

# 基于GPS轨迹的规律路径挖掘算法

何雯<sup>1,2</sup>, 李德毅<sup>1,2</sup>, 安利峰<sup>1</sup>, 张天雷<sup>1</sup>, 郭沐<sup>1</sup>, 陈桂生<sup>2</sup>

(1. 清华大学 计算机科学与技术系, 北京 100084; 2. 中国电子系统工程研究所, 北京 100840)

**摘要:** 基于用户的历史轨迹数据, 对用户的规律路径进行挖掘和提取。在轨迹预处理和聚类的基础上, 定义了支撑路径的概念, 提出了一种基于支撑得分的规律轨迹挖掘算法。并通过规律停止率特征, 提高了对轨迹交通模式识别的准确率。基于178名用户4年的GPS轨迹记录, 以及37名用户的实际轨迹数据, 开展了用户试验。结果表明, 本文算法能够有效地提取用户的规律路径, 并对路径中的干扰具有一定的鲁棒性。

**关键词:** 人工智能; 轨迹挖掘; 规律路径; 交通模式; 位置服务

**中图分类号:** TP18; TP391 **文献标志码:** A **文章编号:** 1671-5497(2014)06-1764-07

**DOI:** 10.13229/j.cnki.jdxbgxb201406035

## Regular route mining algorithm based on GPS trajectories

HE Wen<sup>1,2</sup>, LI De-yi<sup>1,2</sup>, AN Li-feng<sup>1</sup>, ZHANG Tian-lei<sup>1</sup>, GUO Mu<sup>1</sup>, CHEN Gui-sheng<sup>2</sup>

(1. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; 2. Institute of Electronic System Engineering of China, Beijing 100840, China)

**Abstract:** Based on users' historical trajectory data, users' regular routes were mined and extracted. A concept of support route was defined after route pre-processing and grouping. A regular route mining algorithm was proposed based on the support score. A feature of Regular Stop Rate (RSR) was used to improve the accuracy of the transportation mode recognition. The effectiveness of the approach was validated based on the GPS data of 178 users over four years. A real user study was also performed among 37 users. The experiment results demonstrate that the algorithm can effectively extract the regular routes and is robust to slight disturbance in trajectory data.

**Key words:** artificial intelligent; trajectory mining; regular routes; transportation mode; location based service

## 0 引言

用户行为是开展用户推荐和个性化用户服务的基础。用户轨迹是用户行为的重要组成部分, 不仅包含用户的出行信息, 还包含用户的出行习

惯、生活经验等。近年来, 随着GPS智能终端的普及, 越来越多的人开始习惯于记录并分享轨迹日志。基于这些日志, 可以分析并挖掘出用户的一些行为特征, 从而为用户提供个性化的基于位置的智能服务, 有效地提升和改善用户的应用体

**收稿日期:** 2013-04-27.

**基金项目:** 国家自然科学基金重点项目(91120306, 90920305).

**作者简介:** 何雯(1982-), 女, 博士研究生. 研究方向: 智能交通, 空时数据挖掘. E-mail: he-w09@mails.tsinghua.edu.cn

**通信作者:** 李德毅(1944-), 男, 研究员. 研究方向: 不确定人工智能, 智能交通, 网络数据挖掘, 云计算.

E-mail: lidy@cae.cn

验。

规律路径是用户行为中频繁且最具预测性的行为,对理解和预测用户行为具有重要意义。规律轨迹在气象、生态等领域也有着广泛的应用<sup>[1-3]</sup>。

近年来,很多学者对用户的 GPS 轨迹进行了分析和挖掘。例如,对用户个人路径的挖掘和预测<sup>[4-5]</sup>,以及对用户行为目的的推测<sup>[6-7]</sup>。Gidofalvi 等<sup>[8]</sup>基于车辆的 GPS 数据,挖掘了可共享的最长路径。郭黎敏等<sup>[9]</sup>针对轨迹中的不确定性问题,在路网框架下提出了一种路径补全和预测的方法。Lee 等<sup>[10]</sup>提出了一种分割和聚类的方法用于发现轨迹中密度较高的公共子轨迹。Zheng 等基于用户的历史轨迹,分析并挖掘了轨迹中的潜在知识,包括交通模式<sup>[11]</sup>,感兴趣区域及旅行序列<sup>[12]</sup>,以及更高效的出行路径<sup>[13]</sup>等。

