

文章编号: 2095-2163(2024)02-0140-04

中图分类号: TP391.1

文献标志码: A

基于随机森林和BP神经网络的船舶驾驶员疲劳检测算法

张威特, 李俊松, 刘雁飞

(浙江理工大学 计算机科学与技术学院, 杭州 310018)

摘要: 本文针对船舶驾驶场景,提出了一种基于随机森林和BP神经网络的船舶驾驶员疲劳检测算法,及时检测船舶驾驶员的疲劳驾驶行为,对安全航行有重要意义。首先,使用随机森林算法对特征数据按照场景分类;其次,根据场景将数据分发到不同的BP神经网络模型进行预测;最后,使用船舶仿真模拟器进行实验验证。结果显示,算法识别准确率为0.82,召回率为0.65,精确率为0.69,说明本文提出的算法对船舶驾驶员疲劳驾驶行为检测有一定实用价值,且与生理数据监测方法相比,简单方便、成本更低,对驾驶员无干扰。

关键词: 船舶驾驶; 疲劳检测; 随机森林; BP神经网络

Ship pilot fatigue detection algorithm based on random forest and BP neural network

ZHANG Weite, LI Junsong, LIU Yanfei

(School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This paper proposes a ship pilot fatigue detection algorithm based on random forest and BP Neural Network for maritime navigation scenarios, which plays a significant role in ensuring safe navigation by timely detecting fatigue driving behaviors of ship pilots. Initially, the Random Forest algorithm is used to classify feature data according to the scenario. Subsequently, the data is distributed to different BP Neural Network models for prediction based on the scenario. Finally, a ship simulator is used for experimental validation. The results show that the algorithm has an accuracy of 0.82, a recall rate of 0.65, and a precision of 0.69. This demonstrates that the algorithm proposed in this paper has practical value in detecting fatigue driving behaviors of ship pilots. Compared to physiological data monitoring methods, it is simpler, more convenient, lower in cost, and non-intrusive to the pilots.

Key words: ship driving; fatigue detection; random forest; BP neural network

0 引言

水上交通事故具有低概率高风险的特征,极易引发重大人身财产损失^[1]。虽然船舶都配备了先进的自动化设备,但人为因素一直是发生事故的主要原因,研究表明约有75%–96%的海上事故是由人为失误引起或者造成的^[2]。疲劳驾驶的情况并不少见,当驾驶员处于疲劳状态时无法及时应对环境变化,从而造成事故^[3]。因此,针对船舶驾驶员驾驶状态的研究在交通安全中具有重要意义。

目前,对于驾驶员疲劳的客观检测方法主要有3种:

(1)基于生理数据检测,如脑电信号和心电信号等,这种方法能够准确反映驾驶员的实时状态,但

是通常需要专门的设备,测试时需要佩戴设备,且价格昂贵;

(2)基于驾驶员特征的检测,如眨眼频率、瞳孔大小等,该方法基于图像检测技术,但图像的拍摄受设备以及环境的影响较大;

(3)基于驾驶行为特征的检测,如速度、方向盘角度等,分析正常状态以及疲劳状态下的操作特征差异,构建判别模型进行检测。

第三种方法对于驾驶员无干扰,成本更低,鲁棒性强,可以整合于驾驶系统中^[4]。

本文通过综合分析驾驶员操作行为特征,提出了一种基于随机森林和BP神经网络的船舶驾驶员疲劳检测算法,可以有效地检测出驾驶员的疲劳状态,从而避免安全事故的发生。

基金项目: 浙江省自然科学基金(LY12C09005)。

作者简介: 张威特(1996-),男,硕士研究生,主要研究方向:交通安全与驾驶员行为;李俊松(1972-),男,博士,讲师,主要研究方向:智能计算。

通讯作者: 刘雁飞(1968-),男,博士,副教授,主要研究方向:交通安全与驾驶员行为。Email: yliu@zstu.edu.cn

收稿日期: 2023-03-04

1 算法构建

船舶驾驶任务中涉及 3 个主要的场景:速度变化、航向变化和速度航向保持,每个场景中的船舶主要参数变化情况并不一致。神经网络通常会使用损失函数来评估当前预测值的好坏,然后调整模型参数,使得损失函数最小化。损失函数的优化会考虑全局情况来寻求一个“均衡状态”,这个均衡状态通常是全局最优的,但未必是各个局部均为最优。因此,使用分类算法对驾驶数据进行场景分类,然后针对不同场景使用不同的神经网络进行预测,可以使预测结果全局最优的同时在各分类内部也能保持最优。

