

一种基于基准点映射的驾驶行为分析

郭巧丽¹, 曹宏举²

(1.大连东软信息学院, 辽宁 大连 116023;

2.大连外国语学院, 辽宁 大连 116044)

摘要:在汽车驾驶行为分析中,一般通过车载传感器收集数据,并对数据进行处理,但是车载传感器收集的数据具有很多弊端,比如具有数据量大、无法分辨方向等。本文提出一种基于基准点映射的驾驶行为分析算法:根据原始行驶记录,使用一定规则和Geohash算法建立带有方向的基准点,并将行驶记录投影到对应的基准点中,采用k-Means聚类算法对基准点数据进行分类,减少有效数据点个数,为后续危险驾驶行为分析提供基础。

关键词:基准点; 驾驶行为; Geohash算法; k-Means聚类

中图分类号: TP399 **文献标识码:** A

An Analysis of Driving Behavior Based on Reference Point Mapping

GUO Qiaoli¹, CAO Hongju²

(1. Dalian Neusoft University of Information, Dalian 116023, China;

2. Dalian University of Foreign Languages, Dalian 116044, China)

Abstract: In the car driving behavior analysis, data collection and data processing are generally performed by means of vehicle sensors. However, there are many drawbacks for the collected data by vehicle sensors, such as a large amount of data and indistinguishable directions. This paper proposes a driving behavior analysis algorithm based on reference point mapping. According to the original driving record, a certain rule is used to establish the reference point with direction. The driving record is projected into the corresponding reference point classified by k-Means clustering to reduce the numbers of effective points, which is prepared for the subsequent cluster analysis and the dangerous driving behavior analysis.

Keywords: reference point; driving behavior; Geohash; k-Means clustering

1 引言(Introduction)

驾驶人不仅是道路交通系统的信息处理器、决策者,也是调节者和控制者,其驾驶行为直接影响道路交通安全水平。汽车驾驶行为被认为是导致交通事故的重要原因之一,也是交通安全领域重要的研究热点。作为道路交通系统的信息处理器、决策者、调节者和控制者,驾驶人的驾驶行为直接影响道路交通安全水平。据调查,80%—90%的交通事故是人为的因素造成的^[1,2]。汽车驾驶行为是导致交通事故的重要原因之一,也一直是交通安全领域的研究热点^[3]。

针对驾驶行为与交通事故的关联性,国内外均展开了大量研究。武汉理工大学针对大客车的事故行为特点,有针对性地开展了长途客车不良驾驶行为识别的研究^[4]。英国皇家

事故预防协会(The Royal Society for the Prevention of Accidents, RoSPA)的研究表明:20%的重大交通事故是由疲劳驾驶行为导致的^[5];美国公路运输安全管理局(National Highway Traffic Safety Administration, NHTSA)对2009年美国发生交通事故的统计数据进行分析表明,酒后驾驶行为导致的死亡人数相对于交通事故总死亡人数的占比为32%^[6];澳大利亚MONASH大学事故研究中心的研究报告指出,有25%的道路交通事故由驾驶人分神原因导致^[7]。因此,对驾驶人行为的研究,已成为各国道路交通安全研究的核心内容^[8]。

利用车载传感器可以采集车辆的行驶信息(如经度、纬度、时速、加速度、轮转舵角等信息),进而利用这些行驶信息对驾驶员的驾驶行为以及行驶场所进行分析。这些信息一

一般是1s采集一次,或者100ms采集一次。但因为数据量较大以及GPS的不精确性,如果直接将采集的数据标注到地图上或者直接使用,往往使后续的分析无法进行,例如在进行危险驾驶分析时,需要将行驶数据与标准驾驶行为进行对比,这就需要将每一条数据与周边的标准驾驶数据进行对比,计算量非常大(即使每秒采集一次)。如图1所示,GPS不断地采集数据会逐渐把整条路给填满,并且对于后续任何分析的时候,只要涉及比对计算的,都会需要计算距离比较近的点,这个计算量是非常大的。另一方面,由于行驶方向的不同,驾驶行为也是不同的,而直接使用数据是无法分辨行驶方向的。此时,他们的行驶方向是不同,如果直接使用原数据进行分析,是无法获得他们的行驶方向的。



