

基于车辆运行数据的疲劳驾驶状态检测

蔡素贤¹, 杜超坎¹, 周思毅¹, 王雅斐^{*2}

(1. 汉纳森(厦门)数据股份有限公司, 福建 厦门 361000; 2. 上海财经大学 信息管理与工程学院, 上海 200433)

摘要: 疲劳驾驶是导致交通事故的主要原因之一, 及时检测疲劳驾驶, 并提醒驾驶员集中注意力, 对保证安全行车具有重要意义. 本文基于CAN(Controller Area Network)总线采集的车辆运行状态数据, 提取了18项与驾驶行为相关的特征, 并采用随机森林算法对疲劳驾驶进行识别, 结果表明整体的识别准确率为0.785, 其中召回率为0.61, 即61%的疲劳驾驶状态可被识别出来. 实验表明, 基于车辆运行状态的疲劳驾驶检测具有一定的效果, 且与其他客观的疲劳驾驶检测方法(基于驾驶员生理指标和图像面部特征)相比, 具有简单方便, 不影响驾驶, 且成本低的优势.

关键词: 智能交通; 疲劳检测; 随机森林; CAN数据

Fatigue Driving State Detection Based on Vehicle Running Data

CAI Su-xian¹, DU Chao-kan¹, ZHOU Si-yi¹, WANG Ya-fei²

(1. Honorsun(Xiamen) Data Company Limited by Shares, Xiamen 361000, Fujian, China; 2. School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China)

Abstract: Fatigue driving is one of the main causes of traffic accidents. It is of great importance to detect fatigue driving dynamically and remind drivers to concentrate on driving safely. Based on the vehicle running data collected by Controller Area Network (CAN) bus, this paper extracts 18 features relevant to driving behaviors and uses random forest algorithm to identify fatigue driving. The results show that the overall recognition accuracy is 0.785, and the recall rate is 0.61 which means 61% of fatigue driving conditions can be successfully identified. Experiments show that fatigue driving detection based on vehicle running data is effective. Compared with other fatigue driving detection methods (for example, based on driver physiological indicators and image facial features), the proposed method is simple and convenient, without affecting driver's operations and the cost is relatively low.

Keywords: intelligent transportation; fatigue detection; random forest; CAN data

0 引言

公共交通的安全运营很大程度取决于驾驶员, 而疲劳驾驶是导致交通事故的主要原因之一^[1]. 当驾驶员处于疲劳驾驶状态时, 注意力不够集中, 无法快速应对紧急情况, 如突然的行人路过或前车紧急刹车等^[2]. 因此, 分析疲劳驾驶时驾驶员的

行为特征, 及时检测疲劳驾驶, 并提醒驾驶员集中注意力, 对保证安全行车具有重要意义.

目前, 国内外对疲劳驾驶的客观检测方法主要有三种^[3]. (1) 基于驾驶员生理指标, 如脑电信号^[4]、心电信号^[5]、脉搏等; (2) 基于驾驶员行为特征, 如眼动特征^[6]、瞳孔大小、眨眼频率等; (3) 基于

收稿日期: 2020-02-28

修回日期: 2020-05-12

录用日期: 2020-05-14

基金项目: 国家自然科学基金/National Natural Science Foundation of China(61773248); 国家社科基金重大项目/ Major Program of the National Social Science Foundation of China(18ZDA088).

作者简介: 蔡素贤(1988-), 女, 福建厦门人, 工程师.

*通信作者: yafei_wang@163.sufe.edu.cn

车辆运行状态,如车道偏离、转向盘握力^[7]、转向盘转角^[8]、油门刹车等.

