

基于模糊 C 均值和神经网络的驾驶行为评价研究^①

吴紫恒^{1,2}, 吴仲城¹, 张俊¹, 陈松^{1,2}, 陈杰^{1,2}

¹(中国科学院合肥物质科学研究院, 强磁场中心, 合肥 230031)

²(中国科学技术大学, 合肥 230031)

通讯作者: 吴紫恒, E-mail: ziheng88@mail.ustc.edu.cn

摘要: 交通拥挤正成为一个日益严重的问题, 一些不安全的驾驶行为所导致的交通事故是造成拥堵的主要原因之一. 因此, 如何准确评价驾驶员的驾驶行为成为研究的热点. 本文提出了一种基于模糊 C 均值聚类 (Fuzzy C-Means, FCM) 和 BP 神经网络相结合的驾驶行为评价方法, 首先利用 FCM 对驾驶行为进行初始聚类, 基于 FCM 聚类结果, 为了提高 BP 神经网络分类精度, 本文提出了一种自动挑选训练样本即典型样本的方法, 利用 BP 网络进行学习, 最终用训练得到的 BP 神经网络分类器对驾驶行为进行实时分类, 研究结果表明该算法摒弃了人为主观因素, 实现了驾驶行为准确、客观、高效的评价.

关键词: 驾驶行为; 模糊 C 均值聚类; 神经网络; 典型样本

引用格式: 吴紫恒, 吴仲城, 张俊, 陈松, 陈杰. 基于模糊 C 均值和神经网络的驾驶行为评价研究. 计算机系统应用, 2018, 27(3): 263-267. <http://www.c-s-a.org.cn/1003-3254/6256.html>

Study on Driving Behavior Evaluation Based on Fuzzy C-Means and Neural Network

WU Zi-Heng^{1,2}, WU Zhong-Cheng¹, ZHANG Jun¹, CHEN Song^{1,2}, CHEN Jie^{1,2}

¹(High Magnetic Field Laboratory, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China)

²(University of Science and Technology of China, Hefei 230031, China)

Abstract: Traffic congestion is becoming an increasingly serious problem. Traffic accidents caused by risky driving behaviors are one important cause. Therefore, the accurate evaluation of driving behaviors has become a research hotspot. This study puts forward an evaluation algorithm of driving behaviors based on the combination of FCM and BP neural network. Firstly, FCM is used to make initial clusters of driving behaviors. Secondly, in accordance with the results of clusters, an algorithm that the typical samples are automatically selected as the training samples for BP neural network classifier is proposed. Finally, the trained BP neural network is used to classify the driving behaviors. The research result shows that the algorithm can eliminate subjective factors and make accurate, objective and efficient driving behavior evaluation.

Key words: driving behavior; Fuzzy C-Means (FCM) clustering algorithm; neural network; typical samples

随着社会的快速发展, 汽车已经是越来越多人首选交通工具, 给人们出行带来了非常大的便捷. 但另一方面汽车数量的快速增长, 会给公路交通带来一定的压力, 交通拥挤、交通事故等现象对人们的生活造成了不小的影响, 同时也阻碍了交通运输的发展. 驾驶行

为与道路交通安全有着紧密联系^[1]. 规范驾驶人的驾驶行为是保障道路交通安全的重要举措^[2]. 近年来, 有关驾驶行为与道路安全关系的研究备受关注, 牛增良等以大量重大交通事故数据为基础, 通过聚类分析从危险驾驶行为角度研究外部影响因素与驾驶行为的关

^① 基金项目: 国家自然科学基金 (61273323)

