

统计分析与建模

期末项目说明

选题一 基于驾驶任务数据的疲劳定级

• 背景

- 基于司机的驾驶任务统计数据（当前驾驶时长、累计行车告警等数据，详见选题一数据说明），完成对驾驶员的疲劳定级，分为二分类定级和三分类定级

• 建模数据

- 二分类数据集
- 三分类数据集

• 方法

- 逻辑回归模型（二分类）
- 朴素贝叶斯模型（三分类）

• 模型指标

- (1) 准确率 (Accuracy)
- (2) 查全率 (Recall) 查准率 (Precision) F1指标
- (3) ROC曲线 (接受者操作特征曲线) 下的面积AUC (Area Under Curve)

• 评分要点

- 数据预处理
- 建模过程与分析
- 模型指标

speed	fatigue_res	driver_age	age_level	continuuoi	continuuoi	alarm_time	circadian_r	depart_tim	total_alarm	total_alarm	alarm_bet	alarm_bet	alarm_face	alarm_face	spare_time	total_fatig	accumulat	max_spare
40	1	47	B	7537	33.37291	29315	13.62842	21778	21	0	0.00013	0.012346	1	1	8.44E-06	46.85027	0	86400
68	2	53	B	3312	13.90755	46722	40.74108	43410	11	0	0.000329	0.005464	2	0	1.95E-05	54.49756	16799	51360
83	1	41	A	501	0.628664	31256	11.79223	30755	2	0	0.015152	0.026316	2	2	7.63E-06	12.26983	0	86400
50	1	49	B	665	1.405774	45296	32.49836	44631	1	0	0.029412	0.029412	1	1	5.74E-05	33.75308	3895	17412
72	1	40	A	4207	18.09814	49748	58.47664	45541	12	0	0.000417	0.022222	1	1	0.001639	76.42372	11885	73441
54	2	40	A	1692	6.268669	49236	55.65854	47544	4	0	0.002165	0.025641	2	2	8.50E-05	61.77614	13894	47793
79	1	49	B	1197	3.925874	49807	58.79175	48610	7	0	0.000905	0.037037	3	3	0.000267	62.56656	5386	7237
79	2	37	A	1338	4.59347	53848	71.76977	52510	4	0	0.004464	0.013889	4	3	1.42E-05	76.21217	7508	70562
61	0	45	B	1751	6.547722	45408	33.12014	43657	11	0	0.001742	0.017544	4	3	0.000119	39.5168	2816	8423
54	2	52	B	1221	4.03952	45614	34.27746	44393	4	0	0.005181	0.02	4	3	7.41E-05	38.16592	11921	52425
58	1	53	B	995	2.969186	45285	32.43759	44290	6	0	0.001479	0.029412	4	3	0.000119	35.25572	3690	8374
83	1	54	B	2615	10.62776	59229	52.06902	56614	3	0	0.000433	0.003831	1	0	0.000102	62.54573	7332	17759
43	2	44	A	2571	10.42033	24304	32.86144	21733	5	0	0.000402	0.018182	1	1	8.12E-06	43.1307	0	86400
54	2	47	B	2835	11.6643	41361	15.69927	38526	14	0	0.000453	0.000627	0	0	5.98E-05	27.2125	2400	16709
58	1	52	B	2958	12.24331	24513	31.66991	21555	7	0	0.000614	0.002227	0	0	8.04E-06	43.76216	0	86400
40	1	54	B	3048	12.66674	23202	39.43148	20154	4	0	0.000445	0.010989	1	1	8.46E-06	51.94716	0	86400
72	1	38	A	1439	5.071563	54171	71.79976	52732	12	0	0.001344	0.025641	5	3	0.000131	76.72026	6112	17486
65	1	51	B	2874	11.84793	45923	36.04361	43049	15	0	0.000413	0.006173	2	1	0.000128	47.74047	12754	59590
68	1	49	B	1427	5.014765	45323	32.64776	43896	1	0	0.027778	0.027778	1	1	0.00011	37.51146	10100	55768
61	1	44	A	715	1.642684	44310	27.29174	43595	4	0	0.002268	0.006211	2	1	1.44E-05	28.78336	4212	69401
65	2	44	A	1669	6.159873	45141	31.6471	43472	7	0	0.00068	0.018868	3	1	9.60E-06	37.65591	0	86400