图1 在地图上直接标注数据点

Fig.1 Marking data points on maps

针对上述现有研究的不足,本文提出一种基于基准点映射的驾驶行为分析方法,这个算法是通过每天的行驶数据建立出行驶的基准点,然后把行驶数据与标准驾驶的行驶数据映射到对应的基准点上,一个基准点相当于一个集合。

2 基准点的作用(The role of the reference point)

基准点是驾驶行为分析的一个基础,能解决一系列问题,具体如下:

(1)驾驶行为分析的时候,一般的场景是用几辆车在一定区域上行驶,必定会产生同一路段会有多辆车行驶的问题,车载GPS一般1s采集一条数据,这样如果根据经纬度进行在地图上显示时,会把道路填满,在后续的计算时计算量非常大。

(2)在行驶记录显示时,有的车载GPS是不带有方向的,这样就导致在一条马路上看不出车辆行驶方向的。

(3)对危险驾驶进行分析时,往往会采用聚类的算法,聚类的时候,如果不考虑方向的时候会把距离很近但是方向不同的数据归为一类。

(4)一个基准点是一个集合,如果标准行驶的路程覆盖整个行驶区域的话,理论上是在每个基准点中都存在对应的行驶记录。这样在一个基准点中就能看到,普通驾驶员与标准驾驶的行驶数据。将基准点里的数据进行聚类,把每个人的驾驶数据都归到各自的类里,在各自的类里再进行聚类,找到最有效地点,然后与标准驾驶数据进行对比。

3 Geohash算法(Geohash algorithm)

Geohash是一种地理编码,由Gustavo Niemeyer提出,

最初用于geohash.org服务中,目的是为地球上每一个位置提供一条短URL作为唯一标识,之后基于Geo-hash的编码被应用到空间检索中,广泛应用于附近兴趣点(Point of Interest, PoI)的搜索^[9]。它可以把二维的经纬度坐标编码成一维的字符串,作为全球唯一的标识,可以在数据库中表达点数据。通过Geohash编码,空间上相邻的位置在编码上可能具有相同的前缀,使之在解决附近地点搜索的问题上具有明显优势。Geohash已被广泛应用于需要使用一维索引处理空间数据的情况^[10-12]。

Geohash的思想^[13]是把地球沿经度和纬度方向不断地交替进行二分,每次二分称为一次划分,把每五次划分作为一个层次。当层数为奇数时,每次划分的结果与第一层类似;当层数为偶数时,每次划分的结果与第二层类似。一直循环划分到任意精度。用0或1表示每次划分产生的区域,鉴于相邻地点的Geohash编码^[13]通常具有相同前缀的特点,常将Geohash用于附近地点查询,即邻近点查询。基于Geohash的邻近点查询思路是:根据用户查询的点的空间位置和空间查询范围的大小,得到所在的Geohash层级,计算查询点在该层级的Geohash编码,利用该编码在数据库中查询与之具有相同前缀的点数据。

4 基准点创建(The creation of the reference point)

基准点创建的具体算法如下:

(1)建立基准点阶段

①选择计算的开始点,当前地点P1作为初始点。剩余距离D0=0。

②取得当前地点P1的1s后的地点P2,计算P1与P2之间的距离,以及P1与P2之间的夹角 θ 。夹角的定义为:P1与P2连线与正北方向的夹角。如图2所示。

③更新D1,D1=D1+D0。

④基准点建立:

(a)当D1≤5米时,使剩余距离D0=D1。

(b)当D1>5米时:

i.采用Geohash算法计算5米之内存在基准点,如果基准点的方位角与计算完毕的 θ 之间相差在±10度以内时,不生成基准点,并且令D0=D1-5。

ii.否则,从P1开始(P1P2方向)5米的地方建立基准点S,另D0=D1-5,并且S的方位角设为P1到P2的夹角。

⑤将P1的经纬度设为地点P2的经纬度。

⑥返回②。

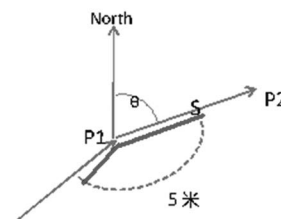


图2 半径为5米进行搜索基准点

Fig.2 Search datum point with radius of 5 meters

(2) 行驶数据映射

基准点做成之后, 选取行驶数据中任意一点P, 当行驶数据P满足下述条件时, 则将行驶数据映射到对应节点。

①地点P的速度在20以上(这个值根据经验获得，取决于具体分析)；

②半径5米以内，存在基准点S：

③基准点的行进方向角与P点的行进方向的差在 $\pm 10^\circ$ 以内。

5 聚类分析(Cluster analysis)

