

网络优先数字出版时间: 2016-08-26

【DOI】10.13657/j.cnki.gxkxyxb.20160826.002

网络优先数字出版地址: <http://www.cnki.net/kcms/detail/45.1075.N.20160826.1048.004.html>

基于 NoSQL 的海量出租车 GPS 数据 Web 服务研究^{*}

Research of Massive Taxi GPS Data Web Service Based on NoSQL

李晓英^{1,2}, 何首武¹, 苏树海¹

LI Xiaoying^{1,2}, HE Shouwu¹, SU Shuhai¹

(1. 桂林理工大学南宁分校, 广西南宁 530001; 2. 武汉大学计算机学院, 湖北武汉 430072)

(1. Campus of Nanning, Guilin University of Technology, Nanning, Guangxi, 530001, China;

2. Computer School, Wuhan University, Wuhan, Hubei, 430072, China)

摘要:【目的】针对传统关系型数据库扩展性差、存取效率低等问题, 提出一种基于非关系型数据库(NoSQL)的海量出租车 GPS 数据 Web 服务方案, 满足海量空间数据高效存储管理与实时在线服务的需求。【方法】利用 MongoDB 地理空间索引实现基于位置的查询, 基于 Node.js 技术提供 Web service 服务, 采用 Express MVC 框架构建 Web 应用。【结果】与 MySQL Spatial 相比, MongoDB 在区域查询方面具有较高的查询效率, 数据量多时, 其性能尤为显著。【结论】基于 NoSQL 的出租车 GPS 数据 Web 服务方案可以支持海量位置数据的实时查询与地图显示, 具有一定的应用参考价值。

关键词: 基于位置的服务 NoSQL Node.js MongoDB Web 服务

中图分类号: TP311 **文献标识码:** A **文章编号:** 1002-7378(2016)03-0231-06

Abstract: 【Objective】In order to meet the demand of massive spatial data storage management and real-time online service, the massive taxi GPS data Web service based on NoSQL was discussed in this paper. The solution is to overcome the problem of poor scalability and low access efficiency in traditional relational database. 【Methods】In our solution, location-based query was realized by the MongoDB geospatial indexes. Web services was built on the Node.js technology. The Express MVC framework was used for building Web application. 【Results】The region query experiments on the MongoDB and MySQL Spatial Extensions was carried out for performance comparisons. The results showed that MongoDB had higher query efficiency. When there was large amount of data, its performance was particularly significant. 【Conclusion】The taxi GPS data Web service based on NoSQL can support the real-time query and display of location data on the digital map. The result has certain reference value

for the similar study.

Key words: location-based services (LBS), NoSQL, Node.js, MongoDB, Web service

收稿日期: 2016-05-25

修回日期: 2016-07-27

作者简介: 李晓英(1981—), 女, 讲师, 武汉大学计算机学院访问学者, 主要从事大数据与数据管理方面的研究, E-mail: 409057436@qq.com.

^{*}2016 年度广西高校中青年教师基础能力提升项目(KY2016YB911), 2015 年度广西职业教育教学改革立项项目“计算机应用技术(Web 方向)‘产品驱动’实践教学体系的改革研究与实践”和桂林理工大学南宁分校课程改革与建设立项建设项目(2014KGB09, 2015KGA06)资助。

0 引言

【研究意义】近年来, 随着基于位置的服务(Location-based Services, LBS)应用的不断发展, 以及云计算、大数据等技术的兴起, 数据正以前所未有的速度增长和累积^[1]。基于这些数据的 Web 服务和

