

基于拓扑数据分析的驾驶疲劳 EEG 数据处理与优化分析研究^①周飞扬^{②*} 柳政卿^{③*} 王秋成* 杨忠**

(* 浙江工业大学机械工程学院 杭州 310023)

(** 深蓝汽车科技有限公司 重庆 400020)

摘要 为提高驾驶疲劳脑电(EEG)数据处理与分析的准确性和鲁棒性,提出一种基于拓扑数据分析(TDA)的驾驶人疲劳脑电分析方法。首先利用汽车性能虚拟仿真平台开展驾驶实验,通过驾驶人状态反馈和面部特征视频,标记脑电数据,形成清醒和疲劳二分数据集。之后利用 EEGLAB 预处理数据,剔除噪声并保留 0.3~30 Hz 频带,直接从时域 EEG 数据中提取拓扑特征。此外还提取了经典频域特征 α 波能量和 α/β 用于对比分析。最后使用支持向量机进行分类。结果表明,基于持久同源(PH)的拓扑特征取得了高达 88.7% 的准确率和 91.4% 的召回率,与经典频域特征性能相当,且对脑电伪影的鲁棒性明显更好,在未剔除 EEG 伪影的情况下仍取得了 87.4% 的准确率和 89.7% 的召回率。综上所述,本文提出的用于驾驶疲劳脑电信号处理与分析的 TDA 方法抗干扰特性好、处理成本低、经济性高,有助于稳定、高效地处理驾驶人脑电数据并检测驾驶疲劳状态,具有较大的科学实际应用价值。

关键词 疲劳驾驶;脑电信号(EEG);拓扑数据分析(TDA);持久同源(PH);支持向量机(SVM)

0 引言

疲劳驾驶是导致交通事故的重要原因之一。据报道,7%的机动车事故和 16%的致命事故都与驾驶人疲劳驾驶有关^[1]。在环境单调的道路上,特别是高速公路上,因疲劳驾驶造成的交通事故比例更是高达 20%^[2]。疲劳驾驶较难通过立法执法进行监管,车载疲劳预警设备是重要的疲劳驾驶预防手段,其核心是驾驶疲劳检测算法^[3],故受到国内外学者的广泛关注。

各国学者从驾驶行为^[4-5]、生理信号^[6-8]、图像信息^[9]等角度切入,提出了不同的疲劳驾驶检测方法,取得了重要研究成果。其中,脑电信号(electroencephalography, EEG)被认为是疲劳检测方法中的“黄金法则”。研究表明,EEG 特征与疲劳的检测密

切相关^[10],作为大脑活动的直接表现形式,其良好的时间分辨率和无创信号采集方法^[11],被频繁地应用于驾驶疲劳的研究中。Lal 等人^[12]利用疲劳过程中 EEG 各频带的变化,建立了检测不同疲劳程度的算法,分析结果显示,受试者疲劳和清醒状态的特征存在显著差异($P < 0.01$)。闵建亮和蔡铭^[13]提取了采集的 EEG 信号的小波对数能量熵(WLE)特征,运用极限学习机(extreme learning machine, ELM)对提取的特征数据集进行分类,其疲劳识别率可以达到 91.8%。Muhammad 等人^[14]从驾驶人脑电数据中提取多种时、频域信号特征,结合心电信号特征来区分清醒与疲劳 2 种驾驶状态,同时对通道简化进行了深入研究,发现只需要结合一个 EEG 与一个眼电图(electro-oculogram, EOG)电极就可以达到 80% 的准确率。经典的时、频域分析往往需要对原始脑电信号进行降噪预处理。基于深度学习(deep

① 浙江省自然科学基金(LQ20E050022)资助项目。

② 男,1997年生,硕士生;研究方向:疲劳驾驶检测,汽车安全技术;E-mail:zhou.uuuuu@qq.com。

③ 通信作者,E-mail:liuzhengqing@zjut.edu.cn。

(收稿日期:2021-07-26)

learning, DL) 的相关方法虽然能够从原始 EEG 数据中直接学习并得到较好的特征信号,但通常需要付出昂贵的训练代价,此外,驾驶人之间的差异性也严重影响着模型的性能表现^[15]。

目前常见的基于 EEG 的驾驶疲劳检测方法,如时域、频域特征和深度学习等往往不能兼顾准确性、鲁棒性、处理速度等方面的性能。由于拓扑数据分析(topological data analysis, TDA)可以在不同尺度下识别数据的形状特征,具有较好的准确性和鲁棒性,近年来在时序数据处理及分析领域受到了学者们的广泛关注。Song 等人^[16]将人体大脑的不同区域映射到网络节点上,从而形成大脑网络拓扑,然后利用拓扑学理论进行情感识别分析。Fernandez-Blanco 等人^[17]将脑电信号的振幅、出现时间等信息编码为不变的星形图,并计算其相关拓扑指数,用于区分病理上的癫痫发作。研究表明,对比时频特征和人工神经网络,使用少量星形图拓扑指标和简单的高斯判别分析(Gaussian discriminative analysis, GDA)模型就能获得较好的预测能力。Das 等人^[18]利用零维和一维同源群的拓扑特征来评估认知疲劳(cognitive fatigue, CF),发现拓扑特征能够有效捕捉 EEG 各频带的变化且对噪声具有较强的鲁棒性,相对利用较少的拓扑特征就能达到与经典功率谱特征基本一致的判别效果,此外,当其与经典功率谱特征相结合进行判别时,性能将优于现有的一些判别方法。文献查阅发现,拓扑数据分析方法目前已经在情感识别^[16]、癫痫检测^[17]、认知疲劳^[18]、利用图像处理检测疲劳驾驶^[19]、大脑网络^[20]、运动想象^[21]、阿尔茨海默病检测^[22]等领域取得了一定的成果。已有文献显示,利用拓扑数据分析处理 EEG 数据用以检测驾驶疲劳方面的相关研究却极少。

