

公交线路选择的相关性规律和换乘总量计算模型

陈 君¹, 杨东援², 赵清梅¹

(1. 西安建筑科技大学 土木工程学院, 陕西 西安 710055; 2. 同济大学 交通运输工程学院, 上海 201804)

摘要: 传统的公交换乘信息是通过居民出行调查方法获取, 具有费用高、数据更新慢等不足. 应用公交 IC 卡数据对乘客连续两次乘车的时间间隔分析, 发现当两次乘车为非换乘关系时, 选择乘坐相同线路的人次与选择乘坐不同线路的人次之间具有强相关性. 通过对相关性规律的内在机理进行理论分析, 证明其具有普适性. 应用发现的规律建立了公交系统换乘总量的计算模型, 并提出了基于公交 IC 卡数据的计算算法. 以大规模实际数据对算法进行了实验和验证, 研究结果表明提出的模型和算法具有良好的实用性和精度.

关键词: 交通工程; 公共交通; 换乘; 相关性分析; 公交 IC 卡数据

中图分类号: U491.1

文献标志码: A

文章编号: 1006-7930(2017)06-0835-06

Correlation law for bus routes selection and total transfer volume calculation model

CHEN Jun¹, YANG Dongyuan², ZHAO Qingmei¹

(1. School of Civil Engineering, Xi'an Univ. of Arch. & Tech., Xi'an 710055, China;

2. School of Transportation Engineering, Tongji University, Shanghai, 201804, China)

Abstract: Traditional public transportation transfer information is obtained by resident travel surveys, which have some shortcomings such as high cost, slow update and so on. This paper analyzed time intervals of two consecutive boardings of bus passengers using smart card data and found that there is strong correlation of person-times between choosing same routes and choosing different routes for two non-transfer transit trips. Through theoretical analysis of the internal mechanism, it was proved that this law is universal. According to the finding, the paper established the calculation model of total transit transfer volume and presented the algorithm based on smart card data. The algorithm was tested and validated using large-scale actual data and the results show that the model and algorithm are practical and accurate.

Key words: Traffic engineering; public transportation; transfer; correlation analysis; bus smart card data

传统的公交换乘信息是通过大规模的居民出行抽样调查来获得, 具有费用高, 数据更新慢等不足^[1]. 近十几年来, 国内、外开始研究通过分析公交 IC 卡数据来获得公交系统的换乘信息^[2-4]. Markus Hofmann 等(2005)^[5]应用爱尔兰公交电子售票数据对公交刷卡乘客的换乘行为进行了研究, 当同一公交 IC 卡连续两次刷卡的时间差小于 90 min, 且两次乘车线路不同时, 将两次乘车判断为换乘关系. Catherine Whitney Seaborn 等(2008)^[6]利用公交 IC 卡数据对伦敦多模式的公共交通换乘问题进行研究, 通过分析得到伦敦市公交换乘所需时间的估计值. Ka Kee Alfred Chu 等(2008)^[7]利用公交乘客连续两次乘车的时间和空间关系,

对乘车记录进行逐一判断, 确定出换乘的记录. 陈君等(2013)^[8]应用公交卡数据和智能调度系统数据建立了公交换乘判断模型, 通过逐个计算任意两次乘车的换乘时间阈值来判断换乘. 这种逐个计算换乘判断阈值的方法^[7-8], 除了需要公交 IC 卡数据以外, 还需要公交 GPS 数据、运行记录数据和站点坐标数据(或者是公交运行时刻表数据). 本文应用公交 IC 卡数据对乘客连续两次乘车的时间间隔分析, 发现连续独立的两次公交乘车(非换乘关系), 在不同乘车时间间隔所选择乘坐相同线路的人次和不同线路的人次之间具有强相关性. 对该规律的内在机理进行了理论分析, 应用这一规律本文提出了一种新的公交乘客换乘总量的计

收稿日期: 2016-11-22 修改稿日期: 2017-11-02

基金项目: 国家自然科学基金项目(51208408); 陕西省自然科学基金基础研究计划项目(2014JQ7279, 2017JM5121); 陕西省教育厅专项科研计划项目(17JK0461)

第一作者: 陈 君(1977—), 男, 博士, 副教授, 主要从事城市交通规划与管理方面的研究. E-mail: chenjuntom@126.com

算模型.

