

硕 士 学 位 论 文

文献综述

**基于海量交通数据挖掘的移动模型的研究与实现**

作 者 姓 名 杨文静

作 者 学 号 ZY1321222

学 科 专 业 软件工程

专 业 方 向 软件工程与管理

学院指导教师 王海泉

企业实习导师 周礼文

企业实习导师单北京三元合众技术有限公司

培 养 院 系 软件学院

论文起止时间 2015年4月~2015年12 月

文献综述提交时间 2015 年4月

##### 摘 要

车辆移动行为会影响了车联网的网络拓扑，从而影响到路由性能、网络延时等指标。同时车辆移动行为还会影响交通规划，路径选择等方面。车辆移动模型定义车辆节点的移动行为。由于智慧城市的建立，海量的出租车数据可以被采集到，但是如何发现和利用从海量数据中体现出的移动行为特征，并建立移动模型是当前研究的热点和难点。本文综述移动模型，首先给出在移动模型中的常见概念和划分，然后分别论述不同类别的移动模型，分析移动模型研究现状和发展趋势。最后指出当前研究中存在的问题。

**关键词：移动模型，轨迹，状态，区域**

##### Abstract

Vehicle movement behavior affects the car networking network topology, thus affecting routing performance, network latency and other indicators. While the vehicle is moving behaviors also affect transportation planning, route selection and other aspects. Moving vehicle model defines the behavior of the vehicle is moving nodes. Due to the establishment of the wisdom of the city's taxi massive data can be collected, but how to find and use the data reflects the massive move from behavioral characteristics, and the establishment of mobile model is hot and difficult current research. This paper reviews the mobile model is given first in the mobile division of common concepts and models, and then were discussed different types of mobile models, analytical models of mobile research status and development trend. Finally conclude the problems in the current study.

**Key words: Mobility model, trace, status, region**

目 录

[第一章 车辆移动模型概述 4](#_Toc415079123)

[1.1 移动模型的产生和意义 4](#_Toc415079124)

[1.2 移动模型分类方法 4](#_Toc415079125)

[1.2 移动模型研究热点问题 5](#_Toc415079126)

[1.1.1 宏观建模 5](#_Toc415079127)

[1.1.1 微观建模 5](#_Toc415079128)

[1.3 出租车轨迹数据集 5](#_Toc415079129)

[第二章 国内外研究现状概述 7](#_Toc415079130)

[2.1 随机移动模型 7](#_Toc415079131)

[2.2 概率约束的移动模型 7](#_Toc415079132)

[2.1 节点相互依赖的移动模型 8](#_Toc415079133)

[2.2 地理受限移动模型 8](#_Toc415079134)

[2.2 基于真实地图的移动模型 9](#_Toc415079135)

[2.2 引入交通特征的移动模型 9](#_Toc415079136)

[2.2 基于实际轨迹数据的移动模型 10](#_Toc415079137)

[本章小结 11](#_Toc415079138)

[第三章 当前存在的问题 12](#_Toc415079139)

# 车辆移动模型概述

## 1.1 移动模型的产生和意义

移动模型决定节点如何移动[1, 2]。车辆的移动模式对车辆自组织网络等以车辆为节点的网络的拓扑有直接影响。因此，在基于移动模型的仿真研究中，移动模型是否符合假设的场景对网络性能的评估影响巨大[3]。对车辆移动模型的研究可以提高仿真的有效性，同时发现车辆移动的规律，帮助提高车辆的运行效率，安全性以及网络通信的效率[4]。

在城市中，车辆移动受到道路限制，车辆节点移动速度较快，节点拓扑变化非常迅速，针对城市场景的仿真研究很难开展。因此从海量交通数据中发现车辆移动规律，建立符合真实场景的移动模型非常必要。

## 1.2 移动模型分类方法

移动模型的划分方法多种多样，根据节点的随机性可以将移动模型分为随机移动模型和受限移动模型。随机移动模型的节点移动时会随机生成方向和速度。然而受限移动模型，节点可以出现几种情况：