为提高 EEG 数据分析的效率和疲劳检测的准确性和鲁棒性,本文在设计驾驶实验,采集受试者 EEG 数据基础上,提出了一种基于 TDA 的驾驶疲劳 EEG 数据处理方法,使用支持向量机对处理结果进行分类检测,并通过对比经典功率谱特征的检测结果,最终得到准确性、鲁棒性和经济性较好的驾驶疲劳 EEG 数据处理方法。

1 实验

本文整体的研究框架如图 1 所示。

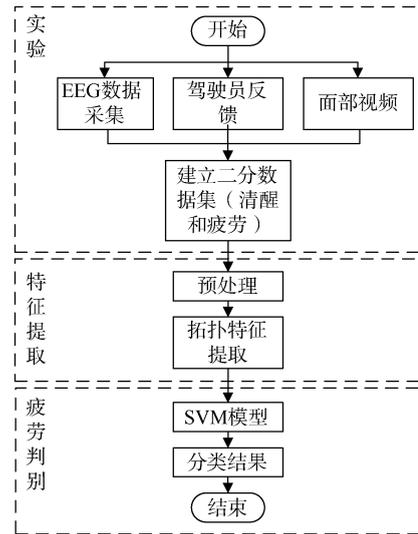


图 1 整体研究框架

实验招募 20 位健康状况良好的受试人员,年龄段为 22 ~ 28 岁。实验期间,受试人员被要求不饮用带有提神效果的饮料并保持正常的作息规律。

驾驶实验所用六自由度汽车性能虚拟仿真实验平台如图 2(a)所示,由六自由度振动台、驾驶仿真系统及模拟驾驶软件组成。模拟驾驶环境设置为天气晴朗的白天,道路设置为闭环高速公路。通过 Emotiv EPOC + 无线便携式脑电仪采集受试人员的脑电信号,采样频率 256 Hz,其电极排布如图 2(c)所示。通过罗技 C270 摄像头实时采集驾驶人面部视频。脑电仪和摄像头由软件统一控制,以确保 2 类数据时域上的对齐。

实验前调查了在正常作息规律情况下,受试人员一天中疲劳或清醒状态明显的时间段。最终选择早上 8:00 - 10:00 开展清醒驾驶实验,中午 12:00 - 14:00 开展疲劳驾驶实验,实验流程如图 3 所示。

实验开始前,每位受试人员有 10 min 的练习和准备时间;实验开始,进行 40 min 模拟驾驶,要求受试者保持正常、安全行车。实验结束,受试者填写疲劳量表,反馈驾驶过程的状态变化。



图2 实验设备

由于驾驶过程中的询问式主观评价,例如 Karolinska 嗜睡量表(Karolinska sleepiness scale, KSS)^[23] 易干扰真实驾驶状态,因此本文选择在实验结束后由驾驶人对疲劳状态进行反馈。此外,考虑到实验后的反馈难以匹配9级KSS疲劳等级,本文对其进行了合理的简化,KSS疲劳等级1~4为清醒状态,5~6为过渡的挣扎状态,7~9为疲劳状态。

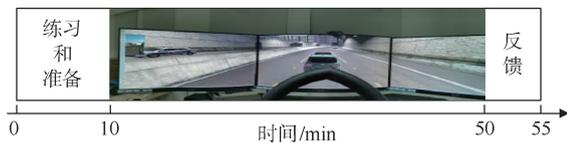


图3 实验流程

2 数据处理

2.1 预处理

使用 EEGLAB 工具箱对实验室采集的 EEG 数据进行预处理,其流程如图4所示。考虑实际应用,应减少通道数量,以保证脑电仪佩戴的舒适性和疲劳检测的处理速度。此外,疲劳相关性较高的 α 波通常在大脑枕区占主导地位^[24]。因此,主要筛选出电极 O1 和 O2 的数据进行分析和研究。在噪声去除阶段,对 EEG 数据进行滤波,滤去与疲劳相关性

不高的频带及 50 Hz 工频噪音等,保留 0.3 ~ 30 Hz 频带。在伪迹去除阶段,通过独立成分分析(independent component analysis, ICA),去除眼电、肌电等伪迹,最终得到“纯净”的 EEG 数据,此外,未进行 ICA 方法处理的 EEG 数据定义为“半纯净”数据。

2.2 拓扑特征提取方法

2.2.1 持久同源性理论

持久同源(persistent homology, PH)是 TDA 的一个强大工具^[25]。其主要目标是利用数据点云中特定结构随空间尺度变化的情况,来区分数据的固有形状与噪声。同源性研究同源结构的集合,其中零维同源特征代表具有相同结构的点的集合,即“集群(clusters)”；一维同源特征代表具有相同结构的线的集合,即“环(loops)”^[26]。持久同源性则进一步研究同源特征在点云空间的持久性,持久性更长的特征被认为是宝贵的数据固有特征而持久性很短的特征则可能是原始数据里的噪声或细节。

