行。这意味着图表对 InceptionE 非常有利。即使在图形处理器上运行，它也比在 CPU 上运行的 RotF 慢。

B. 按数据特征划分的性能

我们根据数据的核心特征对回归器的性能进行分类，以深入了解不同算法在何种情况下表现良好。需要强调的是，这纯粹是探索性的分析：每个类别中的数据集数量相对较少（括号中所示），因此无法进行有意义的显著性检验。表——

表 II 按训练样本数量划分的均方根误差平均排名。

	<300 (15)	300-999 (23)	1000-5000 (13)	>5000 (11)
DrCIF	**2.87**	3.39	*2.85*	3.91
FreshPRINCE	*3.60*	**2.87**	2.92	**2.27**
InceptionE	3.87	*3.17*	**2.69**	3.91
RotF	4.67	4.57	5.46	4.73
MultiROCKET	4.27	4.30	3.85	*3.73*
TSF	5.13	5.48	5.08	5.55
随机森林	6.07	6.04	4.31	4.36
ROCKET	5.53	6.17	5.69	5.55

表 II 展示了按训练样本数量对问题分组后的均方根误差平均排名。其规律是：样本数量较少时 DrCIF 表现更好，而训练集规模较大时 FreshPRINCE 表现更好。表 III 展示了平均

表 III 按通道数量划分的均方根误差平均排名。

	1 (28)	2 (9)	3 或 4 (14)	> 5 (11)
DrCIF	*3.29*	**2.90**	**2.33**	*2.91*
FreshPRINCE	**2.96**	*3.80*	4.00	**2.55**
InceptionE	4.21	**2.90**	3.58	5.09
RotF	4.25	5.50	4.58	*2.91*
MultiROCKET	4.04	4.40	5.75	5.45
TSF	5.50	5.10	*3.33*	5.45
RandF	6.21	5.70	5.92	5.18
ROCKET	5.54	5.70	6.50	6.45

按通道数对问题分组时的均方根误差排名。DrCIF 在包含 2、3 或 4 个通道的多变量问题上表现相对优于 FreshPRINCE。而对于单变量数据集以及通道数超过 5 个的多变量数据集，FreshPRINCE 的结果更好。表 IV 展示了按序列长度对问题分组时的平均均方根误差排名。令人意外的是，基于区间的 DrCIF 在短序列上的表现也相对优于 FreshPRINCE，但在较长序列（长度 >1000）上的表现较差。

最后，我们还评估了不同问题类型下的相对性能，但未发现任何有趣的趋势。

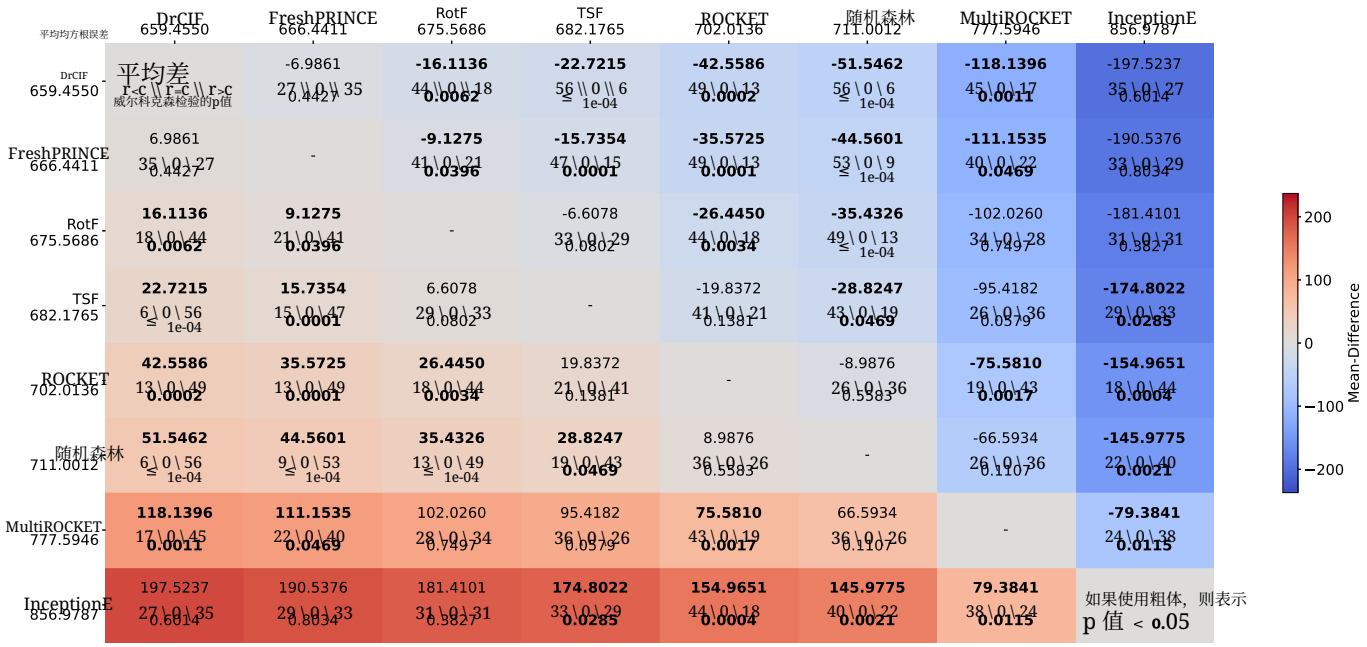


图 8. 八个回归器的最佳性能汇总结果。

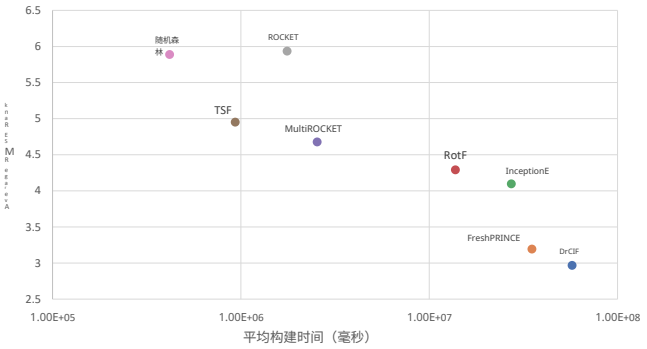


图 9. 以对数刻度表示的运行时间（毫秒，所有数据集的平均值）与均方根误差的平均排名之间的关系。

表 IV 按序列长度划分的均方根误差平均排名（长度为 366–999 的问题不存在）。

	<50 (14)	50–150 (13)	151–365 (16)	>1000 (19)
DrCIF	**2.69**	**2.00**	**2.64**	3.95
FreshPRINCE	3.77	4.25	*3.00*	**2.58**
InceptionE	4.23	5.63	4.64	*2.74*
RotF	*3.69*	4.13	4.05	5.05
MultiROCKET	5.15	*3.88*	5.23	4.05
TSF	4.85	4.63	4.41	5.79
随机森林	5.08	5.88	6.09	6.21
ROCKET	6.54	5.63	5.95	5.63

C. FreshPRINCE 的消融实验

FreshPRINCE 是由 TSFresh 变换和 RotF 回归器组成的流水线。我们探讨该回归器的性能究竟归功于变换、回归器，还是两者兼有。图10总结了 FreshPRINCE、对原始序列使用 RotF，以及使用 TSFresh 变换的性能。随后使用另一种回归器。结果表明，先进行变换再使用随机森林或XGBoost，其效果并不优于直接将 RotF 应用于原始数据。我们得出结论：正是变换与回归器的组合带来了显著更好的性能。几乎没有

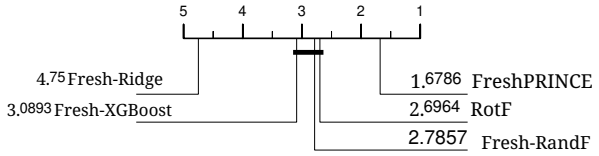


图10. 使用不同回归器时，FreshPRINCE、标准 RotF 和 TSFresh（Fresh）的均方根误差排名。

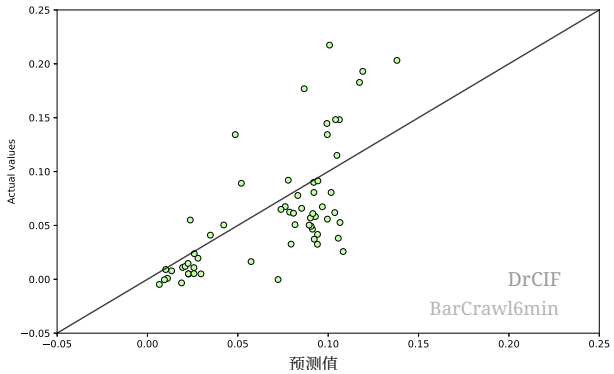


图11. DrCIF 在 BarCrawl6min 上的预测值与实际值散点图。

InceptionE 与其他回归器之间的一致性。我们认为，InceptionE 的性能如此多变，一个原因是它有时会完全无法从某个数据集中找到任何有用的信息，而其他模型至少还能做出一些有预测力的判断。为说明这一点，我们考察了

