

**内蒙古师范大学计算机科学技术学院**

**毕业设计（论文）开题报告**

**题 目： 基于回归算法的公交线路客流预测**

**专 业 计算机科学与技术（非师范）**

**学 生 白璐**

**学 号 20161102910**

**指导教师 李艳玲**

**日 期 2019年11月27日**

**计算机科学技术学院制**

**一、开题报告主要内容**

1．课题来源及研究的目的和意义

课题研究来自于老师推荐的天池比赛。

随着我国经济迅猛发展，城市化建设的加快，城市规模不断扩大，城市人口也不断增加，使得城市居民对交通的需求日益增多，交通拥堵日益严重，城市交通问题也日益凸显。未来学家朱立安曾指出，“作为不发达国家经济发展的条件，交通运输有着多方面的重要作用，如果经济发展的关键因素只有一个，那么它既不是文化，也不是制度和心理，而是交通运输和通讯系统。[1]”明确表明了交通运输在城市中的重要地位，更说明了进行交通规划及发展的必要性。城市公共交通是城市扩大建设和发展的重要基础设施之一，是市民进行生产和活动不可缺少的社会公共设施，也是城市重要规划和社会化生产的首要物质条件。对公交客流信息全面、准确的把握是公交管理工作的基础，它不仅为日常调度提供依据，也为现网优化提供了重要参考。[2]推行公交优先的关键在于及时的公交车辆调度和合理的公交线路规划,而公交调度和线路规划都取决于对公交线路客流量细致的剖析和精准的预测。[3]公交客流是城市公交规划、运营调度和行车组织的基础性数据[4],同时，城市公共交通也是随着客流量、路面状况以及天气情况等发生变化的市民乘车系统。为了制定科学有效并合理的发车计划，实现对现有车辆的高效管理及合理分配，解决乘客对公交公司有限车辆资源的需求在时间上的矛盾，提高城市公共交通的管理水平和服务水平，单从IC卡系统中掌握了客流信息并不能起到重要作用，更重要的是能够对客流数据以及天气等其他因素加以分析并研究人们日常出行的规律，根据客流分布的结果，提高公交运营效率，更好地为人们提供服务。

1. 国内外在该方向的研究现状及分析

国外发达国家尤为重视城市公共公交的发展，所以智能公交IC系统使用较早，主要作为一种新的收费方式和公交公司之间的收益分配，对IC卡所采集的信息在城市公交交通规划方面的应用还未见有关报道[5]。其中法国是最早使用IC卡的国家，主要用于公共汽车、地铁、火车等。在公交数据采集方面，国外发达国家已经形成了较为完整的体系，从数据采集、传输、存储以及挖掘分析这一系列过程已经十分成熟。随着数据挖掘技术的快速发展以及广泛应用，近几年数据挖掘技术越来越多地被用于公交数据的挖掘分析之中。但其研究重心放在了预测的技术和设备上，却很少研究预测方法，在实际中的应用更是少之又少。

目前我国的智能公交系统落后于国外发达国家，国内许多城市正面临着交通拥堵、能源紧缺以及环境污染等一系列严重的交通发展瓶颈问题，因此，公共交通的发展也引起了各相关部门的高度重视。尽管我国智能公交起步晚，但发展较快，国内从1994年开始使用公交IC卡收费系统。目前国内许多大中城市都在推广公交IC卡的使用，并且有些地方已经实现了IC卡的一卡多用功能。近年来，我国在公交客流预测方面向国外借鉴了不少预测模型，尽管预测结果都不够理想，但考虑到公共交通发展的严峻形势，反而越来越多的学者加入此领域的研究中。如董海洋[6]采用传统BP神经网络模型、改进的BP神经网络模型和RBF神经网络模型分别对短时公交客流进行预测并做出评价，并在预测结果中表明改进的BP模型和PBF模型的预测精度要比传统的BP模型有所提高；殷礼胜[7]研究了基于混沌算法的小波神经网络交通流量混沌时间序列预测的问题；邹巍[8]等人提出一种基于遗传算法与小波神经网络的轨道交通短时客流预测方法，利用具有全局搜索最优的遗传算法优化小波神经网络,有效的避免了神经网络易陷入局部最小值的缺陷，在分析轨道交通短时客流的特征上,利用实测数据对模型进行验证；杨志伟等人利用回归分析的方法建立各峰值区间内IC卡乘客量与总体客流量的回归方程，实现通过公交IC卡数据库获取的IC卡时段客流对该峰值区间内的总体客流；何宝民[9]在对空中交通流量短期预测模型研究中，分析和比较了多元线性回归预测模型和SVM预测模型，建立了基于多元线性回归确定组合预测权重系数的空中交通短期流量组合预测模型；杨新苗[10]等人提出了一种基于模糊神经网络的公交线路客流预测模型，并在实验中比较发现此模型具有较强的自适应性，预测结果优于常规时间序列模型AR、ARMA所预测的结果。

1. 主要研究内容

根据广州市内及广佛同城公交线路的历史公交刷卡数据，挖掘固定人群在公共交通中的行为模式。建立公交线路乘车人次预测模型，并用模型预测未来一周（20150101-20150107）每日06时至21时每小时段各个线路的乘车人次。

提供20140801至20141231五个月广东部分公交线路岭南通用户刷卡数据，共涉及近200万用户2条线路约800多万条数据记录。同时提供20140801至20150131期间广州市的天气状况信息。

1. 研究方案
2. 对数据进行整合：原始数据是200多万用户的刷卡信息800万条左右，要整合成2条线路各个时间段的客流，整合之后也就7200条左右，其中有一些缺失的天气数据，取的是前后两天的平均值
3. 构建特征：天气特征类似小雨、晴天，还有线路类型，卡类型这种按照分类进行独热编码处理，时间天数按照1-7分类，小时段按照1-24分类，站点数量，风力，温度这种直接就是数值化，再求均值除方差归一化
4. 分析编写数据模型：通过对回归算法的学习和应用编写算法，构建模型
5. 优化模型：不断实验，减小误差，提高模型的准确性
6. 进行测试：通过测试集，验证程序，将模型进一步完善

5．进度安排，预期达到的目标

预期达到的目标：尽可能完善算法模型，减小误差，提高模型预测的准确性。

6．课题已具备和所需的条件、经费

课题研究目前所具备的开发工具是Anaconda Navigator，Excel

7．研究过程中可能遇到的困难和问题，解决的措施

开发工具可能无法熟练使用，可以通过网上课程学习，多做练习，逐步掌握应用。

8．主要参考文献

[1]王华.城市交通需求特征和公共交通的定位.交通标准化[J].2004,11

1. 王炜、杨新苗等.城市公共交通方法与管理[M].北京科学出版社.2002
2. 陈奇平.基于数据挖掘的公交线路客流预测研究.海南大学.2018
3. 耿学贵.基于IC卡的公交客流时间规律研究[D].昆明理工大学.2012
4. 杨盛华.基于现代信息技术的公交客流数据采集方法及应用研究.吉林大学硕士学位论文.2009,6
5. 董海洋.公交客流实时分析与短时预测研究.大连理工大学硕士学位论文.2013,5
6. 殷礼胜.交通流量时间序列混沌特性分析及预测研究.重庆大学博士学位论文.2007
7. 邹巍、陆百川等.基于遗传算法与小波神经网络的客流预测研究[J].武汉理工大学学报(交通科学与工程版).2014,5
8. 何宝民.空中交通流量短期预测模型研究.西安理工大学硕士学位论文.2012
9. 杨新苗、王炜等.公交线路客流模糊神经网络预测模型.东南大学.2000,8