

学校编码: 10384

分类号_____密级_____

学号: 24320141152407

UDC_____

厦 门 大 学

硕 士 学 位 论 文

车联网海量数据实时处理研究

Research on Massive Data Real-time Processing in
Internet of Vehicles

陈亚婷

指导教师姓名: 林坤辉 教授

专业名称: 软件工程

论文提交日期: 2017 年 4 月

论文答辩日期: 2017 年 5 月

学位授予日期: 年 月

指 导 教 师: _____

答 辩 委 员 会 主 席: _____

2017 年 4 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘 要

作为物联网的重要分支，车联网的研究已得到越来越多的关注。而对于面向企业的车联网平台来说，系统需要收集和分发大量实时信息，系统规模由该企业管辖的车辆规模和消息实时处理性能需求决定。对接收到的信息进行处理，应用领域广泛，如实时监控、驾驶行为分析、告警提醒等。这些数据中有一些需要实时分析，例如实时监控和告警提醒。传统数据处理方式应对高速率的数据实时处理需求时表现出了延迟高，响应缓慢等不足，无法满足车联网性能需求。

本文以某车联网云服务平台项目为背景，研究车联网中海量数据实时处理问题。主要包括实时监控中的海量数据在地图中可视化响应缓慢问题、车辆区域限定判断算法研究处理和车辆最近邻服务站搜索的查询速度问题。

1、用户缩放或拖动地图，以不同视野监控所管辖的车辆分布情况，百度地图提供的聚合功能随着数据量增多响应时间变长，不能满足车辆实时监控的性能要求。论文改进了k-means聚合算法，当用户缩放地图时车辆能快速聚合。通过实验结果表明，论文改进方法能准确反映车辆原始分布情况，响应迅速及时。

2、车辆超出被限定区域时，平台应该及时产生告警。在预处理时，用扫描转换算法将区域转为网格存储，采用莫顿码对每一个网格进行编码，然后用四叉树编码压缩算法进行存储压缩。接收到车辆位置数据时，只需要在四叉树中进行搜索。相比于射线法，改进算法能将判定时间 $O(N)$ 从降到 $O(\log_4^N)$ 。

3、当车辆进入服务站半径范围时，标记该车；根据车辆位置查询最近的服务站，这是个Top-k搜索问题。车辆数据每三十秒传回一次，因此查询非常频繁，遍历系统中所有服务站就相当耗时，效率非常低。论文基于k-means的数据划分，对划分后的簇中心以及每一个簇的数据建立KD树，并通过改进KD树的最近邻搜索提高Top-k查询效率。

论文提出三个改进方法确实可行，并在实际车联网系统中得到应用和验证，该系统基于OpenStack云计算平台，ActiveMQ消息服务器，MySQL和MongoDB数据库，Redis缓存。系统具备支持接入10万台车的能力，并可通过横向扩展快速支撑更多车辆接入能力。

关键词：车联网；海量数据；实时处理

Abstract

As an important branch of the Internet of Things, the research of Internet of Vehicle(IoV) has been paid more and more attention. For enterprise, the IoV system need to collect and distribute large amounts of real-time information, the system scale determined by concurrent access quantity and message processing requirements of the vehicle terminal. Processing the information received from IoV, has a wide range of applications, such as real-time monitoring and analysis of driving behavior, alarm reminder, etc. Some of these data require real-time analysis, such as real-time monitoring and alarm reminder. With the high frequency of real-time data processing, traditional methods have high delay, slow response and other deficiencies, which cannot meet the needs of the performance of the IoV system.

In this dissertation, we research the massive real-time data processing problems in IoV system based on a real project undertaken by our research center. It mainly includes the massive real-time data visualized in the map, the processing speed of the vehicle area limited alarm processing and the time delay of the vehicle service station alarm processing.

1. The user can zoom or drag the map to monitor his vehicle distribution in different field. The aggregation function provided by Baidu maps has long response time, which cannot meet the real-time monitoring performance of IoV system requirements. This dissertation improved k-means aggregation algorithm based on the administrative region partition, which may produce some overlap on map, but it can accurately reflect the original distribution of the vehicle and the response is rapid and timely.