1 研究数据描述

以南宁市 2008 年 12 月 111 条公交线路的 IC 卡数据作为研究数据. 以 12 月 1 日为例, 南宁市公交 IC 卡系统 1 天的数据量为 354 863 条. 应用 SQL-Server 数据库系统导入和存储公交卡数据, 研究数据字段描述, 见表 1.

表 1 研究数据字段描述

Tab. 1 Fields description of research data

字段名	数据示例	数据类型	字段内容
Card_ID	77 009 * * *	字符	公交卡号 (隐去后 3 位数字)
IC_Date	2008-12-03	日期	刷卡日期
IC_Time	14:43:00	时间	刷卡时间
Bus_ID	060 437	字符	公共汽车编号
Route_ID	000 006	字符	公交线路编号
Direction_ID	上	字符	公交运行方向 (上行或下行)

2 公交乘客线路选择的相关性规律

2.1 最大可能换乘时间间隔分析

公交换乘的乘车时间间隔包括 3 个部分: 前一次乘车的车内时间、换乘步行时间和换乘候车时间.

(1) 前一次乘车的车内时间

南宁市公交线路正常运营的全程时间在 30~80 min 之间, 将 80 min 作为最大车内时间.

(2) 换乘步行时间

假定换乘的最大距离为 1 000 m, 按步行速度 1.2 m/s 计算, 最大换乘步行时间需要 13.8 min.

(3) 换乘候车时间

南宁市 111 条公交 IC 卡线路的发车间隔在 25 min 以下的有 107 条, 超过 25 min 的仅有 4 条. 将 25 min 作为最大换乘候车时间.

以上 3 项的最大值之和为 118.8 min, 取 120 min 作为分析南宁市公交换乘的最大可能换乘时间间隔.

2.2 公交线路选择的相关性分析

2.2.1 工作日分析

以南宁市 12 月 2 日(星期二)凌晨 4:00 至 12 月 3 日(星期三)凌晨 4:00 之间 24 h, 111 条公交线路的 362 434 条公交卡数据作为工作日分析对象. 在 SQL Server 数据库中, 将公交卡记录, 按

“公交卡号”、“刷卡日期”、“刷卡时间”的先后顺序依次进行排序. 刷卡时间间隔小于 1 min 的数据作为重复刷卡数据, 将其剔除. 通过 VB.NET 语言编程计算同一卡号相邻刷卡记录的刷卡时间间隔. 将同一乘客连续两次乘车的公交线路分为不同线路、相同线路以及全部线路 3 组, 应用 T-SQL 语言分别统计不同时间间隔的乘车人次 p_{di} 、 p_{si} 和 p_i , 如图 1 所示.

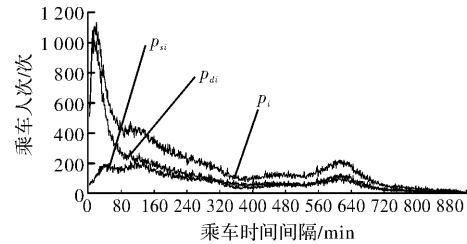


图 1 工作日不同乘车时间间隔的公交乘车人次分布

Fig. 1 Bus travel person times distribution in different interval time at weekday

从图 1 可以看出, 乘车时间间隔小于 T_{\max} (120 min) 时, 选择不同线路的乘车人次明显高于选择相同线路的乘车人数. 这是由于乘车时间间隔小于 T_{\max} 时, 有大量的换乘出行, 对于换乘出行, 一定是选择乘坐不同的线路. 当乘车时间间隔大于 T_{\max} 时, 乘客选择乘坐相同线路和不同线路人次的分布呈现相关关系. 对刷卡时间间隔在 120~920 min 之间选择乘坐相同线路的人次 p_{si} 和不同线路的人次 p_{di} 进行相关性分析, 得到相关系数 R_{sd} 为 0.93, 说明两者之间存在强相关性.

2.2.2 休息日分析

以南宁市 12 月 7 日(星期日)凌晨 4:00 至 12 月 8 日(星期一)凌晨 4:00 之间 24 h 的 351 855 条公交 IC 卡刷卡数据作为休息日的分析对象, 应用 2.2.1 节工作日相同的分析方法, 得到不同时间间隔(0~920 min)的乘车人次的分布曲线, 如图 2 所示.