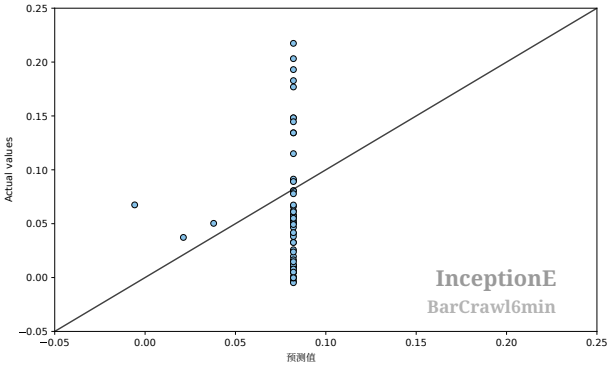


图 12. BarCrawl6min 数据集上 InceptionE 的预测值与实际值散点图。

DrCIF 和 InceptionE 在 BarCrawl6min数据集上的标准化残差。时间序列是加速度计数据，响应变量是受试者的经皮酒精浓度。响应变量的下界为零。在传统

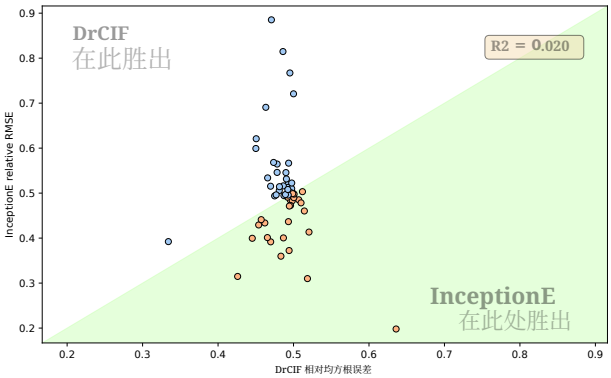


图 13. DrCIF 和 InceptionE 的相对均方根误差散点图。

在回归中，分析人员可能会考虑对响应变量进行变换，例如使用 Yeo-Johnson 变换 [58]。我们关注的是模型在多个数据集上的表现，而不依赖人工定制的变换。DrCIF 在默认训练-测试划分上的均方根误差为 0.0017，InceptionE 则为 0.0045。如果绘制 DrCIF 的预测响应值与实际响应值（图 11），我们会发现，对于较低的实际值，DrCIF 给出了负预测值；对于较高的值，它则低估了响应值，而且残差似乎存在一定的异方差性。尽管如此，它显然已经发现了回归变量序列与响应值之间的某种关系。然而，如果查看 InceptionE 的同类图（图 12），我们会发现 InceptionE 几乎总是预测为同一个值 0.082。仔细配置、调优并变换数据或许能改善 InceptionE 的表现。然而，DrCIF 和 FreshPRINCE 也是如此。我们使用 InceptionE 的方式遵循其创建者的建议 [7]。

此外，图13展示了图7中用于 DrCIF 与 InceptionE 对比的标准化均方根误差散点图。两者之间没有相关性。作为参考，在我们的前四名回归器中，相关性最强的是 DrCIF 与 RotF（R^2 =0.27），而 DrCIF 与 FreshPRINCE 之间的相关性非常弱（$R^2 = 0.13$）。这种多样性表明，或许存在一定价值在集成中。

VII. 结论

我们提出了基于分类器的时间序列外部回归（TSER）新算法，并开展了广泛的实验研究。我们将 TSER 档案库的规模从 19 个数据集扩大到 63 个，引入了先前研究中所用回归器的改进版本，并证明我们对分类算法的新改编显著优于最佳替代方案。本研究存在若干局限。数据并非随机抽样（我们从能够找到数据的任何地方收集问题），某些领域可能占比过高；我们没有对这些回归器进行调优（在 63 个数据集上调优这些算法所需的计算量将难以承受）；我们也没有考察残差分析等更复杂的性能诊断方法。尽管如此，我们相信，我们为推动 TSER 这一新领域的发展作出了重要贡献。RotF 的表现优于此前评估过的所有回归器，而 DrCIF 和 FreshPRINCE 是迄今为止提出的唯一显著优于所有标准回归算法的 TSER 算法。我们发布了结构清晰、兼容标准工具包的源代码、复现实验的指南以及所有实验结果，从而确保所有实验均可复现。

我们认为，TSER 算法仍有进一步改进的空间。借鉴有监督时间序列分类（TSC）方法，或许能进一步发挥这一热门研究主题的潜力。InceptionE 是最有前景的深度学习方法，或许可以通过改进设计，避免它在训练集规模较小时容易出现的灾难性失效。异构集成在 TSC 中表现出色，而最佳算法之间的性能多样性表明，经过谨慎调整后，这种方法或许也能应用于 TSER。我们将继续扩充代码仓库，纳入更多问题，也欢迎大家捐赠。

致谢

本研究部分由“西班牙研究署（Agencia Española de Investigación）”（资助编号：PID2020-115454GB-C22 / AEI / 10.13039 / 501100011033）资助。David Guijo-Rubio 的研究由科尔多瓦大学通过大学部资助的西班牙大学体系重新资格认定项目资助，该项目面向公立大学，资金来自欧盟的 NextGenerationEU（资助编号：UCOR01MS）。本研究还得到英国工程与自然科学研究委员会资助（资助编号：EP/W030756/1）。Guilherme Arcencio 的研究由圣保罗研究基金会（FAPESP）资助（资助编号：#2022/12486-4 和 #2022/00305-5）。Diego Furtado Silva 的研究也得到 FAPESP 资助（资助编号：#2022/03176-1）。部分实验在东安格利亚大学研究与专业计算支持服务所支持的高性能计算集群上完成。我们谨向所有帮助维护时间序列分类档案库的人员，以及为算法开源实现作出贡献的人员表示感谢。

附录 A 数据描述

档案库中全部 63 个数据集的列表见表 V。本节将进一步介绍我们新增的 44 个数据集。

A. 经济分析

1) 石油和天然气价格： Kaggle[3] 发布的一个数据集包含2000年到2022年间布伦特原油、西德克萨斯中质原油、天然气和取暖油的历史价格。我们使用连续30个工作日的西德克萨斯中质原油收盘价和交易量作为预测变量，并将每个30天时间段内天然气收盘价的平均值作为目标变量，创建了**每日油气价格** 。最终数据集包含191条长度为30的二维时间序列，其中70%被随机抽样作为训练数据，其余30%作为测试数据。这类模型可以帮助企业和政府更好地分析和预测与石油和天然气相关的经济形势及其相关性。

B. 能源监测

1) ASHRAE - Great Energy Predictor III：该数据集发布于 Kaggle[4]，，旨在评估提高能源效率的改进措施所带来的价值。为此，确定了四种能源类型：电力、冷冻水、蒸汽和热水。目标是估算以千瓦时（kWh）为单位的能源消耗。

维度对应于气温、露点温度、风向和风速。这些数值在一周内按小时采集，输出则是上述四种能源的仪表读数。据此，我们创建了四个数据集：**ChilledWaterPredictor**、**ElectricityPredictor**、**HotwaterPredictor**和**SteamPredictor**。每个数据集所包含的时间序列数量各不相同，因为这些时间序列对应使用上述能源的不同建筑。具体而言，**ChilledWater-Predictor** 包含 459 条长度为 168、维度为 4 的时间序列，**ElectricityPredictor** 包含 810 条长度为 168、维度为 4 的时间序列，**HotwaterPredictor** 包含 351 条长度为 168、维度为 4 的时间序列，**SteamPredictor**包含 300 条长度为 168、维度为 4 的时间序列。我们随机抽取其中 70% 的时间序列作为训练数据，其余 30% 作为测试数据。

尽管 Kaggle 上有一篇帖子指出，有一栋建筑的仪表读数以千英热单位（kBTU）为单位，但这些读数已相应地转换为千瓦时。

2）占用检测：在 Candanedo 和 Feldheim 的一项研究中，研究人员每分钟采集温度、光照、CO_2,和湿度的 [45], 测量值，用于检测办公室是否有人。这些数据已发布在 UCI 机器学习库[20]中。

