arXiv:2302.09293v1 [cs.CY] 15 Feb 2023

文章

周期性强度揭示时间序列数据的洞见：三个应用案例

Alan F. Smeaton [1]、和 Feiyan Hu [1]

[1] I数据分析洞察中心，都柏林城市大学，Glasnevin，都柏林 9，爱尔兰；alan.smeaton@dcu.ie
* C通信作者：alan.smeaton@dcu.ie

摘要： 周期性现象是许多自然产生的时间序列中存在的振荡信号。周期图可用于测量整个时间序列中不同频率的振荡强度，但有时我们希望测量某一特定频率的周期性强度如何随时间序列变化。具体做法是在一个窗口内计算周期性强度，然后移动窗口并重新计算强度，从而显示某一特定频率的周期性强度如何贯穿整个序列发生变化。我们通过三个应用案例对此进行说明。第一个案例是新生小牛群的活动，我们展示了24小时周期性强度如何在整个牛群中同步增强和减弱。我们还展示了，从家庭传感器中检测到的活动的24小时周期性强度变化，如何反映整体健康状况。对此，我们使用了从23位老年人的住所中分别收集的数周传感器数据。第三个应用案例关注数百名大学生通过虚拟学习环境（VLE）访问在线资源时所表现出的7天周期性强度，以及他们每周学习行为的规律性如何在整个教学学期中发生变化。本文证明，周期性强度能够揭示时间序列数据中通过其他分析形式无法观察到的洞见。

关键词： 周　期性 强度；周期图；昼夜节律 rh　　ythm；

check for updates

C引用： Smeaton, Alan F.; Hu, F.
P周期性强度揭示的洞见
in时间序列数据：三种用
C途。*Preprints* **2023年**，*1*，0。
https://doi.org/
P

出版商说明： MDPI对于已发表地图和机构隶属关系中的管辖权声明保持中立。

版权： © 2023年 作者所有。被许可方为 MDPI，瑞士巴塞尔。本文章根据知识共享署名（CC BY）许可（https://creativecommons.org/licenses/by/ 4.0/）的条款和条件作为开放获取文章发布。

1. 引言

数据分析应用于数据流或时间序列的一个常见用途，是检测以某种方式偏离正常行为的异常行为或异常值，其中正常行为由数据流中的重复模式定义。正如 Li 等人 在[1] 中指出的，这对于欺诈检测、网络入侵检测、医学诊断、视频监控、故障诊断等应用都很有用。根据定义，离群点检测会检测数据中的基线模式，然后检测异常值或基线模式之外的细微变化，并向我们发出警报。

周期性是指某种模式在数据流中以固定间隔重复出现的特征或倾向。该术语用于描述自然界中作为正常行为反复出现的现象的规律性，而偏离这种规律的情况则被视为异常值。这种时间重复是复杂系统所具有的一项重要属性，对这种时间重复的研究有助于深入了解系统的行为[2]。

地球自转、月球周期、地球绕太阳公转、季节、潮汐以及我们每日24小时周期的节律都具有规律的周期性。人类以及大多数其他生物都具有强烈的昼夜节律，这种24小时周期为我们的生活带来规律和结构。时间生物学是生命科学领域的一部分，研究这一24小时周期如何影响生物体的健康、健康福祉和行为[3]。在人类的许多行为中都可以观察到24小时的周期性，因为我们努力维持稳态——即生理机能的平衡，以及面对外部和环境变化时的稳定性[4]。众所周知，当我们以与这种节律高度一致的方式生活和行动时，地

昼夜节律时钟的周期性，这意味着健康状况更佳，整体健康状况得到改善 [5]。

检测数据流中重复出现的一种方法，是通过递归图（RPs）将其可视化 [6]。递归图是对一个矩阵的可视化，其中矩阵元素对应于动态系统中发生重复出现的实例。递归图首先对数据执行聚类分析，然后构建图形模式，以展示数据中的重复出现。递归图已应用于许多领域，例如机械设备的故障诊断[7]以及图像变换 [8]。

递归图适用于分析较短且重复出现次数有限的序列；但对于包含大量重复迭代的较长时间序列，其设计目的在于隐藏迭代次数，并掩盖迭代之间可能出现的变化。有时，当我们在时间序列数据中寻找并发现模式时，我们希望对这些模式进行测量，并比较它们随时间的变化，甚至将其与其他重复模式进行比较。这远不止是利用模式来检测异常值。如果我们能够测量数据集各个位置上的模式及其重复出现的强度，那么这或许能够揭示有关这些模式的洞见，以及对数据更深层次的认识。

在先前的工作中，我们开发了一种用于计算和可视化时间序列中给定频率下周期性强度的技术。该技术应用于美国退伍军人的睡眠模式；这些参与者连续90天佩戴腕戴式加速度计，我们发现，参与者之间的24小时周期性强度存在显著差异 [9]。该研究还发现，较高水平的周期性表明睡眠和活动模式更加规律，并且与较低的低密度脂蛋白胆固醇、甘油三酯和高敏C反应蛋白指标，以及更高的健康相关生活质量指标相关 [9]。同样的、用于可视化周期性强度随时间变化的技术，也被用于分析生活日志数据；生活日志数据被定义为对日常活动进行自动数字记录，该技术揭示了研究对象潜在行为的变化 [10]。

尽管 Chegini 等人 [11] 在其关于机器轴承故障诊断的研究中使用了“周期性强度”这一相同术语，但他们将周期性强度定义为故障机械设备中振动信号自相关函数的局部极大值能量与总能量之比，这与本文所采用的周期性定义大不相同。

我们的研究假设是，针对一系列纵向时间序列数据计算并可视化 24 小时周期性的强度，将揭示通过其他形式的分析无法观察到的数据洞见。在本文中，我们将周期性强度的计算应用于三个不同的应用案例，并展示在每个案例中，分析如何揭示数据中更深层次的洞见。该论文的重要性在于，这项技术及其计算过程相对简单，同时能够为时间序列数据提供强有力的洞见。对于任何底层时间模式会重复出现，且确定这些模式在整个数据中的强度十分重要的数据序列，都可以使用该技术。

2. 材料与方法

2.1. 使用时间滞后重叠窗口计算周期性强度

周期图 [12] 是一种成熟的技术，用于探究时间序列数据中不同频率下重复模式的幅度。周期图利用时间序列数据的样本，测量底层连续函数的振幅与频率特征，从而揭示该底层函数的谱功率密度。如果函数的采样不均匀，或样本中存在缺失数据，则 Lomb–Scargle周期图是一种广为人知的算法，能够处理此类不均匀采样数据 [13]。

Lomb–Scargle周期图已广泛应用于那些关注探索周期性的频率和幅度的领域，例如

天文观测数据。在本文中，我们仅关注数据样本在已知频率（例如24小时昼夜节律）下的大小或振幅。Lomb-Scargle周期图的软件实现可在多个平台上轻松获得，包括MatLab和Python。