应用对后台数据库提出更高的要求,而传统的关系数据库在海量数据的高并发读写、高效率存储访问以及高可扩展性方面已经显得力不从心,非关系型数据库(NoSQL)是目前应对这一问题的有效方案^[2-3]。NoSQL是一个云计算背景下的分布式、非关系型数据库系统,支持半结构化、结构化数据的高并发读写,具有良好的可伸缩性和可扩展性^[4]。MongoDB是一个面向集合的、模式自由的文档型数据库,旨在为Web应用提供可扩展的高性能数据存储解决方案。【前人研究进展】近年来,关于MongoDB的LBS研究和应用日趋深入。在学术研究方面,吴飞^[5]利用MongoDB的地理空间索引,研究基于MongoDB的LBS数据管理系统关键技术;刘卓^[6]利用MongoDB分布式部署方案,实现对存储在MongoDB集群架构上的空间数据进行读写、查询操作,经过对数据存取性能的综合分析,验证此种基于NoSQL的空间数据云存储方案是高效可行的;宋平亮^[7]基于MongoDB和Node.js研究航道数据的Web服务和应用,实现在Web地图上进行指定区域查询、坐标点临近查询、复合地理空间查询和航迹查询等一系列船舶监控功能。在基于MongoDB的LBS应用方面,快的打车在其出租车打车核心业务中使用MongoDB;奇虎360也在MongoDB上构建众多应用,其中包括外部面向用户的服务以及内部商业应用。【本研究切入点】MongoDB原生支持地理位置索引,可直接用于位置距离计算和查询,而且在性能上也能高度优化^[8]。据此,本研究基于MongoDB实现出租车GPS位置数据的高效存储与查询服务。【拟解决的关键问题】结合NoSQL数据库与Node.js Web技术探讨出租车数据的Web服务和应用,提出一种高响应的海量位置数据存储、管理以及Web应用的解决方案。

1 出租车GPS数据Web服务方案体系架构

近年来,位置服务(LBS)需求呈大幅度增长趋势,从地址点导航到兴趣点推荐,不仅可引导用户找到附近的产品和服务,还可获得更高的便捷性和安全性。然而,部分LBS应用仍采用关系型数据库来管理数据。从GPS数据管理角度来看,海量数据无时无刻不在产生,关系数据库难以支撑海量数据的存储和管理,成为制约LBS应用发展的瓶颈。因此,本研究基于NoSQL数据库,提出一种出租车位置数据的存储模型,利用MongoDB的地理空间索引及查询机制,结合Node.js Web服务技术,研究

这些数据的服务和应用模式,以期为海量GPS位置数据的存储、管理以及Web应用提供一种有效的解决方案。

基于上述思路,本研究构建出租车GPS数据的Web服务方案体系架构(图1),从软件层次上划分为以下3层:

(1)表示层:是面向用户的显示和交互界面,可以运行于Web浏览器或移动端。该层负责用户查询交互、从Web服务层接口获取数据并在地图可视化显示等操作。其中,HTML构建Web页面结构,CSS渲染样式,JavaScript完成数据解析以及地图API调用。

(2)Web服务层:基于Node.js技术提供HTTP服务以及REST风格的Web service服务^[9],包括处理路由、模板、数据获取等。该层以Express为开发框架,Embedded JavaScript(EJS)作为模板引擎,通过Mongoose实现MongoDB数据库操作。

(3)数据库层:负责存储出租车GPS数据,包括数据库构建、原始数据的迁移、地理索引的构建等。

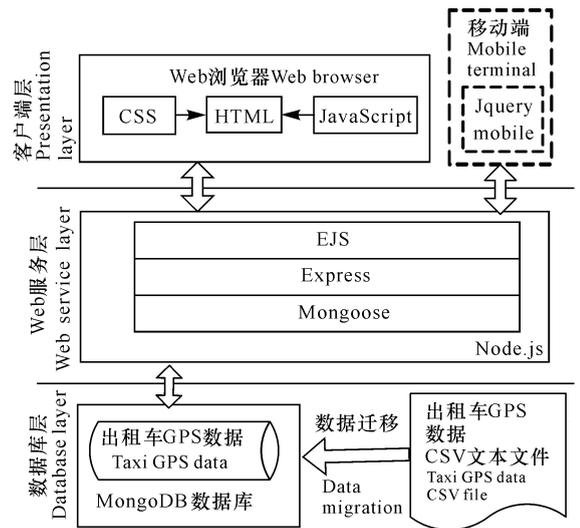


图1 体系架构

Fig.1 System architecture

1.1 关键技术

1.1.1 Node.js技术

Node.js是一个可以让JavaScript运行在服务器端的平台,用于方便地搭建响应速度快、易于扩展的Web应用。Node.js使用事件驱动、异步式I/O模型而得以轻量和高效,同时使用Module模块划分不同的功能,以简化应用开发^[10]。

Node.js可以很方便地通过相应的数据库驱动模块与MongoDB实现数据库连接。Mongoose是MongoDB Node.js数据库驱动,提供一种在Node

环境中操作 MongoDB 数据库的便捷封装。它包装更高层次、更直接的 API 语法,通过 Schema 方式将数据库中的数据转换为 JavaScript 对象以供应用使用^[11]。