本文以用户的历史轨迹数据为基础,对用户的日常规律路径进行挖掘和提取,并识别用户在轨迹中所采用的交通模式。最后基于 178 名用户 4 年的轨迹记录,对算法进行了性能测试。并邀请 37 名用户对算法进行了实测,证明了本文算法的有效性。

## 1 数据准备

### 1.1 异常点预处理

受 GPS 精度以及信号干扰等因素的影响,原始的 GPS 数据中,通常包含一些异常点。这些异常点的存在会影响后续轨迹处理的精度和准确性。

给定一个 GPS 轨迹序列  $T\{p_1, p_2, \dots, p_k\}$ , 每个点  $p(lat, lngt, t)$  是一个 GPS 记录点,其中  $lat$ 、 $lngt$  和  $t$  分别表示记录的纬度、经度和时间。依据常识,序列  $T$  中相邻点的移动距离应小于最大行进距离。为了加快检测效率,适当放宽检测标准为:

$$0 < d(p_{i+1}, p_i) / (p_{i+1}.t - p_i.t) < V_{max} \quad (1)$$

式中:  $d(p_i, p_j)$  为点  $p_i$  和  $p_j$  间的欧氏距离,该距离小于两点间的路网距离;  $V_{max}$  为用户的最大移动速度,本文设置  $V_{max}$  为 160 km/h。

### 1.2 停留区域检测

当用户在某一区域驻留超过一段时间,则该区域可能是用户的一个停留区域。当在一个连续的 GPS 轨迹中,发现了一个停留区域,据此,将轨迹拆分为两段:通往和离开该区域的轨迹,并同时

从 GPS 轨迹中删除停留区域的数据点。

本文采用文献[8]的方法对停留区域进行检测。一个停留区域  $S\{p_m, p_{m+1}, \dots, p_n\} \subset T$  满足:

$$\begin{aligned} d(p_m, p_n) < \delta \text{ 且 } d(p_m, p_{n+1}) > \delta \\ \text{且 } p_n.t - p_m.t > \tau \end{aligned} \quad (2)$$

式中:  $\delta$  为停留区域的距离阈值;  $\tau$  为停留区域的时间阈值。

去除停留区域后,轨迹  $T$  成为  $\{p_1, \dots, p_{m-1}, p_{n+1}, \dots, p_k\}$ , 点  $p_{m-1}$  与  $p_{n+1}$  间存在一个至少为  $\tau$  的时间间隙(以下简称间隙)。

### 1.3 路径分割

当 GPS 轨迹中存在下列两种情形之一时,将其拆分为两条路径:①轨迹中存在停留区域;②相邻 GPS 点的时隙大于一定阈值。第 1 种情形已在 1.2 节中进行了介绍。而第 2 种情形,则可能是用户关闭了 GPS 设备,或 GPS 设备在一段时间内未接收到卫星信号等原因造成的。

由于从 GPS 轨迹中去除停留区域时留下了一个较大的时隙。因而,可以统一检测 GPS 点间的时隙,当大于某一时间阈值时,就进行路径分割。

定义 1 一条路径是一个 GPS 序列,记为:  $R\{p_1, p_2, \dots, p_n\}$ , 满足  $p_{i+1}.t - p_i.t < \epsilon$ , 其中,  $1 \leq i < n-1$ ,  $\epsilon$  为时间阈值。 $R.s$  和  $R.e$  为路径的开始和结束时间。

### 1.4 网格映射

网格映射是 GPS 轨迹预处理中常用的数据压缩和序列化方法。本文将 GPS 点映射入预先划分好的坐标网格中。当连续的 GPS 点映射入同一网格时,压缩记录为一个时序格。

定义 2 一个时序格定义为  $G(g_i, t_u, t_l)$ , 其中,  $g_i$  为网格的标识号;  $t_u$  和  $t_l$  分别为用户进入和离开  $g_i$  的时刻。

这样,一条路径  $R\{p_1, p_2, \dots, p_n\}$  就可以表示为一条时序格链表  $R\{G_1 \rightarrow, \dots, \rightarrow G_n\}$ 。

## 2 规律路径挖掘算法

经过数据准备,用户的轨迹日志被形式化为一系列的时序格链表。本节将从时序格链表中挖掘出用户的规律出行路径,并识别每条路径的交通模式。为了定义规律路径,首先给出有向边的定义。