我们通过重新定义问题创建了**OccupancyDetectionLight**数据集。我们使用一小时内的温度（单位：℃）、湿度比和 CO_2 浓度（单位：ppm）作为预测变量，并将该小时的平均光照（单位：勒克斯）作为响应变量。由此得到 340 条长度为 60 的三维时间序列。我们随机抽取其中 70% 的时间序列作为训练数据，其余 30% 作为测试数据。

针对这些数据构建更好的模型，有助于改进能源消耗分析和预测。

3) 安达卢西亚太阳辐射： 该数据集来自安达卢西亚政府（西班牙）[5]。数据采集自安达卢西亚8个行政区的不同监测站：阿尔梅里亚、加的斯、科尔多瓦、格拉纳达、韦尔瓦、哈恩、马拉加、塞维利亚，时间范围为2000年至2014年2月。该数据集名为**SolarRadiationAndalusia**。维度对应湿度和温度的每日均值，而输出是同一天的太阳辐射。由于时间序列取完整年份的每日数值，因此未使用2014年的数据。

最终的数据集包含 672 个时间序列，具有 2 个维度，长度为 365。训练数据集包含随机抽取的 70%，其余 30 % 构成测试数据集。

4) 得土安能源消耗：此数据集发布于 UCI 机器学习库[21]，，旨在估算得土安三个区域的功耗 [46]。这个新数据集名为**TetuanEnergyConsumption**。数据每十分钟采集一次。因此，时间序列有 144 个值（每小时 6 个值）。共确定了 5 个维度：温度、湿度、风速、总漫射流量和漫射流量。目标是估算得土安三个区域的每日平均电力消耗。

上述数据集包含 364 个长度为 144 的五维时间序列。这 364 个时间序列中有 70% 被随机选入训练集，其余 30% 属于测试集。

5）风力涡轮机发电：Kaggle[6] 上的“风力涡轮机功率（kW）发电数据”数据集包含从风力涡轮机收集的大量数据，包括涡轮机各个部位的温度、每个叶片的角位置以及涡轮机的功率输出。测量值每10分钟采集一次，时间跨度为2019年至2021年。

我们使用这些数据创建了 **WindTurbinePower** 数据集。每个时间序列包含 144 个时间点，即一天的测量值，其目标变量是风力涡轮机当天的平均功率输出。在 76 个可能的特征中，我们仅使用风力涡轮机的扭矩作为预测变量，因为这已足以取得良好效果，而其他特征至多只是次要因素。由于部分样本对应的日期测量值少于 144 个，因此将其移除。最终数据集包含 852 个样本，其中随机抽取 70 % 作为训练数据，其余 30% 作为测试数据。

适用于该数据集的良好回归模型应能通过更准确地预测特定地点和/或季节的风力涡轮机能源输出，改善绿色能源生产的物流。

C. 环境监测

1) 西班牙马德里的噪声污染：该数据集由西班牙马德里政府[7]公开发布。数据由分布在马德里市的多个监测站采集。该数据集每日更新

[20]https://perma.cc/5 5 21 3ER-MXG
https://perma.cc/NX65-A5B4

表 V TSER档案库中的时间序列数据集。

数据集	训练集规模	测试集规模	长度	# 维度	缺失值	类别
AppliancesEnergy	95	42	144	24	No	能源监测
HouseholdPowerConsumption1	745	686	1440	5	Yes	能源监测
家庭用电量2	745	686	1440	5	Yes	能源监测
澳大利亚降雨量	112186	48081	24	3	No	环境监测
北京PM10质量	11918	5048	24	9	Yes	环境监测
北京PM2.5质量	11918	5048	24	9	Yes	环境监测
苯浓度	3349	5163	240	8	Yes	环境监测
洪水建模1	471	202	266	1	No	环境监测
洪水建模2	466	201	266	1	No	环境监测
洪水建模3	429	184	266	1	No	环境监测
活体燃料含水量	3493	1510	365	7	No	环境监测
BIDMC32HR	5550	2399	4000	2	No	健康监测
BIDMC32RR	5471	2399	4000	2	No	健康监测
BIDMC32SpO2	5550	2399	4000	2	No	健康监测
Covid3Month	140	61	84	1	No	健康监测
IEEEPPG	1768	1328	1000	5	No	健康监测
PPGDalia	43215	21482	256-512^{D}	4	No	健康监测
新闻标题情感	58213	24951	144	3	No	情感分析
新闻标题情感	58213	24951	144	3	No	情感分析
每日油气价格	133	58	30	2	No	经济分析
ChilledWaterPredictor	321	138	168	4	No	能源监测
ElectricityPredictor	567	243	168	4	No	能源监测
HotwaterPredictor	245	106	168	4	No	能源监测
OccupancyDetectionLight	237	103	60	3	No	能源监测
SolarRadiationAndalusia	470	202	365	2	Yes	能源监测
SteamPredictor	210	90	168	4	No	能源监测
TetuanEnergyConsumption	254	110	144	5	No	能源监测
WindTurbinePower	596	256	144	1	No	能源监测
马德里噪声污染	166	72	365	1	Yes	环境监测
铝浓度	440	189	2542	1	No	环境监测
BeijingIntAirportPM25Quality	1099	472	24	6	No	环境监测
硼浓度	438	188	2542	1	No	环境监测
钙浓度	444	191	2307	1	No	环境监测
铜浓度	440	189	2542	1	No	环境监测
每日温度纬度	27440	11760	365	1	No	环境监测
达卡每小时空气质量	1447	621	24	1	No	环境监测
铁浓度	427	184	1716	1	No	环境监测
MadridPM10Quality	4845	2077	168	3	Yes	环境监测
镁浓度	1560	669	3578	1	No	环境监测
锰浓度	427	184	1716	1	No	环境监测
地铁州际交通流量	849	365	24	4	No	环境监测
伯明翰停车	1391	597	14-18^{S}	1	No	环境监测
磷浓度	1573	675	3578	1	No	环境监测
钾浓度	1561	669	3578	1	No	环境监测
PrecipitationAndalusia	470	202	365	4	Yes	环境监测
内华达山脉积雪	350	150	30	3	No	环境监测
钠浓度	424	183	1716	1	No	环境监测
硫浓度	444	191	2307	1	No	环境监测
锌浓度	445	191	2307	1	No	环境监测
电动机温度	15503	6645	60	6	No	设备监测
GasSensorArrayAcetone	324	140	7500	1	No	设备监测
GasSensorArrayEthanol	324	140	7500	1	No	设备监测
液化石油气监测家庭活动	2017	865	100	1	No	设备监测
MethaneMonitoringHomeActivity	2017	865	100	1	No	设备监测
波浪数据张力	1325	568	57	1	No	设备监测
BarCrawl6min	140	61	360	3	No	健康监测
COVID-19安达卢西亚	142	62	91	1	No	健康监测
呼吸机压力	52815	22635	80	2	No	健康监测
币安币情感	184	79	24	2	No	情感分析
比特币情感	232	100	24	2	No	情感分析
卡尔达诺情感	74	33	24	2	No	情感分析
以太坊情感	249	107	24	2	No	情感分析
天然气价格情感	65	28	20	1	No	情感分析

对于那些时间序列长度不等的数据集：^{S} 表示时间序列长度不等与样本有关而 ^{D} 表示与维度有关。

to

自 2014 年以来。不过，我们根据截至 2021 年 12 月的数据创建了**马德里噪声污染**数据集。输入时间序列为等效连续A计权声级（LAeq），这是一项基本测量参数旨在用一个数值表示给定时间内不断变化的声源。输出时间序列为 LAS01，即声压级的第一个百分位数，采用 A

频率计权和慢速时间计权，并在相应时段内记录。此类序列和输出示例见图14。

最终数据集包括 238 个单变量时间序列。此外，它们的长度为 365，对应于一年期间获取的每日数值。训练数据集由随机选取的 70% 样本组成，而剩余的 30% 构成测试数据集。

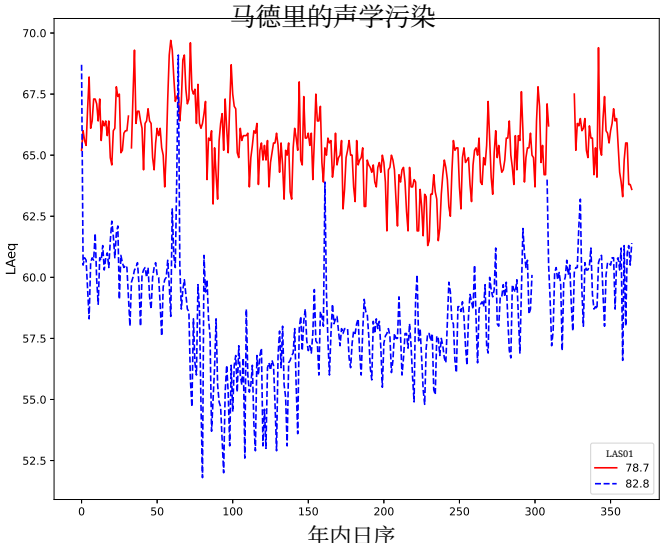


图 14. AcousticContaminationMadrid 数据集中年度噪声污染的两个示例。

2）非洲土壤化学：非洲土壤信息服务（AfSIS）土壤化学[8] 数据集包含大量干化学和湿化学数据，这些数据来自2009年至2013年间撒哈拉以南非洲多个国家采集的土壤样本。干化学分析方法，例如红外光谱法和X射线荧光，其成本相较于湿化学分析低得多。因此，仅使用干化学数据来预测通过湿化学分析测得的特定营养元素含量的模型，在商业上具有较高的应用价值。