2. When the vehicle is out of the restricted area, the IoV system should be alert in time. In the pre-processing, the polygon region is converted to grid storage by using the scan conversion algorithm, and each trellis is encoded by Morton code, then the quadtree compression algorithm is used to compress the storage. When the vehicle position data is received, determine the vehicle is inside or outside the area by

quadtree searching. Compared with ray-crossing method, our improved method can reduce the decision time from $O(N)$ to $O(\log_4^N)$.

3. When the vehicle enters the radius of the service station, IoV system should mark the car. According to the location of the vehicle to query the nearest service station, this is a Top-k search problem. Vehicle data is returned every thirty seconds, so the query is very frequent, traversing all service stations is time-consuming, result in low efficiency. In this dissertation, we propose a data divide based on k-means, set up the KD tree for the data of the cluster centers and the data of each cluster, and then use the KD tree to find the station instead of the Euclidean distance calculation to improve the query efficiency.

The three improved method we proposed is feasible, and has been applied in practical IoV system, the system based on OpenStack cloud computing platform, ActiveMQ message server, MySQL and MongoDB database, Redis cache. The system has the ability to support access to 100 thousand cars, and can support the rapid expansion of more vehicle access through horizontal expansion.

Key words: IoV System; Massive Data; Real-time Processing

目 录

第一章 绪论	1
1.1 引言.....	1
1.2 国内外研究现状	3
1.2.1 国外车联网发展现状.....	3
1.2.2 国内车联网发展现状.....	4
1.3 论文主要研究内容	5
1.4 论文组织结构	6
第二章 相关理论与技术介绍	8
2.1 车联网平台关键技术	8
2.2 车联网中的大数据处理技术	9
2.2.1 大数据主要技术.....	9
2.2.2 云服务平台.....	10
2.2.3 高速缓存服务器.....	10
2.3 数据可视化	11
2.3.1 数据可视化定义.....	11
2.3.2 基于百度地图 API 与 EasyUI 的 GPS 数据可视化.....	12
2.3.3 标记点聚合算法.....	12
2.4 本章小结	17
第三章 车辆实时监控数据聚合展示研究	18
3.1 标记点动态聚合算法改进	18
3.1.1 基于行政单位的标记点动态聚合算法.....	18
3.1.2 基于改进的 k-means 标记点动态聚合算法	20
3.2 车联网中的车辆实时监控	21
3.3 算法实验结果与讨论	22
3.3.1 实验数据及参数设置.....	23
3.3.2 实验结果.....	23
3.4 本章小结	29

第四章 车辆区域限定判断算法研究	30
4.1 区域限定判断方法	30
4.2 区域限定判断方法改进	31
4.3 算法实验结果与讨论	35
4.4 本章小结	39
第五章 车辆最近邻服务站搜索算法研究	40
5.1 Top-k 介绍	40
5.2 Top-k 算法改进	40
5.2.1 基于改进 k-means 的数据划分 (DDK) 算法	40
5.2.2 基于改进 KD 树算法的数据选择	41
5.2.3 基于改进 top-k 算法的服务站进入判断	47
5.3 实验结果与讨论	48
5.4 本章小结	50
第六章 总结与展望	51
6.1 总结.....	51
6.2 展望.....	52
参考文献	53
攻读硕士期间的研究成果	56
致 谢	57