收稿时间: 2017-06-30; 修改时间: 2017-07-14; 采用时间: 2017-07-17; csa 在线出版时间: 2018-02-09

系^[3]; 钱宇彬等以油门踏板行程与车速作为操作行为基本参数, 通过聚类分析实现基于驾驶行为特性的长途客车驾驶时间和换班管理方法^[4]; 庄明科等人探讨了风险驾驶行为与驾驶经验、驾驶技术、驾驶态度以及驾驶员的人格之间的关系^[5]; 孙伟等进行了疲劳驾驶行为的研究^[6]; 任慧君等提出了基于 GPS 数据的驾驶行为特征提取方法, 统计分析了 GPS 轨迹数据, 但缺乏深入挖掘数据背后的信息^[7]; 白云等认为危险驾驶行为可以是失误性驾驶行为、错误性驾驶行为、非故意违规的驾驶行为和过激性驾驶行为^[8]; Tseng 通过对中国台湾地区出租车的超速驾驶数据的研究发现, 驾驶员的行驶风格、工作时长和日行驶距离与超速驾驶之间有着密切的联系^[9]; Herrera 等搭建了基于 GPS 定位手机获取交通数据的系统平台, 但未将其与驾驶行为关联^[10]; Greaves 等采用 GPS 技术获取驾驶数据以弥补问卷调查精度不高的问题, 但其仅挖掘了驾驶人超速驾驶行为^[11].

目前, 对驾驶速度行为的评价一般是通过一系列的驾驶行为减分单元: 急加速减分单元、急减速减分单元、急转弯减分单元等, 根据急加/减速、急转弯次数等数据进行减分, 实现驾驶行为评级分数计算^[12]. 评价过程的主观性较强, 评价结果不够科学、准确. 当前鲜有从数据挖掘的角度出发研究驾驶行为与道路安全的潜在关系. 本文从车联网数据库中查询样本车辆的行驶卫星定位信号集合并提取驾驶行为特征参数, 以此为指标通过模糊 C 均值聚类 (Fuzzy C-Means, FCM) 聚类算法实现驾驶行为的科学聚类, 得到驾驶行为特征参数的聚类结果; 根据 FCM 聚类结果, 选取其中的典型样本作为训练样本, 利用 BP 网络进行学习, 利用训练后 BP 神经网络分类器对经归一化处理的待分类驾驶行为进行分类评价. 本文通过对车联网数据进行聚类分析, 提高了驾驶行为评价的准确性和科学性, 建立了车联网数据、驾驶行为特征、道路安全三者的关联机理.

1 基于模糊 C 均值和神经网络的驾驶行为评价算法

本文所提出的基于模糊 C 均值和神经网络的驾驶行为评价算法流程图如图 1 所示, 其步骤为: (1) 基于车联网大数据选取样本数据, 提取驾驶行为特征参数;

(2) 确定聚类个数, 运行 FCM 聚类算法得到基于车联网大数据的驾驶行为聚类结果; (3) 根据 FCM 算法的聚类结果及隶属度约束条件选取典型样本; (4) 设计并创建 BP 网络, 将典型样本聚类结果归一化作为训练样本, 训练 BP 网络; (5) 待评估的驾驶行为样本归一化, 利用训练完毕的 BP 神经网络进行驾驶行为评价.

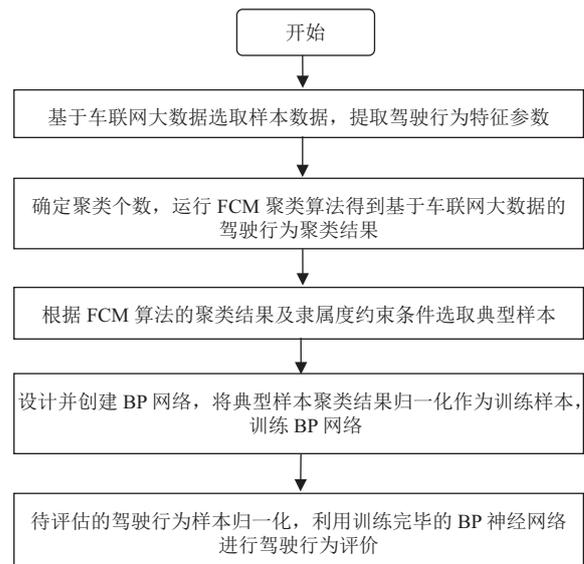


图 1 基于 FCM 聚类和 BP 神经网络的驾驶行为评价方法算法流程

1.1 驾驶行为特征参数提取

车联网数据来自某车辆监控平台, 被监控车辆已装备车载终端设备, 利用北斗卫星导航系统测定车辆的地理位置, 并实时报告车辆地理位置、瞬时速度和运行方位角等信息, k 时刻每个采样点的信息为 $I_k = \{t_k, v_k, j_k, w_k, h_k\}$, 其中 t 为时间, v 为速度, j 为经度, w 为纬度, h 为方向角.