2.2.2 持久同源特征与驾驶状态的关系

图5展示了由驾驶人清醒和疲劳状态 EEG 数据的拓扑特征生成的持久同源性图。对于每个特征,用一对数字(*Birth*, *Death*)表示其持久性。持久性值越大, *Death* 与 *Birth* 的差值越大,在图中表现为越远离对角线。

由于在构建数据点云的时候,使用时序与 EEG 数据值进行配对构成了二维平面点云,因此持久性在一定程度上反映了该特征在时序上的持续时间。由于驾驶人清醒时保持着连续变化的持续注意力,因此由清醒样本提取的特征持久性值均不太大,表现为靠近对角线;而在驾驶人疲劳时,部分持续注意力出现衰退且变化迟缓,因此由疲劳样本提取的特征持久性值整体有变大的趋势,表现为远离对角线。即样本同源特征的持久性值的大小反映了驾驶时注意力变化的快慢。

具体如图 5 所示。对于零维持久同源特征

(H_0),其持久性在清醒和疲劳这 2 类样本上的规律同上文所述:对于绝大部分清醒样本,所有特征的持久性值均小于 2(如图 5 的“清醒 2”)或少量略大于 2(如图 5 的“清醒 1”);对于绝大部分疲劳样本,大量特征的持久性值大于 2(如图 5 的“疲劳 1”)或部分远大于 2(如图 5 的“疲劳 3”)。对于一维持久同源特征(H_1),本文尚未发现可靠规律。此外由于本文建立的是二维点云,故只考虑零维(H_0)和一维(H_1)持久同源特征。

因此本文提取实验样本的 H_0 特征,并将其持久性值作为 SVM 的输入开展进一步的分析研究。

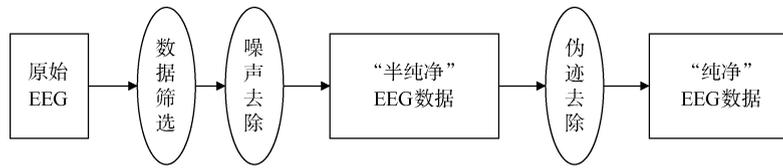


图 4 预处理流程

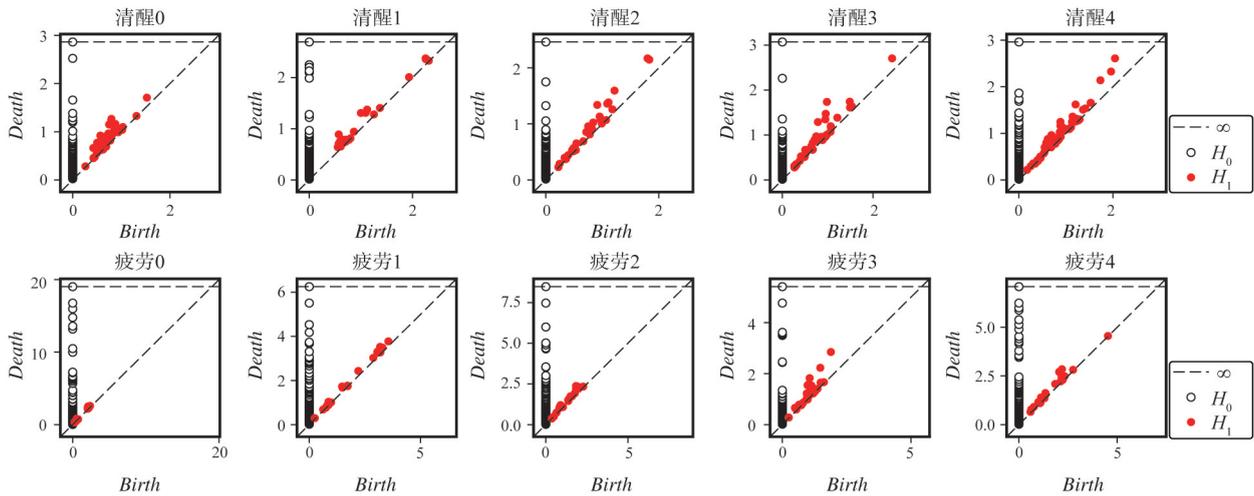


图 5 不同驾驶状态的持久同源性图

3 结果与分析

20 名受试者共完成 65 组实验,不考虑因外界干扰、电极失效等因素造成的失效样本,共计有效样本 52 组。使用 Python 程序对这些有效 EEG 数据样本进行进一步分析。

3.1 疲劳状态分类

“纯净”数据集与“半纯净”数据集训练的具体算法流程如图 6 所示。由于所采集的样本中

清醒类样本数多于疲劳类样本,而这种不均衡将影响模型效果和性能指标的可靠性,因此通过“时间窗重叠”对疲劳样本进行扩充,通过“分层采样法”对部分清醒样本进行删除。随后提取持久同源特征并通过“分层采样法”按照 6:2:2 的比例划分训练集、测试集和验证集。将训练集输入 SVM,得到驾驶状态二分类模型。将测试集输入所得模型,根据性能的优劣对所得模型的参数进行进一步调整。

