arXiv:2304.14853v1 [math.AT] 28 Apr 2023

# 用于儿童阻塞性睡眠呼吸暂停的脑电图信号拓扑数据分析

Shashank Manjunath[1],Jose A. Perea[2] 和 Aarti Sathyanarayana[3]

摘要——**拓扑数据分析（TDA）是一种新兴的生物信号处理技术。TDA 利用信号在度量空间中的不变拓扑特征，即使存在噪声，也能对信号进行稳健分析。本文将 TDA 应用于由脑电图（EEG）信号构建的脑连接网络，以识别患有阻塞性睡眠呼吸暂停（OSA）的儿科患者与未患 OSA 的儿科患者之间的统计差异。我们使用了大量数据，并展示了 TDA 能够揭示两组患者脑动力学之间的统计差异。**

临床意义——**这证实了拓扑数据分析作为一种工具的潜力，可在无需进行完整多导睡眠图检查的情况下识别阻塞性睡眠呼吸暂停，并为开发更简便、更具可扩展性的阻塞性睡眠呼吸暂停诊断方法开展初步探索。**

## I. 引言

信号处理技术的最新进展（包括拓扑数据分析(TDA)）为脑电信号分析提供了强大的方法。TDA借鉴数学拓扑域的思想，并将这些思想应用于现实世界信号的分析[5]。总体而言，TDA使我们能够利用数据中固有的拓扑和几何结构，并借助这些结构研究阻塞性睡眠呼吸暂停 (OSA) 阳性患者与OSA阴性患者脑电信号之间的根本差异。在这项工作中，我们提出了仅使用脑电信号识别OSA的TDA技术。

这项工作的关键假设是，OSA阳性患者的脑连接网络在拓扑结构上与OSA阴性患者的脑连接网络存在根本差异。

目前用于识别儿童阻塞性睡眠呼吸暂停的技术，需要进行一项完整的整夜睡眠研究，称为多导睡眠图（PSG）。患者必须前往医疗机构的睡眠实验室，或预约在家进行夜间睡眠测试，而这两种方式的预约都可能需要等待数月。

*J. A. Perea 的部分研究经费由美国国家科学基金会提供，资助项目包括 CCF-2006661 和 CAREER 奖项 DMS-1943758。

[1]Shashank Manjunath 任职于马萨诸塞州波士顿 02115 的东北大学 Khoury 计算机科学学院manjunath.sh@northeastern.edu

[2]Jose A. Perea 任职于马萨诸塞州波士顿 02115 的东北大学数学系，以及东北大学 Khoury 计算机科学学院
j.pereabenitez@northeastern.edu

[3]Aarti Sathyanarayana 任职于东北大学计算机科学学院 Khoury 学院，地址为马萨诸塞州波士顿市，邮编 02115；同时任职于东北大学 Bouvé 健康科学学院，地址为马萨诸塞州波士顿市，邮编 02115，以及哈佛大学公共卫生学院生物统计学系，地址为马萨诸塞州波士顿市，邮编 02115´东北大学a.sathyanarayana@northeastern.edu

正如本文所展示的，成功应用拓扑数据分析直接通过脑电信号区分阻塞性睡眠呼吸暂停阳性和阻塞性睡眠呼吸暂停阴性患者，是迈向更简便、侵入性更低的儿童阻塞性睡眠呼吸暂停诊断的重要一步，并有助于改善患者的预后。

## II. 背景

### A. 阻塞性睡眠呼吸暂停

儿童阻塞性睡眠呼吸暂停（阻塞性睡眠呼吸暂停）难以诊断，并会导致严重并发症，包括休息不佳和心脏病[3]。尤其对儿童而言，阻塞性睡眠呼吸暂停可能导致行为改变，并进一步降低生活质量。患者通常需要接受夜间多导睡眠监测来识别阻塞性睡眠呼吸暂停 [11]。多导睡眠监测研究包含多种传感器，例如脑电传感器、脉搏血氧测定设备和呼吸频率传感器等。获取这些信息对患者而言并不容易，因为他们必须在陌生的环境中入睡，同时接受医务人员观察，还要在身体上连接多个传感器入睡。多导睡眠监测研究被视为诊断阻塞性睡眠呼吸暂停的金标准，而尝试使用问卷等其他临床方法则会导致诊断准确性较低[11]。