根据任意连续采样点的信息, 可以计算车辆在 k 时刻的加速度值 A_k :

$$A_k = (v_k - v_{k-1}) / (t_k - t_{k-1})$$

根据给定的急加速的加速度阈值和急减速的加速度阈值可以判断车辆是否发生急加速和急减速. 根据给定的急转弯的速度限值和行驶方向改变阈值, 可以判断车辆是否发生了急转弯. 通过以上判断获得车辆的瞬时速度、急变速、急转弯的时间和位置, 并进行相关统计, 可以获得驾驶行为特征参数: 速度 80–120 km/h 时间占比、速度 > 120 km/h 时间占比、100 km 平均急变速次数、100 km 平均急转弯次数.

1.2 模糊 C 均值算法 (FCM)

给定数据 X , FCM 算法^[13]将 X 划分成 C 类, 其聚类中心为 $V = [v_1 \ v_2 \ \dots \ v_c]$, 分类结果用隶属度表示, 其中隶属度集为 U . FCM 的目标函数定义为:

$$J_{FCM}(X, U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d^2(x_k, v_i) \quad (1)$$

$$d(x_k, v_i) = \|x_k - v_i\|^2 \quad (2)$$

其中 U 必须满足:

$$u_{ik} \in \{0, 1\}, \sum_{i=1}^c u_{ik} = 1 \quad (3)$$

由于很难直接求解此类目标函数的最优值, 在实际应用中一般采用交互优化的求解方式即交替迭代, 直至目标函数收敛到指定精度内的最优值. 在迭代过程中, 模糊隶属度和聚类中心的更新式如式 (4) 和式 (5):

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{1/(m-1)}} \quad (4)$$

$$v_c = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m}, \quad 1 \leq i \leq c, \quad 1 \leq k \leq n \quad (5)$$

针对 FCM 算法随机选取初始聚类中心易使算法收敛到陷入局部最小的问题, 本文引入了一种初始聚类中心的选取方法^[14]对传统 FCM 算法进行改进. 初始聚类中心选择的步骤为:

步骤 1. 计算任意两个样本间的距离, 并生成距离矩阵 D , 将距离最近的两个样本定为一类, 并取两个样本点的中点作为第一类聚类中心.

步骤 2. 选定距离阈值 a , 利用距离矩阵 D 计算与第一类中的两个样本距离都大于 a 的所有样本, 并在这些样本选择距离最近的两个点定为一类, 取两个样本的中点作为第二类的聚类中心.

步骤 3. 同理, 在剩下的样本中找到与已经找到的样本距离都大于 a 的样本, 并选择这些样本中的距离最近的两个点定为一类, 并取两个样本点的中点作为其聚类中心.

步骤 4. 重复步骤 3, 直到找到 C 类聚类中心为止.

基于驾驶速度行为数据的 FCM 聚类算法流程图如图 2 所示.