定义 3 有向边是坐标网格间的有向链表,记为  $E_{m,n}$ , 表示用户从  $g_m$  移动至  $g_n$ 。所有包含

$E_{m,n}$  的路径保存至  $E_{m,n}.visit = \{R_i, \dots, R_j\}$ 。  
 $E_{m,n}.f = |E_{m,n}.visit|$ , 其中  $|\cdot|$  表示集合中的元  
 素个数。每条路径经过  $E_{m,n}$  时的速度定义为:

$$E_{m,n}(R).v = \frac{D(R.G_m, R.G_n)}{R.G_n.t_a - R.G_m.t_a} \quad (3)$$

式中:  $D(G_m, G_n)$  表示网格  $g_m$  与  $g_n$  间的路网距  
 离。

定义 4 规律路径是指某一用户经常在同一  
 近似时间所经过的一条完整路线。一条规律路径  
 由一组有向边构成, 记为  $RR\{E_{i,n}, \dots, E_{m,n}\}$ 。

### 2.1 路径聚类

用户每天会产生大量的路径, 要从一段历史  
 时期的所有路径中, 提取出规律路径, 运算量是很  
 大的。但由于规律路径发生在每天的近似时刻,  
 因而, 可以先依据时间对路径进行聚类。

当路径  $R$  与滑动时间窗  $(t - \omega, t + \omega)$  有交集  
 时, 则将该路径记入时段  $t$  的路径集, 表示为:  $\Gamma =$   
 $\{R_m, \dots, R_n\}, R_i \in \Gamma$  满足:

$$(R_i.s, R_i.e) \cap (t - \omega, t + \omega) \neq \Phi \quad (4)$$

式中:  $t$  为时间窗的中心时刻;  $\omega$  为控制窗宽和步  
 进的时间阈值。

由于用户在周末及工作日通常会有不同的生  
 活习惯, 所以本文按照时刻和路径发生的日期分  
 别对路径进行聚类。

### 2.2 规律路径的提取

用户的规律路径是指用户经常访问的完整路  
 线。用户每次经过同一路线, 均会产生一条路径记  
 录, 称为该路线的支撑路径。故支撑路径是对一条  
 路线的一次访问实例。如果一条路线要成为规律  
 路径, 必有多条支撑路径。故规律路径上的每一路  
 段都会被多条支撑路径所访问。简而言之:

If  $E_{m,n}$  为组成规律路径  $RR_j$  的一部分,  
 Then  $RR_j$  的大部分支撑路径会经过  $E_{m,n}$ ;  
 If 路径  $R_i$  为  $RR_j$  的支撑路径,  
 Then  $R_i$  的大多数有向边会被  $RR_j$  的支撑路  
 径频繁访问。

基于此思路, 本文提出了一个基于支撑得分  
 的规律路径提取算法, 基于每组  $\Gamma$  提取规律路径。

#### (1) 计算频繁有向边

为了找到规律路径, 首先要找出支撑路径。  
 而支撑路径的特征是: 它会频繁访问大部分有向  
 边。所以可以先找出被频繁访问的有向边。

定义 5 在一组路径中, 当  $E_{m,n}$  的访问频率  
 ( $E_{m,n}.f$ ) 大于给定阈值  $\alpha$ , 则被称为频繁有向边

(Frequent directed edge, FDE)。

算法 1 给出了计算频繁有向边的伪代码。

```

算法 1: CalFDE
Input: 一组路径  $t_R$ , 频率阈值  $\alpha$ 
Output: 一组 FDE
1 FDE =  $\Phi$ ,  $E_{set} = \Phi$ 
2 Foreach  $R_i$  in  $t_R$  do
3   Foreach  $E_{m,n}$  in  $R_i$  do
4     Eset. Add( $E_{m,n}$ )
5      $E_{m,n}.f++$ 
6      $E_{m,n}.visit. Add(R_i)$  // 添加访问路径
7   Foreach  $E_{m,n}$  in  $E_{set}$  do
8     If  $E_{m,n}.f \geq \alpha$ 
9       FDE. Add( $E_{m,n}$ )
10 Return FDE
  