随机森林(Random Forest, RF)是一种集成学习方法,由 Breiman 于 2001 年提出^[4]。随机森林由多个决策树组成,通过对原始数据集进行有放回的随机抽样来生成多个训练集,然后对每个训练集都建立一棵决策树,多棵树组合在一起形成森林,最终输出结果由多颗决策树投票决定。具体步骤如下:

- (1) 随机选择样本:从原始数据集中使用有放回的采样法(Bootstrapping)选取 n 个样本,用于生成一个训练集;
- (2) 随机选择特征:从所有 m 个特征中随机选择 k 个特征($k < m$)作为该决策树节点的划分依据;
- (3) 建立决策树:基于选定的训练集和特征集,建立一棵决策树;
- (4) 重复步骤(1)~步骤(3),生成多棵决策树,通过设置决策树的数量,来控制随机森林的规模和性能;
- (5) 预测:对于一个新的数据样本,在每棵决策树上进行预测,投票决出最终结果。

常见的分类算法还包括多项式逻辑回归和支持向量机,都具有较高的准确性,但与多项式逻辑回归和支持向量机相比,随机森林算法还具有一些独特

优势:对于数据集中的噪点和离群点具有较好的鲁棒性,数据处理效率更高。因此,本文选择随机森林算法来进行场景分类。

BP 神经网络是一种多层次的人工神经网络,主要包括输入层、隐藏层以及输出层。BP 神经网络能够自动提取输入、输出数据之间的规则,并将学习内容置于网络权重之中,具有较强的非线性映射能力,适合求解复杂问题。本文构造了一个 16-8-2 的神经网络模型,其结构如图 1 所示。

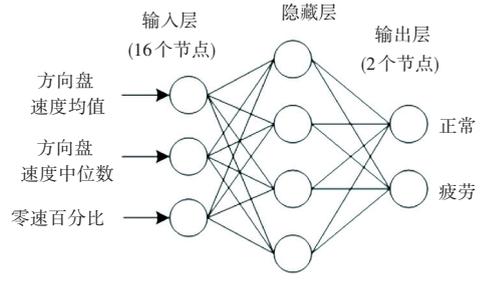


图 1 BP 神经网络结构示意图

Fig. 1 Schematic diagram of BP neural network

隐藏层节点数量采用经验公式(1)确定:

$$s = \sqrt{m + n} + \alpha \tag{1}$$

其中, m 为输入层神经元数量; n 为输出层神经元数量; α 取 $[1, 10]$ 范围的数。

使用 tansig(Tan-sigmoid, 双曲正切 S 型激活函数)作为隐藏层神经元激活函数, logsig(Log-Sigmoid, 对数 S 型激活函数)作为输出层神经元激活函数。学习率 lr 设置为 0.001。当误差函数输出逐渐减小至平缓时,终止训练。

基于随机森林与 BP 神经网络特性构建船舶驾驶员疲劳检测算法主要步骤:首先,提取驾驶行为特征;为了提高准确率,使用随机森林算法对数据进行场景分类,将不同场景的数据送入对应的 BP 神经网络进行判定。算法整体框架如图 2 所示。

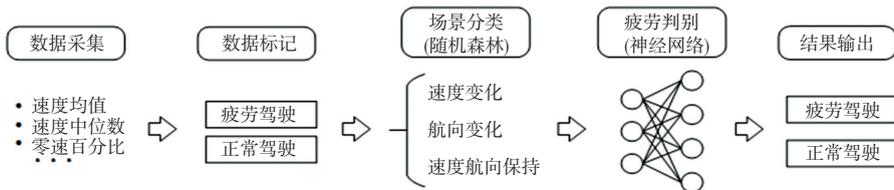


图 2 算法整体框架

Fig. 2 Overall framework of algorithm

2 实验

2.1 数据采集

本文利用船舶仿真系统进行模拟驾驶实验,共招募 12 名被试,年龄 24-32 岁,被试会按要求完

成速度变化、航向变化及速度航向保持(简称保持)任务。在正式实验之前培训所有被试,使被试能熟练掌握船舶驾驶操作。每个被试将进行两次实验,为保证被试每次实验前都处于非疲劳状态,两次实验间隔超过 24 h。

船舶仿真系统每隔 1 s 记录一次船舶相关数据,数据记录主要包括日期和时间、实验编号、口令、驾驶员姓名、船舶经纬度、车钟、速度、加速度、方向盘速度、方向盘角度、航向角、航向角角速度、方向舵舵角等。实验时,被试会循环执行任务,任务具体操作及流程见表 1。