干化学测量采用了多种不同的机器型号，我们从中选取了四种：“Al- pha ZnSe”、“Alpha KBr”、“HTSXT”和“MPA”。每个土壤样本的湿化学数据包含 12 种营养元素的含量：铝、硼、铜、铁、锰、钠、磷、钾、镁、硫、锌和钙。每项干化学测量都是一个时间序列，其中时间轴表示波长，y 轴表示相应的响应值。

我们将每台干化学仪器三分之一的实验与另一种营养元素测量配对，从而创建了 12 个数据集。它们分别是 **铝浓度**（629 个长度为 2542 的样本）、**硼浓度** （626 个长度为 2542 的样本）、**铜浓度** （629 个长度为 2542 的样本）、**铁浓度** （611个长度为 1716 的样本）、**锰浓度** （611 个长度为 1716 的样本）、**钠浓度** （607 个长度为 1716 的样本）、**磷浓度**（2248 个长度为 3578 的样本）、**钾浓度** （2230 个长度为 3578 的样本）、**镁浓度** （2229 个长度为 3578 的样本）、**硫浓度** （635 个长度为 乐彩2307）、**锌浓度** （长度为2307的636个样本），以及**钙浓度**（长度为2307的635个样本）。

在每个数据集中，随机抽取 70% 的样本作为训练数据，其余 30% 作为测试数据。图 1 展示了三个土壤光谱图示例。

3) 北京机场*PM2.5*污染：此数据集来自 UCI 机器学习库[22]。作者 [47], 收集了美国驻北京大使馆的 $PM_{2.5}$ 数据，以及北京首都国际机场的气象数据，频率为每小时一次。此数据集名为**BeijingIntAirportPM25Quality**，包含由 24 个点组成的 6 维时间序列。各维度分别为露点、温度、气压、合并风向，以及北京首都国际机场测得的积雪和降雨累计小时数。输出为 $PM_{2.5}$ 日均数据。

上述数据集包含1571个长度为24的六维时间序列。其中，70%的1571个时间序列被随机选入训练集，其余30%属于测试集。

4) 每日气温与纬度：Kaggle[9] 上发布的一个数据集包含1980年至2020年世界人口最多的1000座城市的每日气温数据，以及这些城市的地理坐标。

我们使用这些数据创建了 **每日温度纬度**数据集。我们将每个城市的气温数据拆分为时长1年的时间序列，即365个时间点。通过对2月 28^{th} 和 29^{th} 的温度取平均值，缩短闰年的长度。预测变量是这一年中记录的每日气温（单位：°C），响应变量是相应城市的纬度。最终数据集包含39200个长度为365的单变量时间序列，其中70%被随机抽样作为训练数据，剩余30%作为测试数据。构建的数据集中的两个样本如图15所示。

通过对该数据集进行探索性分析和回归，可以更好地理解并预测气候变化及其影响在当地的具体表现。

5) 达卡空气质量：数据来源于 AirNow[23] ，并由 Kaggle[10] 提供，包含孟加拉国达卡美国大使馆连续7年的细颗粒物（$PM_{2.5}$）浓度小时测量值，以及相应的空气质量指数（AQI）。

我们使用这些数据创建了**达卡每小时空气质量**数据集，其中预测变量为24小时的$PM_{2.5}$ 浓度，响应变量为当天的平均AQI。这样便能创建并测试更好的模型，以更低的成本、更快地评估不同城市的空气质量。最终数据集包含2068个长度为24的单变量时间序列，其中70%被随机抽样为训练数据，其余30%作为测试数据。

6) 马德里*PM10*污染： 该数据集可在 Kaggle[11]，上公开获取，尽管这些数据最初发布在马德里政府[24]的公共代码仓库中。该数据集称为 **MadridPM10Quality**。其输入数据由二氧化硫、一氧化碳和一氧化氮的测量值构成。时间序列 corre-

[22]https://perma.cc/G8VS-DKKY [23] https://perma.cc/2K2D-LT6T

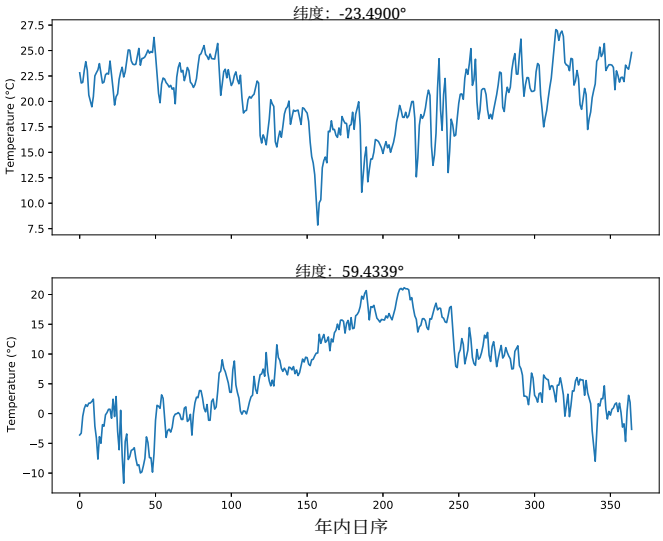


图 15. 每日气温与纬度数据集中两个不同纬度的年度温度曲线示例

对应于一周内测量的每小时数值。输出值为 PM_{10} 的周平均值。

最终数据集包含 6922 条长度为 168 的三维时间序列。训练数据集由随机选取的 70% 样本组成，其余 30% 构成测试数据集。

7) 都市州际公路交通量：该数据集可从 UCI 机器学习库[12]公开获取。数据包含明尼苏达州交通部 ATR 站点 301 的每小时交通量；该站点大致位于明尼阿波利斯和圣保罗之间的中点，数据收集时间为 2012 年至 2018年。该数据集名为**地铁州际交通流量**，旨在估算上述道路的日均交通量。输入维度包括平均温度、降雨量、积雪量和云量百分比。

最终数据集包含 1214 个具有 4 个维度的时间序列。此外，其长度为 24，对应于一天内每小时采集的测量值。训练数据集包含从全部数据中随机选取的 70%，其余 30% 构成测试数据集。

8) 伯明翰停车数据集：该数据集可从 UCI 机器学习库[25]公开获取。作者 [48]从伯明翰市议会运营的国家停车场公司停车场收集了数据。该数据集名为**伯明翰停车**，旨在估算 2016年10月4日至2016年12月19日的占用率。输入时间序列为每小时停放车辆的数量，输出则为占用率。每天测量的总小时数在 14 至 18 小时之间。

上述数据集包含 1988 条长度不等的时间序列（长度在 14 到 18 之间）。其中，70% 的时间序列被随机选入训练集，其余 30% 属于测试集。

9) 安达卢西亚的降水：该数据集来自安达卢西亚政府（西班牙）[13]。数据已被于2000年至2014年2月，从安达卢西亚8个行政区的不同监测站获取：阿尔梅里亚、加的斯、科尔多瓦、格拉纳达、韦尔瓦、哈恩、马拉加和塞维利亚。该数据集名为**PrecipitationAndalusia**，包含日平均温度、湿度、风速和风向作为输入，