```

#### (2) 提取支撑路径

现在可以用 FDE 来提取可能的支撑路径, 但  
 会提取出一些错误的支撑路径, 它们的大部分有  
 向边被路径频繁访问, 但不是支撑路径, 后续的步  
 骤将对此进行修正。

定义频率得分  $f_s$  来衡量路径中 FDE 的比  
 例:

$$f_s(R_i) = m/n \quad (5)$$

式中:  $n, m$  分别为  $R_i$  中有向边和 FDE 的数量。

频率得分可以反映出一条路径被访问的情  
 况。频率得分越高, 说明路径中越多的部分被用  
 户频繁访问。

如果一条路径  $R_i$  的频率得分  $f_s(R_i) \geq \beta$ , 则  
 路径  $R_i$  就可能是某规律路径的一条支撑路径。  
 阈值  $\beta$  同时具有调节算法抗干扰的能力。若设置  $\beta$   
 $= 1$ , 则要求所有的支撑路径完全相同; 若降低  $\beta$   
 的取值, 则每条支撑路径间可以存在一定的不同。

一条规律路径的所有支撑路径集记为:  $RR.$   
 $sup = \{R_i, \dots, R_j\}$ 。算法 2 给出了提取支撑路  
 径的伪代码。

```

算法 2: FindSupR
Input: 一组路径  $t_R$ , 一组 FDE, 频率得分阈值  $\beta$ 
Output: 一组支撑路径  $S_R$ 
1  $S_R = \Phi$ 
2 Foreach  $R_i$  in  $t_R$  do
3    $nf = 0, n = 0$ 
4   Foreach  $E_j$  in  $R_i$  do
5      $n++$ 
6     If  $E_j$  in FDE
7        $nf++$ 
8    $R_i.fs = nf/n$  // 频率得分
  
```

```

9   If  $R_i \cdot fs \geq \beta$ 
10      SR.Add( $R_i$ )
11   Return SR
    
```

(3)提取支撑有向边

有了支撑路径,就可以统计出被支撑路径频繁访问的支撑有向边(Support directed edge, SDE)。

定义 6 在一组路径中,一条有向边  $E_{m,n}$  被支撑路径访问的频率大于给定阈值  $\gamma$ , 则被称为支撑有向边。

SDE 与 FDE 的区别是, FDE 仅被路径频繁访问,而 SDE 被支撑路径频繁访问。定义支撑得分  $E_{m,n}.ss$  来统计经过有向边  $E$  的支撑路径数量:

$$E_{m,n}.ss = \sum_{i=1}^{E_{m,n}.f} (f_s(E_{m,n}.visit_i) > \beta) \quad (6)$$

如果有向边的支撑得分满足  $E_{m,n}.ss > \gamma$ , 则被认为是某规律路径的支撑有向边。所有经过  $E$  的支撑路径集记为:  $E_{m,n}.sup = \{R_i, \dots, R_j\}$ 。则  $E_{m,n}.ss = |E_{m,n}.sup|$ 。算法 3 给出了提取支撑有向边的伪代码。

算法 3: CalSDE

Input: 支撑路径集  $S_R$ , 一组 FDE, 支撑得分阈值  $\gamma$

Output: 一组 SDE

```

1   SDE =  $\Phi$ 
2   Foreach  $E_{m,n}$  in FDE do
3      $E_{m,n}.ss = 0$  //支撑得分
4     Foreach  $R_j$  in  $E_{m,n}.visit$  do
5       If  $R_j$  in  $S_R$ 
6          $E_{m,n}.ss ++$ 
7       If  $E_{m,n}.ss \geq \gamma$ 
8         SDE.Add( $E_{m,n}$ )
9   Return SDE
    
```

(4)更新支撑路径

现在,需要从支撑路径集中去除错误的支撑路径。这些路径经常被 FDE 访问,但不是 SDE。用 SDE 替代 FDE,重复步骤(2)和(3),直到规律路径的支撑路径集( $RR.sup$ )不再发生变化。也就是说在计算式(5)时,  $m$  表示  $R_i$  中 SDE 的数量。

最终,规律路径就是所有 SDE 的集合。算法 4 给出了提取规律路径完整算法的伪代码。

算法 4: RegularRouteMining

Input: 一组路径  $t_R$

Output: 一组 SDE

```

1   FDE = calFDE( $t_R$ )
2    $S_R = \text{FindSupR}(FDE)$ 
3   SDE = FDE //初始化 SDE
    
```

```

4   do
5     SDE = CalSDE( $S_R, SDE$ )
6      $S_R = \text{FindSupR}(S_R, SDE)$ 
7   while no change in  $S_R$ 
8   return SDE
    
```