1.1.2 Express 架构

Express 是一个基于 Node.js 平台的极简、灵活的 Web 应用开发框架^[12]。Express 架构是一个典型的 MVC(Model-View-Controller)框架(图 2),用户通过浏览器发起请求,由路由控制器接受,并根据不同的路径定向到不同的控制器。控制器处理用户的具体请求,可能会访问数据库中的对象,即模型部分,从模型获取数据。然后,调用模板引擎生成 HTML 页面,最后由控制器返回给浏览器,完成一次请求。

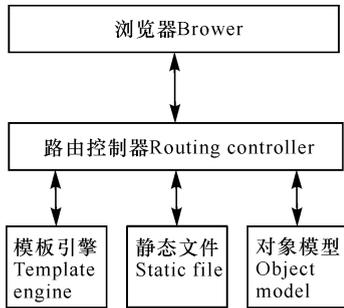


图 2 Express MVC 框架

Fig.2 Express MVC framework

在 Express MVC 架构中,模板引擎以数据和页面模板为输入,生成 HTML 页面。Embedded JavaScript (EJS)是一个 JavaScript 实现的标准模板引擎,具有易于学习、维护性良好的 HTML 代码结构^[13]。

Node.js 使用 JavaScript 作为服务器端的执行语言,可以与 MongoDB 很好地配合开发 Web 应用。因此,本研究结合 MongoDB 与 Node.js 技术开展研究,以 Express 为开发框架、EJS 作为模板引擎、Mongoose 实现数据库连接。

1.2 出租车 GPS 数据库构建

1.2.1 数据模型

出租车 GPS 数据信息包含车牌号、采集时间、采集点经度与纬度、车辆状态、行驶速度、行驶方向等 7 个要素,其详细描述如表 1 所示。

区别于关系型数据库,MongoDB 是一个模式自由的数据库,意味着存入数据之前并不需要事先定义数据结构与类型。MongoDB 的逻辑结构由文档(document)、集合(collection)、数据库三部分组成,其中文档相当于关系数据库中的一行记录,集合

对应于关系型数据库里的表。将关系数据模型映射到 MongoDB 中即为 MongoDB 文档数据模型,其 json 格式描述如下:

```
{
  "_id":ObjectedId("56ee181abb27d9f253cd078-6"),
  "name": "粤 B000H6",
  "time": "2011/04/18 00:08:01 ",
  "jd": 114.118286,
  "wd": 22.573967,
  "status": 0,
  "v": 22,
  "angle": 3
}
```

表 1 出租车 GPS 数据说明

Table 1 Taxi GPS data description

字段 Field name	描述 Field description	类型 Field type	示例 Field value
name	车牌号 Plate number	字符型 Char	粤 B000H6
time	采集时间 Collection time	日期时间型 Datetime	2011/04/18 08:01:00
jd	经度 Longitude	浮点型 Float	114.118286
wd	纬度 Latitude	浮点型 Float	22.573967
status	车辆状态 Vehicle state	整型 Integer	0
v	速度 Speed	浮点型 Float	22
angle	行驶方向 Drive direction	整型 Integer	3

1.2.2 地理索引设计

在 MongoDB 中,建立地理索引的键必须是一对值,即包含两个元素的数组或包含两个键的内嵌文档。在关系型数据模型中,出租车位置由 2 个字段描述,如果将其直接映射到 MongoDB 则会形成 2 个独立的键值对 { "jd": 114.118286, "wd": 22.573967 }, 无法直接创建地理索引,因此需要对该两个字段进行合并 ("gps": [114.118286, 22.573967]), 然后才能创建地理索引。合并后的 MongoDB 文档模型如下所示:

```
{
  "_id": <ObjectedId>,
  "name": "粤 B000H6",
  "time": "2011/04/18 00:08:01",
```

```

“jd”:114.118286,
“wd”:22.573967,
“status”:0,
“v”:22,
“angle”:3,
“gps”:[114.118286,22.573967]
}

```

1.2.3 数据查询方案

MongoDB 支持邻域查询(\$near)和范围查询(\$within)。`$near`命令将指定坐标附近的记录按照由近到远的顺序返回。`$within`区域查询则返回某区域内的所有对象,区域可以是矩形、圆形或多边形。其中,矩形区域查询通过左上点和右下点来指定矩形范围,圆形区域查询通过圆心与半径来指定圆形范围,多边形区域查询则需要指定多边形的各个顶点。