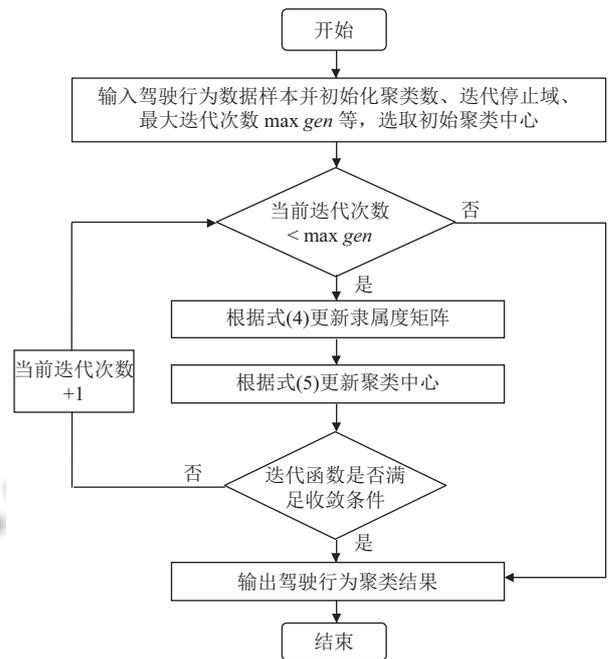


图 2 基于驾驶行为特征数据的 FCM 聚类算法流程图

1.3 典型样本选取

FCM 聚类作为非监督聚类算法, 可以根据预先设定的类别数自动聚类, 并得到每个样本隶属于各个类别的隶属度矩阵. 本文所说的典型样本为仅对某个类别具有较高隶属度而对其他类别隶属度极低的样本. 本文根据 FCM 聚类结果隶属度矩阵 U 自动挑选 BP 神经网络的训练样本 $Y(i)$, 训练样本需要满足如下的约束条件:

$$\begin{cases} U(i)_{1st} > U^* \\ U(i)_{1st} - U(i)_{2nd} > U^\# \end{cases}$$

式中 U 为利用 FCM 算法对驾驶行为数据聚类后得到的隶属度矩阵; $U(i)_{1st}$, $U(i)_{2nd}$ 为样本 $Y(i)$ 的隶属度矩阵中最大值和次大值; U^* 、 $U^\#$ 为设置的用于挑选典型样本的阈值. 本文 U^* 设置为 0.5, $U^\#$ 设置为 0.2.

1.4 BP 神经网络结构设计

神经网络已经被广泛应用于人工智能、机器学习等领域, 其中, BP 神经网络是目前应用最广泛的神经网络模型之一, BP 神经网络是一种按误差逆向传播算法进行训练的多层前馈网络. 已经证明, 具有 sigmoid 非线性函数的三层神经网络可以任意逼近任何连续函数^[15], 因此本文采用三层 BP 神经网络结构. BP 神经网络由输入层、隐含层和输出层组成. 其中输入层作为神经网络的输入对应输入特征向量, 本文中驾驶行为

特征参数为4维,即BP神经网络的输入节点为4个;隐含层由10个神经元组成;输出层对应分类的结果,本文输出层节点数为1,数值为驾驶行为对应的类别.将典型样本聚类结果归一化作为训练样本,训练BP网络,利用训练完毕的BP神经网络进行驾驶行为分类评价.

2 仿真实验

车联网数据来源于某车辆监控平台,共500辆车,一个月的驾驶行为特征提取后的数据如表1所示.

表1 驾驶行为特征参数数据(部分)

车辆编号	80-120 km/h 时间占比(%)	>120 km/h 时间占比(%)	100 km 急变速次数	100 km 急转弯次数
1	1.8	0.14	2.76	2.76
2	12.71	6.35	3.23	1.57
3	2.98	0.33	2.02	0.31
4	2.62	0.21	1.89	0.47
5	1.58	1.05	2.89	1.26
·	·	·	·	·
·	·	·	·	·
·	·	·	·	·

高速行驶往往会造成交通事故,研究表明,车速越高,发生事故的风险越大,微小的车速变化会急剧提升发生事故的概率,在不同的道路环境中,最高安全行车速度是不同的;车辆加速度急剧变化时,会给驾驶员不舒适的感觉,因而急加速或急减速都存在一定的安全隐患;在车辆驾驶中,急转弯不仅会给车辆带来很大的磨损,而且很有可能会造成翻车等交通事故;因此本文选取“速度80-120 km/h时间占比(%)”、“速度>120 km/h时间占比(%)”、“100 km平均急变速次数”、“100 km平均急转弯次数”作为驾驶行为特征参数.

设置聚类类别为4,采用FCM聚类分析,得到聚类中心如表2所示.由表2可知,类别1无明显超速倾向,急变速急转弯次数较小,此类驾驶行为发生交通事故风险最小;类别2无明显超速倾向,急变速急转弯次数较大;类别3有一定的超速倾向,急变速急转弯次数较小;类别4超速倾向严重,急变速急转弯次数较大,此类驾驶行为风险较高,发生交通事故的风险很大,驾驶员应该受到交通运输企业或管理部门的重点监控.