图 1 为从一组包含有 5 条路径的  $t_R$  中提取规律路径的步骤示意图。首先设置  $\alpha = 2$ , 则所有访问超过 2 次的有向边被选为 FDE(图 1(a)中粗线表示)。计算每条路径的频率得分(见图 1(a)左上表)。设置  $\beta = 0.8$ , 则得到 4 条支撑路径,  $R_1, R_2, R_3$  和  $R_4$ 。

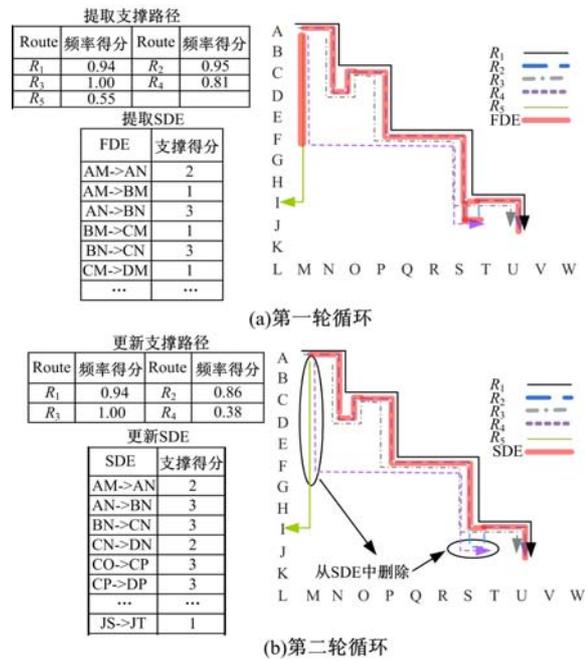


图 1 规律路径提取示意图

Fig. 1 Example of regular routes mining

基于 4 条支撑路径,分别计算每段 FDE 的支撑得分,部分结果见图 1(a)左下表。设置  $\gamma = 2$ , 则所有支撑得分大于等于 2 的 FDE 被选为 SDE。

重复上述步骤,进行第 2 轮循环,并用 SDE 代替 FDE 更新支撑路径集,结果见图 1(b)。该轮循环后,  $R_4$  和 JS->JT 分别被滤出支撑路径和 SDE。

继续执行第 3 轮循环,得到的支撑路径没有发生变化,循环结束。最终得到的所有 SDE(见图 1(b)中粗线表示)就是本组  $t_R$  提取出的规律路径。

2.3 交通模式的识别

文献[8]给出了一种 GPS 轨迹交通模式的识别算法。本文在此基础上,对其中的停止率特征进行了改进。文献[8]关于停止率特征的主要思想是,公交车相对于私家车会有更多的停止。但事实上,公交车还会有更规律的停止。利用这一

特征,可以更好地对交通模式加以区分。

由于在提取规律路径时,已经得到了若干条支撑路径。所以,可以统计各条支撑路径中是否存在一些用户总是低速通过的共同区域,该区域很可能就是规律停止点。如果一条规律路径上分布有较多的规律停止点,那么该路径的交通模式就可能是公共交通,反之则为私人驾驶。

定义 7 停止率是指用户以较低速度通过规律路径有向边的概率,即:

$$P(E_{m,n}) = \sum_{R_i \in E_{m,n}^{sup}} (E_{m,n}(R_i) \cdot v < V_{thre}) / E_{m,n} \quad (7)$$

式中:  $V_{thre}$  为判断停止的速度阈值。如果  $E_{m,n}$  的停止率高于给定阈值  $P_{stop}$ , 则认为是一个规律停止点。

定义 8 规律停止率是指一条规律路径中,单位距离上规律停止点的比例,即:

$$S(RR_j) = \sum_{E_{m,n} \in RR_j} (P(E_{m,n}) > P_{stop}) / RR_j \cdot distance \quad (8)$$