Contents

Chapter 1 Introduction.....	1
1.1 Background.....	1
1.2 Research Status Review.....	3
1.2.1 IoV Research Status at Home	3
1.2.2 IoV Research Status at Abroad	4
1.3 Main Research Contents	5
1.4 Dissertation Structure.....	6
Chapter 2 Overview of the Related Theory and Technologies	8
2.1 Key Technologies of IoV	8
2.2 Big Data Processing Technology in IoV	9
2.2.1 Big Data Technology	9
2.2.2 Cloud Service Platform.....	10
2.2.3 Distributed Cache Server	10
2.3 Data Visualization	11
2.3.1 Data Visualization Definition.....	11
2.3.2 Data Visualization base on Baidu map API and EasyUI	12
2.3.3 Mark Point Aggregation Algorithm	12
2.4 Summary.....	17
Chapter 3 Research on Real Time Monitoring Data Aggregation... 18	
3.1 Dynamic Aggregation Algorithm Improvement.....	18
3.1.1 Aggregation Algorithm based on Administrative Units.....	18
3.1.2 Aggregation Algorithm based on Improved K-means Algorithm.....	20
3.2 Vehicle Real Time Monitoring	21
3.3 Experimental Results and Discussion	22
3.3.1 Experimental Data and Parameter Setting	23
3.3.2 Experimental Result and Discussion	23
3.4 Summary.....	29

Chapter 4 Research on Region Limits Judgment Algorithm	30
4.1 Point-in-Area Algorithm.....	30
4.2 Improvement Algorithm.....	31
4.3 Experimental Results and Discussion	35
4.4 Summary.....	39
Chapter 5 Research on Nearest Service Station Search Algorithm .	40
5.1 Top-k Introduction	40
5.2 Top-k Improvement	40
5.2.1 Data Partitioning Algorithm based on Improved K-means(DDK)	40
5.2.2 Data Selection based on Improved KD- tree Algorithm.....	41
5.2.3 Service Station Entry Judgment based on Improved Top-k Algorithm	47
5.3 Experimental Results and Discussion	48
5.4 Summary.....	50
Chapter 6 Conclusions and Prospect	51
6.1 Conclusions.....	51
6.2 Prospect.....	52
References	53
Research Results in the Period of Postgraduate Study	56
Acknowledgements	57

第一章 绪论

1.1 引言

近年来，随着计算机技术、通信技术、传感技术的发展，物联网受到学术界和工业界的极大关注，信息时代由“互联网时代”到了“物联网时代”^[1]。目前，物联网的研究已经覆盖了很多的领域，例如：家居设备、环境监测、交通、物流运输、医疗、教育、工农业生产等诸多领域等^[2]。车联网是物联网在交通领域的一个重要分支^[3]。

车联网利用智能传感器、通信与网络技术以及数据处理技术等相结合，实现了车与车、车与人等整合到一个整体网络^[4]。在车联网系统中，车辆上安装有各类智能传感器，传感器持续不断的采集格式复杂的各种车辆相关的数据，对这些数据的管理和分析为企业及管理者的车辆管理、市场规划提供支持，也为普通车主日常出行带来便利^[5]。一个完整的车联网包括硬件、网络通讯、车联网服务等，根据 2016 年的数据，车联网价值分布如图 1.1 所示。