而输出是平均降水量。图 16 展示了两个示例。

这产生了 672 个长度为 365 的四维时间序列。我们随机抽取其中 70% 作为训练数据，剩余 30% 作为测试数据。

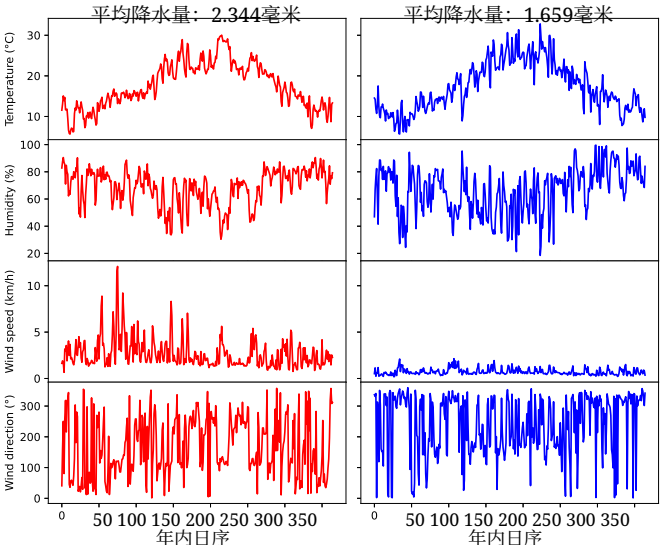


图 16. PrecipitationAndalusia 数据集中的两年不同测量值及其相关平均降水量

10) 内华达山脉测量值：由 Osterhuber 和 Schwartz 发布 [49] 并在 Kaggle[26] 上提供的数据集包含1971年至2019年间在美国内华达山脉的一处野外观测站进行的每日最低和最高气温、降水量及积雪特征测量值。

我们将测量值划分为每组连续 30 天的数据，并使用最低和最高气温（单位：°C）以及降水量（单位：mm），创建长度为 30 的三维时间序列。我们将每个相应 30 天时间段内累积的新降雪量（单位：cm）作为目标变量，构建了**内华达山脉积雪**数据集，共包含 500 个样本。由于存在缺失值，约有 20 个样本被删除。我们通过随机抽取 30% 的数据作为测试集，将数据集划分为训练集和测试集。这些数据可用于训练模型，以预测强降雪，并帮助城市和道路为恶劣天气条件做好准备。

D. 设备监测

1) 电动机温度：Kirchg¨assner、Wallscheid 和 B¨ocker [59] 收集了大量传感器数据，并将其发布在 Kaggle[14], 上。这些数据来自部署在试验台上的一台永磁同步电动机（PMSM）。该数据集由多个测量会话组成，每个测量会话的时长介于一至

[24]https://perma.cc/5QNM-ZWGS [25] https://perma.cc/3LAJ-YTSG

[26]https://perma.cc/2KQT-C5AW

每段记录时长为六小时，采样频率为 2Hz。传感器采集了多种特征，例如电流和电压、环境温度、冷却液温度、转子温度以及电机转速。具体而言，转子温度在商用车辆中难以可靠且经济地测量，因此是一个值得关注的响应变量候选项。因此，这些数据适用于工业过程和监控。

因此，我们创建了**电动机温度**数据集，首先将测量会话划分为每组连续 30 秒，即 60 个时间点。然后，我们使用记录的环境温度和冷却液温度，以及电压和电流的 d 轴和 q 轴分量作为预测变量，构建了长度为 60 的 6 维时间序列。目标变量是各自 30 秒时间段内记录的转子温度最大值。图 17 展示了此类时间序列的一个示例。最终的数据集包含 22148 个样本，其中 70% 被抽样为训练数据，其余 30% 为测试数据。

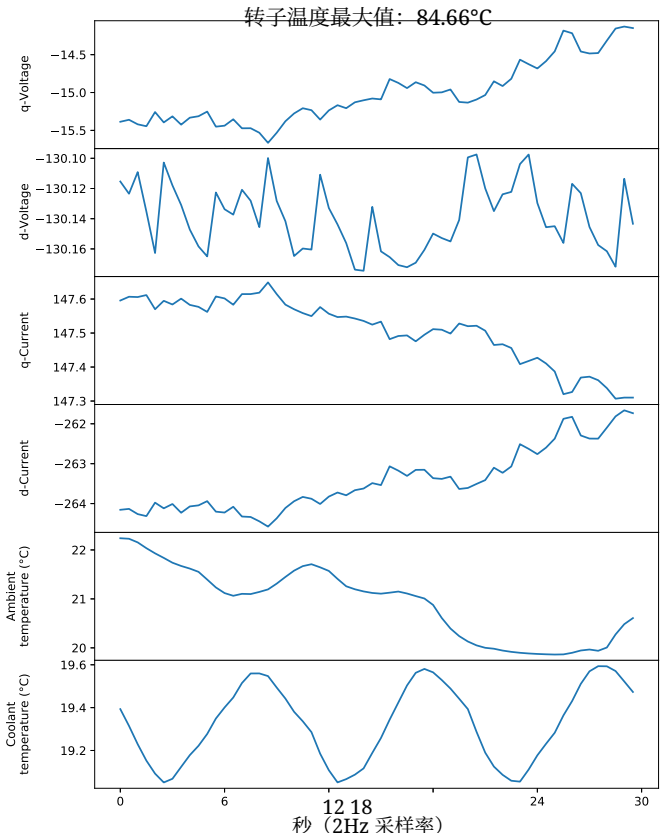


图 17. ElectricMotorTemperature 数据集中用于预测转子最高温度的电流、电压、环境温度和冷却液温度测量值示例

2）家庭气体活动监测：该数据集取自 UCI 机器学习库[27]。作者 [50] 收集了由 8 个金属氧化物气体传感器以及一个温度传感器和一个湿度传感器组成的气体传感器阵列的记录数据。传感器在三种不同条件下进行测试：葡萄酒、香蕉和背景活动。该代码仓库中整理出了两个数据集：**LPGas-家庭活动监测**和 **甲烷家庭活动监测**。第一个数据集根据湿度测量值估算液化石油气浓度。另一方面，第二个数据集根据温度测量值估算甲烷浓度。后者见图 18。

由此得到 2882 个长度为 100 的单变量时间序列。我们随机抽取其中 70% 的时间序列作为训练数据，其余 30% 作为测试数据。

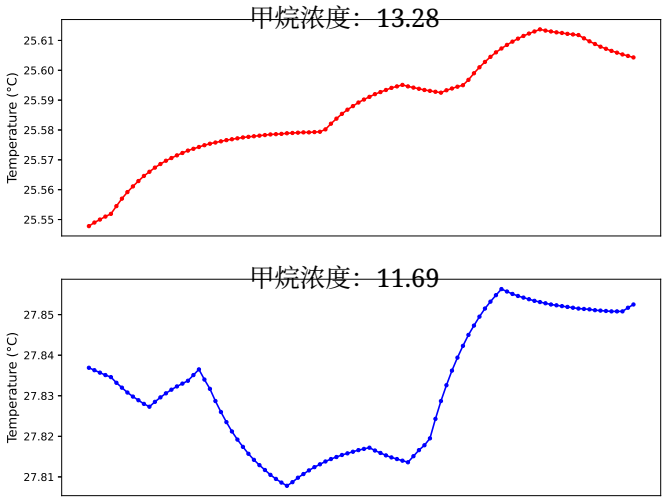


图 18. MethaneMonitoringHomeActivity 数据集中的温度数据及相应的甲烷浓度

3) 流量调制下的气体传感器阵列：Ziyatdinov 等人 [51]将由 16 个金属氧化物气体传感器组成的阵列与一台外部机械通气机结合起来，以模拟生物呼吸周期中的抽鼻行为。该研究从信号中提取了高频和低频特征，并提出了一个回归问题：预测变量为上述任一类特征，响应变量则是用于配制测试气体的两种分析物——丙酮和乙醇的浓度。作者收集的数据已发布在 UCI 机器学习库中[28]。针对这些数据开发更好的回归技术，有望改进化学感知系统中气体的早期检测。

我们使用包含 58 个样本的原始信号数据，每个样本由 16 个传感器组成，共计 928 条时间序列，并据此创建了两个数据集：**GasSensorArrayEthanol**，其目标变量是待测气体中的乙醇浓度；以及**GasSensorArrayAcetone**，其目标变量同样是丙酮浓度。两个数据集均包含 464 条长度为 7500 的单变量时间序列，并随机选取 30% 的样本作为测试集，将数据划分为训练集和测试集。

4）波面升高和缆绳张力：一个发布在 Kaggle 上的数据集[15] 包含一艘船舶的模拟数据。该数据集旨在根据随时间变化的波面升高数据预测弦的张力。虽然这些数据是模拟的，但拟合良好的模型应能提升恶劣条件下对船舶设备的监测效果。

因此，我们通过将源数据划分为长度为 57 的单变量时间序列，创建了 **波浪张力数据** 数据集。

[27]https://perma.cc/69GV-CR7S [28] https://perma.cc/WD2S-KEV9

并使用波高作为预测变量，相应的平均弦张力作为目标变量。最终得到的数据集包含 1893 个样本，其中 70% 被抽样用作训练数据，剩余 30% 用作测试数据。

E. 健康监测

1) Bar Crawl：检测重度饮酒事件：该数据集由作者 [52], 发布于 UCI 机器学习库[29]，旨在通过移动设备数据预测重度饮酒事件。数据采集自 13 名参与者的智能手机。目标是利用加速度计估算经皮酒精含量。因此，我们创建了**BarCrawl6min** 数据集，其中每个输入维度对应加速度计的一个不同轴。此外，尽管数据以 30 分钟为区间进行记录，但为了准确估算饮酒事件，仅保留了每次记录的最后 6 分钟。图 19 展示了两个生成的样本。请注意，所有数据均已完全匿名化，经皮酒精含量读数已由作者预处理和清理。