式中:  $RR_j \cdot distance$  为规律路径  $RR_j$  的长度。

### 3 试验分析

#### 3.1 试验设计

本文采用两套数据集分别对算法的效率以及有效性进行验证。数据集 1 是微软亚洲研究院发布的 Geolife 项目数据集<sup>[11-12,14]</sup>。该数据集记录了 178 名用户,自 2007 年至 2011 年间的生活轨迹。在该数据集上提取了每名用户每个月的规律路径。数据集 2 是来自 37 名用户两周的生活轨迹记录,然后由每名用户分别对提取出的规律路径以及识别的交通模式进行评分。评分内容包括:

召回率:是否检测到了全部规律路径(1~5 分);

准确率:检测出的规律路径是否正确(1~5 分);

交通模式正确性:判断是否正确(1~5 分)。

#### 3.2 试验结果

图 2 是在数据集 1 上得到的所有规律路线。当设置  $\alpha = \gamma = 3, \beta = 0.8$  时,共得到 389 条规律路径,如图 2(a)所示。设置  $\alpha = \gamma = 8, \beta = 0.8$  时,得到 31 条规律路径,如图 2(b)所示(结果中滤除了短于 2 km 的路径)。从结果可以看出,提

取出的规律路径基本覆盖城市的主要街道,并且在城市的西北角较为密集,这里正是本数据集的发布地。

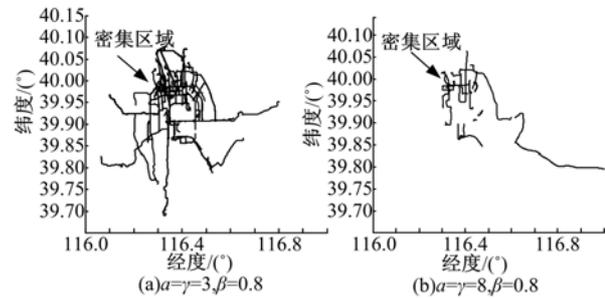


图 2 不同阈值下提取出的规律路径

Fig. 2 Regular routes under the different threshold

在数据集 2 上,挖掘到规律路径 72 条。用户对挖掘结果以及交通模式识别结果给出了评分。各项评分的平均结果如表 1 所示。

表 1 对规律路径挖掘结果的评分

Table 1 Rating result on regular routes mining

	召回率	准确率	交通模式识别率
得分	4.1	4.6	3.8

从表 1 可以看出,用户对挖掘的整体结果表示满意。分析错误案例时发现,有些用户会交替采用私家车或是公交车去上班,尽管出发地和目的地相同,但路线上存在差别,但用户觉得这仍然属于规律路径。此外,在交通模式的识别中,由于忽视了自行车出行,记录中自行车的最大时速达到了 27 km/h,远高于城区道路拥堵时的驾车时速。

#### 3.3 算法性能分析

鉴于目前鲜有基于历史轨迹的规律路线挖掘算法,试验无法与现有方法进行比较,因此只考虑本文所提方法的性能。

表 2 给出了算法各步骤对存储空间的需求变化。第一行为每一步的记录数量。第二行为相对于原始数据的比例。从结果可以看出,提取出的路径相对于原始数据集而言,所需的存储空间非常小。

表 2 算法的存储需求

Table 2 Storage cost of the algorithm

	原始 点数	去除停留 区域后	网格 映射后	规律路径 提取后
存储空间	23667828	16013562	1050313	21544
存储比例	1	0.676	0.044	0.0009

图 3 给出了网格尺寸对算法性能的影响分析。图 3(a)给出了随着网格尺寸的变化,轨迹映射误差和时序格数量的变化趋势。随着网格尺寸的增大,轨迹的映射误差呈线性增加,而时序格数量呈指数下降。图 3(b)给出了时序格数量对算法计算性能的影响曲线。随着时序格数量的增加,规律路径提取所需的计算时间呈明显上升趋势。因而网格尺寸的选择,需要在网格误差和网格数量上做适当的调整。从图 3 中可以看出,当网格尺寸大于  $10''$  后,网格数量的减少趋势明显放缓,而误差仍在持续增加。同时网格尺寸为  $10''$  时,平均映射误差约为 50 m 左右,满足城市大部分道路的分辨距离。综上所述,本文选取网格大小为  $10''$ 。

