Technology & Economy in Areas of Communications

2020年5月

May, 2020

引用著录:丁鹏程,杨明,郑长江,等.基于手机信令数据的城市通勤出行特征研究[J]. 交通科技与经济,2020,22(3):29-34. DOI:10,19348/j. cnki. issn1008-5696. 2020,03,006

基于手机信令数据的城市通勤出行特征研究

丁鹏程1,杨 明1,郑长江2,朱 健

(1. 南京市城市与交通规划设计研究院股份有限公司,江苏 南京 210000;2. 河海大学,江苏 南京 210098)

摘 要:通勤交通出行特征是研究城市交通规划的重要数据基础,如何有效提取通勤出行特征成为研究城市居民通勤出行的关键。文章通过分析手机信令数据特性和通勤出行特征,利用手机信令数据挖掘的相关技术,建立通勤出行特征提取模型并进行实例研究,与传统的居民出行调查分析进行对比,验证利用手机信令数据进行通勤出行特征研究的可行性。

关键词:手机信令数据;通勤;出行特征;数据挖掘

中图分类号:U121

文献标识码:A

文章编号:1008-5696(2020)03-0029-06

Research on urban commuting travel characteristics based on mobile signaling data

DING Pengcheng¹, YANG Ming¹, ZHENG Changjiang², ZHU Jian¹

(1. Nanjing Institude of City & Transport Planning Co., Ltd., Nanjing 210000, China; 2. Hohai University, Jiangsu Nanjing 210098, China)

Abstract: The characteristics of commuting travel are the important data base to study the urban traffic planning. How to extract the characteristics of commuting travel effectively becomes the key to study the commuting travel of urban residents. By analyzing the characteristics of mobile signaling data and commuting trip, this paper establishes a commuting trip feature extraction model based on the technology of mobile signaling data mining and carries out case study. It compares with the traditional residents trip investigation and analysis, and verifies the feasibility of using mobile signaling data to study commuting trip characteristics.

Key words; mobile signaling data; commuting; travel characteristic; data mining

随着大数据时代的发展,手机信令数据为通勤交通出行数据提供了新的采集手段,从而为提取通勤出行特征以及识别城市居民职住分布特征提供了有利条件。如何有效地从手机信令数据中提取城市居民通勤出行特征以及职住地分布特征成为研究城市通勤的关键。目前,利用手机信令数据对城市通勤和职住分布进行研究成为了国内外许多学者关注和研究的热点[1-6],国外有 Gur Y^[7],以以

收稿日期:2019-11-30

第一作者简介:丁鹏程(1995一),男,助理工程师,硕士,研究方向:交 通运输规划与管理

通信作者简介:杨 明(1977一),男,研究员级高级规划师,博士,研究方向:交通运输规划与管理.

色列为例,建立了跨城市范围的手机信令数据交通规划模型,并由此建立以色列城市居民交通 OD 出行计划; V Angelakis [8] 通过对手机信令数据的处理与分析,预测了国家和城市聚集层面的出行轨迹,并用可视化视图,对整体出行轨迹进行表达与展现。最后,以构造 OD 矩阵的方式,来划分交通拥挤路段。国内毕晓萤^[9]从出行次数、出行目的分布、出行方式构成 3 个方面入手,将中小城市出行特征与大城市进行比较分析,同时对中小城市不同时期的出行特征进行对比分析; 刘杰 [10] 等根据手机运营商提供的手机信令数据,建立交通模型得到各个交通小区之间的人员出行 OD 矩阵及其特性; 毛晓

汶[11]通过处理原始手机信令数据,并从中提取手机用户出行特征,用案例的方式分析了区域的交通出行特征;马亮[12]利用轨道交通刷卡数据,提出了绝对通勤圈和相对通勤圈的概念,对轨道交通刷卡数据进行了数据格式转换、数据清洗、数据匹配、空间映射、数据提取;本文通过参考文献[13-15]提出的基于 DBSCAN 的空间停留点提取算法对手机信令数据进行停留点的提取,然后提出一种职住停留点划分方法对停留点进行职住划分,并构建一种通勤出行特征提取模型,最后以案例应用的方式验证此方法与模型的可行性。

