

Analysis of Bus Satisfaction Characteristics Based on Support Vector Machine

Zheng Zhou, Minghua Rao, Fei Cao

Chongqing Public Transport Holding (Group) Co. Ltd., Chongqing
Email: mwei@cauc.edu.cn

Received: Jul. 6th, 2020; accepted: Jul. 20th, 2020; published: Jul. 27th, 2020

Abstract

Public transport satisfaction determines the service level of public transport, which is the guarantee for the implementation of priority public transport development. This paper analyzes many factors that affect bus satisfaction, starting from personal attributes, travel information, bus informatization, supporting facilities and service quality, and based on support vector machine, reveals the relationship between them and bus satisfaction, and calculates the importance of these factors. SPSS Modeler was used to integrate the survey data of Nantong, and it was found that there were significant differences between the socio-economic attributes, travel purpose and bus satisfaction of the new and old urban areas. Finally, by comparing support vector machine with Logistic regression analysis model, it is found that support vector machine has better fitting ability for classification data and higher classification accuracy for samples.

Keywords

Public Transport Satisfaction, Support Vector Machines, SPSS Modeler

基于支持向量机的公交满意度特征分析

周正, 饶明华, 曹飞

重庆市公共交通控股(集团)有限公司, 重庆
Email: mwei@cauc.edu.cn

收稿日期: 2020年7月6日; 录用日期: 2020年7月20日; 发布日期: 2020年7月27日

摘要

公交满意度决定了公交的服务水平, 是实施优先公交发展的保障。本文剖析影响公交满意度的众多因素, 从个人属性、出行信息、公交信息化、配套设施和服务质量等五个方面出发, 基于支持支持向量机, 揭

示它们与公交满意度之间关系，并计算这些影响因素的重要性程度。使用SPSS Modeler对南通调查数据进行整合，发现新城区和老城区的社会经济属性、出行目的和公交满意度具有显著地差异。最后，将支持向量机与Logistic回归分析模型相对比，发现支持向量机对于分类数据有较好的拟合能力，对样本分类的准确率也更高。

关键词

公交满意度，支持向量机，SPSS Modeler

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

优先发展公交是缓解交通拥堵十分有效的手段之一。公交满意度决定公交的服务水平，影响其高低的因素众多，与个体自身所具有的社会经济特性、出行属性及配套设施、服务相关，是一项动态、复杂的活动。因此，亟待研究公交满意度的变化特征及其原因，在对其影响因素分析基础上，揭示公交满意度的内外因素对其影响程度，为如何调整公交服务水平提供科学的依据。辅以政策引导居民公交出行，交通拥堵就会得到缓解。

目前，国内外众多学者关注公交满意度研究，部分学者研究如何建立科学全面实用的乘客满意度评价指标体系，主要采用因子分析方法[1]、四分图模型法[2]、AHP-DEMATEL法[3]；在此基础上，部分学者采用不同评价模型对它们进行综合评估，包括：巩丽媛[4]基于非集计模型构建公交乘客满意度模型，陈明伟[5]、张磊[6]及刘亚坤[7]利用模糊综合评定法对整个城市公交线路逐条分析，欧阳剑[8]基于朴素贝叶斯分类法建立乘客总体满意度模型，霍月英[9]利用 Logistic 分析现在住房位置和长椅的可用性对顾客满意度的影响很大，李琼[10]、胡晓伟[11]等从乘客满意度的角度出发，利用结构方程模型进行评价。由上可知，支持向量机是一种专门研究有限样本情况下机器学习方法，在解决小样本、非线性及高维模式识别中表现出特有的优势，该方法尚未被应用在公交满意度的特征分析领域。此外，现有公交的满意度研究尚未深入揭示居民个人属性与公交信息化、配套设施和服务质量之间关联性，从而不能全面改善公交系统。

综上所述，本文在剖析公交满意度的影响因素基础上，从个人属性、出行信息、公交信息化、配套设施和服务质量等五个方面出发，基于支持向量机，挖掘这些影响因素和公交满意度之间关联性。以南通市为例，利用 SPSS Modeler 分析公交满意度的这些影响因素的重要性程度，给出了相关政策分析。最后，围绕公交满意度的分类正确率，将支持向量机与 Logistic 回归分析模型相对比，从而验证模型的有效性。