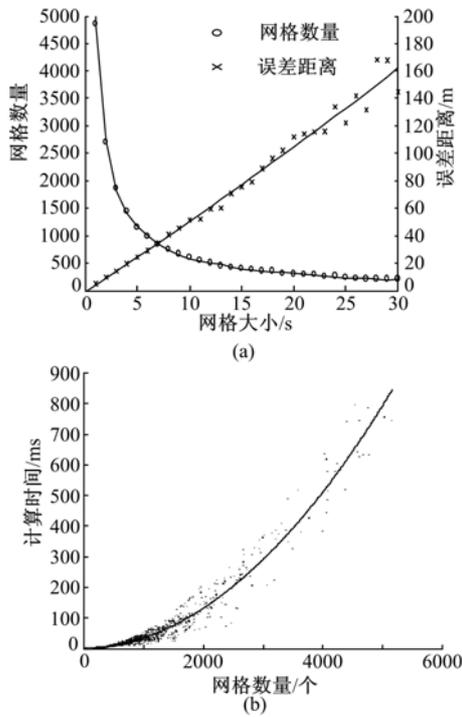


图 3 网格尺寸对计算性能的影响

Fig. 3 Impact of grid size on computing performance

图 4 演示了一个实际的规律路径提取过程。图 4(a)是一组包含有 13 条路径的  $t_R$ 。设置参数  $\alpha=\gamma=3, \beta=0.8$ 。图 4(b)(c)分别是 1~3 轮循环后得到的支撑路径。第 3 轮循环后,支撑路径的数量不再发生变化,运行结束。图 4(d)为最终提取的规律路径。从图 4(c)可以看出,每条支撑路径在细节上都有细微的差别,显示了算法的鲁棒性。