这产生了长度为360的2013维时间序列，对应于连续6分钟的逐秒测量值。我们随机抽取其中70%的时间序列作为训练数据，其余30%作为测试数据。

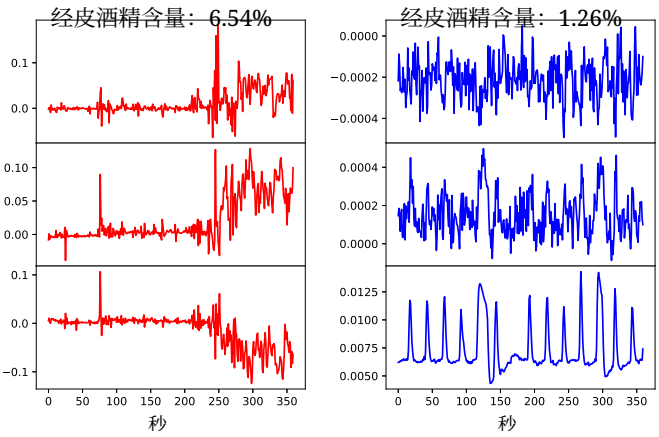


图 19. BarCrawl6min数据集中的两组加速度计数据和酒精浓度样本

2）安达卢西亚的*COVID-19*：该数据集用于估算新冠肺炎各波次期间，西班牙南部安达卢西亚不同地区的死亡率。安达卢西亚是该国面积第二大、人口最多的自治区。输出为该地区死亡人数占感染总人数的比例。该数据集名为**COVID-19安达卢西亚**，由研究 [53], 的作者公开发布，他们从安达卢西亚政府网站[30]获取了相关数据。该数据集涵盖来自34个行政区的6个波次。所有波次的长度均为91个点，因为我们选取了疫情暴发高峰前后各45天的数据；研究表明，这些数据最具参考价值。图20展示了其中两个波次。

[29]https://perma.cc/7VP5-79LN

上述数据集包含204条长度为91的一维时间序列。其中70%的时间序列被随机选入训练集，其余30%属于测试集。

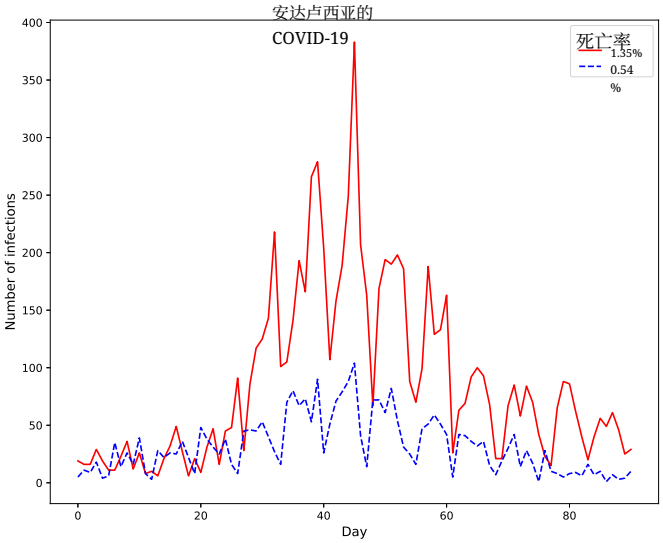


图20. COVID19安达卢西亚数据集中两个不同的新冠肺炎疫情波次及其死亡率

3) 连接镇静患者肺部的呼吸机气压：该数据集由 Kaggle 与 Google Brain[16] 合作提供。本次竞赛使用的数据是将呼吸机通过呼吸回路连接到人工风箱测试肺而生成的。目标是估计气压，该值范围为 0 到 100，表示吸气电磁阀打开的程度，以便让空气进入肺部。所提出的数据集名为 **呼吸机压力**，包含与大约 3 秒呼吸相对应的时间序列。维度是吸气电磁阀的控制输入和输出。

最终数据集包含 75450 条时间序列，每条时间序列具有 2 个维度，长度为 80。训练数据集由随机抽取的 70% 时间序列组成，其余 30% 则属于测试数据集。

F. 情感分析

1) 加密货币情感： 通过将从 EODHistoricalData[31] 提取并在 Kaggle[17]，上提供的 4 种加密货币的历史情感数据，与从 CryptoDataDownload[32]，提取的相同加密货币的历史价格数据相结合，我们创建了**比特币情感**、**以太坊情感**、**卡尔达诺情感**和**币安币情感** 数据集，样本总数分别为 332、356、107 和 263。

在全部四个数据集中，预测变量是每种加密货币一天内每小时的收盘价（以美元计）和交易量，因此形成长度为 24 的二维时间序列。响应变量是由这些时间点涵盖的当天的归一化情感得分。数据集通过从每个集合中随机选取 30% 作为测试集，划分为训练集和测试集

[30]https://perma.cc/U55Q-U4KZ

作为测试数据。利用这些数据，公司可以更好地应对公众对加密货币看法的变化。

2) 天然气价格情绪分析： 天然气价格历史数据取自美国能源信息署[33] ， 同时还获取了通过分析相关主题推文得出的对应情绪分数。基于这些数据，我们创建了 **NaturalGasPricesSentiment** 数据集。

我们首先将数据划分为若干组，每组包含连续20个工作日的数据。随后，我们使用每日天然气价格作为预测变量，并将每个20天天然气价格时间段内的平均情感得分作为响应变量。最终数据集包含93个长度为20的单变量时间序列，其中70%的数据被随机抽样用作训练数据，其余30%用作测试数据。图21展示了其中两个时间序列。同样，企业和地方政府可以利用这些数据分析和预测公众对天然气的看法变化。

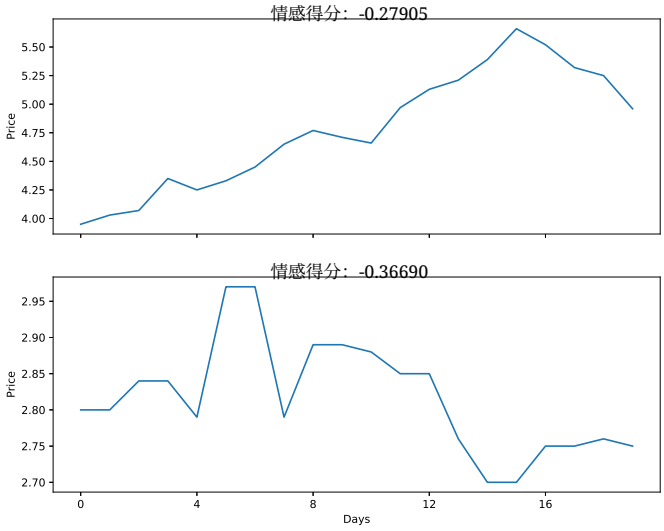


图 21. 天然气价格情感数据集中天然气价格时间序列及其对应情感得分的示例

附录B 回归器配置

表VI展示了所有算法的参数设置。

表VI 我们实验的回归器配置，其中 m 为序列长度， d 为维度数量， rm 为DRCIF表示的长度。

回归器	配置
岭回归	正则化强度： $\in \{1 \cdot 10^{-4}, 1 \cdot 10^{-3}, \ldots, 1 \cdot 10^{4}\}$
网格搜索支持向量回归	核函数 $\in$ {径向基函数,S形函数} 正则化参数 $\in \{1 \cdot 10^{-2}, 1 \cdot 10^{-1}, \ldots, 1 \cdot 10^{2}\}$ 核系数 $\in \{1 \cdot 10^{-4}, 1 \cdot 10^{-1}, \ldots, 1 \cdot 10^{-1}\}$
随机森林	估计器数量：500
RotF	估计器数量：500
XGBoost	估计器数量：500 学习率：0.1
FPCR	函数主成分分析成分数量：10
FPCR-Bs	平滑方法：B样条 函数主成分分析成分数量：10 基函数数量：10 阶数：4
ROCKET	卷积核数量：10000
MultiROCKET	卷积核数量：6250 每个卷积核的最大膨胀率：32 每个核函数的特征数量：4
1NN-ED 5近邻欧氏距离	邻居数量：1 — 5 距离度量：欧几里得距离
1NN-DTW 5近邻动态时间规整	邻居数量：1 — 5 距离度量：动态时间规整
FreshPRINCE	估计器数量：500
TSF	估计器数量：500 每棵树的区间数量： $\sqrt{m}$
DrCIF	估计器数量：500 每种表示的区间数量： $4 + (\sqrt{d}\sqrt{rm})/3$ 每棵树的特征数量：10
CNN	训练轮数：2000 批量大小：16 卷积核大小：7 编号。卷积层数：2
FCN	训练轮数：2000年 批量大小：16
Inception InceptionE	成分数量：1（inceptionE：5） 训练轮数：1500 批量大小：64 卷积核大小：40
残差网络	训练轮数：1500 批量大小：16