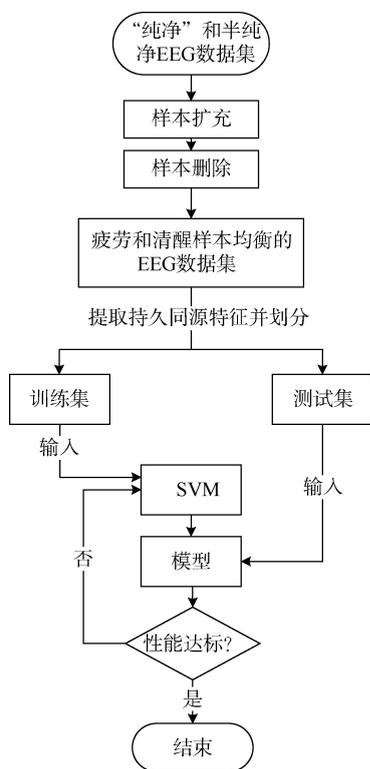


图6 具体算法流程

在参数调整过程中,主要调整了训练集大小、两类样本比例、惩罚系数、核函数等。另外由于在持久同源分析中,持久性值更大的特征通常更具分析的意义,本文选择0维同源特征中持久性值最大的前10%、30%、50%和100%的特征作为SVM的输入,发现使用所有特征作为输入普遍取得了最佳性能。这说明,清醒与疲劳这2类驾驶状态不仅在持久性长的固有结构上存在区别,在持久性短的瞬态结构(如未过滤干净的肢体噪音等)上也存在区分度。

最终,在使用尽可能多的训练集样本(大小约6000例,其中清醒与疲劳样本比例为2:3)、惩罚系数为1、线性核函数的情况下取得了较好的性能。本文考虑的主要性能指标如表1所示。

3.1.1 “纯净”EEG数据集上的性能

图7为拓扑特征与经典功率谱特征在“纯净”EEG数据集上处理结果的对比。结果表明,基于持久同源性的拓扑特征有着与经典功率谱特征相似的性能,在不同EEG片段均取得了较高的准确率和召回率。其中大于2s时,拓扑特征的召回率远高于经典功率谱特征,能够达到更好的性能效果。点云通常需要一定数量的数据点来形成图形结构,基于

表1 性能指标定义

预测结果 \ 实际结果		预测结果	
		预测1(疲劳)	预测0(清醒)
实际1(疲劳)	TP(true positive)	FN(false negative)	
实际0(清醒)	FP(false positive)	TN(true negative)	
对应指标			
性能指标	定义	意义	
准确率 (accuracy)	$Acc = \frac{TP + TN}{TP + FP + FN + TN}$	预测的准确性	
精确率 (precision)	$Pre = \frac{TP}{TP + FP}$	被预测为疲劳的样本中实际疲劳样本的占比	
召回率 (recall)	$Rec = \frac{TP}{TP + FN}$	实际疲劳样本中被预测为疲劳的比例	

持久同源性的拓扑特征在1.5s(384个数据点)以下时性能有所下降。由于仅使用O1和O2电极数据,可以通过增加数据通道数或增大采样频率来取得更好的时间分辨率,即使用更短数据片段来检测是否疲劳。如果继续加长片段至5s以上,数据点云过大,计算分析的经济性不高,且性能提升不大。

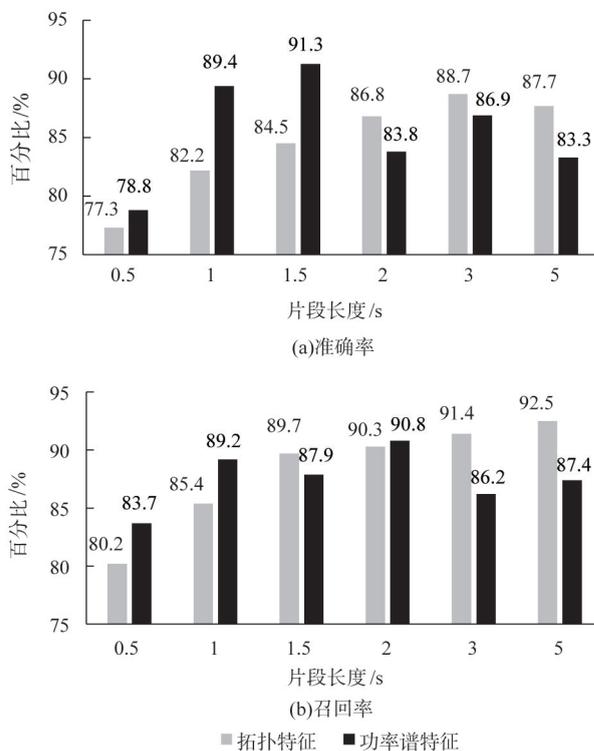


图7 “纯净”EEG数据处理结果

3.1.2 “半纯净”EEG 数据集上的性能

图 8 为拓扑特征与经典功率谱特征在“半纯净”EEG 数据集上处理结果的对比。结果表明,基于持久同源性的拓扑特征在不同片段长度时仍能保持较高的准确率和召回率。准确率最大下跌量为 6.6%,召回率最大下降量为 6.1%;而经典功率谱特征性能则出现了明显的下降,准确率平均下降 19.93%,召回率平均下降 24.98%。伪影去除工作较为复杂,技术要求相对较高。虽然可用自动化算法进行处理,但其准确性有待商榷,算法耗时也高。而拓扑特征对噪音鲁棒性好的特点使其在处理未去除伪影的 EEG 数据时保持了较高的检测性能,且可以使用时域数据作为直接输入,不需要转化到频域,相较于经典功率谱特征,其数据预处理成本低,经济性较好。