1. 节点当前移动受到之前移动的影响，例如高斯-马尔科夫移动模型。
2. 节点移动受到地理的限制，例如Manhattan移动模型。
3. 节点间运动会相互影响，在模型模拟交通流中的车辆，当发生拥堵时，后一辆车会发生减速甚至停车的行为。
4. 节点具有分组的特征，不同组的节点行为有所区别。

根据构建移动模型的方式可以分为轨迹移动模型和合成移动模型。轨迹移动模型通过观测轨迹行为信息，从而模拟真实轨迹生成移动模型。此中移动模型往往需要大量数据。而合成移动模型基于人们主观经验，由于其忽略了很多现实的细节，仅适用于对移动节点真实性较低的移动场景中。

根据节点间的关系也可以讲节点分为独立节点，以及相关节点。根据应用场景的不同也可以将节点分为具有固定节点的移动场景和不具有固定节点的移动场景，前者主要应用于能与周边AP或基站通信的场景中，后者只能节点间相互交互，节点关系较为平等的网络。

## 1.2 移动模型研究热点问题

本节介绍在移动模型建模中的热点问题，可以从宏观和微观两个方面来分析。

### 宏观建模

在宏观建模方面主要有两个方面，用于预测节点移动的宏观方向。一是对区域转移建模，在交通方面的OD矩阵，即出发点、目的点转移矩阵。二是对城市的分层次建模[5]。宏观建模主要涉及设定区域定义，以及区域内移动和跨区域移动模式定义。某些模型还需要确定如何根据当前位置，来确定目的位置。

### 微观建模

微观建模主要关注车辆的速度变化、方向和暂停时间等。简单的微观模型可以按概率设定速度和暂停时间。当地理不受限的移动模型中，方向是从一个角度范围中按均匀分布或按其他概率分布中选择。有些模型还考虑到了节点间的相互影响，节点行为会受其他节点行为的影响，通过反馈来设置自己的行为参数。

## 1.3 出租车轨迹数据集

从“北京智能交通系统关键技术研究与应用示范项目”获取的数据集，该项目以支持智能交通系统工程建设、解决关键技术难题、提升交通科技发展水平和自主创新能力、为实现新北京交通体系和奥运会的顺利召开提供支持与保障为目标。其核心研发内容之一，是实时采集、存储、处理多源异构海量交通数据、形成动态交通信息以及决策支持的分布式处理系统。该项目涉及出租车为12096辆，约占北京市出租车总数的18%，占北京市机动车总数的0.2%，对五环内（含五环）次干路以上路网的覆盖率达到90%以上。通过这些出租车上安装的GPS定位装置，每隔60s上传一次自己的经纬度位置、速度、方向信息到数据中心。每天产生的数据量约1300万条。

数据集包含了北京市区面积为625平方公里内1.22×108条实际移动轨迹记录，涉及出租车车辆12,096辆，数据记录格式如表 2所示。

表 2 智能交通系统的出租车轨迹数据格式

|  |  |
| --- | --- |
| **列名** | **说明** |
| 调度中心ID | 4个ASCII字符； |
| 出租公司ID | 标记出租车为何公司所有，例如：$JYJ |
| 车辆ID | 使用11个ASCII字符； |
| 时间标签 | 使用GMT时间格式共14个ASCII字符，格式为（YYYYMMDDHHMMSS）； |
| 84坐标系经度 | 最长为11个ASCII字符，变长； |
| 84坐标系纬度 | 最长为10个ASCII字符，变长； |
| 02坐标系经度 | 一般为9个ASCII字符，除以3686400（1024\*3600）后变为通用经度坐标； |
| 02坐标系纬度 | 一般为9个ASCII字符，除以3686400（1024\*3600）后变为通用纬度坐标； |
| 速度 | 单位为公里/小时，最长为3个ASCII字符，变长； |
| 方向 | 以正北为0度，顺时针方向增大，为0~360角度，最长为3个ASCII字符，变长; |
| 状态 | 为1个ASCII字符，数值型字符，包括空载、满载等状态。 |
| 事件（event） | 为1个ASCII字符，数值型字符，包括上下车、开锁车门等。 |
| 高度 | 两个ASCII字符，固定值为50，未使用 |

