

Characteristics Analysis of Bus Satisfaction Based on Bayesian Network

Tiandong Liu, Fen Chen, Bo Zhao

Wuhan Planning and Design Co. Ltd., Wuhan Hubei
Email: mingtian911@163.com

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

Abstract

Public transport satisfaction is an important index to measure the rationality of public transport system design and management. This paper starts from personal attributes, travel information, public transport information, supporting facilities and service quality, and based on Bayesian network, reveals the relationship between them and public transport satisfaction, and calculates the importance of these influencing factors. SPSS Modeler was used to analyze the survey data of Nantong's new and old urban areas, and it was found that there were significant differences in their socio-economic attributes, travel purpose and bus satisfaction, and the improvement schemes for different urban areas were proposed. Finally, by comparing bayesian network with TAN, Markov and other models, it is found that the model has better fitting ability for classification data and higher accuracy for sample classification.

Keywords

Public Transport Satisfaction, Bayesian Networks, SPSS Modeler

基于贝叶斯网络的公交满意度特征分析

刘天栋, 陈 芬, 赵 波

武汉市规划设计有限公司, 湖北 武汉
Email: mingtian911@163.com

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

公交满意度是衡量公交系统设计和合理性的管理的重要指标。本文从个人属性、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, 基于贝叶斯网络, 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对南通新城区和老城区调查数据进行分析,发现它们的社会经济属性、出行目的和公交满意度具有显著地差异,提出了针对不同城区的改善方案。最后,将贝叶斯网络与TAN、Markov等模型相对比,发现该模型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, 贝叶斯网络, 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

如何科学提高公交的服务水平,将居民出行从私家车转向公交,是解决城市交通问题的有效途径。城市居民是否优先选择公交出行,主要取决于公交的满意度高低,其因素涉及居民自身的基本属性、公交系统状态、车内状态、车外状态、交通信息化程度和运营状态等各个方面。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。

目前,国内外众多学者关注公交满意度研究,主要如下:巩丽媛[1]运用非集计模型理论建立公交乘客满意度模型,采用模糊评价方法对济南公交服务进行评价。胡国兵[2]等采用四分图模型,从时间调控、线路安排、服务质量等方面的满意度进行打分测评,从而建立公交满意度的评价体系。陈明伟[3]利用层次分析法和模糊综合评定法,对整个城市公交线路逐条分析,评价出城市公交系统的满意度。彭晗[4]等在分析乘客需求特征的基础上,根据指标体系设计原则,构建了科学全面实用的乘客满意度评价指标体系。苟蕾[5]等运用交叉列联表的独立性检验,因子分析方法,得到影响乘客满意度的五个主要因子。张磊[6]及刘亚坤[7]均以乌鲁木齐市公交运营线路为研究对象,使用模糊数学法对调查结果进行分析,对常规公交服务水平进行合理定量评判。胡晓伟[8]等从乘客满意度的角度出发,利用结构方程模型进行评价指标的筛选,从公交运营者和出行者两方面确定评价模型变量。欧阳剑[9]等利用朴素贝叶斯分类法分析调查数据,建立候车时间、步行时间和乘车环境满意度等9个属性与乘客总体满意度之间的满意度模型。霍月英[10]等建立一个快速公交系统的顾客满意度模型,以有序 Logistic 回归作为建模方法,发现在住房位置和长椅的可用性对顾客满意度的影响很大,乘客的社会经济特征对顾客满意度有很大的影响。李琼[11]等利用结构方程模型进行数据分析,建立由感知质量、感知价值和乘客期望等潜变量构成的结构方程模型。由上可知,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,并且已有研究侧重量化这些影响因素与公交满意度之间关联性,没有深入刻画数据不完备与它们之间状态转移耦合关系,与实际应用存在一定差异。

贝叶斯网络是表示变量之间的概率联系的网络模型,将该方法引入公交系统满意度评估中,它能在不确定的因素下利用可视化的图形进行表示影响因素与满意度之间状态转移关系,从而达到预测和决策的目的,不仅能够真正有效地处理不完整的数据,同时它能有效地避免对数据的过度拟合。综上所述,本文在分析公交满意度的影响因素基础上,从出行居民个人状况、居民出行的基本情况、交通信息化水平、公交附属设施和公交服务质量这五个方面出发,基于贝叶斯网络挖掘影响因素和公交满意度之间关联性,并通过比较贝叶斯网络与其它模型分类正确率,从而验证模型的有效性。