1 基于手机信令数据挖掘的关键技术

1.1 手机信令数据的构成

用手机收发短信、主被叫、上网、位置更新、小区切换等,会留下相应的位置信令数据。每一条手机信令数据包括4个字段,分别是:①MSID(匿名加密唯一用户标识):加密之后每位手机用户的唯一识别号;②TimeStamp(时间戳,精确到秒):时间发生时刻;③CellID(基站小区编号):事件发生时刻所处基站小区编号;④EVENT(事件类型):主被叫、发短信、接短信、开关机、小区切换、正常位置更新、周期位置更新等。

1.2 手机信令数据的清洗

手机信令数据具有如下三个特征:①存在大量脏数据②数据量庞大③信令数据时间间隔不均匀。针对以上特性,首先进行相应的数据清洗,包括无效数据的清洗、漂移数据的清洗和"乒乓效应"数据的清洗。

1.3 基于 DBSCAN 的停留点提取算法

本文采用文献[13]提出的基于 DBSCAN 的空间停留点提取算法对手机信令数据进行停留点的提取。此算法在传统的 DBSCAN 空间聚类算法的基础上进行改进,兼顾到空间和时间两个维度。其主要思路是遍历数据源并找出其中的核心对象,然后以核心对象为中心,找出所有与核心对象密度相连的对象,将这些对象圈成新簇,接着找出新簇里所有核心对象的密度相连的对象,并将它们加入到新簇,这样就可以保证簇的边缘都不是核心对象。随后寻找没有被归到簇里的对象,重复以上的步骤,直至遍历整个数据源。遍历整个数据源后,

不被任何簇包含的对象为噪声对象。

1.4 职住停留点划分

本文设计一种职住停留点划分方法,如图 1 所示。考虑到大多数通勤者在早 6:00 是在家中,10:00在单位,所以取一个时间段(6:00—10:00),用此时间段的首尾两个停留点簇分别作为居住停留点簇和就业停留点簇。在得到居民的职住停留点簇后,可以提取出通勤出行距离,并且利用 Arc-GIS 做出居民职住分布图。

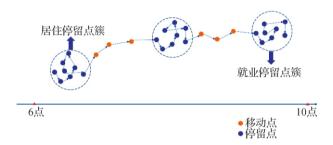


图 1 职住停留点划分示意图

2 基于手机信令数据的通勤出行特征提取 模型

2.1 数据再处理

通过上一节的停留点提取算法,借助 SQLServer2012 数据库处理完手机信令数据后,得到一系列停留点簇以及停留点信息,将每一条手机信令数据添加 Stay 字段,当 Stay=1 时,表示该轨迹点属于停留状态;当 Stay=0 时,表示该轨迹点处于出行状态,然后添加 t_1 和 t_2 , t_1 表示停留点簇中的第一个停留点的发生时间,记为停留事件的起始时间, t_2 表示停留点簇的最后一个停留点的发生时间,记为停留事件的结束时间。在此之后给每一个停留点系加Lon和Lat字段,分别表示该停留点所在簇的经度和纬度,记为停留点簇代表经纬度点,以此点的经纬度代表所在停留点簇中的所有停留点的经纬度,如图 2 所示。新数据表字段如表 1 所示,停留点簇代表经纬度点经度Lon与纬度Lat的计算公式如下

$$\overline{Lon} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{1}$$

$$\overline{Lat} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$
 (2)

表 1 新数据表字段

MSID	TimeStamp	Lon	Lat	Stay	Lon	\overline{Lat}	t_1	t_2	CellID	Event
------	-----------	-----	-----	------	-----	------------------	-------	-------	--------	-------

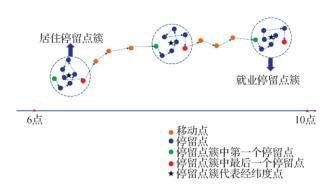


图 2 数据再处理示意图

经过一系列数据处理,得到用户出行识别数据 如表 2 所示。

2.2 通勤出行特征提取

经过数据再处理,得到手机出行用户一天的停留点经纬度(\overline{Lon} , \overline{Lat}),停留开始的时间 t_1 和结束的时间 t_2 ,以及停留点状态 Stay。本小节将在此基础上参考文献[13-15]进行通勤出行特征的提取。