采用k-Means聚类算法对表中记录数据进行分类。

k-Means聚类算法的核心思想是从 n 个数据对象任意选择 k 个对象作为初始聚类中心；而对于所剩下其他对象，则根据它们与这些聚类中心的相似度(距离)，分别将它们分配给与其最相似的(聚类中心所代表的)聚类；然后再计算每个所获新聚类的聚类中心(该聚类中所有对象的均值)；不断重复这一过程直到每个聚类不再发生变化为止。 k 个聚类具有以下特点：各聚类本身尽可能地紧凑，而各聚类之间尽可能地分开。具体定义如下：

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

其中, E 为数据库中所有对象的均方差之和; p 代表对象空间中的一个点; m_c 为聚类 C 的均值 (p 和 m_c 均是多维的)。

6 结果(Result)

每个基准点存储了在该范围内不同车辆的行驶信息，包括行驶时间、经度、纬度、时速、加速度、方向等。

根据地图进行人工标注相关周边信息，例如是否在学校周围、商业街、十字路口等。

直接对采集到的数据进行聚类，聚类的时候因为有了方向这个属性，所以方向不同的数据不会聚为一类。

采用基准点方法后,在地图上有效地数据点减少,并且增加了方向,使在同一范围内的数据可以通过方向区分,图1使用该方法后的效果如图3所示,蓝色点和红色点为不同的驾驶方向。



图3 采用基准点方法后的效果

Fig.3 Effect of datum point method

7 结论(Conclusion)

本文提出一种基于基准点映射的驾驶行为分析算法,算法中使用了Geohash算法和k-Means聚类算法,并根据经纬度计算出驾驶的方向。在驾驶行为分析中,引入该算法,将为危险驾驶、驾驶行为对比,以及地图行驶轨迹展示,提供了分析基础。

参考文献(References)

- [1] Evans L.Traffic Safety Science Serving Society[C]. Mich,2004:461—466.
- [2] NEC HOME.Radar Apparatus of Frequency Modulated Carrier Wave for Vehicle e.g.for Collision Prevention[P].日本:JP9043343—A,1997.
- [3] 严新平,张晖,吴超仲,等.道路交通驾驶行为研究进展及其展望[J].交通信息与安全,2013,31(01):45—51.
- [4] 贺宜,吴超仲,陈志军,等.长途客车事故现状及成因分析[J].交通信息与安全,2011,29(05):78—82;95.
- [5] Royal society for the prevention of accidents(ROSPA).Driver fatigue and road accident Review and position paper[R]. Washington,DC America Royal society for the prevention of accidents,2001.
- [6] Department of transportation,national highway traffic Safety administration (NHTSA) [R].Alcohol impaired Driving traffic safety fact sheet 2009,Washington,D C,America:NHTSA,2010.
- [7] Young K,Renan M,Hammer M.Driver distraction:A review of the literature:Report 206[R].Monash University Accident Research Center,Victoria,Australia,2003.
- [8] 裴玉龙,张需鹏.不良汽车驾驶行为特征分析[J].交通信息与安全,2009,27(03):81—84.
- [9] 叶小榕,邵晴.基于增强现实和位置服务的手机广告系统[J].科技导报,2013,31(4):67—73.
- [10] CHODOROW K,DIROLF M.MongoDB权威指南[M].北京:人民邮电出版社,2011:76—78.
- [11] DIMIDUK N,KHURANA A.HBase in Action[M].New York:Manning Publications Co,2012:203—235.
- [12] The Apache Software Foundation.Apache Lucene 8.1.1Documentation[EB/OL].http://lucene.apache.org/core/8_1_1/index.html.
- [13] 金安,程承旗,宋树华,等.基于Geohash的面数据区域查询[J].地理与地理信息科学.2013,29(05):31—35.

作者简介:

郭巧丽(1980-),女,硕士,副教授.研究领域:图像处理,多媒体技术.

曹宏举(1980-), 男, 硕士, 副教授. 研究领域: 图像处理, 应用数学.