儿童打鼾较为常见，1 至 9 岁儿童的发生率约为 10%。习惯性打鼾的儿童中，阻塞性睡眠呼吸暂停的发生率为 2-3% [3]。近期儿童肥胖人数的增加还会引发进一步的并发症。儿童阻塞性睡眠呼吸暂停通常由腺样体扁桃体肥大引起，并与生长不良和生长发育迟缓相关 [4]。此外，与患有阻塞性睡眠呼吸暂停的成人相比，患有阻塞性睡眠呼吸暂停的儿童较少出现日间嗜睡（嗜睡症），这使阻塞性睡眠呼吸暂停的诊断更加困难。

### B. 拓扑数据分析

拓扑数据分析旨在识别数据结构的定量信息，并利用这些信息开展下游数据分析任务 [5]。假设数据存在于一个度量空间中，因此可以计算不同观测值之间的距离；这些观测值通常被视为环境空间中的点，即构成一个点云。基于这些数据，可以通过改变判断两点是否相连的距离阈值，在输入点云上构建形状（详见第 III-C 节）。

此前， [1], [8]的研究曾使用拓扑数据分析来理解大脑的潜在动力学。然而，这些论文的重点是：Bourakna 等人 [1], 对注意缺陷多动障碍（ADHD）的识别，以及

©2023 IEEE。允许个人使用本材料。所有其他用途均须事先获得 IEEE 许可，无论是在当前还是未来的任何媒介中，包括重印或再版本材料用于广告或宣传目的、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分。

Lella 等人 [8]对睡眠期间脑电信号形态的分类。Lella 等人的研究展现了拓扑数据分析作为单条脑电信号分析技术的优势：与以往研究相比，该研究超越了此前最先进的分类技术，且对固有噪声具有更强的鲁棒性；不过，这项研究聚焦于脑电信号的特定形态。尽管特定的脑电形态具有参考价值，但在阻塞性睡眠呼吸暂停的研究中，我们需要更深入地理解脑动力学，才能恰当地识别阻塞性睡眠呼吸暂停的相关特征。此外，我们需要对所有采集到的脑电图通道进行连贯分析，才能充分利用脑电图数据进行拓扑数据分析。Bourakna 等人的研究为利用拓扑数据分析处理注意缺陷多动障碍数据奠定了坚实基础，同时也运用拓扑数据分析来理解潜在的脑网络动力学。该研究对多个脑电图通道进行了连贯分析，以理解注意缺陷多动障碍患者的脑动力学，而非关注单条脑电信号的特定形态，因此对噪声具有更强的鲁棒性。我们的目标是将这些思路应用于阻塞性睡眠呼吸暂停的识别。

目前不会单独根据脑电信号诊断阻塞性睡眠呼吸暂停，而是结合多导睡眠监测研究期间采集的其他传感器数据进行诊断。在这项工作中，我们对脑电信号应用拓扑数据分析，不纳入其他传感器数据，以识别阻塞性睡眠呼吸暂停阳性组与阻塞性睡眠呼吸暂停阴性组之间具有统计显著性的差异。这项工作为提升阻塞性睡眠呼吸暂停诊断的可扩展性和便捷性迈出了第一步，让临床医生和患者都能从中受益。

## III. 方法

### A. 数据

在这项工作中，我们使用 Nationwide Children's Hospital (NCH) 睡眠数据库 (SDB)[13], [7]。该数据集包含来自 3,673 名儿科患者的 3,984 份多导睡眠监测研究。该数据集包含大量来自多种传感器的数据，其中包括脑电图、肌电图、眼电图、心电图、用于测量气流的鼻部和口腔传感器，以及脉搏血氧测定等。这些丰富的数据和传感器使我们能够针对各种儿科患者研究儿童阻塞性睡眠呼吸暂停。该数据集还包含睡眠事件的标注，例如 NREM1 睡眠、NREM2 睡眠、NREM3 睡眠和快速眼动睡眠，标注间隔为 30 秒。我们重点研究以 256 Hz 采集的脑电信号，使用 7 通道脑电图研究，因此样本范围缩小至 2,883 项睡眠研究。我们根据数据集中患者相应多导睡眠图中是否存在呼吸暂停事件字符串标注，将患者归入“呼吸暂停”或“无呼吸暂停”组。如果患者的多导睡眠图中在睡眠研究期间任何时点出现与呼吸暂停相关的字符串标注，则将整项研究归入“呼吸暂停”组。如果未发现与呼吸暂停相关的字符串标注，则将该研究归入“无呼吸暂停”组。

