

基于归纳式学习法的通勤交通满意度指标研究

赵明宇^{1,2}, 孙立军¹, TYLER Nick², 兰 成³

(1. 同济大学 交通运输工程学院, 上海 200092; 2. 伦敦大学学院 土木环境与地理信息系, 英国 伦敦 WC1E 6BT;
3. 威尼斯 IUAV 大学 建筑学院, 意大利 威尼斯 30135)

摘要: 以评价通勤者对交通系统满意度为目的, 建立通勤交通满意度指标. 以问卷形式采集数据, 采用归纳式学习方法, 分析通勤者满意度等级与通勤时间之间的相关信息, 得到以大于 0.5 证据权的出行时间区间为代表值的满意度指标. 其中, 通勤时间区间 10~30 min 为出行者最满意的通勤时间代表值. 此结论从一个角度说明交通出行行为具有“可达性”与“移动性”的双重价值.

关键词: 通勤交通; 满意度指标; 归纳式学习

中图分类号: U 121

文献标识码: A

Development of Satisfaction Index of Commute Travel Based on Inductive Learning

ZHAO Mingyu^{1,2}, SUN Lijun¹, TYLER Nick², LAN Cheng³

(1. College of Transportation Engineering, Tongji University, Shanghai 200092, China; 2. Department of Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6 BT, United Kingdom; 3. Department of Architecture, University IUAV of Venice, Venice 30135, Italy)

Abstract: A satisfaction index of commute travel was produced in order to evaluate the extent to which the commuters were satisfied with transportation system. Based on inductive learning, an analysis of the survey data was made to identify the relationship between commuters' satisfaction and commute time. The final satisfaction index was conducted with travel time intervals of weight of evidence above 0.5 as indicator. Commute time between 10 and 30 minutes is the indicator of most satisfactory commute time, which proves to some extent that travel has both values of “accessibility” and “mobility”.

Key words: commute travel; satisfaction index; inductive learning

城市交通系统在发展过程中遇到了越来越多亟待解决的问题, 受到了学术界及社会各界的广泛关注. 许多城市都面临供需矛盾所造成的交通拥堵、环境污染与能源消耗、交通安全等严峻挑战. 从 20 世纪 80 年代起, 随着计算机能力的不断提高及经济模型的发展, 交通领域研究的热点转向交通行为. 人们意识到, 要使得以解决交通各种负面问题为目的的政策及技术性措施得到真正有效的实施并达到预期的效果, 首先必须充分了解出行者的心理过程及感受. 一部分研究者开始探讨主观因素与客观交通行为的相互影响, 如 Tardiff 提出出行选择行为不仅受到意识和态度的影响, 而且对其也具有反作用^[1], Golob 由此对传统的叙述性偏好决定显示性偏好的假设提出质疑^[2]. 另一方面, 交通出行本身的价值受到了关注, Mokhtarian, Redmond, Ory 等人的研究成果表明, 交通出行不仅仅是为了到达目的地, 而且包含了可达性与移动性两种价值^[3-6]. 这体现在对各国出行数据的调查和统计中. 在过去的几十年里, 个人每天平均的出行时间稳定在 1 h 左右^[7-9], 虽然在这一时期交通工具得到了飞速发展, 出行效率显著提高. 这与效应理论中将出行时间视为“负效应”、人们出行的选择行为都是为了最小化负效应的假设相矛盾. 由此可以看出, 出行时间不仅从微观角度能够反映个体出行者对出行的感受, 从宏观角度反映交通系统的服务水平, 还提供了一个稳定的视角来评价交通系统的性能.

对交通系统的评价可以从两个角度入手, 一是客观, 二是主观. 本质上来说, 这是一个问题的两个方向, 如图 1 所示. 交通系统越是充分满足出行者的需求, 出行者对交通系统的满意度就越高. 关于出行者对于交通系统满意度的研究较少, 主要集中于对

随着城市规模的不断扩大、交通量的不断增加,

收稿日期: 2010-05-23

基金项目: 国家自然科学基金(60804048)

第一作者: 赵明宇(1983—), 女, 博士生, 主要研究方向为交通系统分析评价. E-mail: 7my_zhao@tongji.edu.cn

通讯作者: 孙立军(1963—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为道路工程、交通系统分析评价. E-mail: ljsun@tongji.edu.cn