1.3 基于 Node.js 的出租车 GPS 数据 Web 服务设计与实现

出租车 GPS 数据 Web 应用采用 Express MVC 框架,分为对象模型层、控制器层与视图层 3 层,下面将逐层介绍 Web 服务的设计与实现。

1.3.1 对象模型层

Mongoose 是 MongoDB 的一个对象模型工具,封装 MongoDB 对文档的增删改查(CRUD)操作。该层主要完成出租车对象模型设计,首先根据出租车数据表创建对象 Schema,然后将其发布为 Model 模型。这里,Schema 仅是一种数据库原型或模型骨架,并不具备数据库的操作能力。Model 则是由 Schema 构造来的,除具有 Schema 定义的数据库骨架以外,还具有数据库操作行为。出租车对象模型关键 javascript 代码如下:

```

var mongoose = require('mongoose');//导入 mongoose 模块
var Schema = mongoose.Schema;
var TaxiSchema = new Schema({
  name:String,
  time:Date,
  jd:Number,
  wd:Number,
  status:Number,
  v:Number,
  angle:Number
});//定义 Schema
exports.Taxi = mongoose.model('Taxi',TaxiSch-

```

```

ema);//将 Schema 发布为 Model

```

1.3.2 控制器层

控制器层负责系统的业务逻辑实现,通过 Express 路由控制实现客户端请求 URL 的路径映射功能。根据系统功能需求与 REST 设计风格,设计路由(表 2)。

表 2 路由规划

Table 2 The plan of Express router

序号 Number	路由名称 Name	方法 Method	路由功能 Routing function
1	/	Get	首页 Index
2	/KNNQuery	Post	出租车邻近点查询 Taxi nearby query
3	/WithinBox	Post	矩形区域查询 Rectangle area query
4	/WithinCenter	Post	圆形区域查询 Circular area query
5	/Track	Post	出租车轨迹查询 Taxi trajectory data query
.....			

下面以“/WithinCenter”的路径处理为例,简述“区域查询(圆形)”的设计过程:首先从客户端获取用户查询条件,包括用户位置信息与圆形区域检索半径,然后执行 mongodb 查询并提取数据,最终返回查询结果到客户端(图 3)。

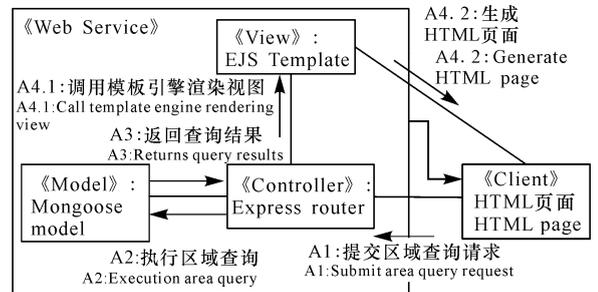


图 3 “区域查询(圆形)”通信图^[14]

Fig. 3 The communication diagram of the circular region query^[14]

1.3.3 视图层

在视图层 EJS 模板中,静态部分由 HTML5 页面实现,动态部分则嵌入 EJS 标签。视图层设计中,地图显示直接调用腾讯地图 API^[15]来实现,时间查询通过简单的时间选择对话框完成,区域设置则通过简单的文本框设置,查询结果 json 数据由 EJS 引擎渲染到页面并在地图标记,其页面设计示例——“区域查询(圆形)”如图 4 所示。



图 4 页面设计示例

Fig. 4 Example of a page view

2 仿真实验及结果分析

2.1 实验环境与数据

本研究搭建 Node.js Web 开发环境,数据库层分别选用 MongoDB 与 MySQL,对 Web 服务在区域查询方面进行性能对比,实验环境如表 3 所示。其中,MySQL 通过 MySQL Spatial Extensions(空间索引)^[16]实现区域查询。实验数据采用 2011-04-18 至 2011-04-26 的深圳市出租车 GPS 轨迹数据,车辆总数为 13 799 台,数据文件约 1.7 亿条记录,数据总量大小约 10 G。

表 3 实验环境

Table 3 Testing environment

硬件环境 Hardware environment	软件环境 Software environment
2 Core Intel Core i3 - 2 310 M 2. 10 GHz CPU; 8 G Memory; 500 GB Hard disk	64 Bit Windows 8;MongoDB 3. 2. 1; Node v5. 8. 0;Express 4. 13. 4;mon- goose 4. 4. 10;Sublime Text 2;Ten- cent map open platform JavaScript API V2;MySQL 5. 5. 38