选题二 基于面部时序特征的疲劳定级

- 背景

- 基于司机的驾驶面部时序特征（详见选题二数据说明），完成对驾驶员的疲劳定级

- 建模数据

- 三分类驾驶眼部时序数据集，每一个事件包含若干秒内的眼部关键点坐标信息

- 方法

- 非时序模型：从眼部关键点中提取出统计特征，如（PERCLOS、左右眼平均睁开百分比等），之后构建分类模型
- 时序分类模型（如RNN、LSTM等）

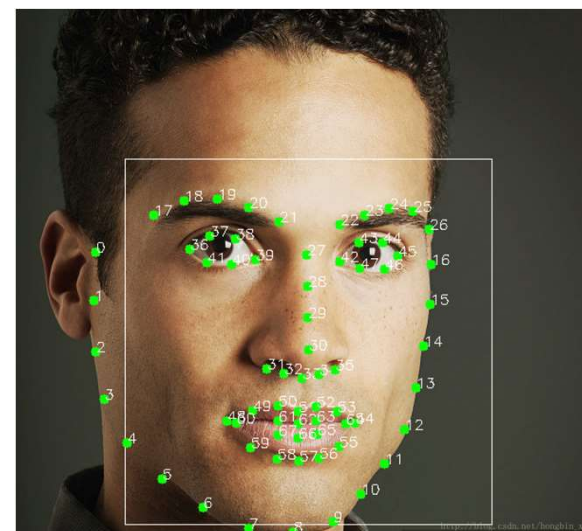
- 模型指标

- (1) 准确率 (Accuracy)
- (2) 查全率 (Recall) 查准率 (Precision) F1指标
- (3) ROC曲线（接受者操作特征曲线）下的面积AUC（Area Under Curve）

- 评分要点

- 特征提取方法
- 建模过程与分析
- 模型指标

	frame	x	y
0	1	227	214
1	1	225	224
2	1	226	234
3	1	229	245
4	1	233	255
5	1	238	264
6	1	243	273
7	1	249	280
8	1	257	282
9	1	268	281
10	1	279	276
11	1	291	270
12	1	301	262
13	1	308	251
14	1	311	238
15	1	313	224
16	1	313	210
17	1	230	193
18	1	231	188
19	1	235	186



PERCLOS介绍

卡内基梅隆研究所提出的度量疲劳/瞌睡的物理量 PERCLOS (Percentage of Eyelid Closure over the Pupil over Time, 简称PERCLOS)。其定义为单位时间内 (一般取1 分钟或者 30 秒) 眼睛闭合一定比例 (70% 或80%) 所占的时间

PERCLOS 的计算公式如下:

$$\text{PERCLOS} = \frac{\text{眼睛闭合帧数}}{\text{检测时间段总帧数}} \times 100\%$$

P70:指眼睑遮住瞳孔的面积超过70%就计为眼睛闭合, 统计在一定时间内眼睛闭合时所占的时间比例。

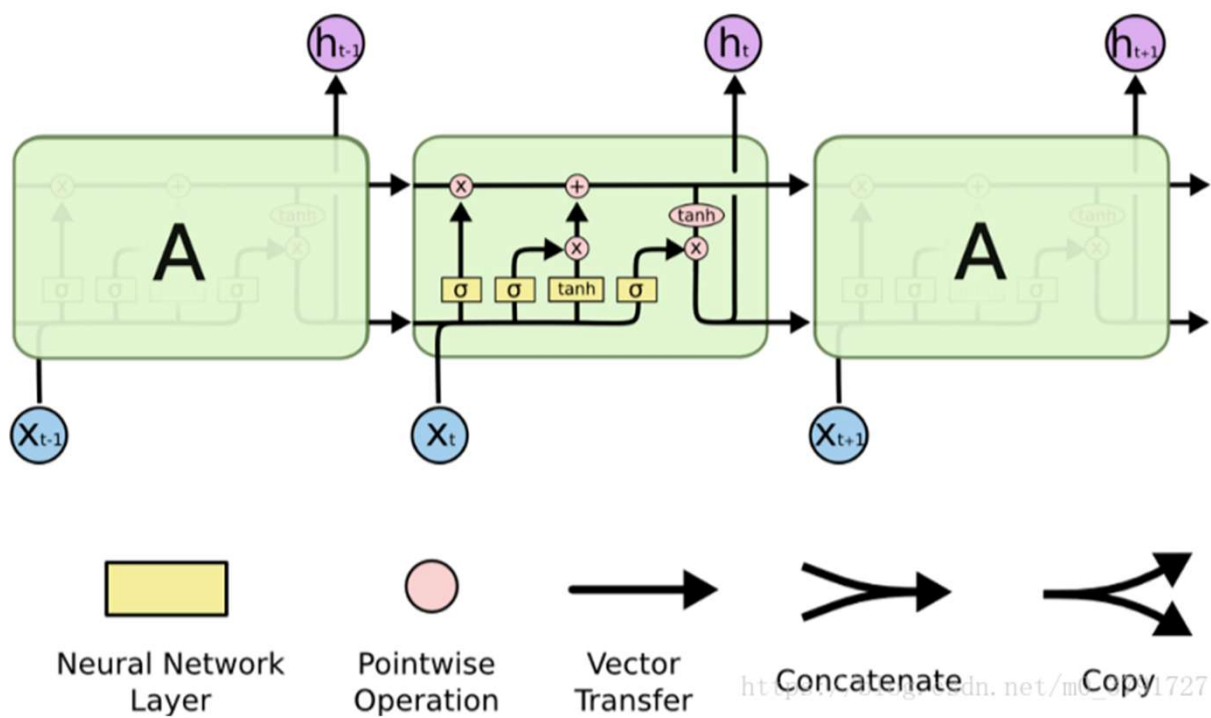
P80:指眼睑遮住瞳孔的面积超过80%就计为眼睛闭合, 统计在一定时间内眼睛闭合时所占的时间比例。