附录C 结果

63个时间序列回归数据集上表现最佳的回归器的均方根误差结果可在随附网站[34]上查看。

参考文献

[1] A. Bagnall、J. Lines、A. Bostrom、J. Large 和 E. Keogh，《时间序列分类大比拼：近期算法进展的综述与实验评估》，数据挖掘与知识发现，第31卷，第3期，第606–660页，2017年。

[2] M. Middlehurst、P. Sch¨afer 和 A. Bagnall，《Bake off redux：近期时间序列分类算法的评述与实验评估》，*arXiv*预印本 *arXiv:2304.13029*，2023年。[3] A. P. Ruiz、M. Flynn、J. Large、M. Middlehurst 和 A. Bagnall，《多变量时间序列分类大评测：近期算法进展的评述与实验评估》，《数据挖掘与知识发现》，第 35 卷第 2 期，第 401–449 页，2021年。[4] S. Makridakis、S. C. Wheelwright 和 R. J. Hyndman，《预测方法与应用》。John wiley & sons，2008年。[5] C. W. Tan、C. Bergmeir、F. Petitjean 和 G. I. Webb，《时间序列外部回归》，《数据挖掘与知识发现》，第 35 卷第 3 期，第 1032–1060 页，2021年。[6] A. Dempster、F. Petitjean 和 G. I. Webb，《Rocket：使用随机卷积核进行极快且准确的时间序列分类》，《数据挖掘与知识发现》，第 34 卷第 5 期，第 1454–1495 页，2020年。[7] H. Ismail Fawaz、B. Lucas、G. Forestier、C. Pelletier、D. F. Schmidt、J. Weber、G. I. Webb、L. Idoumghar、P.-A. Muller 和 F. Petitjean，《InceptionTime：为时间序列分类寻找 AlexNet》，《数据挖掘与知识发现》，第 34 卷第 6 期，第 1936–1962 页，2020年。[8] C. W. Tan、A. Dempster、C. Bergmeir 和 G. I. Webb，《MultiRocket：

[31]https://perma.cc/37GN-BMRL
[32]https://perma.cc/4M79-7QY4
[33]https://perma.cc/8AP5-5R7R
[34]https://tsml-eval.readthedocs.io/en/latest/publications/2023年/tser _ archive expansion/tser archive expansion.html.

用于快速且有效时间序列分类的多种池化算子与变换》，《数据挖掘与知识发现》，第 36 卷第 5 期，第 1623–1646 页，2022年。[9] M. Middlehurst 和 A. Bagnall，《FreshPRINCE：一种基于简单变换的流水线时间序列分类器》，载于《模式识别与人工智能国际会议》。施普林格，2022年，第 150–161 页。[10] M. Christ、N. Braun、J. Neuffer 和 A. W. Kempa-Liehr，《基于可扩展假设检验的时间序列特征提取（TSFresh——一个 Python 软件包）》，《神经计算》，第 307 卷，第 72–77 页，2018年。[11] J. J. Rodriguez、L. I. Kuncheva 和 C. J. Alonso，《旋转森林：一种新的分类器集成方法》，IEEE《模式分析与机器智能汇刊》，第 28 卷第 10 期，第 1619–1630 页，2006年。[12] H. Deng、G. Runger、E. Tuv 和 M. Vladimir，《用于分类和特征提取的时间序列森林》，《信息科学》，第 239 卷，第 142–153 页，2013年。[13] M. Middlehurst、J. Large 和 A. Bagnall，《用于时间序列分类的规范区间森林（CIF）分类器》，载于《2020 IEEE 大数据国际会议（大数据）》。IEEE，2020年，第 188–195 页。[14] M. Middlehurst、J. Large、M. Flynn、J. Lines、A. Bostrom 和 A. Bag[15] nall，《Hive-COTE 2.0：一种新的时间序列分类元集成》，《机器学习》，第 110 卷第 11 期，第 3211–3243 页，2021年。[15] C. H. Lubba、S. S. Sethi、P. Knaute、S. R. Schultz、B. D. Fulcher 和 N. S. Jones，《Catch22：规范时间序列特征》，《数据挖掘与知识发现》，第 33 卷第 6 期，第 1821–1852 页，2019年。[16] J. Goldsmith 和 F. Scheipl，《标量对函数回归中的估计器选择与组合》，《计算统计与数据分析》，第 70 卷，第 362–372 页，2014年。[17] P. T. Reiss、J. Goldsmith、H. L. Shang 和 R. T. Ogden，《标量对函数回归的方法》，《国际统计评论》，第 85 卷第 2 期，第 228–249 页，2017年。[18] H. Drucker、C. J. Burges、L. Kaufman、A. Smola 和 V. Vapnik，《支持向量回归机》，《神经信息处理系统进展》，第 9 卷，1996年。[在线]。可访问：https://proceedings.neurips.cc/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf[19] L. Breiman，《随机森林》，《机器学习》，第 45 卷第 1 期，第 5–32 页，2001年。[20] T. Chen 和 C. Guestrin，《XGBoost：一种可扩展的树提升系统》，载于《第 22 届 ACM SIGKDD 国际知识发现与数据挖掘会议论文集》，2016年，第 785–794 页。[21] Z. Wang、W. Yan 和 T. Oates，《从零开始使用深度神经网络进行时间序列分类：一个强基线》，载于《IEEE 国际神经网络联合会议论文集》，2017年，第 1578–1585 页。[22] K. He、X. Zhang、S. Ren 和 J. Sun，《用于图像识别的深度残差学习》，载于《IEEE 计算机视觉与模式识别会议论文集》，2016年，第 770–778 页。[23] A. Gu、I. Johnson、K. Goel、K. Saab、T. Dao、A. Rudra 和 C. R´e，《通过线性状态空间层结合循环模型、卷积模型和连续时间模型》，载于《神经信息处理系统进展》，M. Ranzato、A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，第 34 卷。Curran Associates, Inc.，2021年，第 572–585 页。 [在线]。可访问：https://proceedings.neurips.cc/paper/2021/file/05546b0e38ab9175cd905eebcc6ebb76-Paper.pdf[24] R. Hasani、M. Lechner、T.-H. Wang、M. Chahine、A. Amini 和 D. Rus，《液态结构状态空间模型》，2022年。[25] S. Bloemheuvel、J. van den Hoogen、D. Jozinovi´c、A. Michelini 和 M. Atzmueller，《用于多变量时间序列回归的图神经网络及其在地震数据中的应用》，《国际数据科学与分析杂志》，2022年。[26] M. A. Siddiquee、V. M. A. Souza、G. E. Baker 和 A. Mueen，《Septor：使用分层神经网络进行地震深度估计》，载于《第 28 届 ACM SIGKDD 知识发现与数据挖掘会议论文集》，2022年，第 3889–3897 页。[27] M. Bayani，《经济学中的机器学习方法研究》，博士学位论文，纽约市立大学，2022年。 [在线]。可访问：https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=6069&context=gcetds[28] M. Shokoohi-Yekta、B. Hu、H. Jin、J. Wang 和 E. Keogh，《将动态时间规整推广到多维情形需要采用自适应方法》，《数据挖掘与知识发现》，第 31 卷第 1 期，第 1–31 页，2017年。[29] C. W. Tan、A. Dempster、C. Bergmeir 和 G. Webb，《MultiRocket：用于快速且有效时间序列分类的多种池化算子与变换》，《数据挖掘与知识发现》，第 36 卷，第 1623–1646 页