### B. 信号预处理

首先，我们对脑电信号进行基本的信号处理。我们使用三阶 Butterworth 带阻滤波器滤除 60Hz 和 120Hz 的电源噪声。

接着，我们借鉴 Bourakna 等人的方法，根据脑电图数据[1]构建距离矩阵。我们逐项处理每项研究的数据。对于单项多导睡眠监测，我们首先提取所有可用的脑电图通道，从而得到一组时间序列信号 $X_i(t)$，对应通道 $i \in V$ 和时间 $t \in \{1, \cdots, T\}$。我们将这些信号切分为 30 秒的片段，以便与清醒期、NREM1、NREM2、NREM3 和快速眼动睡眠等标签对应。随后，我们为每个通道的每个 30 秒片段生成平滑周期图。平滑周期图有助于我们了解信号的频谱，同时通过使用较小的平滑核对频谱进行轻微平滑来消除噪声。

为计算平滑周期图，我们使用以下方程：

$$d(\omega_k) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_i(t) \exp(-i2\pi t \omega_k) \tag{1}$$

$$\hat{f}(\omega_k) = \sum_{\omega} k_h(\omega - \omega_k) d(\omega_k) d(\omega_k)^* \tag{2}$$

$k_h(\omega - \omega_k)$表示以$\omega_k$ 为中心的平滑核， $h$ 是带宽参数。根据这些平滑周期图，我们计算脑电图通道 $i$ 和 $j$之间的平方信号相干性：

$$\mathcal{C}(X_i, X_j, \omega) = \frac{|f_{i,j}(\omega)|^2}{f_{i,i}(\omega) f_{j,j}(\omega)} \tag{3}$$

随后，我们利用一个递减函数，计算给定频率 $\omega$下相干性值之间的距离：

$$\mathcal{D}(X_i, X_j, \omega) = 1 - \mathcal{C}(X_i, X_j, \omega) \tag{4}$$

矩阵 $\mathcal{D}$ 由通道 $i$、通道 $j$和频率 $\omega$共同索引。临床上，脑电图数据按以下频段进行分析：δ频段为 0.5 Hz 至 4 Hz，θ频段为 4 Hz 至 8 Hz，阿尔法频段为 8 Hz 至 12 Hz，贝塔频段为 12 Hz 至 30 Hz，伽马频段为 30 Hz 至 50 Hz。

我们对某一频段内所有频率对应的矩阵取平均值，从而为该频段得到一个单独的 $7 \times 7$矩阵。因此，对于来自 7 个通道的单段 30 秒脑电图信号，我们可得到 5 个独立的距离矩阵，每个频段对应一个。我们针对每个频段和睡眠状态分别处理距离矩阵。这些距离矩阵作为我们拓扑数据分析方法的输入。

### C. 拓扑数据分析方法

为了确定阻塞性睡眠呼吸暂停阳性和OSA阴性患者的脑连接网络在拓扑结构上是否存在根本差异，我们必须了解如何区分拓扑空间。首先，我们从单纯复形的定义开始。

**定义 III.1** （单纯复形 [6]）。单纯复形是由 $K$ 非空有限集 $\sigma \in K$构成的集合，这些有限集称为单纯形，并且 $\sigma \in K$ 和 $\emptyset \neq \tau \subseteq \sigma$ 总是

意味着 $\tau \in K$。如果 $\sigma \in K$ 有 $n+1$ 个元素（即，$\sigma = \{x_0, \ldots, x_n\}$），那么我们称它为 $n$-单纯形。