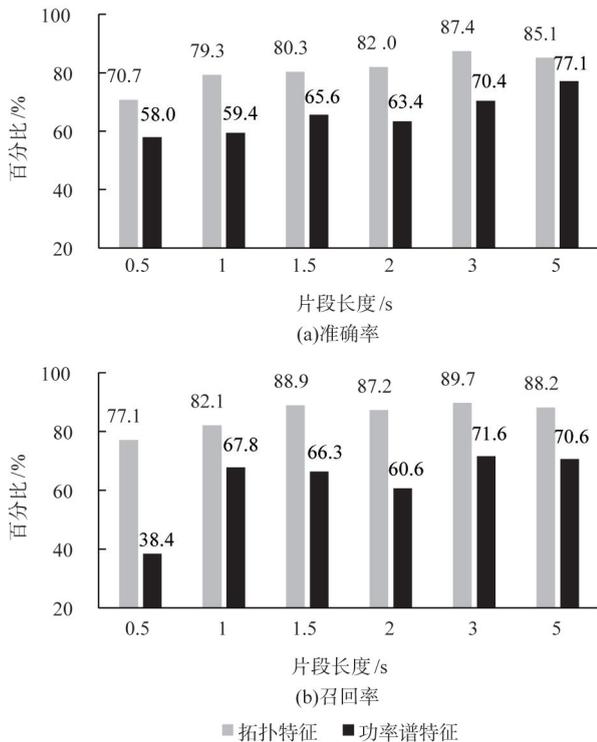


图 8 “半纯净”EEG 数据处理结果

3.1.3 误差分析

由“纯净”EEG 数据训练而来的 SVM 模型,在验证集不同和参数不同的情况下,可以获得较高的召回率,但准确率和精确率则比召回率分别低 3%~6% 和 2%~10%。尤其是精确率,始终无法超过 82%。这意味着,在本文的实验条件和方法下始终存在 $TP > FP > FN > TN$ 的数量关系,即持久同

源性特征对实际疲劳样本有着较高的检测精度,但对实际清醒样本始终存在较低检测精度。

根据疲劳状态产生的原因,驾驶人疲劳可以细分为睡眠相关疲劳 (sleep-related fatigue) 和任务相关疲劳 (task-related fatigue)^[27]。其中任务相关疲劳又可分为被动任务相关疲劳 (通常与轻松的驾驶任务和单调的驾驶环境相关,表现为注意力分散、发呆等) 和主动任务相关疲劳 (通常与艰巨的驾驶任务和复杂的驾驶环境相关,表现为注意力过于集中、紧张等)。在本文设定的实验条件下 (封闭环形高速),被动任务相关疲劳很容易在长时间的清醒驾驶中出现,也易受睡眠相关疲劳的促进而在短时间驾驶后出现。其结果是,清醒状态下连续变化的持续注意力可能转变为与疲劳状态类似的、变化缓慢的状态,并在 PH 上产生持久性较大的特征,并因此导致部分清醒状态会被误检为疲劳状态。

为此,本文提取包含被动任务相关疲劳的样本 (由驾驶人反馈了单调、无聊的情形) 组成“任务相关”验证集,提取不包含被动任务相关疲劳的样本 (驾驶人嗜睡水平较高、从清醒到睡眠相关疲劳迅速转变) 组成“睡眠相关”验证集,观察相同 SVM 模型在上述验证集上的性能变化,结果如图 9 所示。

图 9 中纵坐标表示相对于“纯净”EEG 数据集性能的差值,正值表示性能提升,负值表示性能下降。如图 9 所示,在“睡眠相关”验证集上的性能优于“任务相关”验证集,尤其是精确率差距明显。因此,本文认为,由被动任务相关疲劳引发的特殊疲劳状态是初始模型对实际清醒样本误检率较高的主要原因。这将导致模型预测结果中假正类 FP 增多、真负类 TN 减少,直接降低精确率。此外,由此引发疲劳状态标记不准确、干扰前后 EEG 样本特征等问题,也间接影响了准确率和召回率。

3.2 性能提升方法

本文提出的基于 TDA 的驾驶人疲劳检测方法在 2 s 左右数据片段上取得了优秀的性能。但由于拓扑特征对数据结构完整性有一定要求、特殊疲劳状态引发部分样本结构趋同等原因,仍然存在一些缺陷,因此本文进一步探究了在脑电仪设备限制情况下 (采样频率限制),可能的性能提升方法。