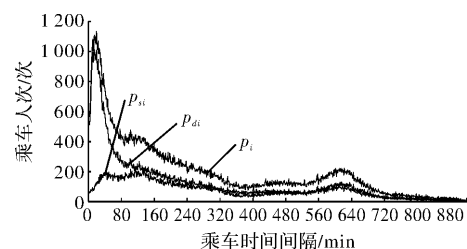


图 2 休息日不同乘车时间间隔的公交乘车人次分布

Fig. 2 Bus travel person times distribution in different interval time at weekend

对休息日不同时间间隔 p_{si} 和 p_{di} 进行相关性分析, 得到两者的相关系数 R_{sd} 为 0.95, 比工作日具

有更高的相关性.

2.2.3 相关性分析结果

乘车时间间隔在 420 min 左右时, 受到集中的通勤返程出行的影响, 相关性有所波动. 因此, 将时间间隔分为 121~420 min、421~920 min、121~920 min 等 3 组, 分别分析工作日和休息日

公交线路选择的相关性, 计算结果见表 2. 根据表 2 分析结果可知, 当乘车时间间隔大于最大换乘时间间隔以后, 乘客不同时间间隔选择乘坐相同线路的人次与选择乘坐不同线路的人次之间具有强相关关系, 该结论本文称之为“公交线路选择的相关性规律”.

表 2 相关性分析结果

Tab. 2 Correlation analysis results

时间间隔		121~420 min	421~920 min	121~920 min
工作日	回归模型	$p_{di}=1.115 \times p_{si}+14.01$	$p_{di}=0.795 \times p_{si}+6.558$	$p_{di}=1.134 \times p_{si}-1.456$
	相关系数	0.95	0.92	0.93
休息日	回归模型	$p_{di}=1.375 \times p_{si}+12.61$	$p_{di}=0.971 \times p_{si}+0.663$	$p_{di}=1.459 \times p_{si}-8.524$
	相关系数	0.95	0.92	0.95

2.3 公交线路选择相关性规律的内在机理

为证明“公交线路选择的相关性规律”具有普适性, 对其内在机理进行理论分析.

2.3.1 变量关系推导

公交乘客线路选择人次的数量关系, 见式(1)~式(5),

$$p_{si}=n_{si}+m_{si} \tag{1}$$

$$p_{di}=n_{di}+m_{di} \tag{2}$$

$$n_i=n_{si}+n_{di} \tag{3}$$

$$m_i=m_{si}+m_{di} \tag{4}$$

$$g_i=\frac{n_i}{m_i}=\frac{n_{si}+n_{di}}{m_{si}+m_{di}}=\frac{e_i \cdot n_{di}+n_{di}}{f_i \cdot m_{di}+m_{di}}=\frac{(e_i+1)n_{di}}{(f_i+1)m_{di}} \tag{5}$$

式中: i 为乘车时间间隔; p_{si} 为选择乘坐相同线路的人次; n_{si} 为返程出行中选择相同线路的人次; m_{si} 为非返程出行中选择相同线路的人次; p_{di} 为选择乘坐不同线路的人次; n_{di} 为返程出行中选择不同线路的人次; m_{di} 为非返程出行中选择不同线路的人次; n_i 为返程出行人次; m_i 为非返程出行人次; g_i 为返程出行人次与非返程出行人次的比例; e_i 为返程出行中选择相同线路与不同线路的人次比例; f_i 为非返程出行中选择相同线路与不同线路的人次比例.

由式(5), 可以得式(6);

$$n_{di}=\frac{g_i(f_i+1)}{e_i+1}m_{di} \tag{6}$$

由式(1)~式(6), 可得式(7);

$$c_i=\frac{p_{si}}{p_{di}}=\frac{n_{si}+m_{si}}{n_{di}+m_{di}}=\frac{e_i \cdot n_{di}+f_i \cdot m_{di}}{n_{di}+m_{di}}=\frac{e_i \cdot \frac{g_i(f_i+1)}{e_i+1}+f_i \cdot m_{di}}{\frac{g_i(f_i+1)}{e_i+1}m_{di}+m_{di}} \tag{7}$$

式中: c_i 为选择乘坐相同线路人次与选择乘坐不同线路人次的比例.