第一种方法依赖于直接接触驾驶员,需驾驶员佩戴设备进行疲劳检测.通过驾驶员处于疲劳状态时在生理指标(如脑部组织含氧情况、心电信号、肌电信号等)特征上的变化,检测驾驶员是否处于疲劳状态^[9].这种方法检测疲劳驾驶准确度高,但是成本高且有可能影响驾驶.第二种方法依赖于图像检测,根据驾驶员的眼动特征对驾驶员的疲劳状态做出判断^[10].这种方法受周围环境(如光暗)及摄像头位置的影响较大.第三种方法中,通过分析正常与疲劳驾驶时,车速、发动机转速、方向盘转角、刹车、油门、车道偏离等数据的差异,生成用于检测疲劳驾驶的模型^[11].目前已有许多学者研究从车辆运行时的方向盘转角速度、车辆行驶速度、发动机转速等数据判断疲劳驾驶员的疲劳状态.如利用BP神经网络对驾驶员疲劳驾驶时的车辆行驶特性进行训练,并建立疲劳驾驶行为的检测模型^[12].其中,车道偏离依赖于道路车道情况(清晰度)的影响,较难采集.而转向、油门刹车、车速等数据采集较为简单,同时较为稳定可靠.研究表

明,当驾驶员处于疲劳状态时,对油门刹车、换挡及转向盘等操作能力会下降^[13].

当前主流的疲劳驾驶检测是通过防疲劳监控设备对司机面部图像进行判别,这种方法在实际应用中受到周围环境及司机眼部大小的影响较大,同时设备较为昂贵.而通过车辆运行数据进行判别,成本较低,且数据较为可靠.目前,基于车辆运行数据的疲劳驾驶研究大部分采用特定的实验数据,与实际司机的驾驶有一定的偏差.本文以CAN(Controller Area Network)总线采集的真实的车辆运行数据为基础,综合分析疲劳驾驶的数据规律模式,其结论更符合实际情况,可有效用于司机的疲劳检测与预警.

1 车辆运行数据与疲劳驾驶检测

本文基于CAN总线采集实际的车辆运行数据和防疲劳监控设备上报的疲劳预警数据,将车辆运行数据切分为清醒状态样本与疲劳状态样本,构建判断疲劳驾驶的模型,实现通过车辆运行数据检测疲劳驾驶状态的目的.整体流程如图1所示.

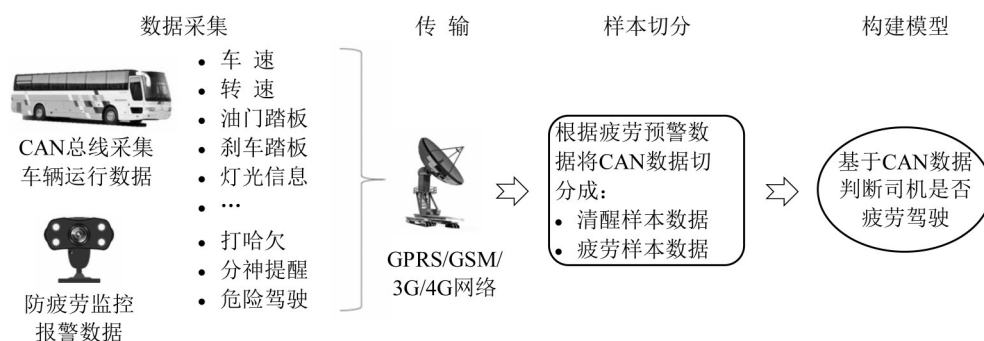


图1 基于车辆运行数据的疲劳驾驶检测流程

Fig. 1 Fatigue driving detection process based on vehicle operating data

(1) 数据采集与传输.

数据采集部分包括实时采集车辆的运行数据和疲劳预警的数据,并通过GPRS/GSM/3G/4G网络将数据传输到数据库.其中,CAN总线采集的车辆运行数据包括的维度主要有记录日期和时间、车辆ID、驾驶员、位置经度、位置纬度、CAN速度、上下行方向、电机转速、发动机转速、档位、电子刹车、手刹、油门踏板百分比、点火信号、司机制动踏板开度、开关状态(转向灯信号、门开关信号).防疲劳监控报警数据主要包括记录时间、线路ID、车牌

号、驾驶员ID、预警的行为(疲劳相关的行为包括打哈欠、分神提醒和危险驾驶)及视频图片.其中,分神提醒:当发现驾驶员进入瞌睡状态,或者低头看手机1~3 s等分神动作,设备将发出分神提醒;危险驾驶:触发分神提醒后,仍然没有恢复正常的驾驶姿势,则触发危险驾驶报警.

(2) 疲劳驾驶与清醒状态的数据切分.