# 国内外研究现状概述

## 2.1 随机移动模型

随机移动模型是指节点运动方向速度随机的移动模型。由于其模型简单，依赖信息少可以简单描述节点移动，被广泛应用到基于移动模型的仿真研究中。

随机游走模型（Random walk mobility model, RW）[6]是由Einstein提出的用于模拟物理粒子的布朗运动。节点在每段独立时间t内，随机选择移动方向和移动速度，当到达边界时，遵循反射定律按照一定角度反弹后回到场景中继续移动。随机游走模型在仿真中经常被用于模拟人类在无障碍的场景中的移动行为。

随机路点模型（Random Waypoint model, RWP）[7]与RW的不同之处在于不是固定时间，而是在一段路径开始时，随机选取下一步的目标点，并随机生成速度。到达目的点后，会定义一个暂停时间，暂停时间也是从一个范围内随机生成的。因此在随机路点模型中不用考虑边界问题。随机路点模型被广泛用于车载网络的仿真中。随机路点移动模型节点都在初始位置周围运动，但是节点可能无法遍布仿真区域，节点分布不均匀。

随机方向模型(Random Direction model, RD)[8]被应用与多种自组网协议中。它定义节点运动为匀速，节点方向从[0,2ᴨ]角度内均匀选择一个方向，直到运动到区域边界点D。然后保持静止时间t, 然后以此点为初始节点，重复以上运动[1]。但是RD模型在空间分布上表现出不均匀的特质，中心概率密度小，边界概率密度大。

以上三种模型是经典的移动模型，在移动方式方面没有利用任何先验知识，适用于简化的移动场景中。

## 2.2 概率约束的移动模型

基于概率矩阵的随机漫步模型（Probabilistic version of random walk mobility model）采用三个状态来决定移动方向，1是开始位置，0是之前的位置，2是下一步的位置。通过三个状态的转移矩阵来确定下一步的方向。

2002年，Hsu等人提出了加权路径点模型（Weighted Way Point model）[9]。该模型对较高访问频率的热点区域，目的地的选择与当前时间和位置有关，每个位置有不同的暂停时间。

## 2.1 节点相互依赖的移动模型

高斯-马尔科夫模型(Gauss-Markov model)[10]采用马尔科夫链模型，认为第n次的运动与前一次运动的速度、方向等有关。其速度和方向是符合高斯分布的随机变量。

2004年，Musolesi, M.等[11]提出了基于社会关系的移动模型。采用加权图表示社会关系网络，权值用来衡量个体之间交互关系的强弱。每个节点都具有社交因子（Social Factor, SF）用来衡量与其他节点的态度。

2001年，Bettstetter等人[12]提出平滑随机移动模型（Smooth Random Mobility model）,该模型是在RD基础上的改进。节点的速率和方向是和时间相关的使得模型速率和方向上的改变时平滑的。速率控制的思想是设定一个目标速率，使得加速度线性变化，最后达到目标速率。这个过程符合泊松分布。节点的方向变化与速率相关，节点速率小轨迹的半径小，速度偏转大，反之，节点速度大时，轨迹的半径也随之变大。

另一类移动模型中节点以组的形式移动。队列移动模型（Column Mobility Model）[13]定义一组节点以统一方向移动。追逐移动模型（pursue mobility model）[14]模拟节点追逐某物时的移动模式，按照加速函数来确定节点的速度。游牧团体移动模型（Nomadic Community Mobility Model）定义一组节点共同从一个区域移动到另一个区域的运动模式。通过定义参考点，组内节点在参考点附近随机移动。当参考点运动时，组内节点会向参考点附近游走。

## 2.2 地理受限移动模型

曼哈顿模型（Manhattan mobility model）[15]是通过对曼哈顿街区建模，并规定节点按照街道模型移动的移动模型。节点在城区地图中沿着垂直或水平的方向移动。在十字路口节点可以选择直行，或者转向。类似的移动模型有市区移动模型[16]。该模型将对街道的建模精细化，由简单的网格街道变为模拟市区内的公路或街道的移动模型。文献[17]结合了交通理论，将地理信息分为市区、地区和街区，并依据此建模。另一类移动模型，关注在地理上的限制区域。文献[5]对建筑物等障碍物建模，提出了障碍物移动模型。