，2022年。

[30] H. Ismail Fawaz、G. Forestier、J. Weber、L. Idoumghar 和 P.-A. Muller，《时间序列分类的深度学习：综述》，《数据挖掘与知识发现》，第33卷，第4期，第917–963页，2019年。[31] B. Zhao、H. Lu、S. Chen、J. Liu 和 D. Wu，《用于时间序列分类的卷积神经网络》，《系统工程与电子技术》，第28卷，第1期，第162–169页，2017年。[32] A. Bostrom 和 A. Bagnall，《用于多分类时间序列分类的二元形状片段变换》，《大规模数据与知识中心化系统汇刊》，第32卷，第24–46页，2017年。[33] L. Ye 和 E. Keogh，《时间序列形状片段：一种可实现准确、可解释且快速分类的新技术》，《数据挖掘与知识发现》，第22卷，第1-2期，第149–182页，2011年。[34] P. Sch¨afer，《BOSS关注含噪声情况下的时间序列分类》，《数据挖掘与知识发现》，第29卷，第6期，第1505–1530页，2015年。[35] P. Sch¨afer 和 U. Leser，《Weasel 2.0——一种用于快速、准确且受内存限制的时间序列分类的随机膨胀字典变换》，arXiv预印本 arXiv:2301.10194，2023年。[36] M. Middlehurst、J. Large、G. Cawley 和 A. Bagnall，《用于时间序列分类的时间字典集成（TDE）分类器》，载于《欧洲机器学习与数据库知识发现原理与实践会议论文集》，丛书《计算机科学讲义》，第12457卷，2020年，第660–676页。[37] M. Christ、N. Braun、J. Neuffer 和 A. W. Kempa-Liehr，《基于可扩展假设检验的时间序列特征提取（tsfresh——一个Python软件包）》，《神经计算》，第307卷，第72–77页，2018年。[38] A. Bagnall、M. Flynn、J. Large、J. Line、A. Bostrom 和 G. Cawley，《对于具有连续特征的问题，旋转森林是最佳分类器吗？》，arXiv预印本 arXiv:1809.06705，2018年。[39] C. Pardo、J. F. Diez-Pastor、C. Garc´ía-Osorio 和 J. J. Rodr´íguez，《用于回归的旋转森林》，《应用数学与计算》，第219卷，第19期，第9914–9924页，2013年。[40] L. Breiman、J. Friedman、R. Olshen 和 C. Stone，《分类与回归树》，1984年。[41] N. Cabello、E. Naghizade、J. Qi 和 L. Kulik，《通过监督区间搜索实现快速准确的时间序列分类》，载于《2020年IEEE国际数据挖掘会议（ICDM）》。IEEE，2020年，第948–953页。[42] M. Flynn、J. Large 和 A. Bagnall，《契约随机区间谱集成（c-RISE）：对分类器施加契约对准确率的影响》，载于《国际混合人工智能系统会议》。施普林格，2019年，第381–392页。[43] E. J. Keogh 和 M. J. Pazzani，《导数动态时间规整》，载于《2001年SIAM国际数据挖掘会议论文集》。SIAM，2001年，第1–11页。[44] B. D. Fulcher 和 N. S. Jones，《hctsa：一种利用大规模特征提取进行自动化时间序列表型分析的计算框架》，《细胞系统》，第5卷，第5期，第527–531页，2017年。[45] L. M. Candanedo 和 V. Feldheim，《利用统计学习模型，根据光照、温度、湿度和 CO_2 测量值准确检测办公室房间占用情况》，《能源与建筑》，第112卷，第28–39页，2016年。[46] A. Salam 和 A. El Hibaoui，《用于电力消耗预测的机器学习算法比较：得土安城市案例研究》，载于《2018年第六届国际可再生与可持续能源会议（IRSEC）》。IEEE，2018年，第1–5页。[47] X. Liang、T. Zou、B. Guo、S. Li、H. Zhang、S. Zhang、H. Huang 和 S. X. Chen，《评估北京PM2.5污染：严重程度、天气影响、亚太经合组织和冬季供暖》，《皇家学会会刊A辑：数学、物理与工程科学》，第471卷，第2182期，第20150257页，2015年。[48] D. H. Stolfi、E. Alba 和 X. Yao，《预测智慧城市中的停车场占用率》，载于《智慧城市：第二届国际会议，Smart-CT 2017》。施普林格，2017年，第107–117页。[49] R. Osterhuber 和 A. Schwartz，《中央内华达山脉雪研究实验室1971年至2019年水文年的积雪层、降水和温度测量值》，2021年，Dryad。[50] R. Huerta、T. Mosqueiro、J. Fonollosa、N. F. Rulkov 和 I. Rodriguez-Lujan，《化学传感器中湿度与温度的在线去相关及连续监测》，《化学计量学与智能实验室系统》，第157卷，第169–176页，2016年。[51] A. Ziyatdinov、J. Fonollosa、L. Fern´andez、A. Gutierrez-G´alvez、S. Marco 和 A. Perera，《通过化学感知系统中的气流调制实现仿生早期检测》，《传感器与执行器B辑：化学》，第206卷，第538–547页，2015年。[52] J. A. Killian、K. M. Passino、A. Nandi、D. R. Madden、J. D. Clapp、N. Wiratunga、F. Coenen 和 S. Sani，《学习如何——

检测重度饮酒事件所使用的智能手机加速度计数据。”发表于 *KHD@ IJCAI*，2019年，第 35–42 页。 [在线]。可访问：https://ceur-ws.org/Vol-2429/paper6.pdf[53] M. D´ıaz-Lozano、D. Guijo-Rubio、P. A. Guti´errez、A. M. G´omez-Orellana、I. T´u˜nez、L. Ortigosa-Moreno、A. Romanos-Rodr´ıguez、J. Padillo-Ruiz 和 C. Herv´as-Mart´ınez，“基于曲线分解和进化人工神经网络的新冠肺炎传染预测框架：安达卢西亚、西班牙的案例研究”，《专家系统与应用》，第 207 卷，第 117977 页，2022年。[54] R. Ghosh，“带有 Twitter 情感得分的天然气价格”，2022年，Kaggle。[55] J. Dem˘sar，“多个数据集上分类器的统计比较”，《机器学习研究杂志》，第 7 卷，第 1–30 页，2006年。[56] S. Garc´ıa 和 F. Herrera，“对‘多个数据集上分类器的统计比较’的扩展：适用于所有成对比较”，《机器学习研究杂志》，第 9 卷，第 2677–2694 页，2008年。[在线]。可访问：https://www.jmlr.org/papers/v9/garcia08a.html[57] A. Benavoli、G. Corani 和 F. Mangili，“我们真的应该使用基于平均秩的事后检验吗？”《机器学习研究杂志》，第 17 卷，第 1–10 页，2016年。 [在线]。可访问：https://jmlr.org/papers/v17/benavoli16a.html[58] I. Yeo 和 R. A. Johnson，“用于改善正态性或对称性的新型幂变换族”，《生物计量学》，第 87 卷，第 4 期，第 954–959 页，2000年。[59] W. Kirchg¨assner、O. Wallscheid 和 J. B¨ocker，“利用深度残差机器学习估计电动机温度”，*IEEE*《电力电子学汇刊》，第 36 卷，第 7 期，第 7480–7488 页，2021年。



David Guijo-Rubio 于2016年获得计算机科学理学学士学位，并于2017年获得人工智能理学硕士学位。2021年，他从西班牙东安格利亚大学获得计算机科学博士学位。他目前在科尔多瓦大学和英国东安格利亚大学担任博士后研究员。他目前的研究兴趣包括序数分类、模式识别，以及应用于时间序列的不同任务（分类——名义分类和序数分类、聚类与外部回归）。



Matthew Middlehurst 于2018年获得计算机科学理学学士学位，并于2023年从英国诺里奇的东安格利亚大学（UEA）获得计算机科学博士学位。他目前在UEA的时间序列机器学习小组担任高级研究助理。他的研究兴趣主要与时间序列分类相关，近期扩展至时间序列聚类和时间序列外部回归。他致力于为时间序列相关领域的研究人员提供和维护开源软件。



Guilherme Arcencio 目前正在巴西圣卡洛斯联邦大学（UFSCar）攻读计算机工程理学学士学位。˜圣卡洛斯、巴西。他曾在东安格利亚大学（UEA）时间序列机器学习组担任研究实习生。他目前的研究兴趣包括时间序列机器学习，重点关注外部回归，以及开源软件的开发与维护。



Diego Furtado Silva 获得了圣保罗大学（USP）数学与计算机科学研究所（ICMC）的计算机科学理学学士学位，以及计算机科学与计算数学理学硕士和博士学位，研究方向为时间序列与数字信号分析。目前，他是圣保罗大学（USP）ICMC的助理教授，并与计算智能实验室（LABIC）以及圣卡洛斯联邦大学（MIDAS-UFSCar）的数据挖掘与应用组合作。他的主要研究研究兴趣包括机器学习基础及其应用。



Anthony Bagnall 于2001年获得英国诺里奇东安格利亚大学（UEA）的计算机科学博士学位。自2018年起，他担任UEA计算机科学正教授，并领导时间序列机器学习组。他的主要研究兴趣是时间序列机器学习，早期主要关注分类，近来则更多研究聚类和回归。他还对集成设计抱有兴趣。