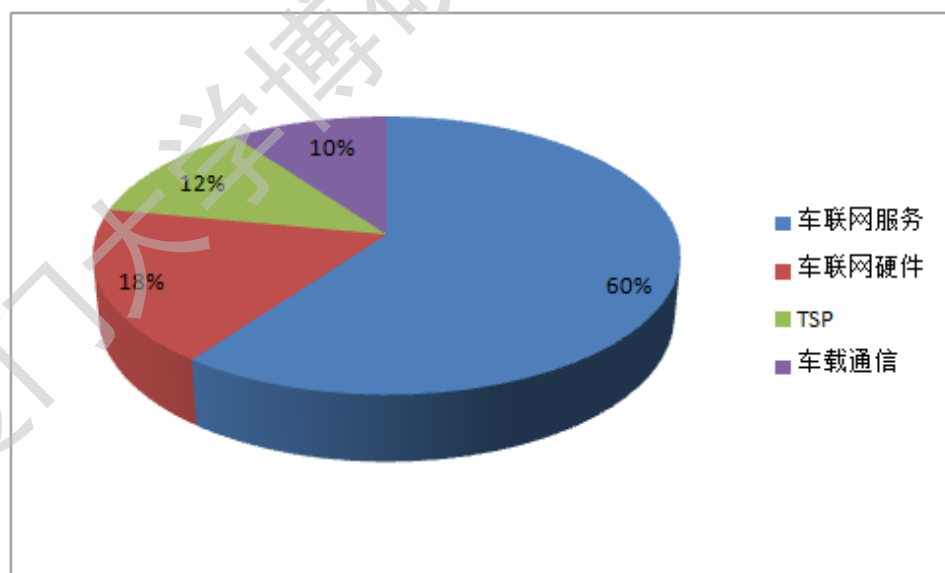


图 1.1 车联网组成价值分布

由图 1.1 可知，车联网应用服务的价值达到 60%。其中，车联网应用服务主要分为三类：

1、娱乐相关应用。利用蓝牙、显示器、网络接入等设备，为用户提供多种不同类型的娱乐相关服务，例如：Wifi 热点、音乐、电影、小游戏、语音播报、视频聊天等^[4]。

2、安全相关的应用。根据车辆速度、方向信息、转弯曲率等，提高车辆行驶安全和危险警示提醒。例如：倒车时避免碰撞、转弯时的盲点警告等。或者根据其他采集到的数据，提醒用户车辆可能存在的不易发现的车辆安全隐患，例如油量不正常、冷却液过少、漏油、刹车失灵等^[4]；

3、便利相关的应用。利用网络、定位、交互技术等，提供导航、语音播报、交通状况提醒、自动缴费、停车信息、行车规划等，以提高用户出行的便利^[5]；

根据汽车工业协会的统计，中国 2016 年的汽车销量 1379 万辆。而这些车 80%以上由 10 家生产企业生产。根据中国汽车工业协会数据，2016 年中国分车型前十家汽车生产企业销量排名表^[6]，如表 1-1 所示。

表 1-1 2016 年中国分车型前十家汽车生产企业销量排名表

排名	汽车		乘用车		商用车	
	企业名称	销量（万辆）	企业名称	销量（万辆）	企业名称	销量（万辆）
1	上汽集团	255.22	上汽集团	221.89	北汽集团	48.49
2	中国长安	171.81	中国长安	135.41	东风集团	48.10
3	东风集团	137.84	长城汽车	96.89	中国长安	36.41
4	北汽集团	135.88	东风集团	89.64	上汽集团	33.34
5	长城汽车	107.45	北汽集团	87.39	安徽江淮	27.10
6	吉利控股	79.92	吉利控股	79.92	一汽集团	22.86
7	安徽江淮	63.83	奇瑞汽车	61.13	中国重型	20.35
8	奇瑞汽车	63.60	比亚迪汽车	49.66	重庆广帆	17.15
9	一汽集团	50.18	广汽集团	37.62	华晨汽车	12.52
10	比亚迪汽车	49.66	安徽江淮	36.73	陕汽集团	11.59
10 家企业合计		1115.29	896.28		277.91	
中国品牌企业合计		1397.58	1052.86		344.72	
占品牌企业比重		79.80%	85.13%		80.62%	

随着汽车工业技术的发展,全球汽车销量不断增长,有些机构特别是大型汽车企业对车联网系统提出了更高的要求,他们急需一款为自己量身定做的车联网系统,具备实现对其管辖的车辆进行注册管理、定位监控与告警等功能。由表 1-1 可知,大部分汽车由主要十家生产企业销售,因此对于这些企业来说,有必要对于所生产运营车辆进行简单而有效的监控,把已售出车辆的实时信息传送给用户,对未售出的车辆的实际位置监控、调度等。随着统一监管的做法的普及,车载终端传回越来越多的车辆数据、GPS 数据、驾驶行为数据等等,对于车辆终端传回的数据报文进行清洗、分析、存储的需求越来越强烈^[7]。

然而,随着车联网应用圈不断丰富,面对新的需求,车联网技术受到了新的挑战^[5]。例如:现有数据不能实现自动驾驶功能,需要加入激光雷达等更为敏感的传感器;现有的数据处理方式不能满足大数据环境下车联网中车辆监控、告警等的实时性需求。因此,本文基于研发中心承担的某车联网云服务平台,针对项目中车联网数据实时处理存在的问题,改进传统数据处理方法。解决当前车联网系统中实时数据处理慢、内存积压的问题,并根据实际数据进行实验验证。

