附件4： 北航软件学院研究生开题报告内容要求



硕 士 学 位 论 文

开题报告

|  |  |
| --- | --- |
| 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 副教授 |
| 企业实习导师 | 周礼元 |
| 企业实习导师单位 | 北京三源合众科技有限公司 |
| 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 |
| 学院指导教师签字： |  |

基于海量交通数据挖掘的移动模型的研究与实现

目 录

[1 课题来源和意义 4](#_Toc415554982)

[1.1 课题来源 4](#_Toc415554983)

[1.2 课题背景 4](#_Toc415554984)

[1.3 选题意义 4](#_Toc415554985)

[2 国内外研究现状 5](#_Toc415554986)

[2.1 现状分析 5](#_Toc415554987)

[2.1.1 随机移动模型 5](#_Toc415554988)

[2.1.2 概率约束的移动模型 6](#_Toc415554989)

[2.1.3 节点相互依赖的移动模型 6](#_Toc415554990)

[2.1.4 地理受限移动模型 6](#_Toc415554991)

[2.1.5 引入交通特征的移动模型 7](#_Toc415554992)

[2.1.6 基于实际轨迹数据的移动模型 8](#_Toc415554993)

[2.1.7 小结 9](#_Toc415554994)

[2.2 当前存在的问题 10](#_Toc415554995)

[3 论文研究方案 10](#_Toc415554996)

[3.1 研究目标 10](#_Toc415554997)

[3.2 研究内容 10](#_Toc415554998)

[3.3 拟解决的关键问题 10](#_Toc415554999)

[3.4 拟采用的研究方法和技术路线 11](#_Toc415555000)

[3.5 可能的创新之处 11](#_Toc415555001)

[4 预期达到的目标和成果 11](#_Toc415555002)

[5 论文工作计划 11](#_Toc415555003)

[参考文献 12](#_Toc415555004)

# 课题来源和意义

## 课题来源

项目来源是网络技术北京市重点实验室。

## 课题背景

车载自组网（ad hoc）是通过测量互联或者测量与周边基础设施通信建立的网络，由于车辆具有主动连接成网的特征，因此被称为自组织网络。在ad hoc 网络中，车辆可以通过车载通信设备，交换包括车辆信息（如车速、加速度、位置和方向等）以及路况信息（如交通拥塞情况、交通灯信息和道路实时状态等），以提高车辆运行效率，较少交通拥塞，提高出行的安全性。因此Ad hoc主要应用于智能交通系统中的安全预警、协助驾驶、分布式交通信息发布、基于通信的车辆控制及办公与娱乐化五个方面[1]。

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。

## 选题意义

移动模型对仿真验证的有效性有显著影响，然而对大规模的车载网络建立真实的移动模型十分困难，因此，本课题的意义在于：

1. 发现海量数据中表现出的出租车行为特征
2. 设计和实现移动模型，用于仿真研究
3. 促进对出租车行为模式的认识，利于仿真研究、路径规划、车载网络通信等方面。

# 国内外研究现状

## 现状分析

本章主要对国内外移动模型现状进行综述，针对移动模型的约束信息和依赖信息，从随机移动模型、概率约束移动模型、节点关系依赖的移动模型、地理受限移动模型、基于真实地图的移动模型、引入交通特征的移动模型和基于真实轨迹的移动模型几方面进行综述。

### 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）[5]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）[6]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)[7]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。

### 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）[8]。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。

### 节点相互依赖的移动模型

高斯-马尔科夫模型(Gauss-Markov model)[9]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[10]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[11]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）[12]定义一组节点以统一方向移动。追逐移动模型（pursue mobility model）[13]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。

### 地理受限移动模型

1. 人为规定的地理受限移动模型

曼哈顿模型（Manhattan mobility model）[14]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[15]。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[16]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[17]对建筑物等障碍物建模，提出了障碍物移动模型。

1. 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。

### 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。

### 基于实际轨迹数据的移动模型

2006年，文献[27]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[28]。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[29]。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[30]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[30]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。

### 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

# 论文研究方案

## 研究目标

## 研究内容

* 发现海量交通数据中的出租车移动特征
* 建立移动模型
* 验证模型的有效性

## 拟解决的关键问题

* 数据清洗和规整

由于GPS数据偏差和采样间隔不统一，需要去掉错误数据，并且对数据进行规整，使其更符合语义。

* 对载客平均速度和空车平均速度建模

因为速度字段是GPS采样数据，不准确，为了减小误差，我们采用一段重车（空车）距离除以时间的方法获取速度，然后研究其速度的分布情况。

* 定义区域以及区域划分

最简单的区域定义是将场景划分成规则的四边形，我们认为这样不能体现出不同区域在出租车在地理上的不同点。

因此，我们将区域划分为细粒度的网格，对相邻的具有相似上客（下客）量的网格聚类为区域。

* 区域转移概率的计算
* 目的地点选取以及寻径

## 拟采用的研究方法和技术路线

## 可能的创新之处

# 预期达到的目标和成果

预期达到的研究成果包括毕业论文、移动模型的实现算法、关于移动模型的论文一篇

# 论文工作计划

论文的工作计划如表1所示。

表 1 论文工作计划表

|  |  |  |
| --- | --- | --- |
| 序号 | 起止时间 | 工作内容 |
| 1 | 2015年3月15日-2015年4月1日 | 文献整理及阅读，撰写开题报告及文献综述 |
| 2 | 2015年4月1日-2015年4月15日 | 数据清洗以及速度建模 |
| 3 | 2015年4月15日-2015年5月1日 | 区域定义，区域划分，区域转移概率矩阵计算 |
| 4 | 2015年5月1日-2015年5月15日 | 引入地图，将地图处理为连通地图便于仿真 |
| 5 | 2015年5月15日-2015年6月15日 | 在仿真平台上设计和实现移动模型 |
| 6 | 2015年6月15日-2015年7月1日 | 选取比较对象和比较指标，设计仿真实验 |
| 7 | 2015年7月1日-2015年7月15日 | 整理资料撰写中期报告 |
| 8 | 2015年7月15日-2015年10月1日 | 对比分析，验证模型的有效性 |
| 9 | 2015年10月1日-2015年12月 | 毕业论文撰写 |

参考文献

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

[6]. 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.

[7]. 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.

[8]. 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.

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

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

[11]. 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.

[12]. 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.

[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]. 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.

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

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

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

[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.

[30]. Huang, H., et al., META: A Mobility Model of MEtropolitan TAxis Extracted from GPS Traces. 2010. p. 1-6.

校对报告

当前使用的样式是 [Numbered(Multilingual)]

当前文档包含的题录共33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;