2. 数据来源及描述性统计

以 2013 年南通新、老城区公交满意度调查问卷为例，从个人特征及出行特征、居民对公交系统的感受两个视角出发，包括出行活动、交通信息化水平、附属设施、服务质量等 22 个指标总共收集 1216 个有效样本，其中新城区：669 个；老城区：547 个。新、老城区的出行特征总体上相似，它们差异在于新城区的乘客年轻化、高学历化，这对新、老城区公交满意度有一定的影响。因此，应该区分新老城区居民的公交满意度差异，据此改善公交服务水平。

3. 基于支持向量机的公交满意度分析

支持向量机(Support Vector Machine, SVM)定义为特征空间上间隔最大的线性分类器,从现有样本数据出发,运用统计学的方法,对样本数据规律进行学习,寻找能够将全部训练样本点正确分类的最优分类面,同时保证距离该分类面最近的样本点与其间隔最大。

由上可知,公交满意度的特征分析主要围绕揭示个人基本属性、出行活动、交通信息化水平、附属设施和服务质量五方面与满意度评价之间关联性,输入变量 x 是五个方面涉及 21 个影响因素,决策变量 y 是公交满意度价,将问卷样本数据离散化后,将支持向量机应用挖掘它们之间因果关系过程如下:

已知观测样本为 $[x_i, y_i]$, 首先进行线性分类,即找到一个最优分类平面 $\omega \cdot x + b = y$, 其中 $\omega, b \in R$ 待定,满足如下条件:

$$\begin{cases} (\omega \cdot x_i) + b = 1, y_i = 1 \\ (\omega \cdot x_i) + b = 0, y_i = 0 \\ (\omega \cdot x_i) + b = -1, y_i = -1 \end{cases}$$

要使样本总体到分类面的距离最大,则有

$$\max \frac{2}{\|\omega\|} \Rightarrow \min \frac{1}{2} \|\omega\|^2$$

公交系统满意度问题属于非线性可分问题,而目前尚无能够直接处理非线性数据分类的 SVM 算法,我们需要将原来的坐标空间进行变换,应用映射函数将其变为新的坐标空间,以便可以使用线性支持向量机方法进行分类。要将非线性空间映射为线性空间,从技术上就需要利用核函数的支持。

核函数算法实现如下:若在原始空间中的简单超平面不能得到满意的分类效果,利用核函数将输入样本空间映射到高维特征空间,令:

$$K(x_i, x_j) \leq \phi(x_i) \cdot \phi(x_j)$$

用核函数 $K(x_i, x_j)$ 代替最优平面中的点积 $x_i^T x_j$, 就相当于把原特征空间变换到了某一新的特征空间,此时优化函数变为:

$$Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

而相应的判别式则为:

$$g(x) = \text{sgn} \left[(w^*)^T \varphi(x) + b^* \right] = \text{sgn} \left(\sum_{i=1}^n a_i^* y_i K(x_i, x) + b^* \right)$$

应用核函数将数据的非线性纬度变换为线性纬度后,再使用前面的线性支持向量机方法对数据进行分类。SVM 中几种常见的核函数包括线性核函数、多项式核函数、径向基核函数、Sigmoid 核函数(两层神经网络)等。核函数决定了支持向量机的性能,核函数的确定是建立在样本的基础上,同时受到各参数的影响。如表 1 所示,针对四种常见的核函数,基于问卷调查数据,建立基于不同核函数的公交系统满意度预测模型,确定核函数中参数,检验模型的预测效果,以回代训练样本准确率和预测检验样本准确率来综合评价预测模型的性能,并最终确定合适的核函数作为预测模型的核函数。

Table 1. Test results of different kernel functions**表 1.** 不同核函数测试结果

核函数名称	训练集准确率	测试集准确率
线性核函数	80.98%	82.05%
多项式核函数	100%	79.49%
径向基核函数	100%	83.76%
Sigmoid 核函数(两层神经网络)	77.16%	82.91%