## 2.2 基于真实地图的移动模型

1999年，Hong, X.[18]等人提出了Reference Point Group Mobility (RPGM)用于表现移动主机的关系。实验验证了移动模型对网络协议和性能的影响，验证了移动性对分簇和网络性能的影响。研究发现移动性增加，连接变化更为频繁，移动性增加，簇头变化更为频繁，移动性增加，负载变化。RPGM也是一种群主移动模型。

2004年，Bhattacharjee, D.等人提出了混合移动模型[19]。该模型在加州大学圣迭戈分校校园内采集AP数据，收集用户轨迹，并通过统计分析来建立模型，可以用来预测节点的位置。

2004年，Saha, A.K[20]等引入了实际道路，对车载网络建模。实验验证从路由性能等方面与RWP模型进行了比较，发现两者的不同。

2005年，Jain R等人[21]采用Dartmouth 大学校园收集的真实的轨迹数据

2005年，文献[22]提出了车载无线网络综合模型和交通模型。该模型基于由真实路网简化的道路模型来验证ad hoc网络性能。提出了基于道路的移动模型STRAW. 验证发现该模型的网络性能与RWP模型差别非常大。此外，文献还验证了不同城市环境中的网络性能，发现节点增多的时候芝加哥和波士顿的场景中平均速度会减小，和投递率。该模型基于真实的道路信息，验证了RWP模型无法表现出城市车载网络的特征。

这些移动模型探究在加入了地理信息后，移动模型在网络性能、移动特征等方面的影响，在验证方面没有与真实的数据比较，因此难以说明模型的真实性。

## 2.2 引入交通特征的移动模型

从对车流的刻画粒度上，可以将移动模型分为宏观、微观和宏观微观结合的移动模型。宏观模型将车流看作连续流体，忽略车辆的细节行为。微观模型则对节点行为进行细粒度的刻画。关注节点速率、方向、加速度、交通灯情况等。但是一个真实的移动模型需要同时考虑车辆宏观和微观的特征。车辆移动模型需要关注车辆密度、车流等宏观特征，也需要考察车辆的间距、加速度、刹车、超车等微观特征。

1992年，Seskar, Ivan等人在文献[23]中讨论了移动网络中轨迹跟踪的问题，并基于真实的交通参数的关系建立了车辆移动模型（Fluid Traffic Model，FTM）。该模型给出了车辆速度与车流密度之间的关系，即当车流密度增加时，车辆速度会相应地减慢, 并在实验验证中观察了该模型的适用场景。

2000年，Treiber, M.从德国高速公路采集到的数据中发现，几种不同类型的拥塞发生在不同的地段，分别是关闭通道的道路、十字路口和上坡路等，并分析了各种不同拥塞以及混合型拥塞发生的状况，在文献[24]中提出智能驾驶员模型（Intelligent Driver Model，IDM），它是一个连续的微观单通道模型，将相邻车辆行为进行关联，建立车辆跟随模型，实验结果表明，该模型可以用来发现图瓶颈的更通用方式。VanetMobiSim[25]优化了 IDM 模型，引入了对十字路口和多车道的控制，设计了具有十字路口管理的移动模型（Intelligent Driving Model with Intersection Management，IDM-IM）和可切换车道的移动模型（Intelligent Driving Model with Lane Change，IDM-LC）。

2011年，Helbing, D等人[26]对“幻像交通拥塞”、走走停停（stop-and-go）交通现象的产生机制、不同拥塞的产生原因和相关之处、当将要达到道路交通能力时交通堵塞最为严重、暂时减少交通量是否会造成交通堵塞等现象和问题进行了解释，并依据此为自动（self-driven）多组分（many-particle）系统建立了一个通用的模型框架。

引入交通特征的移动模型，对车辆间的关系以及关系对车辆微观行为的影响较为关注。文献[24]等研究者同时也引入了交通数据，用于研究真实场景中的拥塞分类和发生的时间和地点。但是关于交通特征的研究过于偏向微观，对宏观现象的解释和建模较为缺乏。

## 2.2 基于实际轨迹数据的移动模型

2006年，文献[28]采用移动用户数据，对无线网络中的移动节点建模。他们从实际收集的13个月的轨迹数据中发现节点的速度和暂停时间符合对数正太分布，并且节点的运动方向收到道路方向的影响。在此基础上，他们提出了关注热点区域的移动模型。模型的平均相对误差为17%.