公交服务的评价研究. 如 Stuart 对纽约地铁服务满意度指标的研究^[10], Stradling 直接采用经济学的消费者满意度体系六步法评价公共交通^[11]. 以上方法均是基于经济学中消费者对商品和服务的满意度指标体系, 对评价体系主体没有任何改动, 只针对公共交通系统特点, 研究主客观指标. 评价数据的收集均依赖于问卷形式的调查, 问卷体系较复杂^[10-12]

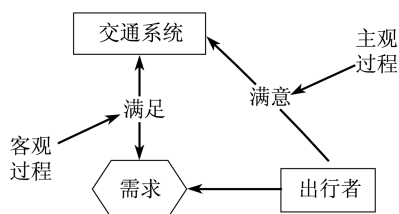


图 1 交通系统评价的主客观过程

Fig. 1 Subjective & objective process of evaluation of transportation system

在诸多交通出行目的中, 以通勤为目的的交通一直是交通领域的研究重点. 其原因可以总结为三点: ①以通勤为目的的交通量占总交通量的比例较高(英国 2008 年调查数据显示: 通勤交通次数占交通出行总次数比例为 27%, 通勤交通公里数占交通总公里数 23%^[13]); ②由于通勤交通的时间较为集中, 造成城市交通的早晚高峰, 给城市交通系统带来巨大压力; ③其他目的的交通出行行为总是围绕着通勤交通展开, 即通勤交通构成了交通出行选择行为的主干^[5].

综合以上原因, 笔者选择以出行时间为指标、以通勤交通为出行目的, 研究通勤交通满意度指标, 从主观上评价交通系统性能.

1 归纳式学习基本理论

归纳式学习方法在 20 世纪 80 年代开始应用于专家系统的信息获取及模式识别, 取得了广泛的发展与应用. 在专家系统的建立过程中, 属性与类之间关系的建立是关键之一. 归纳式学习方法的一个主要应用是通过对属性样本连续值的观测, 根据一定的检验原则, 将属性值连续区间划分为有序类别; 通过建立属性类之间的对应规律, 有效分类新的观测值. 其中使用较为广泛的方法是由加拿大滑铁卢大学 Andrew Wong 教授主持的模式识别与机器智能研究室(pattern analysis and machine intelligence research group, PAMI research group)开发的基于熵最大化原则的聚类离散化归纳式学习算法^[14]. 一个

由 M 个样本点组成的集合, 每个样本点属于事先指定的 S 个类别中的一类 c_s , 同时样本点具有属性 A_j , 将 A_j 的值域划分为 j 个属性区间 L_j , B_j 表示边界值集合, $B_j = \{e_0, e_1, \dots, e_{L_j}\}$. 用 q_{sr} 表示观测样本中属于类 c_s 、属性值 A_j 属于区间 $[e_{r-1}, e_r]$ 的样本量. A_j 的值属于区间 $[e_{r-1}, e_r]$ 的概率估计为

$$P(C = c_s, A_j \in e_r) = P_{sr} = q_{sr}/M \quad (1)$$

同样, 可以得到边际概率的估计值

$$P(C = c_s) = P_{s+} = q_{s+}/M \quad (2)$$

$$P(A_j = e_r) = P_{+r} = q_{+r}/M \quad (3)$$

类-属性之间的关联信息值, 由下式计算:

$$I(C : A_j) = \sum_C \sum_{A_j} p_{sr} \log(p_{sr}/p_{s+} p_{+r}) \quad (4)$$

其中: p_{sr} 表示 A_j 的观测值 v_{jk} 满足 $e_{r-1} < v_{jk} < e_r$, 且属于类 c_s 的概率. p_{+r} 表示 $A_j \in e_r$ 的边际概率.

假设类别 C 与属性 A_j 都是随机变量, 则联合熵 $H(C, A)$ 与关联信息 $I(C : A)$ 可以规范化为以下形式:

$$R_{C, A_j} = I(C : A_j)/H(C, A_j) \quad (5)$$

其中

$$H(C, A_j) = - \sum_C \sum_{A_j} p_{sr} \log p_{sr} \quad (6)$$

式中, R_{C, A_j} 称为类-属性关联度. R_{C, A_j} 取值区间为 $[0, 1]$. 若类与属性完全依赖, $R_{C, A_j} = 1$; 若完全独立, $R_{C, A_j} = 0$.

2 基于归纳式学习方法的数据分析

2.1 满意度等级

满意度指标是反映出行者对出行行为的满意程度. 满意度的建立是一个主观过程, 受到人类主观理性和判别力的限制. 因此, 等级的建立基于两个原则:

①对于交通使用者, 其间的等级差异易于区分.

②对于交通管理者, 可以充分反映使用者对交通系统的满意度.