直观地说，0-单纯形 $\{x_0\}$ 表示一个顶点，1-单纯形 $\{x_0, x_1\}$ 表示连接 $x_0$ 和 $x_1$的边，2-单纯形$\{x_0, x_1, x_2\}$ 表示由这些顶点张成的三角形，更高维的情形依此类推。接下来要定义的同调，是描述单纯复形拓扑性质的一种量。它衡量连通分支的数量、孔洞和空腔是否存在，以及它们在更高维中的类似形式。具体来说，取一个域 $\mathbb{F}$（例如有理数域或模素数的整数），并令 $C_n(K)$ 表示由 $K$的 $n$-单纯形生成的$\mathbb{F}$-向量空间。也就是说，$J$每个 $\gamma \in C_n(K)$ 都可以唯一地写成 $\gamma = \sum a_j \sigma_j{}^{j=0}$的形式，其中系数为标量 $a_j \in \mathbb{F}$，而 $n$为 $\sigma_j \in K$-单纯形。将 $n$-单纯形 $\sigma = \{x_0, \ldots, x_n\} \in K$的边界定义为$_n\partial_n(\sigma) = \sum(-1)^i$ $(\sigma \setminus \{x_i\}) \in C_{n-1}(K)$。由于 $K$ 的 $n$-$_{i=0}$单纯形构成了 $C_n(K)$的一组基，因此我们得到线性变换 $\partial_n : C_n(K) \longrightarrow C_{n-1}(K)$；可以验证，对每个 $n \in \mathbb{N}$，它们都满足 $\partial_n \circ \partial_{n+1} = 0$。

**定义 III.2**（单纯同调群 [5]）。单纯复形 $K$ 的 $n$ 阶同调群是商向量空间

$$H_n(K) = \mathsf{Ker}(\partial_n)/\mathsf{Img}(\partial_{n+1})$$

$K$ 的同调可以解释如下：$H_0(K)$ 的维数恰好等于 $K$ 的连通分支数；$\dim(H_1(K))$ 统计 $K$ 中围成空区域的本质不同的闭环数；$\dim(H_2(K))$ 统计空腔数；类似地，$\dim(H_n(K))$ 表示 $n \geq 1$。

在本研究中，我们尝试了解阻塞性睡眠呼吸暂停阳性和 OSA 阴性患者脑连接网络的每个同调群的拓扑特征（例如连通分支、孔洞、空腔等）。有关拓扑概念的进一步介绍，请参见 [6]。

为了计算离散数据点集的同调特征，我们对第三节 B 部分所述的距离矩阵应用持久同调。实际上，一旦获得了一个 30秒信号的距离矩阵，我们就构建一个单纯复形，特别是 ***Rips*复形**，并进一步构建***Rips*滤过**，使用 Python 包 Ripser.py[12]。

设 $\mathcal{S}$ 为我们的点云，其成对距离由前文的距离矩阵 $\mathcal{D}$ 描述。由 $\mathcal{S}$ 和距离 $r$ 构成的 Rips 复形 $\mathcal{R}$，是由 $\mathcal{S}$ 中直径不超过 $r$的所有子集构成的单纯复形，即

$$\mathcal{R}(\mathcal{S}, r) = \{\sigma \subset \mathcal{S} | \mathrm{diam}(\sigma) \leq r\}$$

请注意，$\mathcal{R}$ 本质上与给定的 $r$ 值相关。我们构造一个滤过，称为 Rips滤过，它是通过改变自由参数 $r$ 而生成的一组 Rips复形。我们从一个较小的 $r$（$r = 0$）开始，连续增大 $r$，以获得我们单纯复形[6]的一组子复形：

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m$$

我们利用 Ripser.py 计算 Rips滤过。当我们将 $r$ 从 $K_{i-1}$ 改变到 $K_i$时，拓扑特征会发生变化，我们也会获得或失去拓扑特征[9]。我们将某个拓扑特征在特定同调群中出现时的 $r$，以及同一拓扑特征消失时的 $r$，量化为一对（出生、死亡）值，这对值存在于 $\mathbb{R}^2$中。我们可以将这些点绘制在图上，生成持久图，如图1所示。