为了说明我们如何使用时滞重叠窗口，计算时间序列中给定频率下周期性变化的幅度，我们将通过图1所示的示意图进行说明。

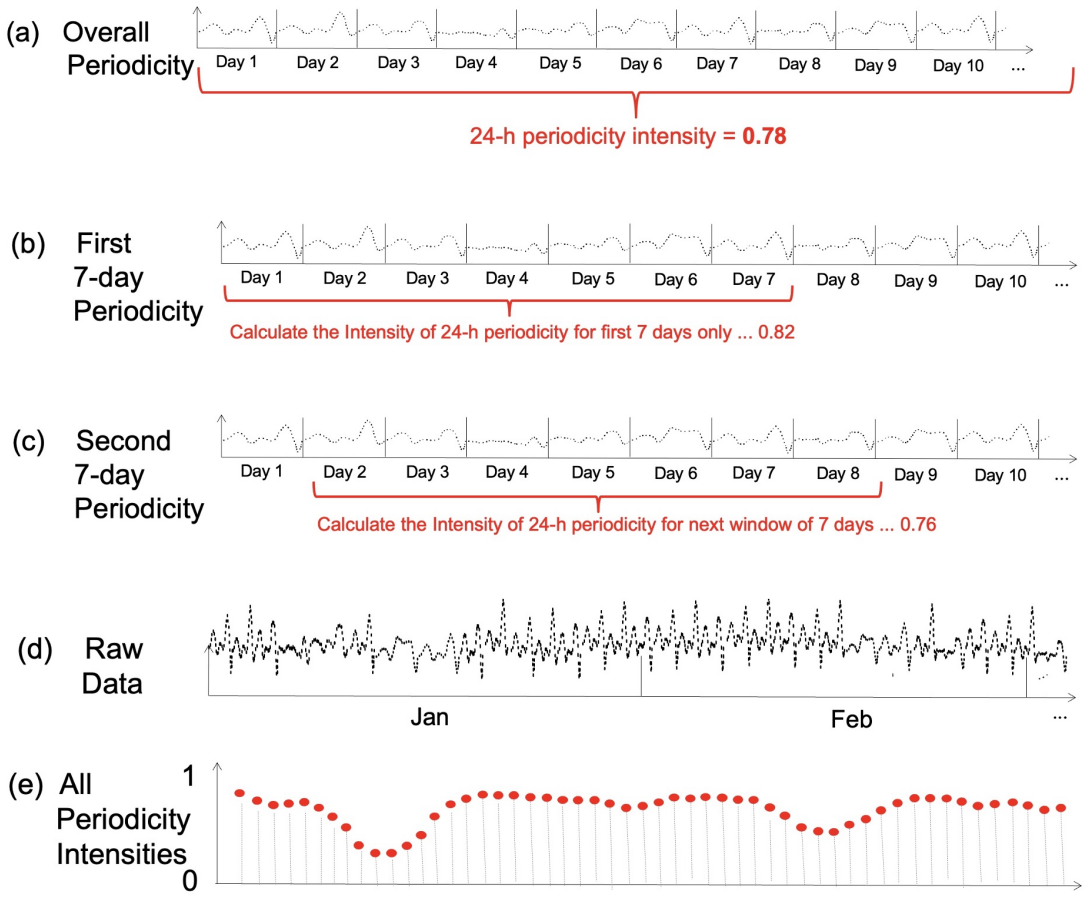

图1。 使用时滞重叠窗口计算时间序列中周期性强度的示意图。

图1(a) 显示了一个较长的（合成）数据点序列的前10天绘图。从可视化结果中可以看出，存在明显的重复性日模式，即频率为24小时的模式；不过，仔细观察会发现，每一天与其他天相比都略有不同。如果我们计算整个时间序列中这种24 小时周期性的幅度，其值可以假定为0.78，取值范围为0到1。

虽然这种分析具有一定价值，但如果我们拥有很长时间跨度的数据（例如数周、数月或数年），那么在整个时间段内计算一个代表 24 小时周期性的单一数值（图中的 0.78），将掩盖数据中的任何变化细节。相反，如图1(b)所示，我们仅使用最初 7 天数据的窗口来计算 24 小时周期性，由此得到一个数值，例如 0.82。然后，我们将这个 7 天窗口沿时间向前移动一个时间间隔，例如完整的 1 天，并重新计算 24 小时周期性，如图1(c)所示

这样得到的数值例如为 0.76。我们重复移动 7 天窗口并重新计算的过程，直到到达时间序列的末尾。

这一过程中有三个重要参数：(1) 重复模式Fo的频率；如果我们关注的是昼夜节律，通常为 24 小时，但并非总是如此；(2) 计算周期性时所使用的窗口大小m；以及 (3) 窗口之间的重叠l，也称为步长，即在计算周期性窗口的新值之前，窗口移动的距离。

对于一系列 N 个数据点（如图 1(d) 所示），若窗口大小为 m，每次重新计算周期性时的移动距离或步长为 l，则会生成一个包含 $\lfloor \frac{N-m+1}{l} \rfloor$ 个数据点的时间序列，如图 1(e) 所示。这能够捕捉随时间采样的数据规律性逐渐发生的行为变化，而不是个别特别异常的数据值。周期性强度图只是时间序列数据分析的众多可能形式之一，但它与其他形式各不相同。虽然我们确实会对数据进行谱分析 [14]，但并不是在完整频谱范围内进行。相反，我们关注 24 小时昼夜周期，因为先前研究表明，昼夜周期性在纵向加速度计数据中承载了大部分能量 [15]。同样，我们也不会像 [16] 对蛋白质序列所做的那样，将时间序列数据映射到另一个空间，例如嵌入空间。离散小波变换 [17] 可用于时间序列数据的噪声滤波和数据降维，尤其用于检测异常值等奇异点 [18]，但这不同于我们的用例，因为我们旨在识别长期行为变化。希尔伯特变换（经验模态分解） [19] 通常用于从时间序列中获取振幅和相位等频率特征信息，但不能用于识别时间序列特征的长期变化。过零率（ZCR） [20] 广泛用于语音和音乐音频处理，对应于信号从正值转为负值或从负值转为正值的转换速率；而历史波动率 [21] 是经济学和金融学中的一种统计度量，用于衡量时间序列中相邻值的变化程度。后面这些分析形式都能测量时间序列中的动态变化，但无法像我们的度量一样，给出 24 小时周期性强度的纵向变化。

这些信号处理技术广泛应用于各种场景，例如用于支持深度学习的数据增强 [22]，以及 [23]，其中利用从机械设备中轴承的振动和声发射信号导出的时域、频率域和谱功率域特征向量来检测故障。经验模态分解（EMD）是一种将信号分解为有用成分的技术，类似于傅里叶变换和小波等分析方法，尤其适用于数据的探索性分析。EMD的基本原理是将原始时间序列分解为一组平滑曲线或本征模态函数。在故障诊断中，它通常用于将振动信号分解为用于训练分类器的成分，且通常具有很高的准确率。然而，众所周知，EMD对噪声较为敏感，并且需要手动选择参数[24] 以支持其探索性特征。过零密度分解（ZCD）是一种较新开发的技术 [25]，其基于过零点概念构建，在计算效率方面优于EMD。然而，在处理非平稳信号时，就准确性和鲁棒性而言，EMD比DD更有效 [26]。我们的算法不会像EMD或ZCD那样将时间序列分解为各个组成部分；由于我们使用Lomb-Scargle周期图 [13] 来计算周期性，因此我们的方法对噪声不那么敏感。Lomb-Scargle周期图能够容忍不规则采样和缺失数据。