图 5 给出了在不同的停止率阈值下,交通模式识别的准确率曲线。在数据集 1 中,有 69 名用

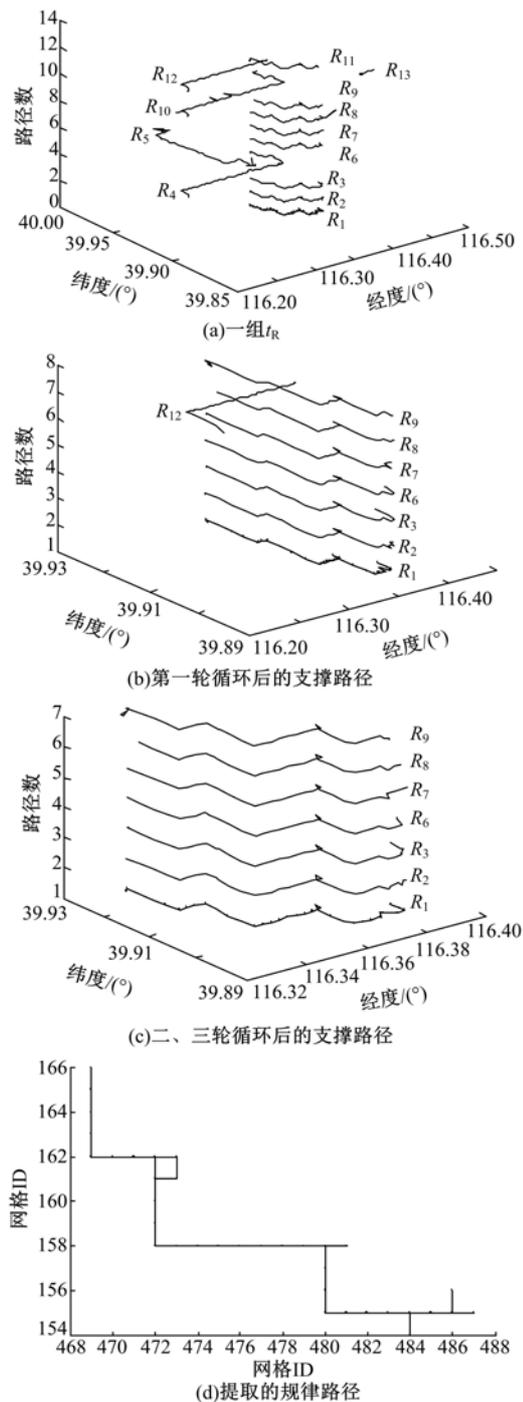


图 4 规律路径提取示例

Fig. 4 Example of Regular route mining

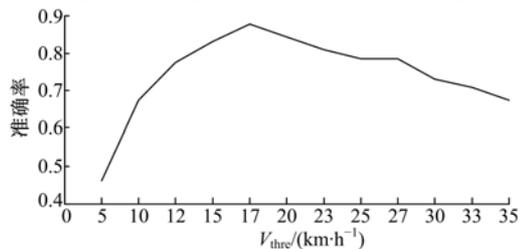


图 5 不同停止率阈值的识别准确率

Fig. 5 Recognition accuracy of different threshold for stop rate

用户对 GPS 轨迹的交通模式进行了标记。规律路径 89 条,并对交通模式进行了识别。从结果可以看出,设置适当的低速阈值后,方法的准确率可以达到 0.876,比文献[11]仅采用停止率,在该数据集上所得到的 0.6 的准确率有了较大提升。

#### 4 结束语

基于用户的历史轨迹记录,提出了一种规律路径的提取方法,并通过规律停止率特征提高了交通模式识别的准确率。通过在这两套数据集上的测试,表明了该方法的有效性和鲁棒性。

#### 参考文献:

- [1] 袁冠,夏士雄,张磊,等. 基于结构相似度的轨迹聚类算法[J]. 通信学报,2011,32(9):103-110.  
Yuan Guan, Xia Shi-xiong, Zhang Lei, et al. Trajectory clustering algorithm based on structural similarity[J]. Journal on Communications, 2011, 32(9):103-110.
- [2] Lee J G, Han J, Li X, et al. Mining discriminative patterns for classifying trajectories on road networks[J]. IEEE Trans Knowledge and Data Eng, 2011, 23(5):713-726.
- [3] 赵越,刘衍珩,余雪岗,等. 基于模式挖掘与匹配的移动轨迹预测方法[J]. 吉林大学学报:工学版,2008,38(5):1125-1130.  
Zhao Yue, Liu Yan-heng, Yu Xue-gang, et al. Method for mobile path prediction based on pattern mining and matching[J]. Journal of Jilin University (Engineering and Technology Edition),2008,38(5):1125-1130.
- [4] Chang K, Wei L, Yeh M Y, et al. Discovering personalized routes from trajectories[C]//Pro 3rd ACM Sigspatial International Workshop on Location-Based Social Networks,New York, 2011:33-40.
- [5] Chen L, Lü M, Ye Q, et al. A personal route prediction system based on trajectory data mining[J]. Information Sciences, 2011,181(7):1264-1284.
- [6] Zheng J, Ni L. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data[C]//Pro 2012 ACM Conf on Ubiquitous Computing, New York, 2012:153-162.
- [7] Farrahi K, Daniel G. What did you do today? Discovering daily routines from large-scale mobile data [C]//Pro of the 16th ACM Conference on Multimedia, New York, 2008:849-852.
- [8] Gidofalvi G, Pedersen T B. Mining long, sharable patterns in trajectories of moving objects[J]. GeoInformatica,2009,13(1):27-55.
- [9] 郭黎敏,丁治明,胡泽林,等. 基于路网的不确定性轨迹预测[J]. 计算机研究与发展,2010,47(1):104-112.  
Guo Li-min, Ding Zhi-ming, Hu Ze-lin, et al. Uncertain path prediction of moving objects on road networks[J]. Journal of Computer Research and Development,2010,47(1):104-112.
- [10] Lee J G, Han J, Whang K Y. Trajectory clustering: a partition-and-group framework[C]//Pro of the 2007 ACM SIGMOD International Conference on Management of Data, New York, 2007:593-604.
- [11] Zheng Y, Li Q, Chen Y, et al. Understanding mobility based on GPS data[C]//Pro of ACM Conference on Ubiquitous Computing, New York, 2008:312-321.
- [12] Zheng Y, Zhang L, Xie X, et al. Mining interesting locations and travel sequences from GPS trajectories [C]//Pro of International Conference on World Wild Web, New York, 2009:791-800.
- [13] Yuan J, Zheng Y, Xie X, et al. T-drive: enhancing driving directions with taxi drivers' intelligence[J]. IEEE Transactions on Knowledge and Data Engineering,2013,25(1):220-232.
- [14] Zheng Y, Xie X, Ma W. GeoLife: a collaborative social networking service among user, location and trajectory [J]. IEEE Data Engineering Bulletin, 2010,33(2):32-40.