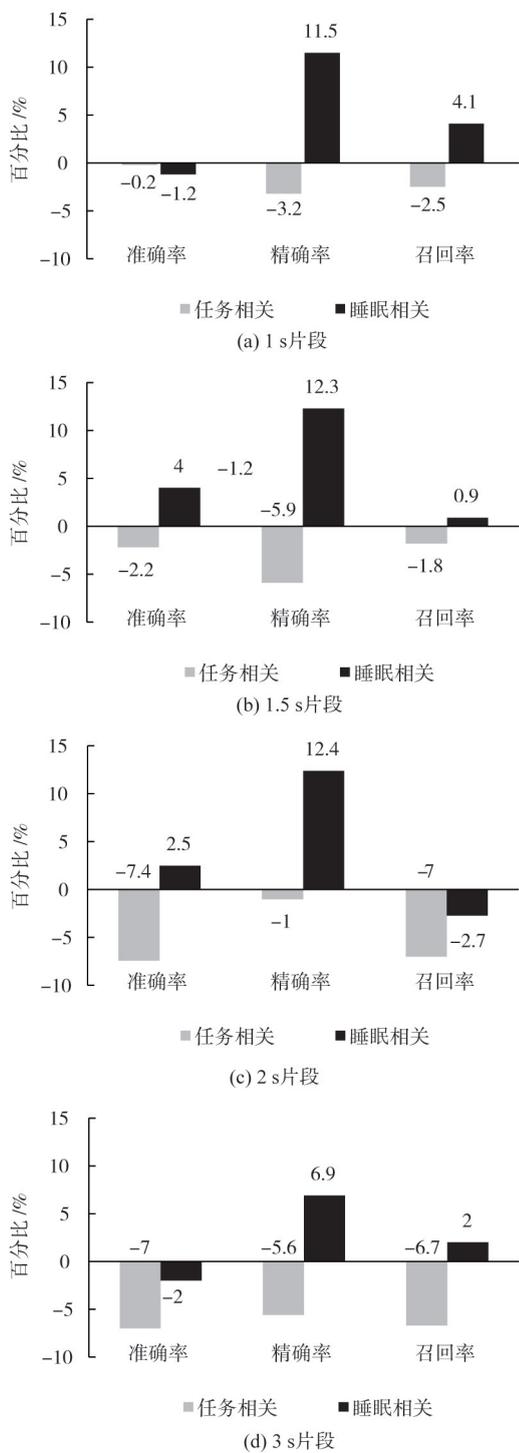


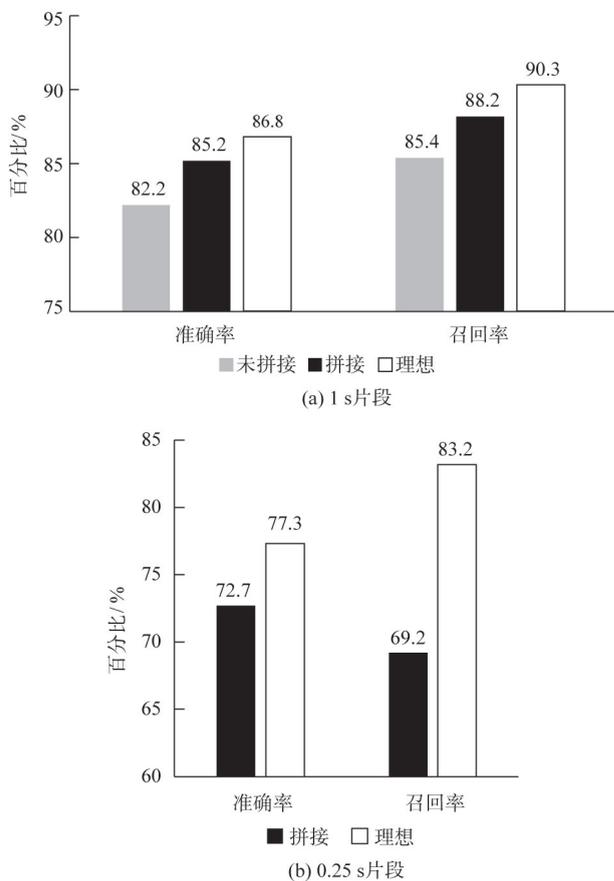
图9 不同验证集结果

3.2.1 增加 EEG 电极以提高时间分辨率

增加的 EEG 电极数据可以通过 2 种方式扩充至数据点云。

方式 1 增大数据维度,每个电极为一维,形成高维数据点云。然而此方式的处理速度显著变慢,极大伤害了实时性,因此不予考虑。

方式 2 保持二维点云,各电极数据直接进行首尾拼接。方式 2 的优势在于没有增大数据维度。结果表明:方式 2 可以提高时间分辨率和检测性能,但由于拼接可能产生的幅值跳跃(见图 10(c)虚线处),干扰拓扑形状特征,电极数量的加倍并不能获得加倍的时间分辨率性能。以图 10(a)为例,2 段 1 s 片段拼接为 1 段 2 s 片段后,性能获得提升,但仍无法达到理想的 2 s 片段性能;随着时间片段的变短,固有结构的破坏程度加剧,如图 10(b)所示,时间分辨率在 0.5 s 以下,利用多电极拼接得到较好的性能难度较大。



(c) 拼接 EEG 图像

图 10 时间分辨率提升结果

3.2.2 结合经典特征以提高精确率

由于 PH 可以从高维数据中直接提取结构特

征,因而具有一定的“黑盒性质”,难以可视化具体结构从而区分被动任务相关疲劳与睡眠相关疲劳。因此考虑结合经典功率谱特征。图 11 展示了一个包含被动任务相关疲劳的样本,在 20 min 之后,驾