算法1 展示了计算周期性强度的方法。我们假设已知最优窗口大小m、步长l 以及已识别且感兴趣的频率Fo。在本文的案例研究中，Fo 可以是昼夜节律（24 小时）或每周周期性。X_w 是窗口化输入数据，而Sf 是通过快速傅里叶变换计算得到的周期图。强度函数通过以下方式计算强度：

对窗口化/局部时间序列数据和一个关注频率*Fo*进行周期图计算。对于能量泄漏的情况，我们汇总*Fo* 周围所有相近频率所携带的能量。

算法1： 计算周期性强度

输入： 时间序列数据 $x_n \in X$，其中 $n = 1$、2、3.....N，包含 N 个数据点。最优窗口大小 m 和步长 l。对于 $\in \{124_h$、$1168_h\}$——

输出： 时间序列数据 $y_t \in Y$，其中 $t = 1$、2、3、..... $\lfloor \frac{N-m+1}{l} \rfloor$。

对于 $t \leftarrow 0$ **到** $\lfloor \frac{N-m+1}{l} \rfloor$，**执行**

  $X_w = x_{k\times l} \ldots x^{k\times l+m}$

  Sf = 快速傅里叶变换（X_w）

  y_t = 强度函数（Fo，Sf）**其中**

  在本研究中，强度函数 = $\sum_{f \in \hat{} Fo} S \quad f$其中 $\hat{Fo}$ 是满足频率条件的Fo 的邻近频率其频率满足 $|{}^{1}Fo _ \ {}^{1\hat{}}Fo| < 0。01。$ —

  IntensityFunc 也可以采用 [10] 中定义的相同形式

结束

2.2. 计算周期性强度的应用案例

2.2.1. 老年人家庭传感器数据中的周期性强度

NEX 项目 [27] 开发了一个集成式物联网系统，为老年人提供非侵入式健康与健康福祉监测，以支持他们独立居家生活。该项目在住宅各处以非侵入方式部署环境传感器，用于检测各种家务和居家活动。通过汇总和分析每个传感器采集的时间序列数据，可以建立个人日常活动模式的模型。2022 年上半年，24 名年龄在 60 岁及以上且独立居家生活的健康老年人参与了 NEX 项目的试验。参与者以女性为主（81%，n=21），总体平均年龄为 73.2 岁。大多数参与者具备独立生活能力且功能状况良好，仅有 8%（n=2）报告在完成日常生活活动（ADLS）[28]（如穿衣等）时存在困难；另有 4%（n=1）报告在完成被定义为工具性日常生活活动（IADLS）的更复杂任务（如购买杂货等）时存在困难。

在每位参与者的家中都安装了一系列家庭传感器，包括：（1）安装在橱柜门或抽屉上的接触传感器，这些橱柜或抽屉用于存放厨房餐具、刀叉餐具、餐盘、主食和锅具；（2）安装在冰箱门上的接触传感器；以及（3）安装在水壶、微波炉和烤面包机上的电源插头传感器。此外，卧室衣柜门和抽屉上也安装了接触传感器，浴室内安装了用于监测温度、湿度、光照和存在感应的六合一环境传感器，前门、露台门或后门上也安装了接触传感器。总体而言，这些传感器采集的数据反映了参与者有目的的行为，即表明某种刻意的动作，例如开门或关门、人员移动，以及电器设备的开启或关闭。传感器数据来自24位参与者家中的家庭传感器，每个家庭的数据采集时长从6周到6个月不等 [27]。

通过计算家中多个传感器设备所采集的全部数据中的 24 小时周期性强度，可以展示 24 小时昼夜节律的重要性，以及 NEX 项目每位参与者对该 24 小时节律的遵循程度。融合来自多个 Sensors 的多条数据流，可以捕捉包括食物准备和进食、个人清洁、清洁、放松等在内的日常生活活动。周期性强度强调这些活动在家中的规律性及其定期重复发生的情况。

周期性强度可视化被呈现给 NEX 项目的参与者，用于展示他们的长期行为及其变化，以及这些趋势随时间推移所呈现的规律性。周期性强度图是一条连续时间线，其中高值表示家庭传感器检测到的规律行为时段，低值表示不规律活动时段。这些不规律活动是规律习惯发生变化的指标，并可能指向以下方面的变化：睡眠模式（入睡时间、持续时间、夜间醒来、日间小睡）；进出住所以及在住所周围活动的情况（夜间保持清醒、白天睡觉、在家停留时间比正常更长或更短）；个人清洁（在不规律的时间使用卫生间）；以及进食（白天或夜间吃零食，而不是规律用餐）。

对于 NEX 试验，我们融合了从参与者家中 Sensors 收集的数据，并为每位参与者计算出一条总体周期性强度时间线；本文将在结果部分展示其中一些时间线。

2.2.2. 学生学习习惯中的周期性强度

许多大学都设有健康促进项目，以促进教职员工和学生的整体健康与福祉 [30]。这些项目的益处包括提高健康素养、形成更健康的心态、改善与他人的关系，以及在锻炼、营养等生活方式方面带来积极变化。在我们大学，我们创建并开设了一门名为 FLOURISH 的本科生选修课，主题是健康福祉。该课程还包含数据素养方面的内容，通过向学生展示如何将从用于测量睡眠、锻炼、营养等情况的设备中收集的个人数据善加利用，帮助学生自行监测这些健康福祉因素。

这门健康福祉课程完全通过大学的虚拟学习环境（VLE）Loop 在线授课。Loop 是 Moodle 的一种变体，并获得了大学认证；成功完成课程的学生可获得 5 个欧洲学分转换与累积系统学分（ECTS）。对于 10 个健康福祉主题中的每一个，学生都必须收集与自身相关的个人数据，但这些数据不会与课程教师或大学共享。然而，作为课程参与的一部分，为展示大学在记录学生在线活动过程中收集的个人数据同样可以得到善加利用，该课程的 169 名学生均授权我们下载他们在该学期所有授课课程中的虚拟学习环境（VLE）访问日志。这些日志包括每次学生访问虚拟学习环境（VLE）上的任何资源时所对应的日期、时间以及 Loop 资源的 URL。

在教学学期结束时，我们使用所有学生访问其所有课程中的所有 Loop 资源的日志，为每名学生生成一张涵盖全部在线访问的周期性强度图，并将该图私下在线分享给他们。生成的图表能够帮助每名学生了解，在此前的授课学期中，他们通过访问在线资源所体现出的学习习惯在哪些时期更具结构性和规律性。这可能有助于他们回顾整个学期的学习情况。本文的结果部分展示了一些用于说明这一点的示例图表，以及进一步的分析结果。

2.2.3. 小牛运动中的周期性强度

睡眠-觉醒周期的规律性不仅对人类很重要，对其他有生命的动物同样重要。我们研究了商业农场中新生犊牛的昼夜周期性强度，以及这种周期性强度在其出生后的最初 6–8 周内如何变化 [31]。包括牛在内的许多动物在行为上都与牛群同步，牛群会协调对外部授时因子的反应 [32]，因此，我们对这一用例的兴趣在于，既关注整个牛群的昼夜节律周期性强度，也关注单个小牛的周期性强度。