2. 数据来源及描述性统计

数据来自南通市人民政府于 2013 年委托地方院校调查新城区和老城区公交满意度所得问卷。调查问卷设计如表 1 所示, 调查内容包括个人特征及出行特征、居民对公交系统的感受两部分, 涉及个人基本属性、出行活动、交通信息化水平、附属设施、服务质量以及满意度评价六方面, 总共 22 个调查指标。

Table 1. Questionnaire design and basic attributes

表 1. 调查问卷设计及基本属性

类别	因素	取值范围
个人状况	性别	男、女
	年龄	18岁以下、18~25岁、26~35岁、36~45岁、46~60岁、60岁以上
	教育状况	小学及小学以下、初中、高中、中专、大专及以上
	职业	学生、工人、公务员、事业单位职员、公司职员、私营个体、无固定职业及离退休人员、其他
	收入	无收入、2000元以下、2000~5000元、5000~10,000、10,000元以上
出行状况	一周内乘坐次数	5天或以上、3~4天、1~2天、好几周才、一次、其他
	换乘次数	0次、1次、2次、3次、3次以上
	实际步行时间	0~5分钟、5~10分钟、10~20分钟、20分钟以上
	实际候车时间	0~5分钟、5~10分钟、10~20分钟、20分钟以上
交通信息化水平	实际乘车时间	0~15分钟、15~30分钟、30~45分钟、45分钟以上
	线路信息获取难易	非常容易、一般、不容易
	线路信息获取方式	朋友咨询、网络咨询、公交查询系统、电话相关部门、其它
	电子站牌利用率	非常大、一般、不大
公交附属设施	站牌清晰程度	非常清晰、一般、不清晰
	公交专用道利用率	高、中、低
	站台方便程度	高、中、低
公交服务质量	驾驶员服务态度	很满意、一般、不满意
	车厢设施	很满意、一般、不满意
	车内卫生	很满意、一般、不满意
	车内治安	很满意、一般、不满意
公交系统满意度	高峰拥挤	很满意、一般、不满意
	总体评价	很满意、一般、不满意

通过对调查数据校对核验和筛选后, 最终获得了 1216 个有效个体样本, 其中新城区有效个体样本 669 个, 老城区有效个体样本 547 个(图 1)。对样本进行整理分析后, 发现新城区和老城区所获得的样本

在总体上保持着高度的一致性，但在某些因素方面也存在着一一定的差异，如：新城区乘客较老城区相比向年轻化发展，大多集中在 18~25 岁；同时，新城区乘客学历较老城区有了较大的提高，有 70% 的乘客学历集中在大专及大专以上。这些属性属于乘客的个人状况的因素，对于公交系统满意度也有着一定的影响。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。