由此可见，径向基核函数具有学习性能好，偏差小的优点。在这里我们该核函数作为核函数，其核函数公式为：

$$K(x_i, x_j) = \exp\left\{-\frac{\|x_i - x_j\|^2}{\sigma^2}\right\}$$

4. 数据分析与对策

4.1. 结果分析

利用 SPSS Modeler，采用 SVM 模型对调查问卷数据进行分析，具体情况如表 2 所示，研究表明：影响公交满意度评价的因素重要性大于 5%，从高到低依次为驾驶员服务态度、公交专用道利用率、公交站牌标识清晰度、车内卫生、车内治安、乘客从事职业、站台上下车方便程度和实际步行站点时间，总体情况如下：

1) 乘客的职业、月收入、性别、教育状况等个人情况对公交的满意度均有较大的影响，但是这些因素在客观上难以改变，需要乘客从自身出发。

2) 驾驶员服务态度最为重要，其重要性为 12.92%，为提高满意度公交公司需要提高驾驶员素质，对驾驶员开展定期培训，同时制定严格的奖惩措施用以规范驾驶员的言行。

3) 公交专用道利用率、站台上下车方便程度和实际步行站点时间也是影响公交满意度的重要因素，重要性分别为 8.74%、5.64%和 5.54%，这些因素体现了公交设施的设置合理性的影响。交通部门可以在公交专用车道的设置以及公交站点的设置方面做适当的调整，使乘车更加便捷从而提高满意度。

4) 随着科技的不断发展，乘客对于车辆信息的获取也有了越来越高的要求，不仅要求公交站牌标识清晰，车辆信息准确、可靠，还需要为乘客提供一些便捷的信息获取方式，如现在已投入使用的公交实时到站的电子站牌及公交 APP 等工具。

5) 除此之外，乘客对于车内的卫生以及治安情况也较为重视，公交公司也需要加强车内管理，努力为乘客营造一个良好、舒适的乘车环境。

Table 2. Importance of factors influencing public transport system satisfaction**表 2.** 影响公交系统满意度各因素重要性

因素	重要性	老城区	新城区	因素	重要性	老城区	新城区
驾驶员服务态度	0.1292	0.0451	0.0662	公交专用道利用率	0.0874	0.0244	0.042
公交站牌标识清晰度	0.0789	0.037	0.0309	车内卫生	0.0755	0.0363	0.0743
车内治安	0.0733	0.0289	0.0525	从事职业	0.0711	0.0302	0.0175
站台上下车方便度	0.0564	0.0284		实际步行站点时间	0.0554	0.0727	0.0219

Continued

一次换乘次数	0.0496	0.0376	0.0214	线路信息获取难易	0.0496	0.0266	0.0149
线路信息获取方式	0.0417	0.1924	0.0334	月收入	0.0388	0.0322	0.0412
公交电子站牌利用率	0.0367	0.0308	0.0746	性别	0.0348	0.0384	0.0486
一周乘坐次数	0.0314	0.0545	0.0382	车内设施	0.0275	0.0485	0.0728
居住地	0.0271			旅途乘车时间	0.0208	0.0373	0.0559
教育状况	0.0159	0.1137	0.0549	站点实际候车时间	0.0007	0.0287	0.0602

此外，新、老城区公交满意度的影响因素重要性也存在差异，如图 1 所示，具体表现在：

- 1) 线路信息方式以及实际步行至站点时间等因素对老城区的影响更大，这主要取决于老城区年长人口较多，对于乘车的便捷程度更为重视，交通部门及公交公司需要从线路和站台设置方面入手。
- 2) 车内设施、车内卫生、车内治安以及高峰拥挤等对新城的影响更大，主要取决于新城青年人口较多，对于公交的乘车体验更为重视，交通部门及公交公司需要改善车辆状况和车内环境，合理安排车次。