持久图可以帮助我们直观呈现特定样本的拓扑特征，但它无法扩展到可分巴拿赫空间，因此无法通过随机变量[2]的视角进行分析。不过，我们可以将持久图转换为持久景观。持久景观是一种函数，且确实存在于可分巴拿赫空间中。构造持久景观时，我们以持久图中的点为中心绘制等腰三角形。发生交点时，取最高的函数作为持久景观。

### *D.* 置换检验

现在我们已获得可用的拓扑数据分析特征，即持久景观，接下来需要确定一种检验方法，以判断阻塞性睡眠呼吸暂停阳性儿科患者的拓扑数据分析特征是否不同于OSA阴性患者的拓扑数据分析特征。为此，我们采用 [10]中所述的置换检验。在这种方法中，我们的零假设是阻塞性睡眠呼吸暂停不会对脑连接网络产生影响。为进行检验，我们实现以下算法：

1) 计算阻塞性睡眠呼吸暂停阳性持久景观的平均值 $\{\lambda_1^{(1)}, \cdots, \lambda_{n_1}^{(1)}\}$，以及阻塞性睡眠呼吸暂停阴性持久性景观的平均值 $\{\lambda_1^{(2)}, \cdots, \lambda_{n_2}^{(2)}\}$。2) 计算阻塞性睡眠呼吸暂停阳性持久景观的平均值与阻塞性睡眠呼吸暂停阴性持久性景观的平均值之差，并找出该差值景观中最显著的值 $F_T$。3) 对两组的标签进行置换，使其中 $n_1$ 个持久景观被随机标记为阳性，其余 $n_2$ 个景观被标记为阴性。4) 计算置换后阳性组与置换后阴性组之间的差值，并找出该差值景观中最显著的值 $F_p$。5) 如果 $F_p \geq F_T$，则计为一次显著置换。6) 重复步骤 3，共进行 4 $B$ 次。本分析中，我们将这些步骤重复 $B = 1000$ 次。7) 计算 $p$ 值，即显著置换次数 $S$ 除以置换总次数 $B$。

我们专门对按照第 III-C 节所述生成的零阶同调群（$H_0$）持久景观进行此置换检验。

## IV. 结果

我们在图1和图2中展示了一名阻塞性睡眠呼吸暂停阳性患者的持久图和持久性景观示例。

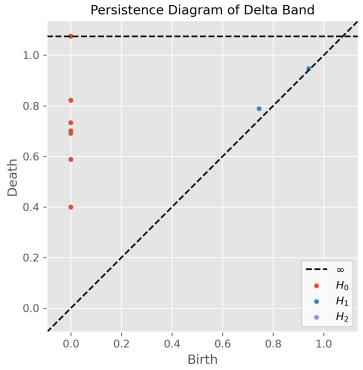



图1. 阻塞性睡眠呼吸暂停阳性患者 δ频段的持久图示例

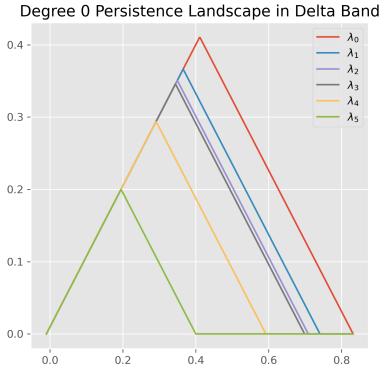



图2. 阻塞性睡眠呼吸暂停阳性患者 δ频段的零度持久景观示例

表I列出了 III-D节所述拓扑数据分析置换检验的 $p$ 值。我们将 $p$ 值为 $p < 0.05$ 的结果视为显著。我们将距离矩阵按信号发生时的睡眠状态分为 NREM1 睡眠、NREM2 睡眠、NREM3 睡眠和快速眼动睡眠。

表I 使用拓扑数据分析的各睡眠状态 $p$ 值

| **脑电频段** | NREM1 | NREM2 | NREM3 | REM |
|---|---|---|---|---|
| δ频段 | 0.000 | 0.001 | 0.000 | 0.040 |
| θ频段 | 0.000 | 0.000 | 0.000 | 0.000 |
| 阿尔法频段 | 0.000 | 0.000 | 0.000 | 0.000 |
| 贝塔频段 | 0.000 | 0.000 | 0.000 | 0.000 |
| 伽马频段 | 0.000 | 0.000 | 0.000 | 0.000 |