表 1 列出讨论建立的等级划分, 从非常满意到非常不满意分为 5 个等级, 每个等级都可以清晰表达和描述.

表 1 满意度等级分类等级

Tab. 1 Classification set of satisfaction index

等级	出行者对此时间长度通过程的心理描述
5 非常满意	激动、满足、愉悦
4 满意	称心、肯定
3 中立	没有明显情绪
2 不满意	抱怨、懊恼
1 非常不满意	愤慨、难以容忍

2.2 数据收集

数据通过网上问卷调查的形式收集.样本量为 4 790,且覆盖各个年龄层(0~20,1%;21~35,41%;36~50,40%;51~65,19%).通勤交通指住所与工作地点或全日制学习地点之间的往返行为,本调查

对象包含在读大学生,因此具有 1%的 20 岁以下的样本对象.性别比例基本平衡(男性 43%,女性 57%).通过对不同时段出行时间进行满意度评价,得到表 2 的初步分布结果.

表 2 通勤时间与满意度等级调查结果
Tab.2 Survey result of commute time and satisfaction level

满意度等级	通勤出行时间 /min									人数
	0~5	5~10	10~20	20~30	30~40	40~60	60~90	90~120	>120	
5	0	0	320	390	0	0	0	0	0	
4	149	112	0	0	230	72	96	0	0	
3	38	17	54	126	157	360	116	42	56	
2	63	63	99	72	30	62	513	230	155	
1	61	61	91	62	20	40	208	320	305	

2.3 属性区间合并

属性与类之间的关系建立,本质上是对样本数据的不断迭代计算、使类-属性关联度最大化的过程.样本点所属的类别一般事先给定,而属性值的区间划分是可以调整的,因此,计算目的是在一个相对于给定类别的区间划分,使得类-属性关联度最大.关键的解决手段是对区间合理地合并,减少冗余信息,并提高属性与类之间的相关性.归纳式学习方法提供了判断两个相邻区间与属性是否显著相关的公式

$$R_{C,A_j} \geq \frac{\chi^2_{(S-1)(L_j-1)}}{2M \cdot H(C,A_j)} \tag{7}$$

其中:M 为总的判别样本量;S 为类别数;L_j 为 A_j 区间数;χ² 为卡方检验值.

如果判别为真,则认为 C 与 A_j 显著相关.否则,则认为两相邻区间与属性之间非显著相关,可以合并.

研究的样本以 C 为满意度等级,A_j 为通勤出行时间.根据式(7)检验相邻区间,得到区间[0,5]与[5,10]与属性的相关性信息判别式为

$$R_{C,A_{j1}} \cdot 2M_1 \cdot H(C,A_{j1}) = 0.002\ 9 \times 2 \times 564 \times 0.839 = 2.74$$

区间[10, 20]与 [20,30]与属性的相关性信息判别式为

$$R_{C,A_{j2}} \cdot 2M_2 \cdot H(C,A_{j2}) = 0.009\ 2 \times 2 \times 1\ 214 \times 0.785 = 17.54$$

区间[90,120] 与[120-]与属性的相关性信息判别式为

$$R_{C,A_{j3}} \cdot 2M_3 \cdot H(C,A_{j3}) = 0.003\ 4 \times 2 \times 1\ 108 \times 0.691 = 5.206$$

判别条件中,S = 5,L_j = 2.由于本例中检验样本量均大于 500,因此取显著性 α = 0.001 水平下的临界值 χ₄² = 18.47.对以上两个相邻区间判断结果为非真,说明两个相邻区间与类别非显著相关.因此,合并三个时间区间:[0,5]与 [5,10],[10,20]与 [20,30],[90,120]与 [120,∞].

样本最初的关联度指数由式(1)——(6)计算得到: R_{C,A_j} = 0.174 4.通过合并增加关联度指数,R_{C,A_j} = 0.197 4.合并后的属性与类关系见表 3.