2.2 区域查询性能分析

试验分别对 MongoDB 与 MySQL 所存储的相同数据集的不同量级(从 50 MB 到 500 MB)进行半径为 0.5 km 和 1 km 的地理范围查询,连续多次测试并取平均值。当数据量小于 150 万条,两者的 Web 响应时间差别并不明显;随着数据规模增加,MySQL Spatial 的查询耗时剧增,接近直线趋势;而 MongoDB 耗时很少,并且增幅不大(图 5)。

将查询区域半径扩大至 2.5 km 的几何区域时,实验发现,基于 MySQL Spatial 的查询响应时间已远超出可等待的范围,因此只测试 MongoDB 的查询性能。从图 6 可以看出,随着数据量增大, MongoDB Web 响应时间有所增加,但整体耗时很

少,可以支持地图实时浏览。

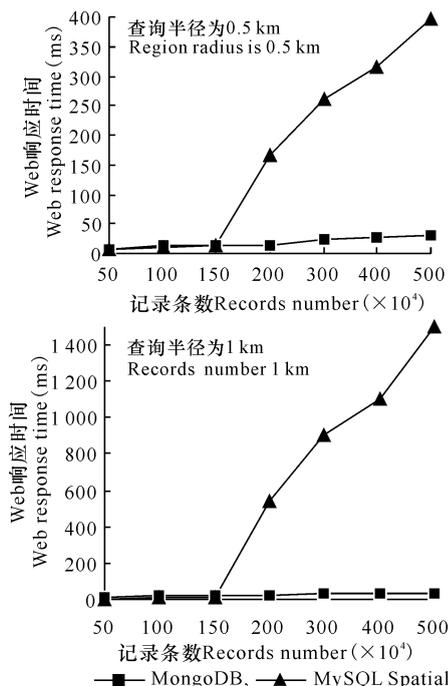


图 5 MongoDB 与 MySQL 区域查询性能对比

Fig. 5 Performance comparison for regional query based on MongoDB and MySQL

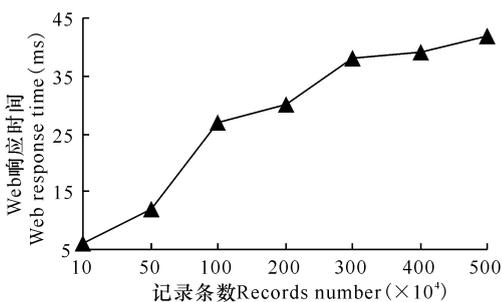


图 6 区域半径为 2.5 km 时 MongoDB 区域查询响应时间

Fig. 6 Response time of regional query based on MongoDB when region radius is 2.5 km

3 结论

位置数据的高速增长态势,给海量空间数据分布式存储、管理与网络在线服务提出了巨大的挑战。MongoDB 数据库原生支持地理位置索引并且性能高度优化,Node.js 可以方便地搭建轻量、高效和易于扩展的 Web 服务,本文结合 MongoDB 数据库与 Node.js Web 技术,探讨一种出租车 GPS 数据的 Web 服务方案。实验结果验证该方案具有较好地时效性,为更高层次的应用提供参考。

参考文献:

[1] 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. 计算

- 机研究与发展,2013,50(1):146-169.
- MENG X F, CI X. Big data management: Concepts, techniques and challenges[J]. Journal of Computer Research and Development, 2013, 50(1):146-169.
- [2] BANKER K. MongoDB 实战[M]. 丁雪丰,译.北京:人民邮电出版社,2012.
- BANKER K. MongoDB in Action[M]. DING X F, trans. Beijing: Posts & Telecom Press, 2012.
- [3] 申德荣,于戈,王习特,等.支持大数据管理的 NoSQL 系统研究综述[J].软件学报,2013,24(8):1786-1803.
- SHEN D R, YU G, WANG X T, et al. Survey on NoSQL for management of big data[J]. Journal of Software, 2013, 24(8):1786-1803.
- [4] CATTELL R. Scalable SQL and NoSQL data stores [J]. ACM SIGMOD Record, 2010, 39(4):12-27.
- [5] 吴飞.基于 MongoDB 的 LBS 数据管理系统关键技术研究[J].测绘通报,2014(7):121-124.
- WU F. Key technology research of LBS data management system based on MongoDB [J]. Bulletin of Surveying and Mapping, 2014(7):121-124.
- [6] 刘卓.基于 NoSQL 的空间数据云存储的研究[D].开封:河南大学,2014.
- LIU Z. Research of Spatial Data Cloud Storage Based on NoSQL Database[D]. Kaifeng: Henan University, 2014.
- [7] 宋平亮.基于 MongoDB 的航道数据 Web 服务研究[D].大连:大连海事大学,2015.
- SONG P L. Research on Channel Data Web Service based on MongoDB[D]. Dalian: Dalian Maritime University, 2015.
- [8] STEVIĆ M P, MILOSAVLJEVIĆ B, PERIŠIĆ B R. Enhancing the management of unstructured data in e-learning systems using MongoDB [J]. Program Electronic Library and Information Systems, 2015, 49(1): 91-114.
- [9] 黄扬子.基于 NodeJS 平台搭建 REST 风格 Web 服务 [J]. 无线互联科技, 2015(16):57-59.
- HUANG Y Z. Build a REST-style web services based on Node JS [J]. Wireless Internet Technology, 2015 (16):57-59.
- [10] 郭家宝. Node.js 开发指南[M].北京:人民邮电出版社,2012.
- GUO J B. Node.js Development Guide[M]. Beijing: Posts & Telecom Press, 2012.
- [11] Mongoose Document [EB/OL]. [2015-05-07]. <http://www.nodeclass.com/api/mongoose.html>.
- [12] STRONGLOOP. Express [EB/OL]. [2016-05-20]. <http://www.expressjs.com.cn/>.
- [13] BITOVI. EJS [EB/OL]. [2016-05-11]. <http://www.embeddedjs.com/>.
- [14] GOMAA H. 软件建模与设计:UML、用例、模式和软件体系结构[M].彭鑫,吴毅坚,赵文耘,等译.北京:机械工业出版社,2014:96-100.
- GOMAA H. Software Modeling & Design: UML, Use Cases, Patterns, & Software Architectures[M]. PENG X, WU Y J, ZHAO W Y, et al, trans. Beijing: China Machine Press, 2014:96-100.
- [15] 腾讯地图开放平台 JavaScript API V2 [EB/OL]. [2016-05-21]. http://lbs.qq.com/javascript_v2/index.html.
- Tencent Map Open Platform JavaScript API V2 [EB/OL]. [2016-05-21]. http://lbs.qq.com/javascript_v2/index.html.
- [16] 张明波,申排伟,陆锋,等.空间数据引擎关键技术与应用分析[J].地球信息科学,2004,6(4):80-84.
- ZHANG M B, SHEN P W, LU F, et al. Analysis and discussion on spatial data engine technologies [J]. Geo-information Science, 2004, 6(4):80-84.

(责任编辑:米慧芝)

(上接第 230 页 Continue from page 230)

- [3] 朱留华,陈时东,孔令江,等.过路收费站对高速公路交通的影响[J].物理学报,2007,56(10):5674-5678.
- ZHU L H, CHEN S D, KONG L J, et al. The influence of tollbooths on highway traffic[J]. Acta Physica Sinica, 2007, 56(10):5674-5678.
- [4] 田欢欢,薛郁,康三军,等.元胞自动机混合交通流模型的能耗研究[J].物理学报,2009,58(7):4506-4513.
- TIAN H H, XUE Y, KANG S J, et al. Study on the energy consumption using the cellular automaton mixed traffic model [J]. Acta Physica Sinica, 2009, 58(7): 4506-4513.
- [5] 梁玉娟.从动能损失看弯道路段的通行能力[J].四川师范大学学报:自然科学版,2011,34(3):355-359.
- LIANG Y J. Investigation of the traffic capacity in the bend by the loss of kinetic energy [J]. Journal of Sichuan Normal University: Natural Science, 2011, 34 (3):355-359.
- [6] 梁玉娟.基于坡道元胞自动机模型的能耗研究[J].广西科学,2011,18(1):44-47.
- LIANG Y J. Investigation of the energy dissipation based on cellular automatic model of slope [J]. Guangxi Sciences, 2011, 18(1):44-47.
- [7] 韦兰香,梁玉娟.收费站对高速公路交通流能耗的影响[J].广西科学,2016,23(1):92-96.
- WEI L X, LIANG Y J. Study on the influence of toll-booth on highway traffic flow's energy loss [J]. Guangxi Sciences, 2016, 23(1):92-96.

(责任编辑:尹 闯)