疲劳驾驶数据的判断规则为:当某段时间持续上报疲劳驾驶相关的行为预警时,专家根据上报的图片和视频再判断该段时间司机是否为疲劳

驾驶,若是,则为这段时间打上疲劳驾驶的标签.若未出现预警信息,则为这段时间打上清醒状态的标签.基于时间段标签,将车辆运行状态数据切分为疲劳状态数据和清醒状态数据.不同状态数据最后再以1分钟的时长切割为一个样本.

(3) 构建疲劳驾驶检测模型.

基于疲劳驾驶样本与清醒状态样本,构建二分类模型,使得可基于CAN采集的车辆运行数据识别出疲劳驾驶,从而可在司机驾驶过程中,对疲劳状态进行预警,以提高驾驶安全.



图2 构建疲劳驾驶检测模型

Fig. 2 Building a fatigue driving test model

2.1 数据样本描述

数据来自某公交公司同一条线路,通过询问3名公交司机来人工筛选出其疲劳驾驶和清醒状态驾驶的CAN数据(车辆运行状态数据),数据采集间隔为0.2 s.时间窗口设定为1 min,收集到的样本数如表1所示.

表1 疲劳驾驶和清醒驾驶的样本数量
Table 1 Number of samples for fatigue driving and conscious driving

样 本	驾驶员 A	驾驶员 B	驾驶员 C	总 共
疲劳驾驶	59	14	40	113
清醒驾驶	103	42	93	238

疲劳驾驶样本数量占比32.19%,将近1/3的比例.

2.2 驾驶特征提取

当司机处于疲劳驾驶时,其对于车辆的驾驶预判能力和感知外界环境的能力会有所下降,导致司机对于车速、油门踏板、刹车踏板、转弯等的控制稳定性和灵活度与清醒状态有一定的差异.因此,我们可从这些维度中提取与疲劳驾驶相关的特征.

CAN采集的与司机驾驶行为相关的车辆运行数据维度包括:记录时间、车牌号、驾驶员ID、档位、速度、加速度、油门踏板开度、刹车踏板开度、转向灯信号、门开关信号以及GPS经纬度.根据业务理解和数理统计的知识,对收集的驾驶数据进行特征提取,包括以下特征:

2 求解算法

本文基于车辆运行数据,构建疲劳驾驶检测模型,主要分为以下几个步骤:首先从样本数据中提取驾驶行为特征,包括司机对车速、油门踏板开度、刹车踏板开度、转弯等的操作特征;再通过分析各个特征之间的相关关系以及特征与是否疲劳的关系;最后,采用随机森林(Random Forest)算法对疲劳驾驶数据进行判别,达到预警的目的.整体的模型框架如图2所示.

(1) 车速的均值、中位数和标准差:车速均值表示车速的快慢、中位数描述车速的集中趋势,标准差表现司机对车速控制的平稳性.

(2) 车速的样本熵.样本熵(SampEn)是一种用于度量时间序列复杂性的方法.

(3) 加速度的均值、中位数和标准差.

(4) 加速度绝对值 $>2\text{ m/s}^2$ 的比例:统计每分钟记录的加速度绝对值 $>2\text{ m/s}^2$ 的记录次数占总记录数的比例.加速度绝对值超过阈值 2 m/s^2 说明司机急加速或急刹车.

(5) 油门踏板开合度的均值、中位数、标准差、最大值和样本熵.

(6) 刹车踏板开合度的均值、中位数、标准差、最大值和样本熵.

一共提取了18个与司机驾驶行为相关的特征.

2.3 特征分析

通过分析驾驶特征与是否疲劳驾驶之间的关系,可用于获取与疲劳驾驶最相关的特征,为判别疲劳驾驶提供数据基础.所提取的特征数量共18个,它们之间与是否疲劳的相关关系可采用皮尔森相关系数表示.

皮尔森相关系数(Pearson Correlation Coefficient)^[14]是一种线性的相关系数,可以用来体现两个变量之间的相关程度.两个特征之间的皮尔逊相关系数,为两者的协方差和标准差的之商:

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y} \quad (1)$$

式中: X, Y 表示两个变量的特征值所构成的向量; μ 为特征的均值; σ 为特征的标准差. ρ 的值范围在 $[-1, 1]$ 之间, 越接近 1, 则两个特征正相关性越强, 越接近 -1, 则两个特征负相关性越强. 各驾驶行为特征与是否疲劳驾驶的相关性程度如表 2 所示.