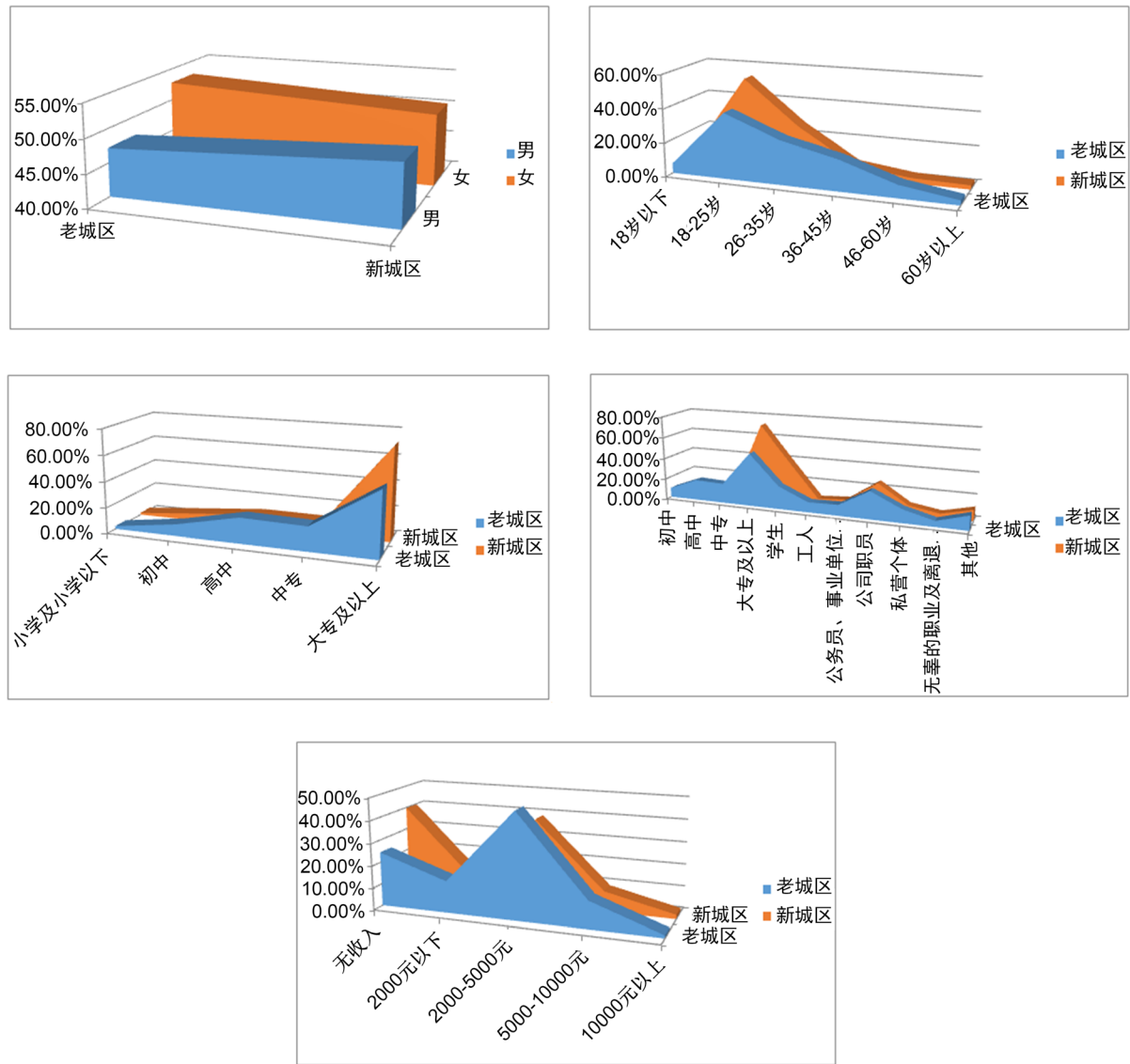


Figure 1. General characteristics of residents in new and old urban areas
图 1. 新老城区居民的总体特征

3. 公交满意度特征分析的贝叶斯网络模型构建

根据贝叶斯网络的基本原理，构建公交满意度特征分析模型的核心在于节点变量定义、确定各条件和决策变量的取值范围及先验概率分布、结构学习和参数学习四部分，主要步骤如下。

步骤 1：变量节点定义。共有 $n+1$ 个节点变量 $X = \{X_1, X_2, \dots, X_n\} \cup \{X_{n+1}\}$ ，分为 n 个条件和 1 个决策变量节点，其中条件变量为 22 个调查指标，决策变量是公交满意度的调查结果。

步骤 2: 汇总 n 个条件和 1 个决策变量节点的取值范围及它们之间先验概率分布。

步骤 2.1: 查找任意变量 $\forall X_i \in X$ 的取值范围 $Val(X_i) = [X_{\min}^i, X_{\max}^i]$, 根据专家知识, 对之进行离散化处理, 共有 K 个特征取值状态空间为 $(x_1^i, x_2^i, \dots, x_K^i)$, 其中: $x_j^i = X_{\min}^i + (j-1) * [X_{\max}^i - X_{\min}^i] / K$ 。