我们收集了来自 Axivity AX3 可穿戴式加速度计的数据 [33] 该设备固定在项圈上，项圈佩戴于 24 头奶牛犊的颈部，从它们出生后不久便开始记录。小牛每天喂食两次，每天铺设垫料，并在这些活动期间与人接触。由于它们食用全脂牛奶，喂食时间与挤奶时间相近，因此它们的日常活动会非常有规律且结构化。加速度计检测到的单个小牛运动会因它们玩耍、互动以及探索周围环境而有所不同，因此，尽管喂食和睡眠具有规律的结构，它们之间仍可能存在很大的个体差异。

加速度计的数据以 12.5Hz 的频率进行采样，每头小牛最多收集 8 周。数据按照 [34] 中所述的方法进行处理，首先预处理为信号矢量幅值（SVM），这是一种与传感器方向无关的时间序列，因此不受项圈在颈部周围移动的影响。随后应用巴特沃斯四阶带通滤波器 [35]，其频率范围为 0.5 Hz 至 20 Hz，以去除白噪声，并将负值转换为绝对值。用于分析周期性的运动值，是在不重叠的 60 秒窗口上计算得到的 SVM 聚合平均值。

3. 结果

3.1. 老年人住所中的周期性强度结果

我们首先考察并讨论对老年人住所传感器数据进行周期性强度计算的结果。NEX 项目参与者数据的周期性检测输出是在实时计算后在线提供的。这些可视化结果由其临床护理人员和项目研究人员在与参与者进行总结说明时使用，并辅以一段简短的信息性 YouTube 视频，用于解释图表并确保参与者理解如何解读这些图表。

图 2 中展示了三个示例输出，其中每个图表中的数值都归一化到一个 [0,1]范围内，因为不同参与者的实际周期性强度值会有很大差异，而且每位参与者关注的是自己图表中的相对变化，而不是绝对值或与他人的比较。经过实验确定最佳参数后，对于每个图表，周期性计算的频率为 24 小时，这反映了对昼夜节律的关注；窗口为 7 天，从而消除周末与工作日之间的变化；重叠部分或步长为 1 小时，使计算更加快速。每个图表表示 3 个月期间的周期性强度。

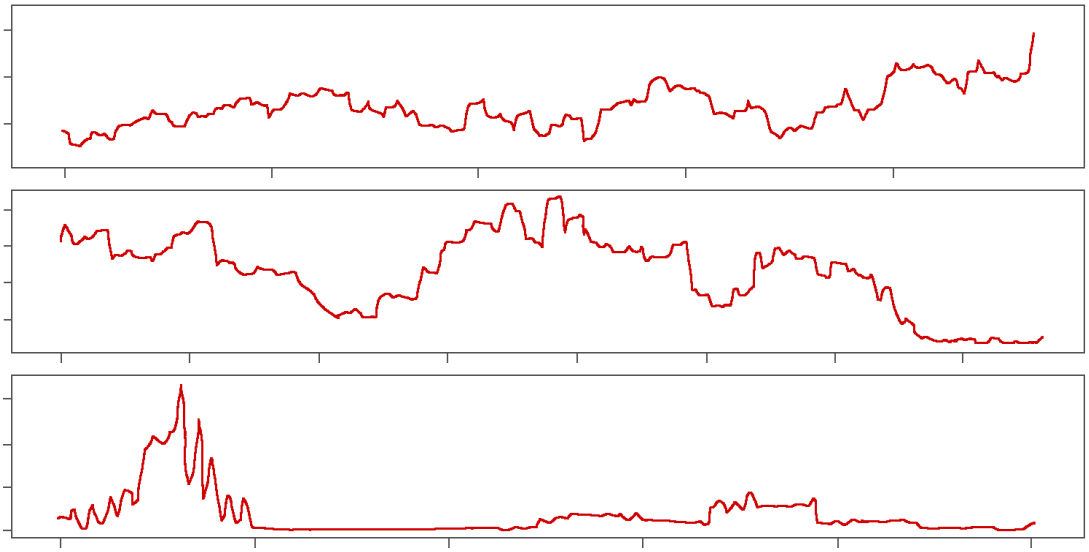

图2。 3名ARC试验参与者为期3个月的归一化周期性强度。频率为24小时，窗口大小为7天，窗口移动步长为1小时。横轴数值跨度为3个月，纵轴数值已归一化至0到1的范围。

图 2 中的第一个图表显示，一名参与者的生活方式相对稳定且规律，周期性强度仅有小幅变化。在 3 个月末期，周期性强度逐渐上升，表明生活方式规律性略有改善；整个 3 个月期间没有出现重大扰动。已知该参与者拥有健康、规律且舒适的生活方式。

图2 中的第二个图表显示，一名参与者的生活方式起伏明显更多。监测期中段生活方式规律性的提高，与该参与者因感染新冠肺炎而自我隔离、康复期间一直在家，并保持规律行为和习惯相对应。在此期间，传感器激活次数远低于前一阶段，因为该参与者在家中除了卧床休息外几乎没有进行其他活动；但促成周期性强度的并不是传感器激活的次数，而是其发生时间。在为期3个月的监测期接近结束时，该参与者搬离住所，前往家人处休养，因此末段呈现平线。

图 2 中的第三个图表显示，一名参与者在为期 3 个月的记录期间前段也感染了新冠肺炎，随后不久还搬离家中，与家人一起疗养，这一点通过图表中的持平线表现出来。3 个月后半段周期性强度的轻微上升表明，家人会以半规律的频率到其住所取邮件，并检查住所的安全状况。

这项分析的结果是一种对个人的自我监测，但能够更深入地了解其整体行为和活动。当研究对象试图在生活中保持较高水平的节律性时，这种方法可用于促进或维持长期行为改变。

3.2. 学生学习习惯中的周期性强度结果

现在，我们将深入分析和讨论将周期性强度应用于学生在线学习习惯所得的结果。在教学学期结束时，每位学生都可以查看自己的周期性强度图；同时，我们还提供了一段简短的 YouTube 讲解视频，用于解释这些图表，并确保他们理解如何对其进行解读。图 3 展示了五个示例图表。在每个图表中，周期性强度值均被归一化到 [0,1] 范围，以平滑不同学生之间的方差，并使每个图表都具有同等的可读性。这与前文研究中的成年参与者情况类似，因为每位学生关注的将是自己图表中的相对变化，而不是绝对值或与他人的比较。

研究表明，每位学生都会关注自己图表中的相对变化，而不是绝对值或与他人的比较。我们还在每个图表中添加了 3 条彩色竖条：红色表示周期性强度最低的时期，绿色表示最高的时期，黄色表示变化幅度最大（最陡峭）的时期。计算图表所使用的参数为：频率 24 小时，窗口大小 7 天，窗口移动步长或步长为 3 小时。