表 2 驾驶行为特征与疲劳驾驶的相关系数
Tabel 2 Correlation coefficient of driving behavior characteristics

特 征	是否疲劳驾驶
车速中位数	0.30
车速均值	0.35
车速标准差	0.05
车速样本熵	-0.09
加速度中位数	0.03
加速度均值	-0.02
加速度标准差	-0.02
加速度绝对值大于 2 的比例	0.11
油门踏板开度中位数	0.10
油门踏板开度均值	0.09
油门踏板开度标准差	0.09
油门踏板开度最大值	0.12
油门踏板开度样本熵	0.14
刹车踏板开度中位数	-0.08
刹车踏板开度均值	-0.19
刹车踏板开度标准差	-0.20
刹车踏板开度最大值	-0.15
刹车踏板开度样本熵	-0.14

从相关系数看来, 单个特征与疲劳驾驶的相关度不高.

2.4 疲劳驾驶识别(随机森林)

随机森林(Random Forest, RF)^[15]是一种基于分类树的集成学习算法, 由 Breiman 于 2001 年提出. 随机森林是由随机子空间算法和装袋算法集成的一种算法, 其基本原理是通过随机采样特征和样本, 生成很多决策树, 每一颗决策树是不相关的, 将多棵决策树组合在一起就形成森林. 通过各决策树进行投票决策, 最终选择多数投票(Bagging)的策略来决定结果^[16]. 具体的算法步骤如下:

(1) 记原始训练集中有 M 个特征, 样本总数为 N . 采用 Bootstrap 抽样技术, 从训练集中抽取 N 个样本形成训练子集.

(2) 随机选取 m 个特征作为特征子集 ($m \leq M$), 从这 m 个特征中选择最优的切分点再

做节点分裂, 直到节点的所有训练样例都属于同一类. 节点通常按基尼指数、信息增益率、均方差等规则分裂, 且在分裂过程中完全分裂不剪枝.

(3) 重复(1)、(2)步骤 k 次即可得到由 k 棵决策树构建而成的随机森林.

(4) 使用随机森林进行决策. 若设 x 代表测试样本, k 代表决策树数量, h_i 代表单棵决策树, $i \in \{1, \dots, k\}$, Y 代表输出变量即分类标签, I 为指示性函数, H 为随机森林模型, 则决策公式^[16]为

$$H(x) = \arg \max_Y \sum_{i=1}^k I[h_i(x) = Y] \quad (2)$$

基于采集的车辆运行数据(档位、速度、油门踏板开度、刹车踏板开度、时间等), 共提取了驾驶行为特征 18 项. 随机森林在处理高维特征的样本数据时, 通常能得到极好的准确率. 且在训练完之后, 随机森林能够给出特征重要度的排名. 随机森林算法还具有很强的抗干扰能力. 如果数据有很大一部分的特征遗失, 用随机森林算法仍然可以维持准确度. 此外, 该算法的抗过拟合能力也很强^[17].

3 模拟计算与结果分析

本文集中于分析疲劳状态下的驾驶行为特征, 并通过差异性来检测疲劳. 通过随机分配的方式, 将数据集的 70% 拆分为训练集, 30% 拆分为测试集. 本次实验使用了随机森林(RF)对数据进行训练, 算法经过剪枝和参数调优^[18], 得到最终的实验结果如表 3 所示.

表 3 实验结果
Table 3 Experimental result

样本数量		实际类别	
		清醒驾驶	疲劳驾驶
预测	清醒驾驶	64	13
类别	疲劳驾驶	10	20

考虑到本文所使用数据的样本标签分布并不均匀, 因此在评价实验结果所采用的指标中除了判断算法的准确率以外, 也使用了召回率和精确率来进一步对有偏数据的实验结果进行评价^[19]. 具体指标说明如下.

Recall: 召回率, 表示疲劳驾驶被正确检测出来的概率, 即预测为疲劳驾驶实际也为疲劳驾驶

的样本数与总的实际疲劳驾驶的样本数的比值,即 $20/(20+13) = 0.61$.