表 1 任务具体操作及流程

Table 1 Specific operations and processes in the task

序号	口令
1	任务开始,准备。
2	速度 10 kn
3	航向 030.0°或航向 330.0°(随机)
4	保持航向、航速 1 min
5	航向 000.0°
6	保持航向、航速 1 min
7	速度 0 kn

为了获取被试状态,每执行两次任务(约 15 min)会对其状态进行一次自我评估,评估结果分为正常与疲劳两个状态。当被试感到疲劳,会继续执行至少 2 次任务,随后由被试自主决定是否要继续执行实验。随着时间的变化,被试疲劳情况也在发生变化,疲劳人数随时间变化情况见表 2。

表 2 疲劳人数随时间变化情况

Table 2 The change of number of fatigue members with time

时间段/min	疲劳人数
0-15	0
15-30	1
30-45	2
45-60	6
60-75	2
75-90	1

由于疲劳状态的检测是每 15 min 一次,因此在被试从正常状态到疲劳状态之间的这段时间内,无法准确确定被试的驾驶状态,因此将该时间段的数据排除。同时,并非被试报告疲劳之后所有的驾驶行为都属于疲劳驾驶,会综合驾驶视频、图像以及被试自我评估结果来判断是否是疲劳驾驶。以 1 min 为时间窗口截取数据,并对数据打上疲劳驾驶或正常驾驶的标签,共计收集到有效样本数据 284 条,其中正常驾驶数据 188 条,疲劳驾驶数据 96 条,疲劳驾驶样本占比 33.8%。实验中共计有 3 个驾驶场景:速度变化、航向变化以及保持,不同场景样本数量见表 3。

表 3 不同场景样本数量

Table 3 Number of samples in different scenarios

场景	正常状态样本数	疲劳状态样本数
速度变化	22	11
航向变化	99	51
保持	67	34

2.2 特征提取

驾驶员在疲劳状态下,反应和信息处理能力会受到影响,导致对速度、方向等维度的控制能力下降,因此与正常状态下的驾驶员在许多方面存在差异^[5]。

驾驶员在疲劳驾驶与正常驾驶下方向盘的典型操作波形如图 3 所示。图 3(a)为驾驶员正常驾驶的方向盘操作波形,可以看出驾驶员方向盘修正频率高,修正幅度小;图 3(b)是疲劳驾驶方向盘操作波形,方向盘修正频率减小,同时修正幅度增大。

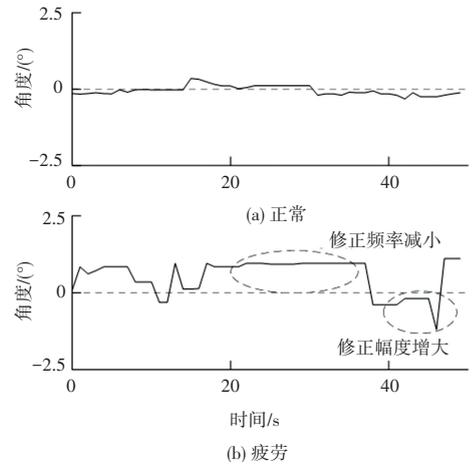


图 3 不同状态下方向盘典型角度波形

Fig. 3 Typical angle waveform of steering wheel in different states

这两种情况下的方向盘操作特征,可以使用方向盘零速百分比进行描述。零速百分比定义如式(2)所示:

$$P = \frac{n}{N} \quad (2)$$

其中, P 代表零速百分比; n 是单位时间内方向盘角速度在 $\pm 0.2^\circ/\text{s}$ 之间的采样数; N 为单位时间内总采样数。

零速度百分比用来衡量一段时间内方向盘不动的程度。为了更全面地评估驾驶员行为,共收集 16 个相关特征:

- (1) 方向盘速度均值、中位数、零速百分比;
- (2) 方向盘角度均值、中位数、标准差、最大值;
- (3) 航向角角速度均值、中位数、标准差;
- (4) 船舶速度的均值、中位数、标准差;
- (5) 船舶加速度的均值、中位数、标准差。

2.3 特征分析

取得 16 项数据特征,每个特征与疲劳驾驶之间的关系可以用皮尔逊相关系数表明。皮尔逊相关系数是一种用于衡量两个变量之间线性相关程度的指标,定义为两个变量之间协方差和标准差的商^[6],定义如式(3)所示:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