1)出行次数。在上面得到用户出行识别表中,当 Stay 等于 0 时,说明手机用户处于移动之中;当 Stay 等于 1 时,说明手机用户处于停留状态。当 Stay 值从 1 变成 0,又从 0 变成 1 时,表明用户完成一次出行。因此可以利用 Stay 值变化的次数来表征用户一天之内的出行次数,计算式如式(3)所示:

$$F_{j} = \frac{1}{2} \sum_{i=1}^{n} |\operatorname{Stay}_{i+1} - \operatorname{Stay}_{i}|,$$
 (3)

表 2 用户出行识别表

	MSID	TimeStamp	Stay	\overline{Lon}	\overline{Lat}	t_1	t_2
1	635	2:25:53	1	119. 377 6	32, 366 99	2:25:53	Null
2	635	3:32:42	1	119. 377 6	32, 366 99	Null	Null
3	635	4:12:26	1	119. 377 6	32, 366 99	Null	Null
4	635	5:26:06	1	119. 377 6	32, 366 99	Null	Null
5	635	6:12:44	1	119. 377 6	32, 366 99	Null	Null
6	635	7:28:16	1	119. 377 6	32, 366 99	Null	7:28:16
7	635	8:56:14	0	Null	Null	Null	Null

式中: F_j 为手机出行用户j 一天之内的出行次数; Stay $_i$ 为手机出行用户 $_j$ 第 $_i$ 条信令数据的 Stay 值。

进而,可以用式(4)计算一天之内所有人的平均出行次数:

$$F = \frac{1}{2M} \sum_{i=1}^{M} \sum_{i=1}^{n} (Stay_{i+1} - Stay_i), \qquad (4)$$

式中:F 为 1 d 之内所有人的平均出行次数;n 为每一个手机用户的所有手机信令数据总条数;M 为手机出行用户总数。

2)通勤出行距离。本文用停留点间的直线距离作为出行距离,对于用户,第j个停留点与第j+1个停留点之间的距离可用式(5)计算:

$$D_{ij} = R * \frac{\pi}{180} *$$

 $\sqrt{(\overline{Lon_j} - \overline{Lon_{j+1}})^2 + (\overline{Lat_j} - \overline{Lat_{j+1}})^2}$, (5) 式中·R 为赤道半径,取 6 138 km。

因此,手机用户i一天内的出行距离就可以用式(6)计算:

$$D = \sum_{j=1}^{n-1} R * \frac{\pi}{180} * \sqrt{(\overline{Lon_{j}} - \overline{Lon_{j+1}})^{2} + (\overline{Lat_{j}} - \overline{Lat_{j+1}})^{2}}, \quad (6)$$

式中: n 为手机用户 i 全天内停留点的数量。

可以利用上面的职住停留点的划分方法,提取出居住停留点和就业停留点,利用式(5)计算居住停留点簇代表点和就业停留点簇代表点间的直线距离,即表征为通勤出行距离。

进而可以计算所有手机用户的平均通勤出行距离:

$$D = \frac{1}{m} \sum_{i=1}^{m} R * \frac{\pi}{180} *$$

 $\sqrt{(\overline{Lon_j} - \overline{Lon_{j+1}})^2 + (\overline{Lat_j} - \overline{Lat_{j+1}})^2}$, (7) 式中:m 为手机出行用户的数量;R 为赤道半径,取 6 138 km:

3)通勤出行时耗。居民出行时间可以通过出行过程中起止停留点之间的时间差来表征,即用出行开始时之前的停留点簇中最后一个停留点时间记为出行开始时间 t_f ,用出行结束之后的停留点簇中第一个停留点记为出行的结束时间 t_f ,则手机用户 i 通勤出行时耗 t_c 即可用上午就业停留点簇中第一个停留点时间 t_f 与居住停留点簇中最后一个停留点时间 t_f 之差来表征,计算式如式(8)所示。

$$t_c = t_l - t_f, \tag{8}$$

进而,居民平均通勤出行时耗以式(9)计算:

$$\frac{-}{t_c} = \frac{1}{m} (t_l - t_f). \tag{9}$$

其中, m 为手机出行用户总数。

4)通勤出行速度。经过上面分析,得到通勤出行距离和通勤出行时耗的计算式,则通勤出行速度的计算式如式(10)所示。

$$S = \frac{R * \frac{\pi}{180} * \sqrt{(\overline{Lon_f} - \overline{Lon_l})^2 + (\overline{Lat_f} - \overline{Lat_l})^2}}{t_l - t_f}.$$