2007年，Zhang, X等人提出了基于公交车容断络移动模型[29]。该模型采用UMass DieselNet数据集，该数据集由安装了WiFi设备的公交车组成网络收集而来。通过对数据的研究发现节点对级别的接触并没有明显的规律，然而在路由级别的节点对接触表现出明显规律。基于此发现，他们提出了接触间隔时间生产模型。实验证明基于路由级别的接触生产模型能更准确符合实际路由性能。

同年，Hsu, W等人提出了随时间变化的用户移动模型[30]。该模型采用WLAN轨迹数据。他们利用不均匀的访问地点偏好将社区定义为经常被节点访问的地区，并采用不同参数的时间段来发现在某区域周期性出现的节点。实验结果显示其在hitting time和meeting time两个指标上与实际的相对误差不高于20%.

2010年，Hongyu Huang等人[31]提出基于上海市出租车GPS数据的城市移动模型。该模型分为宏观和微观两个方面，宏观方面，他们对区域进行了划分，探究区域间的转移概率。微观方面，他们对车辆速度和运行方向进行建模。该模型中引入了城市路网，因此在出租车方向的选择方面，他们对路口的转向进行了建模。验证结果表明，该模型与实际具有很好的相似性。

由于数据集难以获取，并且难以达到统一，因此现有的研究者从真实数据集出发建立的模型较少，模型验证也偏向于说明模型是否达到预期的指标，仿真现象是否符合常识推断，难以与真实数据集相比较。基于大规模真实数据集的移动模型由于数据的准确性、难获取性建模较为困难。在出租车数据集上建立的移动模型更为少见，文献[31]中建立的模型忽略了出租车在空车和重车时的不同点，会影响移动模型的真实性。

## 本章小结

本章综述了各类移动模型，通过分类比较现有的移动模型，我们发现1）经典的随机移动模型和基于概率的随机移动模型具有简单、易于构造和使用的特点，但是由于过于简化并且是依据人们经验得出的，因此不太符合真实的移动场景。在车载网络中情况更为复杂，与道路车辆间的影响都十分相关，因此用此类移动模型进行仿真必然会造成实验结果上的误差。2）地理受限的移动模型考虑到节点运动与道路、地区的相关性，但是对车辆自身具有的主观性等考虑较少。3）引入交通特征的移动模型，比较关注车辆微观行为、交通灯影响、车辆拥塞等情况，对宏观情况的研究比较缺乏。4）基于真实轨迹的移动模型从真实数据的统计结果中发现车辆行为规律，反过来建立模型会较为符合真实情况。但是数据难以获取，使得此类模型较少。

# 当前存在的问题

存在的问题有：

* 现有的移动模型很少有反映真实的移动环境

移动模型研究的目的是模拟真实的移动场景。但是现有的模型主要探究节点移动模式对路由协议的影响，缺乏真实数据支持。首先，真实统计数据难以获取。

* 在真实环境中节点分布的非均匀性没有考虑
* 在现有研究中忽视出租车状态和行为的关系
* 对大规模的出租车数据没有很好的应用

**References:**

[1]. 高小鹏, 童.A.牛.A., 移动模型研究综述. 计算机科学, 2009(10): 第5-10+85页.

[2]. Camp, T., J. Boleng and V. Davies, A survey of mobility models for ad hoc network research. Wireless communications and mobile computing, 2002. 2(5): p. 483--502.

[3]. 林航, 车载自组织网络移动模型研究, 2013, 西安电子科技大学.

[4]. 王沿锡, 车载自组网移动模型研究, 2013, 吉林大学.

[5]. Jardosh, A., et al., Towards realistic mobility models for mobile ad hoc networks. 2003, ACM. p. 217--229.