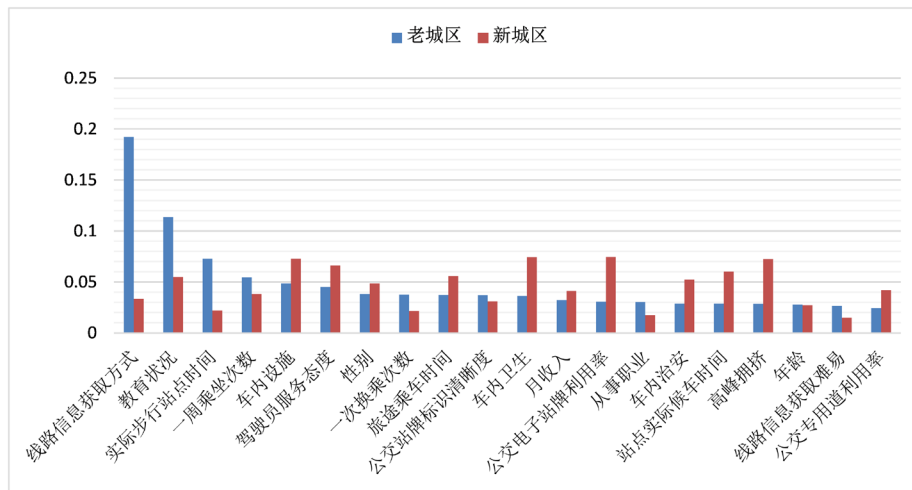


Figure 1. Comparison of the importance of each factor between new and old urban areas
图 1. 新、老城区各因素重要性比较

4.2. SVM 模型与 Logistic 回归算法模型比较分析

将支持向量机和 Logistic 回归算法进行比较分析，二者的模型分类正确率如表 3 所示，从中可知：不管是支持向量机还是 Logistic 回归分析算法，对公交满意度样本数据都起到了很好的分类作用；但是支持向量机的分类正确率高于 Logistic 回归分析算法，特别是在对训练样本的分类上，支持向量机有着很大的优势。

Table 3. Classification accuracy of SVM model and logistic regression algorithm model

表 3. SVM 模型与 Logistic 回归算法模型分类正确率

算法	支持向量机		Logistic 回归算法	
	训练	测试	训练	测试
正确	100%	83.76%	80.35%	81.2%
错误	0	16.24%	19.65%	18.8%

5. 总结

本文利用支持向量机挖掘公交满意度的变化特征及其原因, 在对其影响因素分析基础上, 以南通市的问卷调查数据分析为对象, 揭示公交的满意度与驾驶员服务态度、公交专用道利用率、公交站牌标识清晰度等因素有较大的联系, 对比了新城区和老城区之间各因素如何影响公交满意度的差异, 并给出了相关对策分析。最后, 通过将支持向量与 Logistic 回归算法的比较, 从而验证模型合理性。本研究工作为如何调整公交服务水平提供科学的依据, 辅以政策引导居民公交出行, 交通拥堵就会得到缓解。

参考文献

- [1] 胡国兵, 孙倩, 毛瑞婷, 段晓达. 公交乘客满意度评价体系的探讨[J]. 江苏商论, 2011(18): 96-98.
- [2] 苟蕾, 汪文俊, 唐晓龙, 石连生. 海口市公交乘客满意度统计分析[J]. 海南师范大学学报(自然科学版), 2014(2): 138-142.
- [3] 彭晗, 吴金妹, 姚林晓. 基于 AHP-DEMATEL 法的公交换乘系统乘客满意度模糊评价[J]. 重庆交通大学学报(社会科学版), 2014(1): 19-22.
- [4] 巩丽媛. 基于乘客满意度模型的城市公交服务评价与改进系统研究[D]: [硕士学位论文]. 济南: 山东大学, 2011.
- [5] 陈明伟. 城市公交系统满意度模糊多目标综合评价研究[D]: [硕士学位论文]. 成都: 西南交通大学, 2006.
- [6] 张磊. 基于模糊数学的城市公交服务满意度评价研究[J]. 交通与运输(学术版), 2014(2): 98-102.
- [7] 刘亚坤. 乌鲁木齐市常规公交乘客满意度评价模型[J]. 交通科技与经济, 2015(1): 32-34.
- [8] 欧阳剑, 胡郁葱, 李橘云, 李健行. 公交乘客满意度相关因素的影响程度分析[J]. 深圳大学学报(理工版), 2016(3): 288-292.
- [9] 霍月英, 李文权, 陈茜. 快速公交乘客满意度建模研究——以常州快速公交为例(英文) [J]. Journal of Southeast University (English Edition), 2016(2): 233-239.
- [10] 李琼, 马兴慧, 朱顺应. 基于 SEM 的公交乘客满意度测评模型[J]. 物流技术, 2016(10): 78-81+97.
- [11] 胡晓伟, 魏艳波. 基于 DEA 的寒区城市冬季公交线路服务满意度评价[J]. 北京工业大学学报, 2015(10): 1566-1573.