驶人处于清醒状态但反馈了被动任务相关疲劳的发生,此时经典功率谱特征没有将其误判为疲劳状态,即没有出现 FP 型误判,说明经典功率谱特征对被动任务相关疲劳的干扰保持了鲁棒性。

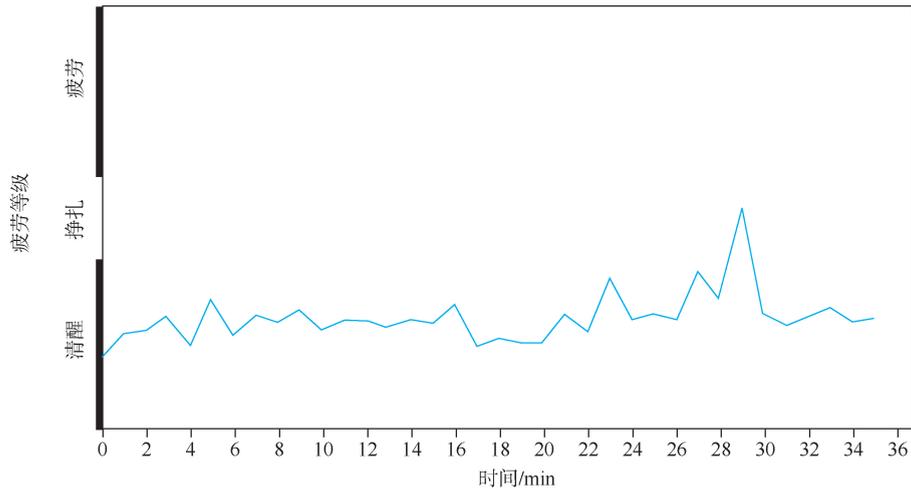


图 11 包含被动任务相关疲劳的清醒状态

因此,本文设计了如表 2 所列的机制用于结合二者优点,并应用在“任务相关”验证集上。结果表明,该机制可以减少 FP 型误判。以 2 s 片段为例,疲劳检测的准确率提升了 2.3%,精确率提升了 6.1%。

表 2 结合机制

功率谱	PH	预测 1(疲劳)	预测 0(清醒)
	预测 1(疲劳)		睡眠疲劳
预测 0(清醒)		被动任务相关疲劳	清醒

4 结论

本文研究拓扑数据分析在 EEG 数据处理和疲劳驾驶领域的应用,组织 20 名受试者开展模拟驾驶实验,采集 EEG 数据。根据实验数据和分析结果,研究了基于 PH 的 EEG 数据分析方法,并与经典功率谱分析方法作对比,得出以下研究结论。

(1) 基于 PH 的 EEG 特征提取方法,避免了 EEG 数据点云的“可视化”或降维,保留所有原始信息,直接从原始的高维特征空间中分析出数据的结构特征,从而达到经济性分析的效果。

(2) 基于 PH 的 EEG 特征提取方法的准确率和时间分辨率都较高,性能与经典功率谱特征相当;对噪音的鲁棒性极强,在未去除伪影的 EEG 数据集上仍然保持了较高的性能,对比经典功率谱特征有着明显优势。

(3) 在本文所提的方法中,增加电极数量、扩充点云和结合经典功率谱特征均可进一步提高 EEG 数据处理的效果和疲劳驾驶检测的性能。

本文提出的方法和结论可对 EEG 等序列数据分析提供参考,也可为疲劳驾驶检测提供新的研究思路。例如可以使用基于 PH 的 TDA 方法简化分析高维脑网络,或解决在实际驾驶场景下 EEG 数据包含过多复杂噪声的问题。未来应考虑:(1) 开展真实驾驶实验,扩大受试者数量,研究本文所提出的基于 PH 的 TDA 方法在真实驾驶场景中的应用;(2) 继续探索数据点云构建方法,提升数据分析效率和性能;训练深度学习模型,优化模型性能。

参考文献

- [1] TEFFT B C. Acute sleep deprivation and culpable motor vehicle crash involvement[J]. Sleep, 2018, 41(10): 1-11.
- [2] 张旭欣,王雪松,马勇,等. 驾驶行为与驾驶风险国际研究进展[J]. 中国公路学报, 2020, 33(6): 1-17.