其中,S 为用户的出行速度; t_f 、 Lon_f 、 Lat_f 为出行前的最后一个停留点时间、经度和纬度; t_i 、 Lon_i 、 Lat_i 为出行后的第一个停留点时间、经度和纬度;R 为赤道半径,取 6 138 km。

进而可计算得到所有用户的平均出行速度,计 算式如式(11)所示。

$$S = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \bullet$$

$$\frac{R * \frac{\pi}{180} * \sqrt{(\overline{Lon_f} - \overline{Lon_l})^2 + (\overline{Lat_f} - \overline{Lat_l})^2}}{t_l - t_f}.$$
(11)

其中,n 为用户的出行次数总和,m 为手机出行用户总数,S 为所有手机出行用户的平均出行速度。

3 实例分析

本文利用扬州移动某一工作日内的手机信令数据,通过上文的算法进行清洗与挖掘,过滤了无效数据 10.523 条,占 12%;过滤乒乓数据 3.142 条,约占全部数据的 3.6%;过滤漂移数据 1.136 条,约占 1.3%。经过信令数据的清洗后,剩余 72.891 条有效信令数据,约占全部数据的 83%。经过数据处理后的数据表如表 3 所示,表中的字段含义如表 4 所示。

表 3 数据处理结果示例

MS ID	Time Stamp	Lon	Lat	delta T	delta D	delta V	stay	Ts	Time Point	center Lon	center Lat
562 112	11:26:32	119. 460 1	32, 414 1	3 156	578. 326 5	0. 562166	0	295			
562 112	12:47:22	119. 458 2	32, 413 1	3 536	358. 309 5	0. 108 572	0	300			
562 112	13:42:52	119. 466 3	32, 420 5	213	0	0	1	5 850	13:42:52	119. 166 45	32, 420 63
562 112	15:20:46	119. 466 5	32, 420 8	233	463. 742 6	1. 9823	1	5 850	15:20:46	119. 166 45	32, 420 63
562 112	15:23:13	119. 491 2	32, 444 3	80	463. 742 6	5. 796 562	0	0			

表 4 数据处理结果字段含义表

MSID	TimeStamp	Lon	Lat	deltaT	deltaD	
匿名加密唯一用户标识	时间戳	经度	纬度	相邻点时间差	相邻点距离差	
deltaV	stay	Ts	timePoint	centerLon	centerLat	
 相邻点速度差	停留点状态	停留时间	停留开始与结束时间	停留点簇的中心经度	停留点簇的中心纬度	

下面利用通勤出行特征提取模型提取出行次数、通勤出行距离、通勤出行时耗和通勤出行速度,然后与居民出行调查数据进行对比分析,验证此模型的可行性。

1)出行次数。根据国内外研究对"手机出行"的定义,将停留距离的阈值设置为 500 m,将停留时间的阈值设置为 10 min,根据出行次数计算式(3)可以算出该工作日扬州市居民平均出行次数:

$$F = \frac{1}{2M} \sum_{j=1}^{M} \sum_{i=1}^{n} (Stay_{i+1} - Stay_{i}) = 2.61$$
次 /(人・日)

扬州居民出行次数统计表如表 5 所示,所得结果可以看出,利用本文的出行特征提取模型对手机

信令数据进行提取,可以得出居民出行次数为 2.61 次/(人•日),与居民出行调查分析所得的 2.72 次相比较为接近,误差在 4%左右,说明本文的出行特征提取模型在出行次数的提取上具有可行性。

表 5 扬州居民出行次数统计表

年度	2014	2015	2016
出行次数(次/人/日)	2. 72	2. 71	2. 72
出行总量(万人次/日)	422. 4	421. 7	422. 6

2)通勤出行距离。根据平均通勤出行距离计算式(7)可以计算出扬州居民的平均通勤出行距离:

$$D = \frac{1}{m} \sum_{i=1}^{m} R * \frac{\pi}{180} *$$

$$\sqrt{(\overline{Lon_i} - \overline{Lon_{i+1}})^2 + (\overline{Lat_i} - \overline{Lat_{i+1}})^2} = 3.11 \text{ km}.$$