1.2 国内外研究现状

1.2.1 国外车联网发展现状

在 20 世纪 80 年代早期日本就开始对车联网进行摸索研究,后来加州的 PATH, 欧盟的 Chauffeur 等组织也使用无线通信进行了智能交通探索。而美国最早提出“车联网”的概念。

1996 年, PERRETT.K E 等人作了一个关于道路交通远程信息处理(transport telematics)潜在利益的报告,称交通远程信息处理是一个快速发展的领域,并且交通远程信息处理技术可以确保大幅度改进交通效率、安全、环境和我们的生活质量^[8]。

2000 年, Turner. J D 在对当前传感器在汽车领域和交通控制方面的实践进行了研究综述^[9]。2002 年, Sastry Duri 等人在论文中阐述了车辆远程信息处理概念,可以定义为车辆与电子通讯和计算机技术融合的信息应用,车辆远程信息处理系统本身要求捕获传感数据,数据的存储和数据交换^[10]。此时的 Telematics 系统已经是多项技术的集成,远程要求车辆上安装传感器,收集传感数据,车辆和

系统保持通讯和数据传输，系统进行数据存储、处理^[11]。

车联网的明确概念是随着物联网概念的提出而进一步形成。2005年，“物联网”概念由国际电信联盟提出^[12]。2005年，Matthias Rebel等提出，车内、车与车之间联网是将许多技术集成、整合到一个整体网络中，并且能确保有效且可持续的通讯。2006年，C. K. Toh讨论了未来无线网络在自主智能交通系统的潜在应用，如自主车队、本地智能交通导航系统、交通流和拥堵控制、追踪被盗车辆等，甚至实现“目标驱动”和“no-hassle”驾驶^[13]。2009年，美国、日本等国的车联网产品开始进入中国的市场。2010年，欧盟推广提供自动紧急呼叫等功能eCall车联网系统^[14]。2014年，美国在汽车上加入电子记录设备以减少车祸。

2015年，美国斯特拉电动车实现自动泊车。谷歌自动驾驶理论在全世界广泛讨论。2016年，欧盟HIGHTS团队研发智能交通系统，防止交通事故等。德国将激光雷达传感器等加入车辆感知层，提高自动驾驶车安全^[15]。

目前，以欧美、日韩和加拿大等为代表的发达国家在车联网，特别是车联网中的车辆监控系统领域的研发和应用实践上已经投入了巨大的人力和物力，并且取得了相当可观的成效，获得了海量有价值的数据库。

1.2.2 国内车联网发展现状

受工业化进程的影响，当前中国的车联网研究稍落后于欧美日。然而随着“互联网+”的东风和车联网市场的巨大需求，中国的车联网技术的研究以几何级的增长速度蓬勃发展^[16]。

2010年，“车联网”第一次在中国提出^[17]。中国智能交通研讨会讨论物联网技术在车联网中的应用，包括：超声检测器、交通信息采集、电子车牌等技术应用展示。

2011年，中国交通运输部提出：物联网技术是“车联网”落地的突破口^[18]。

2012年，武晓钊等人提出海量数据处理技术在车联网中的应用^[19]。

2013年，“车联网产业技术创新联盟”在北京成立^[20]。张洪超等人将云数据处理模块加到车联网系统中，将车联网系统分为三个主要模块，分别是车载智能终端、无线通信网络以及数据处理云端^[21]。

从2010年至2015年，我国车联网用户数量发展速度迅猛，5年时间用户量从350万增至1700多万，年复合增长率达到37%^[22]。

2015年,作为“互联网+汽车”的关键产业之一^[23],车联网得到越来越多的互联网企业的关注和技术研究投入。随着互联网企业不断进军车联网行业,应用从最早的车载终端诊断系统到各种车联网应用解决方案,形式越来越多种多样^[24]。