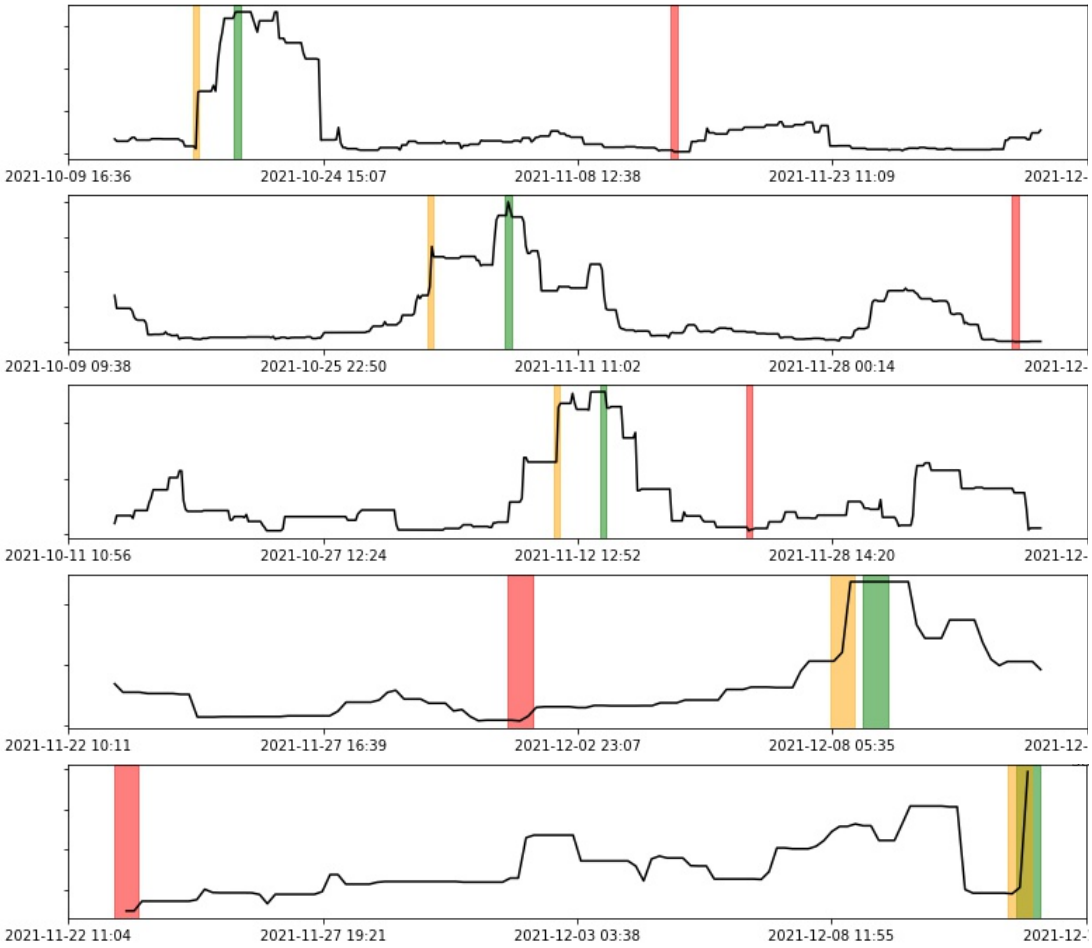

图3。 五名学生参与者样本的标准化周期性强度。频率为24小时，窗口大小为7天，窗口移动步长或步长为3小时。纵轴数值已标准化至0到1的范围。

图 3 所示的五个图表具有不同的起始日期，反映了学生开始学习的时间各不相同；这些图表也反映出完全不同的学习模式：就在线访问的结构化和规律性而言，第一、第二、第三和第四名学生在学期中的不同时间达到峰值。第一名学生在学期最早达到峰值，而第四名学生最后达到峰值。第五名学生的图表显示，该学生在整个学期中的学习规律性逐渐提高，并在课程考试前夕，也就是学期末，出现了一段规律学习的高峰期。

虽然每位学生的独立周期性强度图都具有一定价值，但如图 4 所示，通过对这 169 张学生独立图表进行时间对齐并堆叠，我们可以从这项分析中获得更多价值。为此，我们在堆叠之前设置一个固定的开始和结束日期/时间

图表。由于周期性计算所使用的窗口为7天，因此图表的开头和结尾各有3.5天的空缺。

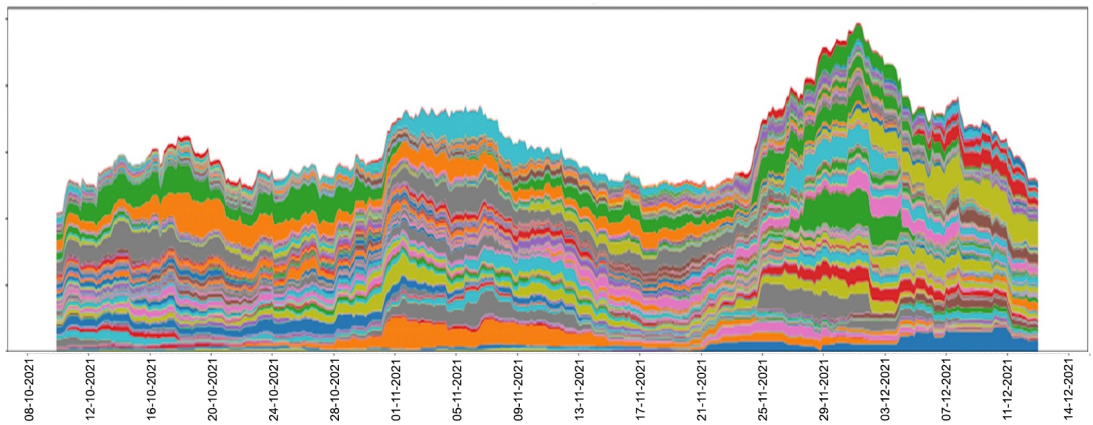

图4。 全部169名学生参与者的累积周期性强度堆叠折线图。纵轴数值已标准化至0到1的范围。

图 4 展示了169名学生这一学生群体的汇总周期性强度，其中纵轴是每名学生周期性强度的堆叠，每名学生以不同颜色显示。整个图表的顶端折线展示了所有学生的汇总周期性，并显示出三个明显的峰值，分别对应学期初学生适应大学生活的阶段、学期中学生利用学期中阅读周补上学习进度的阶段，以及学期末考试前学生准备考试的阶段。

正如 Byron和Wattenberg 在 [36], 中所示，堆叠折线图可能会造成峰值与谷值的错觉，而这些峰值与谷值可能是由部分学生造成的，并非所有学生都同时出现。因此，整体堆叠折线图可能会因少数学生的峰值或谷值而被人为抬高或压低。为评估这一点，我们将学生随机分成若干子组，并为每个子组绘制堆叠折线图。由于这些子组的堆叠折线图与整个学生群体的堆叠折线图呈现出相同的形状，因此不存在由某个学生子组造成的人为峰值或谷值，这一模式在整个学生群体中具有普遍性。

我们可以从图 4 中看到，11月1日之前出现了一条断层线。该大学的学术日历显示，这一天是学生课程选择登记截止日期。11月25日之前又出现了第二条值得关注的断层线。该学期的授课阶段于11月27日结束，考前学习阶段为11月29日至12月5日。我们可以看到，随着学生提交作业并复习学习材料，为考试做准备，虚拟学习环境活动呈现出更强的结构化特征；随着每名学生完成考试，这种趋势逐渐减弱。