步骤 2.2: 计算变量 $\forall X_i \in X$ 取值状态 x_j^i 的概率 $p(x_j^i) = Count(x_j^i) / \sum_{j=1}^K Count(x_j^i)$, 以及两个变量 $\forall X_i \in X$ 和 $\forall X_j \in X$ 取它们不同状态值之间的先验概率分布 $p(x_j^i / x_j^j) = Count(x_j^i) / Count(x_j^j)$, 其中 $Count(\bullet)$ 表示事件 \bullet 在样本集 D 出现的次数。

步骤 3: 构建公交满意度特征分析的贝叶斯网络各节点之间拓扑结构, 描述其涉及个人基本属性、出行活动、交通信息化水平、附属设施、服务质量以及满意度评价六方面和公交满意度之间因果关系。

步骤 3.1: 采用 K2 算法, 在训练集上进行无监督的机器学习, 得到初始网络结构;

步骤 3.2: 利用专家的先验知识, 基于条件独立性检验方法, 若任意两个结点 X_i 和 X_j 之间相互依赖, 存在有向边相连接 $p(X_i / X_j) \neq \emptyset$, 对网络结构进行微调。

步骤 3.3: 检测获得调整后的网络结构是否符合要求, 若满足要求, 输出公交满意度特征分析的贝叶斯网络结构图 S ; 否则返回步骤 2.3.2, 继续微调网络结构。

步骤 4: 在上述网络结构基础上, 利用最大似然估计方法, 估计公交满意度的影响因素和评价结果之间条件概率分布表。

步骤 4.1: 将先验分布和似然函数相结合, 估计参数 $\theta = \prod_{i=1}^n p(X_i | X_{n+1})$ 。

步骤 4.2: 令 θ 的分布为 Dirichlet 函数, θ 的似然函数为

$$L(\theta | X) = P(X | \theta) = \prod_{k=1}^{r_i} \prod_{i=1}^n P(X_i | \theta_i) = \prod_{i=1}^n \prod_{j=1}^K P(x_j^i | \theta_{ij}), \text{ 又由于 } L(\theta_{ij} | X) = P(X | \theta_{ij}) = \prod_{k=1}^{r_j} \theta_{ijk}^{N_{ijk}}, \text{ 根据}$$

$\frac{\partial(L(\theta | X))}{\partial \theta} = 0$, 计算 $\theta = N_{ijk} / \sum_{k=1}^{r_j} N_{ijk}$, 其中: r_i 为 X_i 的取值数目; N_{ijk} 为节点 X_i 取值第 k 个值时, 在父节点取值第 j 个值的数量。

步骤 4.3: 根据上述公式推导过程, 可计算它们之间条件概率。

$$p(X_{n+1} / X_1, X_2, \dots, X_n) = \frac{p(X_{n+1}) \times \prod_{i=1}^n p(X_i | X_{n+1})}{p(X_1, X_2, \dots, X_n)}$$

在上述基础上, 根据预测客流或行驶时间的概率高低, 预报公交环境变化, 并进行反向推理其外部环境可能的变化原因, 为公交动态调度提供决策数据支撑。

4. 实证分析与政策反馈

采用 SPSS Modeler 的贝叶斯网络工具箱对调查问卷数据进行分析, 选择合适的贝叶斯网络模型, 计算影响公交满意度的重要性程度, 得出这些影响因素与公交满意度之间概率推理关系, 为公交系统改善提供政策反馈。

4.1. 模型选取

贝叶斯网络模型的种类众多, 选取 TAN 算法、Markov 算法和 Markov-FS 算法三种模型进行评估进而选择最佳预测模型, 由图 2 可知: TAN 模型与 Markov、Markov-FS 模型都生成了相似的结果, 从训练数据来看 TAN 模型具有较好的准确率, 从测试数据来看 Markov 模型具有很好的准确率, 而