选题三 碰撞事故的时序数据分析预测

- 背景
 - 基于两份时序数据（详见选题三数据说明），完成对时序数据的分析与预测
- 建模数据
 - 上海快速路交通事故数据集
 - 污染物浓度监测数据
- 方法
 - ARIMA
 - 其他时序预测模型（LSTM等）
- 模型指标
 - （1）均方误差（MSE）
- 评分要点
 - 数据预处理
 - 时序数据分析（两份时序数据对比）
 - 建模预测

day	crash_number
2011/10/1	62
2011/10/2	47
2011/10/3	56
2011/10/4	59
2011/10/5	59
2011/10/6	50
2011/10/7	45
2011/10/8	63
2011/10/9	64
2011/10/10	87
2011/10/11	72
2011/10/12	109
2011/10/13	91
2011/10/14	87
2011/10/15	130
2011/10/16	54
2011/10/17	85
2011/10/18	98
2011/10/19	97
2011/10/20	96
2011/10/21	108
2011/10/22	88

LSTM介绍



- 遗忘门
- 记忆门
- 输出门

LSTM代码示例

Pytorch

```
class LSTM(nn.Module):#注意Module首字母需要大写
    def __init__(self, input_size=1, hidden_layer_size=100, output_size=1):
        super().__init__()
        self.hidden_layer_size = hidden_layer_size

        # 创建LSTM层和Linear层, LSTM层提取特征, Linear层用作最后的预测
        # LSTM算法接受三个输入: 先前的隐藏状态, 先前的单元状态和当前输入。
        self.lstm = nn.LSTM(input_size, hidden_layer_size)
        self.linear = nn.Linear(hidden_layer_size, output_size)

        #初始化隐含状态及细胞状态c, hidden_cell变量包含先前的隐藏状态和单元状态
        self.hidden_cell = (torch.zeros(1, 1, self.hidden_layer_size),
                             torch.zeros(1, 1, self.hidden_layer_size))
```

Keras

```
model = Sequential()
model.add(LSTM(32, input_dim=64, input_length=10, return_sequences=True))
# note that you only need to specify the input size on the first layer.
# for subsequent layers, no need to specify the input size:
model.add(LSTM(16, return_sequences=True))
model.add(LSTM(10))
```

项目要求

- 自由组队，3-4人为一组
- 字体、字号按模板要求
- 不超过10页

提交方式与给分

- 分组（共享文档，截止日期：本周末12.12 23:59前）
- 小组成员按百分比分配（上限为40分满分）
- 每个选题评优数目上限：小组数/3
- 评分标准
 - 文献检索 25%
 - 数据分析与建模 25%
 - 评估与解读 25%
 - 答辩 25%
- 参与荣誉计划的同学需额外采用一种扩展方法



统计分析与建模期末项目...

扫一扫二维码打开或分享给好友



- 腾讯文档 -

可多人实时在线编辑，权限安全可控