- [3] 《中国公路学报》编辑部. 中国交通工程学术研究综述·2016[J]. 中国公路学报, 2016, 29(6): 1-16.
- [4] 蔡素贤, 杜超坎, 周思毅, 等. 基于车辆运行数据的疲劳驾驶状态检测[J]. 交通运输系统工程与信息, 2020, 20(4): 77-82.
- [5] 陈志勇, 杨佩, 彭力, 等. 基于BP神经网络的驾驶员疲劳监测研究[J]. 计算机科学, 2015, 42(S1): 67-69, 93.
- [6] TJOLLENG A, JUNG K, HONG W, et al. Classification of a driver's cognitive workload levels using artificial neural network on ECG signals [J]. Applied Ergonomics, 2017, 59: 326-332.
- [7] 李鑫, 张晖, 吴超仲, 等. 基于脉搏波特征融合的驾驶疲劳检测方法[J]. 中国公路学报, 2020, 33(6): 168-181.
- [8] 张兆瑞, 赵群飞, 张朋柱. 基于GRNN的可穿戴式脑电仪EEG疲劳检测[J]. 高技术通讯, 2019, 29(3): 266-273.
- [9] 潘剑凯, 柳政卿, 王秋成. 基于眼部自商图—梯度图共生矩阵的疲劳驾驶检测[J]. 中国图象图形学报, 2021, 26(1): 154-164.
- [10] BORGHINI G, ASTOLF L, VECCHIATO G, et al. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness[J]. Neuroscience and Biobehavioral Reviews, 2014, 44: 58-75.
- [11] BALANDONG R P, AHMAD R F, SAAD M N M, et al. A review on EEG-based automatic sleepiness detection systems for driver[J]. IEEE Access, 2018, 6: 22908-22919.
- [12] LAL S, CRAIG A, BOORD P, et al. Development of an algorithm for an EEG-based driver fatigue countermeasure [J]. Journal of Safety Research, 2003, 34(3): 321-328.
- [13] 闵建亮, 蔡铭. 基于前额脑电多尺度小波对数能量熵的驾驶疲劳检测分析[J]. 中国公路学报, 2020, 33(6): 182-189.
- [14] MUHAMMAD A, NASREEN B, MICHEAL D. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability [J]. Sensors, 2017, 17(9): 1-16.
- [15] ROY Y, BANVILLE H, ALBUQUERQUE I, et al. Deep learning-based electroencephalography analysis: a systematic review[J]. Journal of Neural Engineering, 2019, 16(5): 1-37.
- [16] SONG T F, ZHENG W M, SONG P, et al. EEG emotion recognition using dynamical graph convolutional neural networks [J]. IEEE Transactions on Affective Computing, 2020, 11: 532-541.
- [17] FERNANDEZ-BLANCO E, RIVERO D, RABUNAL J, et al. Automatic seizure detection based on star graph topological indices [J]. Journal of Neuroscience Methods, 2012, 209(2): 410-419.
- [18] DAS A K, KUMAR K, GAVAS R D, et al. Cognitive fatigue detection from EEG signals using topological signal processing [C] // The 28th European Signal Processing Conference. Amsterdam: IEEE, 2021: 1313-1317.
- [19] CHANG Z, BAN X J, WANG Y. Fatigue driving detection based on HAAR feature and extreme learning machine [J]. The Journal of China Universities of Posts and Telecommunications, 2016, 23(4): 91-100.
- [20] LEE H, KANG H, CHUANG M K, et al. Persistent brain network homology from the perspective of dendrogram [J]. IEEE Transactions on Medical Imaging, 2012, 31(12): 2267-2277.
- [21] ALTINDIS F, YILMAZ B, BORISENOK S, et al. Use of topological data analysis in motor intention based brain-computer interfaces [C] // The 26th European Signal Processing Conference. Roma: IEEE, 2018: 1695-1699.
- [22] KUANG L Q, HAN X, CHEN K W, et al. A concise and persistent feature to study brain resting-state network dynamics: findings from the Alzheimer's disease neuroimaging initiative [J]. Human Brain Mapping, 2019, 40(4): 1062-1081.
- [23] AKERSTEDT T, GILLBERG M. Subjective and objective sleepiness in the active individual [J]. International Journal of Neuroscience, 1990, 52(1/2): 29-37.
- [24] 刘晓燕. 临床脑电图学 [M]. 北京: 人民卫生出版社, 2006.
- [25] EDELSBRUNNER H, LETSCHER D, ZOMORODIAN A. Topological persistence and simplification [J]. Discrete and Computational Geometry, 2002, 28(4): 511-533.
- [26] MUNCH E. A users guide to topological data analysis [J]. Journal of Learning Analytics, 2017, 4: 47-61.
- [27] MAY J F, BALDWIN C L. Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies [J]. Transportation Research Part F: Psychology and Behaviour, 2008, 12(3): 218-224.

Analysis of driving fatigue EEG signals based on topological data analysis

ZHOU Feiyang*, LIU Zhengqing*, WANG Qiucheng*, YANG Zhong**

(* College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310023)

(** Deepal Automobile Technology Co. Ltd, Chongqing 400020)

Abstract

In this paper, topological data analysis(TDA) is used to enhance the accuracy and robustness of driving fatigue Electroencephalogram(EEG) signals processing and data analysis. The experiment of occupant driving fatigue is conducted by using a driving simulator. The experimental obtained EEG data are labeled by combining the fatigue

self-evaluation results and recorded driver's facial video, to form a binary data set of wake-up and fatigue. The EEG data are preprocessed by using EEGLAB, and the noises are carefully eliminated and a 0.3 ~ 30 Hz frequency band is retained. The classical frequency domain characteristics α wave energy and α/β are extracted for comparative analysis and research. Finally, the extracted features are classified by support vector machine. The results show that the accuracy of 88.7% and the recall rate of 91.4% are obtained based on the topological characteristics of persistent homology (PH). The performance agrees well with the classical frequency domain features. The topological features are significantly more robust to EEG artifacts, and the accuracy of 87.4% and the recall rate of 89.7% are achieved in the EEG data which has not eliminated artifacts. The proposed TDA method for driving fatigue EEG signals process and analysis has good anti-interference characteristics, low data processing cost, and high economy. It is helpful for stable and efficient detection of driver fatigue state and has a great scientific-practical application value.

Key words: fatigue driving, electroencephalogram (EEG) signal, topological data analysis (TDA), persistent homology (PH), support vector machine (SVM)