表 3 修正后属性与类的关系矩阵
Tab.3 Improved class-attribute matrix

满意度等级	通勤出行时间 /min						人数
	0~10	10~30	30~40	40~60	60~90	>90	
5	0	710	0	0	0	0	
4	261	0	230	72	96	0	
3	55	180	157	360	116	98	
2	126	171	30	62	513	385	
1	123	153	20	40	208	625	

2.4 证据权分析

为了进一步明确指标等级与通勤出行时间之间的对应关系,分别计算通勤出行时间属性区间与每个等级之间的证据权.证据权的定义为:在情况 1 下,结果 2 的发生存在证据权 W,为建立属性区间与类别等级的直接联系提供依据.证据权的计算公式为^[15]

$$W(I = c_p / I \neq c_p | v_{jk}) = \log \frac{O(I = c_p / I \neq c_p | v_{jk})}{O(I = c_p / I \neq c_p)} \tag{8}$$

其中

$$O(I = c_p/I \neq c_p | v_{jk}) = \frac{\Pr(I = c_p | v_{jk})}{\Pr(I = c_p | v_{jk})} \quad (9)$$

$$O(I = c_p/I \neq c_p) = \frac{\Pr(I = c_p)}{\Pr(I = c_p)} \quad (10)$$

其中, $I = c_p$, 表示类别等级属于区间 c_p ; v_{jk} 为代表属性; O 为类别等级属于区间 c_p 的可能性, 用属于区间 c_p 与不属于区间 c_p 的概率之比表示。根据以上公式, 计算每个通勤出行时间的属性区间与满意度等级类别之间的证据权, 得到结果如表 4 所示。

表 4 满意度等级与出行时间区间证据权矩阵
Tab.4 Weight of evidence matrix of satisfaction level and commute time interval

满意度等级	通勤出行时间 /min					
	0~10	10~30	30~40	40~60	60~90	>90
5	—	0.908 230	—	—	—	—
4	0.732 368	—	0.842 927	- 0.010 140	- 0.143 280	—
3	- 0.368 810	- 0.161 710	- 0.346 282	- 0.913 294	- 0.250 220	- 0.415 550
2	- 0.106 240	- 0.350 430	- 0.697 610	- 0.446 690	- 0.521 730	- 0.161 184
1	- 0.067 560	- 0.349 520	- 0.827 600	- 0.600 160	- 0.050 770	- 0.603 438

注: “—”表示空值。

证据权小于零, 说明属性与类之间的关系假设不成立. 通过证据权分析, 可以解读通勤出行时间与满意度等级之间的相关性: 如通勤时间为 30~40 min, 则可认为满意度为等级 4, 满意的可信度为 0.842 927.

3 通勤交通满意度指标

采用归纳学习法分析数据样本的结果首先表明, 以通勤为目的的出行时间与出行者满意度之间具有显著的相关性; 其次, 一定的出行时间区域对某一满意度等级的证据权较高, 存在对应关系及规律。

最终, 为建立直观清晰的通勤交通满意度等级指标, 选取对满意度等级的证据权大于 0.5 的时间区间为代表值, 得到如表 5 所示的满意度指标. 满意度等级为 5 的代表值区间为 10~30 min, 等级为 4 的代表值区间为 0~10 min 及 30~40 min. 由此向下, 等级越低, 出行时间代表值越高. 由等级 5 和 4 的结果可以看出, 通勤者并非希望通勤出行时间越短越好, 一定的出行时间具有非到达目的地以外的效用价值。

表 5 通勤交通满意度指标

Tab.5 Satisfaction index of commute travel

满意度等级	出行时间 /min
5	10~30
4	0~10, 30~40
3	40~60
2	60~90
1	>90

4 结论

基于 4 792 个样本的问卷调查结果, 采用归纳式学习方法, 分析出行者满意度等级与通勤时间之间的关系, 得到大于 0.5 证据权的通勤时间区间与满意度等级的指标体系. 时间区间为 10~30 min 为通勤者非常满意的通勤时间代表值. 这说明, 人们对于出行时间的感觉并不是越短越好, 即不仅仅是效用理论中所假设的负效用. 一定的出行时间给人们提供了一个“缓冲”的空间. 以通勤为目的的出行为例, 在紧张的学习工作地点与繁忙的家庭生活中间, 人们从心理上需要一定的个人时间来“缓冲”. 因此, 在分析交通出行行为时, 即要考虑出行的目的地, 即“可达”的需求, 也要考虑对出行本身的需求, 即“移动”的需求。

以通勤时间为指标的出行者满意度体系能从微观反映出行者的心理状态, 也能从宏观反映交通系统满足以通勤为目的的交通需求的能力. 与传统的商品服务满意度指标体系相比, 其数据便于收集, 表达更简明, 可为城市规划和交通需求分析提供参考。

由于研究的数据是基于网上调查的结果, 因此, 对地区的适用性还需要进一步地检验和修正。

参考文献:

[1] Tardiff T J. Causal inferences involving transportation attitudes and behavior[J]. Transport Research, 1977, 11: 397.

(下转第 1344 页)