Precision: 精确率,表示正确预测出疲劳驾驶的比例,即预测为疲劳驾驶实际也为疲劳驾驶的样本数与总的预测为疲劳驾驶的样本数的比值,即 $20/(20+10)=0.67$.

Accuracy: 准确率,识别正确的概率,实际为清醒驾驶预测也为清醒驾驶的样本数与实际为疲劳驾驶预测也为疲劳驾驶的样本数之和,再除以总的样本数,即 $(64+20) / (64+20+13+10) = 0.785$.

将逻辑回归、支持向量机和决策树算法应用于本次实验的数据,实验结果如表4所示.

表 4 各类算法实验结果				
Table 4 Experimental result on models				
算 法	实际分类	逻辑回归	支持向量机	决策树
疲劳驾驶	33	11	12	15
清醒驾驶	74	96	95	92

将各类算法的结果与随机森林做比较,如图3所示.

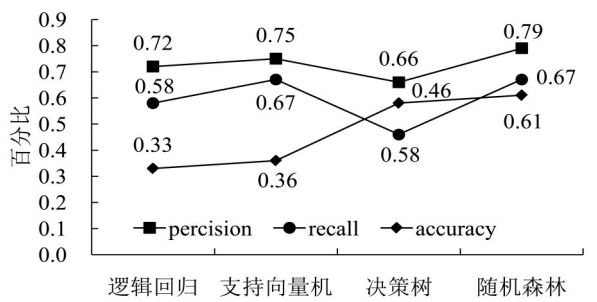


图3 疲劳驾驶检测模型指标对比
Fig. 3 Comparison on fatigue driving test models

可以看出随机森林的准确率和召回率均大于其他三种算法,说明随机森林算法相比较更适用于对疲劳驾驶的判别.

4 结 论

本文基于车辆运行时采集的车速、油门踏板开合度、刹车踏板开合度等数据,提取与驾驶行为相关的18项特征,对驾驶员的疲劳状态进行了分析和检测,与基于驾驶员生理指标和图像面部状态的疲劳驾驶检测相比,具有方法简单且成本低的优势.本文实验采用了随机森林、逻辑回归、支持向量机和决策树这几种算法对样本数据进行训练并测试训练结果.实验表明,随机森林算法对疲

劳状态的识别效果最好,其中:召回率为0.61,即表明其中有61%的疲劳状态能被识别出来;精确率为0.67,即识别出来的疲劳驾驶状态有67%的概率是真的疲劳驾驶状态;总体的准确率为0.785,表明随机森林算法的识别准确率为78.5%.结果表明通过车辆运行状态数据检测驾驶员疲劳状态具有一定的可行性和实用性.

论文在基于车辆运行数据对驾驶员疲劳状态监测方面进行了有价值的探索,涉及更大规模样本数据,考虑疲劳等级因素与个体差异化因素,可进一步提高疲劳驾驶检测准确率与可信度的相关实验与方法,尚有待进一步深入研究.

参考文献:

[1] 彭明,杨雪峰,应江婷,等.疲劳驾驶警示系统在公交车的应用探讨[J].公路交通科技(应用技术版),2011,7(6): 291-293. [PENG M, YANG X F, YING J T, et al. Discussion on the application of fatigue driving warning system in bus[J]. Highway Traffic Technology (Applied Technology Edition), 2011, 7(6): 291-293.]

[2] 李都厚,刘群,袁伟,等.疲劳驾驶与交通事故关系[J].交通运输工程学报,2010,10(2): 104-109. [LI D H, LIU Q, YUAN W, et al. Relationship between fatigue driving and traffic accident[J]. Journal of Transportation Engineering, 2010, 10(2): 104-109.]

[3] 游俊.驾驶员疲劳驾驶检测系统研究[D].哈尔滨:哈尔滨工程大学,2016. [YOU J. Research on driver fatigue driving detection system[D]. Harbin: Harbin Engineering University, 2016.]

[4] 王斐,吴仕超,刘少林,等.基于脑电信号深度迁移学习的驾驶疲劳检测[J].电子与信息学报,2019,41(9): 2264-2272. [WANG F, WU S C, LIU S L, et al. Driver fatigue detection through deep transfer learning in an electroencephalogram-based system[J]. Journal of Electronics and Information Technology, 2019, 41(9): 2264-2272.]