Markov-FS 算法模型具有更高的预测准确率。因此，本文选取 Markov-FS 分析公交满意度，具体的正确率结果如表 2 所示。

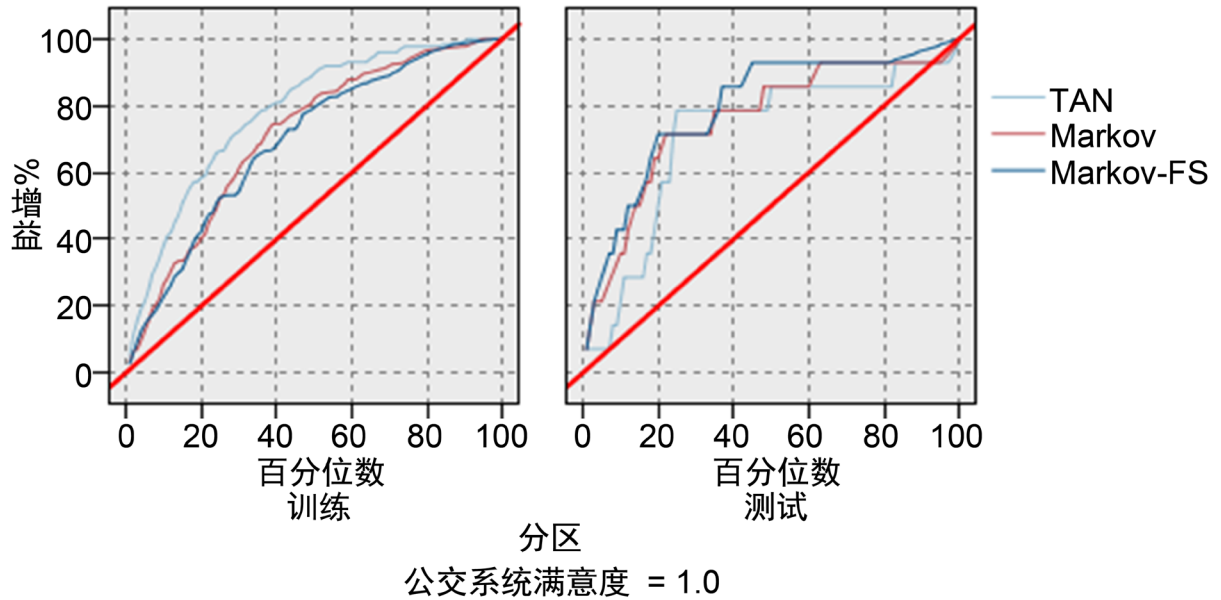


Figure 2. Comparison of prediction accuracy of each model
图 2. 各模型预测准确率比较图

Table 2. Prediction accuracy and consistency of each model
表 2. 各模型预测准确率及一致性

	TAN		Markov		Markov-FS		一致性	
	训练	测试	训练	测试	训练	测试	训练	测试
正确率	81.62%	75.21%	78.53%	80.34%	76.89%	83.76%	81.26%	87.18%
错误率	18.38%	24.79%	21.47%	19.66%	23.11%	16.24%	18.74%	12.82%

4.2. 影响公交满意度的因素重要性程度分析

如图 3 所示，公交满意度的影响因素重要性程度从高到低前 4 名依次为公交电子站牌利用率、公交专用道利用率、站台上下车方便度、车内设施以及车内卫生，计算结果符合直观分析，因为：

1) 城市电子公交站牌投入使用以来为乘客提供了大量的便利，给乘客合理安排乘车提供了基础，因此公交电子站牌利用率成为影响公交系统满意度的重要因素。

2) 公交专用道利用率是影响公交满意度的重要因素，重要性高达 26.51%。站台上下车便捷程度也起到了重要作用，其重要性为 17.3%。

3) 车内卫生以及车内设施也是影响乘车体验的重要因素，公交公司要提高自己的服务意识，紧抓卫生，组织公交驾驶员的培训活动，及时更新车内设施，从而更加直接的提高满意度。

此外，由于南通市新老城区居民组成差异较大，故它们关注的公交满意度影响因素也稍微不同，图 3 中的两者公交满意度的因素重要性差异决定了改善公交满意度的政策反馈也应该稍微差异，主要如下：

给出了针对不同城区的解决措施。最后,将 TAN 作为贝叶斯分类器的情况和将 Markov 相结合的情况作了比较,从而验证模型合理性。本研究可以为交通管理部门以及公交公司等提出合理的决策方案,对提高公共交通选择率,缓解城市交通压力有较大的帮助。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [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] 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.
- [9] 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.
- [10] 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.
- [11] 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.