扬州居民通勤出行距离分布图如图 3 所示,根据居民出行调查,居民平均通勤出行距离为 3.46 km,大部分集中在 5 km 以内。

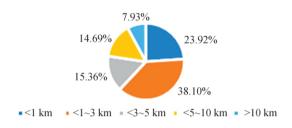


图 3 扬州居民通勤出行距离分布图

提取所得结果与居民出行调查分析所得的 3.46 km 相比,误差在 10% 左右,误差原因主要为提取手机信令数据得出的出行距离是直线距离,与实际的路径距离相比有所减小,但在可接受范围内,大体能表征居民的出行距离,说明本文的通勤出行特征提取模型在通勤出行距离的提取上具有可行性。

利用式(5)对扬州居民通勤出行距离进行计算,可以得到居民通勤出行距离分布图,如图 4 所示。

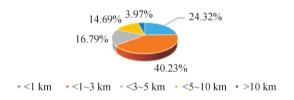


图 4 由手机信令数据提取的居民通勤出行距离分布

3)通勤出行时耗。根据通勤出行时耗计算式 (8)和式(9)可以分别计算出每个手机用户的通勤 出行时耗和所有用户的平均通勤出行时耗。平均通勤出行时耗为:

$$\overline{t_c} = \frac{1}{m}(t_l - t_f) = 1 \ 209 \ \text{s} = 20.15 \ \text{min}$$

扬州居民通勤出行时耗折线图如图 5 所示,根据居民出行调查,全目的方式的平均出行时耗为 22,06 min,主要集中在 15 min 及 30 min 左右。

提取所得结果 20. 15 min 与居民出行调查数据中的 22. 06 min 相比,误差在 8. 6%左右,在可接受范围内。提取居民通勤出行时耗后,可以得到居民通勤出行时耗分布图,如图 6 所示,可以看出居民通勤出行时耗主要分布在 $5\sim20$ min 和 $25\sim35$ min 内,与居民出行调查分析大体一致。说明本文的通

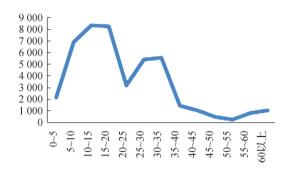


图 5 扬州居民通勤出行时耗折线图

勤出行特征提取模型在通勤出行时耗的提取上具 有可行性。

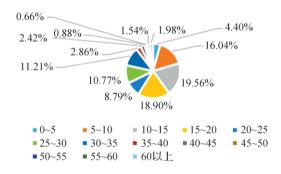


图 6 由手机信令数据提取的居民通勤出行时耗分布

4)通勤出行速度。按照平均出行速度计算式(11),可以计算得出扬州市居民出行的平均速度:

$$\overline{S} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \bullet$$

$$\frac{R * \frac{\pi}{180} * \sqrt{(\overline{Lon_f} - \overline{Lon_l})^2 + (\overline{Lat_f} - \overline{Lat_l})^2}}{t_l - t_f} =$$
3, 89 m/s = 14, 004 km/h

利用出行速度计算式(10)计算得出的出行速度分布如图 7 所示。

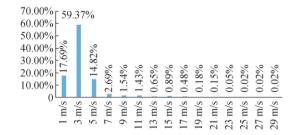


图 7 由手机信令数据提取的居民通勤出行速度分布

4 结 论

本文选取扬州市居民通勤出行特征分析作为 实例,利用基于手机信令数据的通勤出行特征提取 模型对扬州居民手机信令数据进行数据挖掘,提取 出相应的通勤出行特征,包括出行次数、通勤出行 距离、通勤出行时耗以及通勤出行速度,并与居民 出行调查分析进行对比验证,结果较为接近。说明 本文基于手机信令数据的通勤出行特征提取方法 具有可行性,为以后利用手机信令数据进行通勤出 行特征的研究提供了理论依据。

参考文献:

- [1] 唐语灏. 基于手机信令大数据获取城市居民 OD 特征的新方法[J]. 信息通信,2016(2):267-269.
- [2] 李淑庆,石路源.基于手机信令数据的区域通道出行特征研究[J].公路,2019,64(11);167-172.
- [3] 肖志权,张子民,毛曦,等.基于手机信令数据居民出行链提取算法[J].北京测绘,2019,33(10):1192-1195.
- [4] 管娜娜,田苗.基于大数据的综合交通规划创新应用实践[J].智能城市,2019,5(15):14-17.
- [5] 黄建华,孟伟强,吴飞霞.基于信令数据的人流时空分布和移动模式研究[J].计算机工程与应用,2019,55(23):53-63.
- [6] 罗名海,谭波,秦思娴,等. 大数据视角下的武汉市人口格局探析[J]. 地理空间信息,2019,17(10);1-6,10.
- [7] GUR Y, BEKHOR S, SOLOMON C, et al. Intercity Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone

- Data[J]. Transportation Research Record: Journal of the Transportation Research Board, 2009, 2121: 145-151.
- [8] ANGELAKIS V, GUNDLEGÅRD D, RYDERGREN C, et al. Mobility modeling for transport efficiency: Analysis of travel characteristics based on mobile phone data[J]. Media & Information Technology, 2013.
- [9] 毕晓萤,罗崴.中小城市居民出行特征分析及交通改善策略研究[J].交通科技与经济,2018,20(3): 28-31,17.
- [10] 刘杰,胡显标,傅丹丹,等.基于无线通信网络的人员出行信息分析系统设计与应用[J].公路交通科技,2009,26(增1):151-154.
- [11] 毛晓汶. 基于手机信令技术的区域交通出行特征研究 [D]. 重庆: 重庆交通大学, 2014.
- [12] 马亮. 基于轨道交通刷卡数据的城市通勤圈范围研究 [J]. 城市轨道交通研究,2017,20(8):80-84.
- [13] 苗壮. 基于手机信令数据的数据清洗挖掘与居民职住空间分析[D]. 成都: 西南交通大学, 2017.
- [14] 张振. 基于手机信令数据的区域通道出行特征研究 [D]. 重庆: 重庆交通大学, 2017.
- [15] 张维. 基于手机定位数据的城市居民出行特征提取方法研究[D]. 南京:东南大学,2015.

[责任编辑:王文福]

(上接第16页)

- [9] 郝媛,徐天东,干宏程,等.城市快速路交通流特性研究 [J].交通运输工程与信息学报,2006(4):21-27.
- [10] 谢劲松,范炳全,干宏程.一致拥挤快速路交通系统稳定性分析[J].交通科技与经济,2008,10(1):66-68.
- [11] 李涵. 高速公路互通立交分流区交通冲突预测模型 [D]. 南京:东南大学,2017.
- [12] 林兰平. 双向八车道高速公路合流区交通特性与交通 冲突特性研究[J]. 公路与运,2019(2):34-39.
- [13] 庄焰,吕慎. 城市道路交通流三参数关系研究[J]. 深圳 大学学报,2005(4):95-98
- [14] 石建军,刘晨强. 快速路出入口组合类型适用性研究 [J]. 交通科技与经济,2018,20(6):17-21,80.
- [15] 王振报,郭南,李金山. 辅路对城市快速路出口交通流特性的影响研究[J]. 交通运输工程与信息学报,2008 (1):31-37.
- [16] 郝志国. 高速公路换道冲突预测与安全评价研究[D]. 长春: 吉林大学, 2019.
- [17] 王立明. 快速路匝道与衔接交叉口触发式协调控制方法研究[D]. 长春:吉林大学,2009

- [18] 王胤皓. 城市快速路出口交通冲突影响及建模研究 [D]. 西安: 长安大学, 2018.
- [19] 吴玲涛. 交通冲突技术在城市快速路出口安全性评价中的应用研究[D]. 北京:交通部公路科学研究院,2008.
- [20] 余思远,杜豫川. 城市快速路瓶颈拥堵分析[J]. 交通科 学与工程,2018,34(1):93-98.
- [21] 薄雾. 城市快速路匝道交通特性分析与优化控制研究 [D]. 成都: 西南交通大学, 2013.
- [22] 李英帅,姚红云,吕乔. 城市快速路匝道交通特性分析 [J]. 交通信息与安全,2010,28(5):21-24,38.
- [23] ZENG Jun-Wei, QIAN Yong-Sheng, YU Sen-Bin, et al. Research on critical characteristics of highway traffic flow based on three phase traffic theory [J]. Physical A: Statistical Mechanics and its Applications, 2019,530.

[责任编辑:王文福]