对周期性强度进行分析的结果，是为每名学生提供有关其学习习惯的个性化反馈，并深入洞察整个学生群体的学习行为和活动。这不仅有益于每名学生，也有益于大学和课程负责人。

3.3. 犊牛运动中的周期性强度结果

我们最后深入讨论的是应用于新生小牛加速度计数据的周期性强度算法。小牛运动研究中的24头小牛出生日期各不相同，因此它们的传感器数据也分别从不同日期开始。对于年龄最大的 小牛（传感器ID 20714），加速度计数据于3月2日09:00开始；而对于年龄最小的小牛，

加速度计数据于4月5日10:00开始。尽管开始日期各不相同，但牛群中的所有小牛到4月5日都已开始贡献加速度计数据，因此从该日期起，我们获得了完整牛群周期性的图景。在24头小牛中，有5头被视为异常值，未参与整体牛群的分析，因为它们的项圈曾在数据记录期间某个时间点脱落；如果重新佩戴项圈，数据中就会产生间断，从而影响超过2周时间段内周期性强度的计算。其余19头小牛被用于分析。单个小牛的周期性强度图表见图5，其横轴已按时间对齐。对于这些小牛，周期频率设为24小时，窗口大小为7天，时滞重叠窗口的步长或移动量为15分钟。

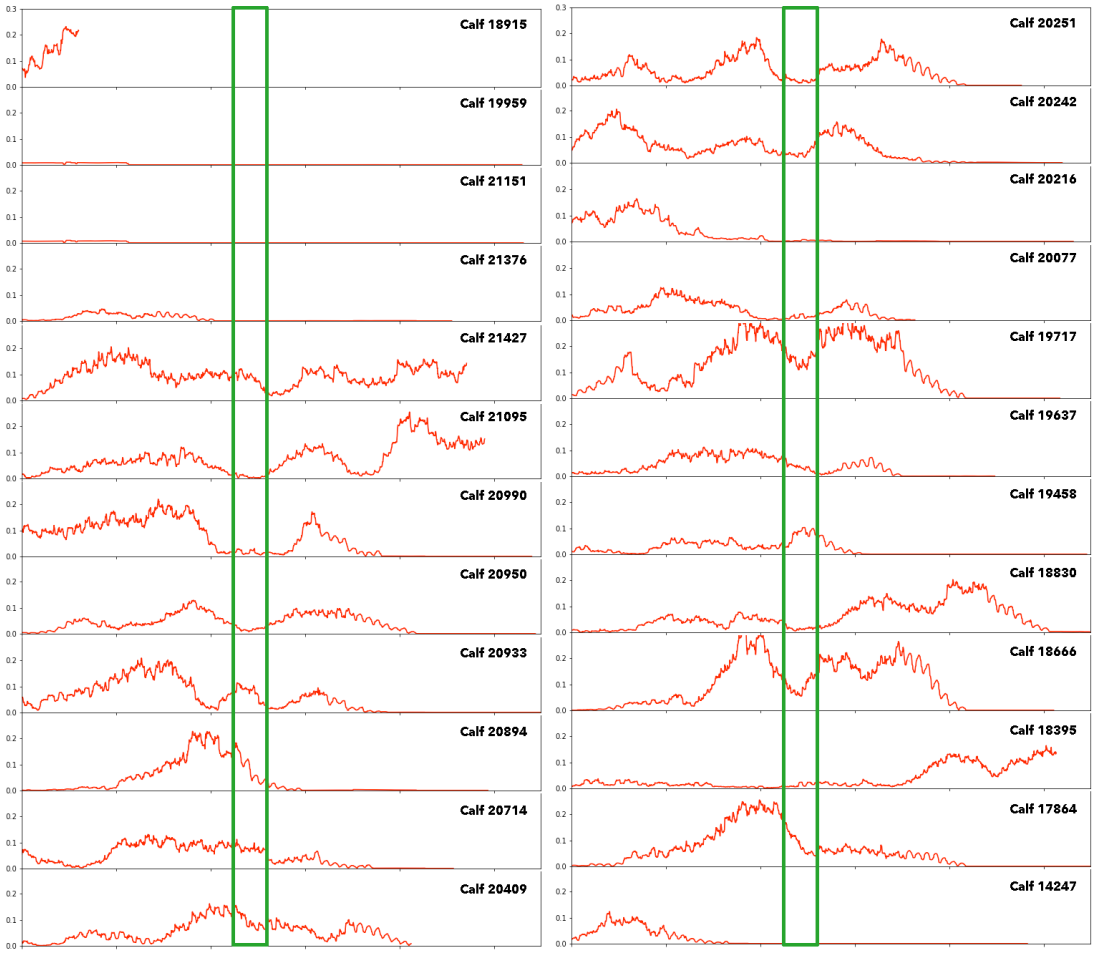

图5。 24头小牛各自的周期性强度折线图。频率为24小时，窗口大小为7天，窗口移动步长为15分钟。绿色竖线表示去角芽的日期。

图 5 展示了24个单独的周期性强度图表，其中左列的前4个和右列的第3个因前文所述原因被剔除。图表中的绿色竖线表示去角芽的日期；去角芽是在小牛长大并对其他小牛和农场主造成危险之前，通过外科手术去除角芽的过程 [37]。

纵观这组由19个图表组成的数据，很难辨认出牛群中存在任何协调一致的模式。当我们以与处理学生虚拟学习环境访问记录相同的方式，生成时间对齐的周期性强度堆叠图时，得到的图表如图 6所示，其中每种颜色代表一头不同犊牛的周期性强度。

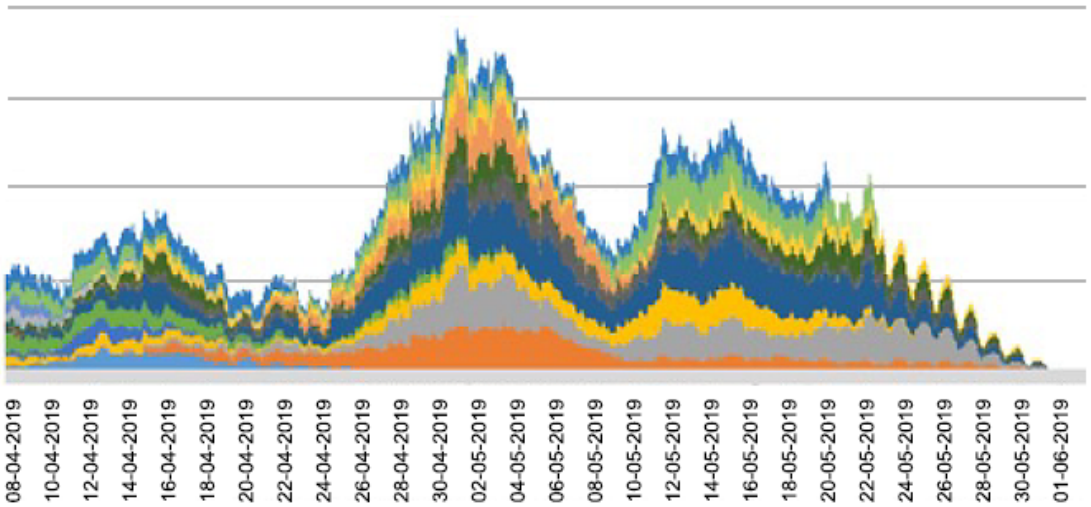

图6 . 来自24头小牛中的19头的小牛周期性强度累积堆叠折线图，来自 [31]。纵轴数值已归一化至0到1的范围。

周期性强度的堆叠折线图显示，5月初左右出现峰值，随后牛群的24小时周期性强度明显下降，表明牛群福利同步恶化。福利在大约1周后，即5月9日左右降至最低点，之后恢复到较好的水平，但仍未达到5月2日左右的水平。由于我们的计算采用了7天窗口化，5月9日左右达到最低点的下降，很可能是由5月2日左右发生的压力性或创伤性事件引起的，而这一时间窗口正是去角芽过程进行的时期。