其中, ρ 表示皮尔逊相关系数; $\text{cov}(X,Y)$ 表示 X,Y 之间的协方差; σ 表示特征值标准差; μ 表示特征值均值。

通过相关系数来获取最有可能与疲劳相关的特征。通过计算可得不同驾驶特征与疲劳驾驶的相关系数, 见表 4。从特征与疲劳驾驶行为的相关系数来看, 单一特征很难判别具体状态, 需要多个特征进行联合判断。

表 4 不同驾驶特征与疲劳驾驶的相关系数

Table 4 Correlation coefficient between different driving characteristics and fatigue driving

特征	相关系数
方向盘速度均值	0.22
方向盘速度中位数	0.15
方向盘零速百分比	0.33
方向盘角度均值	0.16
方向盘角度中位数	0.10
方向盘角度标准差	0.29
方向盘角度最大值	0.19
航向角速度均值	0.12
航向角速度中位数	0.15
航向角速度标准差	0.08
船舶加速度均值	-0.08
船舶加速度中位数	-0.07
船舶加速度标准差	0.08
船舶速度均值	0.02
船舶速度中位数	0.04
船舶速度标准差	0.09

2.4 模型验证

将数据集分为训练集和测试集, 其中 70% 用于训练, 30% 用于测试, 验证采用十折交叉法。判定过程与模型构建过程类似, 首先通过随机森林进行场景分类, 然后输入到对应的神经网络, 随后由神经网络给出判断结果。实验结果见表 5。

表 5 实验结果

Table 5 Experimental results

样本数量	实际类别	
	正常驾驶	疲劳驾驶
预测类别	正常驾驶	8
	疲劳驾驶	18

由于样本标签分布并不均匀, 除常见的准确率以外, 还引入召回率以及精确率来辅助分析。准确率是指所有样本中被正确预测样本所占的比例; 召回率指的是所有实际为正例的样本中, 被正确预测为正例的样本数所占的比例; 精确率指的是所有被预测为正例的样本中, 实际为正例的样本数所占的

比例。

将逻辑回归、支持向量机、未分类-神经网络算法以及本文方法应用于本次实验的数据, 实验结果如图 4 所示, 可见进行场景分类后的神经网络算法在准确率、召回率和精确率上均有较好表现。

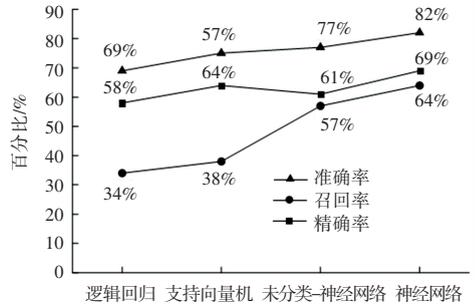


图 4 各个算法实验结果

Fig. 4 Experimental results of different algorithms

3 结束语

本文通过采集船舶仿真系统的船舶运行数据, 提取了包括方向盘零速百分比等在内的 16 项特征数据; 根据具体场景, 使用随机森林算法将数据分为速度变化、航向变化和速度航向保持 3 个场景, 然后根据场景分类将数据送入对应的神经网络, 判断驾驶员的驾驶状态。

通过船舶仿真模拟器进行船舶驾驶实验的结果表明, 基于随机森林和 BP 神经网络的船舶驾驶员疲劳检测算法相较于逻辑回归、支持向量机和未分类的神经网络算法的船舶驾驶员驾驶状态识别的准确率、召回率和精确率都具有一定优势, 且检测过程对驾驶员无干扰。基于随机森林和 BP 神经网络的船舶驾驶员疲劳检测算法在船舶驾驶员疲劳驾驶检测方面具有一定实用意义。

参考文献

- [1] 刘清, 王志远, 杨柳, 等. 基于 EEG 的船舶驾驶员工作负荷识别研究[J]. 武汉理工大学学报(交通科学与工程版), 2022, 46(5): 759-763.
- [2] HANZU PAZARA R, BARSAN E, ARSENIU P, et al. Reducing of maritime accidents caused by human factors using simulators in training process[J]. Journal of Maritime Research, 2008, 5(1): 3-18.
- [3] 靳明浩. 基于人脸识别的驾驶人员疲劳检测系统设计[J]. 智能计算机与应用, 2020, 10(2): 172-173, 179.
- [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] 毛喆, 严新平, 吴超仲, 等. 疲劳驾驶时的车速变化特征[J]. 北京工业大学学报, 2011, 37(8): 1175-1183.
- [6] 徐维超. 相关系数研究综述[J]. 广东工业大学学报, 2012, 29(3): 12-17.