[6]. Einstein, A., Investigations on the Theory of the Brownian Movement. 1956: Courier Corporation.

[7]. Bettstetter, C., H. Hartenstein and X. P E Rez-Costa, Stochastic properties of the random waypoint mobility model: epoch length, direction distribution, and cell change rate. 2002, ACM. p. 7--14.

[8]. Royer, E.M., P.M. Melliar-Smith and L.E. Moser, An analysis of the optimum node density for ad hoc mobile networks. 2001, IEEE. p. 857--861.

[9]. Hsu, W., et al., Weighted waypoint mobility model and its impact on ad hoc networks. ACM SIGMOBILE Mobile Computing and Communications Review, 2005. 9(1): p. 59--63.

[10]. Liang, B. and Z.J. Haas, Predictive distance-based mobility management for PCS networks. 1999, IEEE. p. 1377--1384.

[11]. Musolesi, M., S. Hailes and C. Mascolo, An ad hoc mobility model founded on social network theory. 2004, ACM. p. 20--24.

[12]. 徐鑫鑫, 无线移动AD Hoc网络移动模型研究, 2008, 国防科学技术大学.

[13]. Camp, T., J. Boleng and V. Davies, A survey of mobility models for ad hoc network research. Wireless communications and mobile computing, 2002. 2(5): p. 483--502.

[14]. Camp, T., J. Boleng and V. Davies, A survey of mobility models for ad hoc network research. Wireless communications and mobile computing, 2002. 2(5): p. 483--502.

[15]. Bai, F., N. Sadagopan and A. Helmy, IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of RouTing protocols for Adhoc NeTworks. 2003, IEEE. p. 825--835.

[16]. Davies, V.A. and Others, Evaluating mobility models within an ad hoc network. 2000, Citeseer.

[17]. Markoulidakis, J.G., et al., Mobility modeling in third-generation mobile telecommunications systems. Personal Communications, IEEE, 1997. 4(4): p. 41--56.

[18]. Hong, X., et al., A Group Mobility Model for Ad Hoc Wireless Networks, in MSWiM '99. 1999: New York, NY, USA. p. 53--60.

[19]. Bhattacharjee, D., et al., Empirical modeling of campus-wide pedestrian mobility observations on the USC campus. 2004, IEEE. p. 2887--2891.

[20]. Saha, A.K. and D.B. Johnson, Modeling Mobility for Vehicular Ad-hoc Networks, in VANET '04. 2004: New York, NY, USA. p. 91--92.

[21]. Jain, R., D. Lelescu and M. Balakrishnan, Model T: an empirical model for user registration patterns in a campus wireless LAN. 2005, ACM. p. 170--184.

[22]. Choffnes, D.R. and F.A.N.E. Bustamante, An Integrated Mobility and Traffic Model for Vehicular Wireless Networks, in VANET '05. 2005: New York, NY, USA. p. 69--78.

[23]. Seskar, I., et al., Rate of location area updates in cellular systems. 1992.

[24]. Treiber, M., A. Hennecke and D. Helbing, Congested traffic states in empirical observations and microscopic simulations. Physical Review E, 2000. 62(2): p. 1805.

[25]. VanetMobiSim, http://vanet.eurecom.fr..

[26]. Helbing, D., Traffic and related self-driven many-particle systems. Reviews of modern physics, 2001. 73(4): p. 1067.

[27]. Kim, M., D. Kotz and S. Kim, Extracting a Mobility Model from Real User Traces. 2006. p. 1-13.

[28]. Zhang, X., et al., Study of a Bus-based Disruption-tolerant Network: Mobility Modeling and Impact on Routing, in MobiCom '07. 2007: New York, NY, USA. p. 195--206.

[29]. Hsu, W., et al., Modeling Time-Variant User Mobility in Wireless Mobile Networks. 2007. p. 758-766.

**校对报告**

当前使用的样式是 [Numbered(Multilingual)]

当前文档包含的题录共34条

有1条题录存在必填字段内容缺失的问题

参考文献 [25] ：字段(作者)内容缺失;

字段(年份)内容缺失;