与虚拟学习环境中学生学习习惯的用例一样，我们将小牛划分为随机子组，并为每个子组绘制堆叠折线图。这些图表与整个牛群的堆叠折线图呈现出相同的形状，因此，某个小牛子组不会产生人为的峰值或低谷，该模式在整个牛群中具有普遍性。

这种周期性强度分析的结果与对学生访问其虚拟学习环境的分析结果相似。我们根据每头小牛的活动，为其提供了单独的福祉指标反馈，同时也能更深入地了解整个牛群的福祉状况，而这对农场主很有意义。

4. 讨论

本文展示了三个应用案例：在三个截然不同的现实世界用例中，计算时间序列的周期性强度。在每个案例中，我们都能比未进行周期性分析时获得更深入的数据洞见。对于任何将内在周期性作为固有特征的数据源，都可以从其底层时间序列数据中获得这些有力的洞见。

这种数据分析形式的价值在于它是无监督的，不需要训练数据，因此具有普适性。就经济影响而言，这使该技术颇具吸引力，因为它可以应用于任何包含纵向时间序列数据的用例，以检测随时间推移逐渐出现的变化或转变。

在实施方面，三个参数的取值并不难确定。要研究的重复模式的频率 Fo通常为 24 小时，但并非总是如此，而用例通常已经知道该参数应取何值。计算周期性的窗口大小 m 通常也可预先确定。例如，如果 Fo 为 24 小时，那么将 m 设为 7 天，就可以消除底层数据中周末与工作日行为转换所产生的影响。窗口之间的重叠程度或步长 l 是

由用例的需求和可用计算资源决定，而不受任何算法限制的影响。该技术在所使用的时间序列数据的数量甚至类型方面都十分灵活，并且计算速度很快。由于我们使用 Lomb-Scargle算法 [13] 来计算窗口中的周期性，因此它能够处理不规则采样数据，例如学生访问其虚拟学习环境的情况，也能处理适量的缺失数据；但无法处理完全缺失的数据，例如在我们的第三个用例中，那些项圈脱落但很快得到更换的小牛的数据。最后，用于计算周期性强度的算法在数学上是确定性的，因而对相同输入重复运行时会产生相同的输出。

周期性强度是一项复杂的指标，要完全理解它并不容易，因为它新颖且不具传统性。在我们涉及人类参与者的两个应用案例中——老年人家中的家庭传感器，以及学生访问其虚拟学习环境（VLE）——由于我们的用户是医疗保健专业人员和老年人，因此我们需要借助动画辅助视频来说明周期性强度的含义。

未来的工作将集中于探索如何通过将周期性强度图中的变化映射回底层数据，使该指标更易于直观理解，从而准确展示哪些数据导致了周期性强度的变化；这对于涉及人类受试者的应用案例将非常有用。我们还将通过引入噪声来考察算法在不同数据质量下的稳健性，以了解数据质量下降会如何影响生成的输出。最后，我们将重点实现 Lomb-Scargle算法，仅根据 24小时频率确定周期性，因为根据我们的经验，大多数应用案例最关注的正是这一频率设置。

作者贡献： 概念化，A.S.和F.H.；方法论，A.S.和F.H.；软件，F.H.；验证，A.S.和F.H.；形式分析，A.S.和F.H.；调查，A.S.和F.H.；数据整理，A.S.；撰写——初稿，A.S.；撰写——审阅与编辑，A.S.和F.H.。所有作者均已阅读并同意发表的稿件版本。

资助： 本研究由爱尔兰科学基金会资助，资助编号为SFI/12/RC/2289_P2；该项目由欧洲区域发展基金和由爱尔兰企业局管理的颠覆性技术创新基金共同资助，项目资助编号为DT-2018-0258；此外还获得了UCD威康机构战略支持基金的资助，该基金由都柏林大学学院与SFI-HRB-Wellcome生物医学研究合作伙伴关系共同出资（编号204844/Z/16/Z）。

机构审查委员会声明： 本研究遵循《赫尔辛基宣言》开展；其中，（1）涉及从老年人住所获取传感器数据的研究，经DCU研究伦理委员会批准（DCUREC202221）；（2）涉及学生访问在线虚拟学习环境的日志数据的研究，经DCU数据保护部门批准（07042021-DPO），并获得计算机学院研究伦理委员会的伦理批准。由于（2）的伦理批准被认定为仅需备案，根据机构政策，只需获得学院层面的伦理批准。涉及牛群运动与行为传感器数据的动物研究方案是在获得UCD动物研究伦理委员会伦理豁免的情况下收集的，批准编号为AREC-E-19-46-McAloon。

知情同意声明： 已获得所有参与研究的人类受试者的知情同意。

数据可用性声明： 研究中涉及来自老年人住宅的传感器数据，这些数据已在 FigShare 上公开提供，地址为https://doi.org/10.6084/m9.figshare.21415836.v2。研究中涉及学生访问在线虚拟学习环境的数据已在 FigShare 上公开提供，地址为https://doi.org/10.6084/m9.figshare.20288763.v2。研究中涉及牛群运动与行为的数据已在 FigShare 上公开提供，地址为 https://doi.org/10.6084/m9.figshare.20039486.v1。

利益冲突： 作者声明不存在利益冲突。资助方未参与研究设计、数据收集、分析或解释、稿件撰写，也未参与决定是否发表研究结果。

缩略语

本文使用了以下缩略语：

ADL 日常生活活动 ECTS 欧洲学分转换与累积系统 IADL 工具性日常生活活动 RP 递归图 SVM 信号向量幅度 VLE 虚拟学习环境 EMD 经验模态分解 ZCR 过零率 ZCD 过零密度分解