2016年,是车联网快速发展的一年,人工智能、自动驾驶等技术在车联网中的应用研究势头正猛^[24]。其中,安吉星推出“虹膜识别启动车辆”。腾讯路宝APP2.0支持驾驶员疲劳监测、偏离车道提醒。长安汽车成功完成2000km无人驾驶。搜狗地图结合语义分析技术提供全语音智能导航^[25]等等。全国最大电动汽车车联网平台在北京上线运行;摩拜单车在上海上线运行;车联网公交在青海试运行。城市智能交通、人工智能、大数据各种技术在车联网这个框架中得到了新的应用。

目前,国内大部分企业都将车联网定位在了自己熟悉的领域,而且产品面向个人用户,较少面向大型汽车企业。对于企业平台来说,随着统一监管的需要,车联网系统规模随着其所管辖车辆的规模和数据处理需求而不断增加。车辆车载终端发送的数据有增长迅速、数据类型繁多且数据格式多变的特点。而大部分汽车企业的对这些数据的处理方式还停留在传统阶段。

1.3 论文主要研究内容

车联网系统持续不断的接收海量数据,并且要对这些数据进行处理与分析,这是一个典型的大数据应用场景。前置机(或车载终端)连接车辆内部CAN(Contoller Area Network, CAN)控制总线以及各类传感器,实时收集车辆数据,例如GPS数据、CAN数据、驾驶行为数据、状态数据、摄像头数据等。数据通过无线通讯方式及时传给车联网平台,平台对这些数据进行聚集、清洗、存储、处理、分发以满足对车辆监控与告警等功能。对车联网中接收到的数据进行处理,应用领域广泛,如决策系统、交通诱导系统、实时监控、驾驶行为分析、告警提醒等。这些数据有一些需要实时分析,例如实时监控和告警提醒。

在大数据背景下,车联网数据的实时处理中的准确率和实时性是值得研究的问题。本论文主要研究海量数据实时处理在车联网系统中的应用,通过开发基于分布式技术与云计算的车联网平台,并对海量GPS数据可视化和实时数据处理进行研究以及改进。

主要包括以下研究内容:

1、实时监控数据的聚合展示

在车联网中,主要应用于车辆实时监控可视化显示,或称为标记点聚合显示,在线地图提供的 API,也很难在大数据量的情况下达到高效的显示和较好的用户体验。为了解决这些问题,论文提出了基于行政单位的 k-means 标记点动态聚合算法,在服务器端对接收到的车辆位置数据进行聚合,然后推送到客户端进行显示。

2、区域限定判断算法研究

车联网中,车辆制造商或者经销商等用户可以对其管辖的车辆进行运营区域限定,当车辆超出被限定的区域时,平台要及时告警。当运营区域用封闭多边形表示时,区域限定判断实际上是点在多边形内的判定问题。一般的射线法或者凸多边形法虽然算法简单,但是处理速度比较慢,而且受多边形形状的影响。论文结合扫描线算法,基于莫顿码编码的四叉树压缩算法实现了改进的区域限定判断算法,比其他点在多边形内判定算法速度更快且不需要太多的辅助存储。

3、最近邻服务站搜索研究

当故障车辆进入服务站或者维修站半径范围时,车联网平台标记该车,并取出该车最近故障码信息,根据故障码信息计算维修时间。判断车辆是否进入服务站半径范围分两步:首先,根据接收到的车辆位置数据搜索最近邻的服务站;其次,判断是否进入服务站。根据 GPS 数据搜索最近邻服务站,这是一个 Top-K 问题,论文研究了车联网中 Top-k 搜索存在的问题,并从数据集划分和搜索算法两方面改进了 Top-k 算法,并将改进 Top-k 算法应用到车联网中的服务站超时告警模块中。

本论文主要涉及车辆实时监控和告警管理都是对消息处理实时性要求非常高的模块。如果消息处理不及时,则监控和告警都失去意义。基于此,本文改进现有方法,使得车辆监控和告警满足实际需求。实时数据通过 ActiveMQ 消息服务器进行交互,而查询频繁、经常改变的数据保存在分布式缓存 Redis 中。

1.4 论文组织结构

本文分为六章:

第一章 论文研究背景,车联网目前国内外研究发展水平,论文的主要研究内容和各个章节的安排;

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库