## 五、讨论

我们的结果表明，拓扑数据分析可以识别阻塞性睡眠呼吸暂停阳性队列和阻塞性睡眠呼吸暂停阴性队列之间具有统计学显著性的差异。所有频段均达到统计学显著性，适用于在所有睡眠阶段中，这些结果有力地表明，仅使用脑电信号即可辅助诊断儿童患者的阻塞性睡眠呼吸暂停，从而实现侵入性更低的诊断方式。

## VI. 结论

在这项研究中，我们发现，使用拓扑数据分析进行分析时，阻塞性睡眠呼吸暂停阳性儿科患者的脑依赖网络与阻塞性睡眠呼吸暂停阴性儿科患者的脑依赖网络存在统计学上的显著差异。我们的研究为利用拓扑数据分析进行队列级阻塞性睡眠呼吸暂停分类奠定了基础，并迈出了通过单次脑电图诊断阻塞性睡眠呼吸暂停的第一步。未来还需开展研究，以实现临床应用价值，并确定拓扑数据分析所识别的脑动力学与阻塞性睡眠呼吸暂停诊断中更常用的呼吸测量指标之间的关系。

## 致谢

NCH 睡眠数据库由美国国立卫生研究院美国国立生物医学成像与生物工程研究所通过资助编号 R01EB025018 提供支持。国家睡眠研究资源由美国国立卫生研究院美国国家心肺血液研究所提供支持（R24 HL114473，75N92019R002）。

## 参考文献

[1] Anass El Yaagoubi Bourakna、Moo K. Chung 和 Hernando Ombao。《多变量时间序列数据的拓扑数据分析》，2022年4月。[2] Peter Bubenik。使用持久性景观进行统计拓扑数据分析。机器学习研究杂志，16 102，2015 3 77( ): – 。[3] Oscar Sans Capdevila、Leila Kheirandish-Gozal、Ehab Dayyat 和 David Gozal。儿童阻塞性睡眠呼吸暂停。美国胸科学会会议录，5(2):274–282，2008年2月。[4] James Chan、Jennifer C. Edman 和 Peter J. Koltai。儿童阻塞性睡眠呼吸暂停。美国家庭医生，69(5):1147–1155，2004年3月。[5] Fr´ed´eric Chazal 和 Bertrand Michel。《拓扑数据分析导论：数据科学家的基础与实践方面》。人工智能前沿，4，2021年。[6] Herbert Edelsbrunner 和 John Harer。计算拓扑学：导论。美国数学学会，2010年。[7] Harlin Lee、Boyue Li、Shelly DeForte、Mark L. Splaingard、Yungui Huang、Yuejie Chi 和 Simon L. Linwood。大量真实世界的儿科睡眠研究数据集。科学数据，9(1):421，2022年7月。[8] Paolo Lella。脑电图中睡眠阶段和 K复合波检测的持久同调与分形维数。硕士论文，米兰理工大学，2021年。[9] Jose A. Perea。拓扑时间序列分析。美国数学学会通报，66(05):1，2019年5月。[10] Andrew Robinson 和 Katharine Turner。拓扑数据分析的假设检验。应用与计算拓扑学杂志，1(2):241–261，2017年12月。[11] Michael S. Schechter 和儿科肺病学分会、阻塞性睡眠呼吸暂停综合征小组委员会。技术报告：儿童阻塞性睡眠呼吸暂停综合征的诊断与管理。《儿科学》，109(4):e69，2002年4月。[12] Christopher Tralie、Nathaniel Saul 和 Rann Bar-On。Ripser.py：一个精简的 Python 持久同调库。开放源代码软件杂志，3(29):925，2018年9月。[13] Guo-Qiang Zhang、Licong Cui、Remo Mueller、Shiqiang Tao、Matthew Kim、Michael Rueschman、Sara Mariani、Daniel Mobley 和 Susan Redline。国家睡眠研究资源：迈向睡眠数据共享平台。美国医学信息学会杂志：*JAMIA*，25(10):1351–1358，2018年10月。