[5] JEONG I C, LEE D H, PARK S W, et al. Automobile driver's stress index provision system that utilizes electrocardiogram[C]. 2007 Ieee Intelligent Vehicles Symposium, Istanbul, 2007: 652-656.

[6] 牛清宁,周志强,金立生,等.基于眼动特征的疲劳驾驶检测方法[J].哈尔滨工程大学学报,2015,36(3): 394-398. [NIU Q N, ZHOU Z Q, JIN L S, et al. Detection of driver fatigue based on eye movements[J]. Journal of Harbin Engineering University, 2015, 36(3): 394-398.]

- [7] 沙春发, 李瑞, 张明明. 基于方向盘握力的疲劳驾驶检测研究[J]. 科学技术与工程, 2016, 16(30): 299–304. [SHA C F, LI R, ZHANG M M. Detecting fatigue driving based on steering wheel grip force[J]. Science Technology and Engineering, 2016, 16(30): 299–304.]
- [8] 李作进, 李仁杰, 李升波, 等. 基于方向盘转角近似熵与复杂度的驾驶人疲劳状态识别[J]. 汽车安全与节能学报, 2016, 7(3): 279–284. [LI Z J, LI R J, LI S B, et al. Driver fatigue recognition based on approximated entropy and complexity of steering wheel angle[J]. Journal of Automotive Safety and Energy, 2016, 7(3): 279–284.]
- [9] 薛春杰. 驾驶员疲劳驾驶的生理信号研究[J]. 纳税, 2017(17): 197. [XUE C J. Research on physiological signal of driver fatigue driving[J]. Ratepaying, 2017(17): 197.]
- [10] 柳龙飞, 伍世虔, 徐望明. 基于人脸特征点分析的疲劳驾驶实时检测方法[J]. 电视技术, 2018(12): 27–30. [LIU L F, WU S Q, XU W M. Real-time fatigue driving detection based on analysis of facial landmarks[J]. TV Technology, 2018(12): 27–30.]
- [11] 万蔚, 王振华, 王保菊. 基于驾驶行为的疲劳驾驶判别算法研究[J]. 道路交通与安全, 2016(6): 21–24. [WANG W, WANG Z H, WANG B J. Research on detection of fatigue driving based on driving behaviors [J]. Road Traffic and Safety, 2016(6): 21–24.]
- [12] 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 (s1): 67–69, 93. [CHEN Z Y, YANG P, PENG L, et al. Fatigue driving monitoring based on BP neural network[J]. Journal of Computer Science, 2015 (s1): 67–69, 93.]
- [13] 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. [MAO J, YAN X P, WU C Z, et al. Analysis of velocity changing rules under driving fatigue[J]. Journal of Beijing University of Technology, 2011, 37(8): 1175–1183.]
- [14] 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. [XU W C. A review on correlation coefficients[J]. Journal of Guangdong University of Technology, 2012, 29(3): 12–17.]
- [15] CUTLER A, CUTLER D R, STEVENS J R. Random forests[M]. Ensemble Machine Learning. Springer, Boston, MA, 2012: 157–175.
- [16] 杜续, 冯景瑜, 吕少卿, 等. 基于随机森林回归分析的PM2.5浓度预测模型[J]. 电信科学, 2017(7): 20172111–20172110. [DU X, FENG J Y, LV S Q, et al. PM2.5 concentration prediction model based on random forest regression analysis[J]. Telecom Science, 2017(7): 20172111–20172110.]
- [17] 毛树华, 王先朋, 文江辉, 等. 基于自回归条件持续期模型的疲劳驾驶研究[J]. 交通运输系统工程与信息, 2018, 18(3): 85–91, 111. [MAO S H, WANG X P, WEN J H, et al. Fatigue driving detection based on autoregressive conditional duration model[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(3): 85–91, 111.]
- [18] LIUZ K, HWARD D. Bondell. Binormal precision–recall curves for optimal classification of imbalanced data[J]. Statistics in Biosciences, 2019, 11 (1): 141–161
- [19] RAND W M. Objective criteria for the evaluation of clustering methods[J]. Journal of the American Statistical Association, 1971, 66(336): 846–850.