基于FCM聚类结果,根据本文所提出的自动挑选BP神经网络训练样本的方法挑选典型样本.设计BP神经网络,将典型样本的驾驶行为特征数据归一化

作为BP神经网络的输入,典型样本的类别数值作为BP神经网络的输出.BP网络最大训练次数为5000,目标误差为0.001.利用BP网络进行学习,可以得到对驾驶行为进行实时分类的BP神经网络分类器.待分类的驾驶行为特征数据为[1.33 0.35 1.45 0.66],将待分类驾驶行为特征数据归一化后作为BP神经网络分类器的输入,输出为1.09,即认为此驾驶行为特征数据属于类别1.

表2 FCM聚类得到的4个类别的聚类中心

类序号	8-120 km/h 时间占比(%)	>120 km/h 时间占比(%)	100 km 急变速次数	100 km 急转弯次数
1	1.26	0.19	1.63	0.71
2	1.20	0.16	3.35	2.13
3	5.12	0.87	1.66	0.76
4	11.64	4.21	2.04	1.23

3 结论

本文根据车辆监控平台获取的车联网数据,提取并量化了驾驶行为特征的参数指标,采用模糊聚类方法对驾驶行为特征数据进行聚类,基于FCM聚类结果,为了提高BP神经网络分类精度,提出了一种自动挑选训练样本即典型样本的方法,利用BP网络进行学习,最终用训练得到的BP神经网络分类器对驾驶行为进行实时分类.本文通过驾驶行为安全等级分类,挖掘出驾驶行为习惯,建立了车联网数据、驾驶行为特征、道路安全三者的关联机理,为交通管理部门重点监控较高风险驾驶人提供了参考依据,为UBI(Usage Based Insurance)驾驶行为评价提供一种新的方法.

参考文献

- 鲁光泉,宋阳. 车路协同环境下驾驶行为特性与交通安全综述. 交通信息与安全, 2014, 32(5): 13-19.
- 陶瑞华,魏恒. 机动车道的非直接左转交通措施对驾驶员行驶特性和安全的影响. 交通运输系统工程与信息, 2009, 9(3): 55-63.
- 牛增良,李海斌,王文峰,等. 基于聚类分析的营运驾驶人危险驾驶行为研究. 山东交通学院学报, 2014, 22(1): 19-23.
- 钱宇彬,刘浩学,吴长水,等. 夜间长途客车驾驶员操作行为差异性识别研究. 中国安全科学学报, 2014, 24(3): 9-15.
- 庄明科,白海峰,谢晓非. 驾驶人员风险驾驶行为分析及相关因素研究. 北京大学学报(自然科学版), 2008, 44(3): 475-482.

- 6 孙伟, 张为公, 张小瑞, 等. 疲劳驾驶检测方法的研究进展. 汽车技术, 2009, (2): 1-5, 17.
- 7 任慧君, 许涛, 李响. 利用车载 GPS 轨迹数据实现公交车驾驶安全性分析. 武汉大学学报 (信息科学版), 2014, 39(6): 739-744.
- 8 白云, 石京. 北京市驾驶行为与影响因素分析. 交通信息与安全, 2010, 28(2): 114-119.
- 9 Tseng CM. Operating styles, working time and daily driving distance in relation to a taxi driver's speeding offenses in Taiwan. *Accident Analysis & Prevention*, 2013, 52: 1-8.
- 10 Herrera JC, Work DB, Herring R, *et al.* Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 2010, 18(4): 568-583. [doi: [10.1016/j.trc.2009.10.006](https://doi.org/10.1016/j.trc.2009.10.006)]
- 11 Greaves SP, Ellison AB. Personality, risk aversion and speeding: An empirical investigation. *Accident Analysis & Prevention*, 2011, 43(5): 1828-1836.
- 12 彭江琴, 刘南杰, 赵海涛, 等. 智能 UBI 系统研究. 计算机技术与发展, 2016, 26(1): 142-146.
- 13 Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- 14 张慧哲, 王坚. 基于初始聚类中心选取的改进 FCM 聚类算法. 计算机科学, 2009, 36(6): 206-209.
- 15 熊桢, 郑兰芬, 童庆禧. 分层神经网络分类算法. 测绘学报, 2000, 29(3): 229-234.