参考文献

1. Li, J.; Zhang, J.; Bah, M.J.; Wang, J.; Zhu, Y.; Yang, G.; Li, L.; Zhang, K. 采用遗传算法的高维数据自编码器：实现准确且可解释的离群点检测。Algorithms 2022, 15. https://doi.org/10.3390/a15110429。2. Fisher, D.N.; Pruitt, J.N. 从复杂系统研究中获得的关于动物种群生态学与进化的见解。Current Zoology 2019, 66, 1–14. https://doi.org/10.1093/cz/zoz016。3. Kuhlman, S.J.; Craig, L.M.; Duffy, J.F. 时间生物学导论。Cold Spring Harbor Perspectives in Biology 2018, 10, a033613。4. Cohen, M. 健康生活方式的健康福祉与热力学。Asia-Pacific Journal of Health, Sport and Physical Education 2010, 1, 5–12。5. Farhud, D.; Aryan, Z. 昼夜节律、生活方式与健康：叙述性综述。Iranian journal of public health 2018, 47, 1068。6. Marwan, N.; Romano, M.C.; Thiel, M.; Kurths, J. 用递归图分析复杂系统。Physics reports 2007, 438, 237–329。7. Nath, A.G.; Udmale, S.S.; Raghuwanshi, D.; Singh, S.K. 使用多传感器模糊递归图和分类器融合改进结构转子故障诊断。IEEE Sensors Journal 2021, 21, 21705–21717. https://doi.org/10.1109/JSEN.2021.3079366。8. Li, X.; Li, T.; Wang, Y. GW-DC：利用二维图像变换与增强的深度聚类模型。Algorithms 2021, 14. https://doi.org/10.3390/a14120349。9. Buman, M.P.; Hu, F.; Newman, E.; Smeaton, A.F.; Epstein, D.R. 基于24小时腕部加速度测量的行为周期性检测及其与心脏代谢风险和健康相关生活质量的关联。BioMed Research International 2016, 2016, 485–506。10. Hu, F.; Smeaton, A.F. 用于指示生活日志数据中行为变化的周期性强度。载于：2016年IEEE生物信息学与生物医学国际会议（BIBM）论文集。IEEE，2016年，第970–977页。11. Chegini, S.N.; Manjili, M.J.H.; Bagheri, A. 检测轴承轻微退化的新型故障诊断方法。Meccanica 2020, 55, 261–286。12. Bartlett, M.S. 周期图分析与连续谱。Biometrika 1950, 37, 1–16。13. VanderPlas, J.T. 理解Lomb–Scargle周期图。The Astrophysical Journal Supplement Series 2018, 236, 16. https: //doi.org/10.3847/1538-4365/aab766。14. Azar, Y.; Fiat, A.; Karlin, A.; McSherry, F.; Saia, J. 数据谱分析。载于：第三十三届ACM计算理论年会论文集；Association for Computing Machinery：美国纽约州纽约，2001年；STOC '01，第619–626页。https://doi.org/10.1145/380752.380859。15. Montaruli, A.; Castelli, L.; Mulè, A.; Scurati, R.; Esposito, F.; Galasso, L.; Roveda, E. 生物节律与生物钟型：健康领域的新视角。Biomolecules 2021, 11, 487。16. Ho, Q.T.; Phan, D.V.; Ou, Y.Y.; 等。使用词嵌入技术高效表示蛋白质序列，以识别转运体的底物特异性。Analytical biochemistry 2019, 577, 73–81。17. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. 小波变换在非平稳时间序列分析中的应用：综述。Applied Sciences 2019, 9, 1345。18. Chaovalit, P.; Gangopadhyay, A.; Karabatis, G.; Chen, Z. 基于离散小波变换的时间序列分析与挖掘。ACM Comput. Surv. 2011, 43. https://doi.org/10.1145/1883612.1883613。19. Shukla, S.; Mishra, S.; Singh, B. 结合希尔伯特变换的经验模态分解用于电能质量评估。IEEE transactions on power delivery 2009, 24, 2159–2165。

20. Lartillot, O.; Toiviainen, P. 用于从音频中提取音乐特征的MatLab工具箱。载于：国际数字音频效果会议论文集。波尔多，2007年，第237卷，第244页。21. Hong, Y.; Lee, Y.J. 时间序列波动率模型检验的一般方法。Journal of Management Science and Engineering 2017, 2, 1–33。22. Abayomi-Alli, O.O.; Sidekerskien˙e, T.; Damaševi˘cius, R.; Siłka, J.; Połap, D. 基于经验模态分解的数据增强：使用NARX网络进行时间序列预测。载于：人工智能与软计算；Rutkowski, L.; Scherer, R.; Korytkowski, M.; Pedrycz, W.; Tadeusiewicz, R.; Zurada, J.M.，编；施普林格国际出版公司：沙姆，2020年；第702–711页。23. Altaf, M.; Akram, T.; Khan, M.A.; Iqbal, M.; Ch, M.M.I.; Hsu, C.H. 基于统计特征的轴承故障诊断新方法：使用振动信号。Sensors 2022, 22. https://doi.org/10.3390/s22052012。24. Cohen, M.X. 神经时间序列数据分析：理论与实践；麻省理工学院出版社，2014年。25. Sidekerskien ˙e; Damaševi˘cius, R.; Wo´zniak, M.，过零密度分解：一种新型信号分解方法。载于：数据科学：新问题、挑战与应用；Dzemyda, G.; Bernatavi˘cien˙e, J.; Kacprzyk, J.，编；施普林格国际出版公司：沙姆，2020年；第235–252页。https://doi.org/10.1007/978-3-030-39250-5_13。26. Uzuno˘glu, C.P. 高压放电中经验模态分解与变分模态分解的比较研究。Electrica 2018, 18, 72–77。27. Timon, C.M.; Heffernan, E.; Kilcullen, S.M.; Lee, H.; Hopper, L.; Quinn, J.; McDonald, D.; Gallagher, P.; Smeaton, A.F.; Moran, K.; 等。物联网技术平台（NEX系统）的开发，以支持老年人独立生活：开发与可用性研究方案。JMIR研究方案 2022, 11, e35277。https://doi.org/10.2196/35277。28. Katz, S.; Ford, A.B.; Moskowitz, R.W.; Jackson, B.A.; Jaffe, M.W. 老年人疾病研究：日常生活活动指数：生物和心理社会功能的标准化测量。Journal of the American Medical Association 1963, 185, 914–919。29. Lawton, M. 工具性日常生活活动（IADL）量表。Psychopharmacol Bull 1988, 24, 785–787。30. Baik, C.; Larcombe, W.; Brooker, A. 大学如何提升学生心理福祉：学生视角。高等教育研究与发展 2019, 38, 674–687。31. Rhodes, V.; Maguire, M.; Shetty, M.; McAloon, C.; Smeaton, A.F. 新生犊牛24小时昼夜节律的周期性强度可作为牛群福利指标。Sensors 2022, 22, 5843。32. Conradt, L.; Roper, T.J. 动物的群体决策。Nature 2003, 421, 155–158。33. De Craemer, M.; Verbestel, V. 比较ActiGraph GT3X+与Axivity AX3加速度计所得结果，以客观测量成年人的24小时运动行为：一项横断面研究。International Journal of Environmental Research and Public Health 2021, 19, 271。34. Riaboff, L.; Aubin, S.; Bedere, N.; Couvreur, S.; Madouasse, A.; Goumand, E.; Chauvin, A.; Plantier, G. 用于根据加速度计数据预测牛行为的预处理方法评估。Computers and Electronics in Agriculture 2019, 165, 104961。35. Fridolfsson, J.; Börjesson, M.; Buck, C.; Ekblom, Ö.; Ekblom-Bak, E.; Hunsberger, M.; Lissner, L.; Arvidsson, D. 频率滤波对基于加速度计的身体活动测量中强度和噪声的影响。Sensors 2019, 19, 2186。36. Byron, L.; Wattenberg, M. 堆叠图——几何学与美学。IEEE可视化与计算机图形学汇刊 2008, 14, 1245–1252。37. Espinoza, C.; Lomax, S.; Windsor, P. 局部麻醉对接受去角处理的小牛皮质醇反应的影响。Animals 2020, 10, 312。