由式(7), 可以推导出式(8), 得到 c_i 值与 e_i 、 f_i 、 g_i 等 3 个值之间的数量关系.

$$c_i=\frac{g_i \cdot g_i \cdot f_i+e_i \cdot g_i+f_i \cdot e_i+f_i}{g_i \cdot f_i+g_i+e_i+1} \tag{8}$$

2.3.2 e_i 、 f_i 和 g_i 对 c_i 的影响分析

式(8)中, g_i 为返程出行与非返程出行人次的比例, 当存在较为集中的返程出行时, 会造成 g_i 值产生较大变化, 从而引起 c_i 值也发生较大变化. 工作日晚高峰时段通勤的返程出行较为集中, c_i 波动较大. 在休息日, 通勤交通较少, 返程出行时间较为随机, 全天 g_i 值变化很小, 从而 p_{si} 与 p_{di} 相关性很强.

式(8)中, 当乘车人数不断增大时, e_i 和 f_i 值分别趋于常量, p_{si} 与 p_{di} 趋于完全线性相关.

通过对以上分析可知: 当公交出行量越大、返程出行时间越随机, 公交线路选择的相关性越强.

3 公交乘客换乘总量计算模型建立

应用上文得到的“公交线路选择的相关性规律”, 建立公交乘客换乘总量的计算模型.

3.1 建模思路

不同时间间隔的换乘人次, 可由式(9)计算得到,

$$p_{ii}=p_{di}-p_{ai} \tag{9}$$

式中: p_{ii} 为换乘人次; p_{ai} 为选择乘坐不同线路, 且为非换乘关系的乘车人次; 其余符号同上.

由“公交线路选择的相关性规律”可知, 当 $i>T_{\max}$ 时, p_{si} 与 p_{di} 具有相关关系, 通过相关分析可

得两者关系模型 $p_{di} = f(p_{si})$ 。又由于当 $i > T_{\max}$ 时, $p_{ai} = p_{di}$, 则可得到 p_{ai} 和 p_{si} 的关系模型 $p_{ai} = f(p_{si})$, 再将该关系模型推论到在 $i \leq T_{\max}$ 的区间, 则可计算得到 p_{ai} 值, 从而得到换乘人次 p_{ti} 。

3.2 模型建立

根据以上思路, 建立公交换乘总量计算模型, 见式(10),

$$P_t = \sum_{i=T_{\min}}^{T_{\max}} [p_{di} - f(p_{si})] \quad (10)$$

式中: P_t 为换乘总量; P_{\max} 按 2.1 节的方法确定; T_{\min} 为最小可能换乘时间间隔, 一般取 0~2 min; 其余符号同上。

由式(10), 可得式(11)~式(12),

$$M = N - P_t \quad (11)$$

$$\beta = N/M = N/(N - P_t) \quad (12)$$

式中: M 为公交出行总人次(指目的出行); N 为乘车总人次; β 为换乘系数; 其余符号同上。

4 算法提出

根据以上模型, 应用 IC 卡数据计算公交换乘总量的算法如下:

Step0: 计算同一卡号连续两次乘车的刷卡时间间隔。

Step1: 确定最大可能换乘时间间隔 T_{\max} 和最小可能换乘时间间隔 T_{\min} 。

Step2: 统计不同刷卡时间间隔(1 min 为单位)的 p_{di} 、 p_{si} 值。

Step3: 对 p_{di} 和 p_{si} 值进行回归分析, 得到 $p_{ai} = f(p_{si})$ 。(休息日采用 $i > T_{\max}$ 的数据; 工作日为减少通勤交通对相关性的影响, 可采用 $T_{\max} < i < 420$ min 的数据)

Step4: 根据 $p_{ai} = f(p_{si})$, 计算 $T_{\min} \sim T_{\max}$ 之间的 p_{ai} 值。

Step5: 根据 $P_{ti} = p_{di} - p_{ai}$, 计算 $T_{\min} \sim T_{\max}$ 之间, 不同时间间隔的换乘量 P_{ti} 。

Step6: 对不同时间间隔的 P_{ti} 求和, 得到换乘总量 P_t 。

Step7: 根据式(11)~式(12)计算公交出行总量 M 和换乘系数 β 。

5 算法试验和验证

5.1 算法试验

以南宁市 111 条公交卡线路 12 月 2 日(星期二)凌晨 4:00 至 12 月 3 日(星期三)凌晨 4:00 之间, 24 h 的 362 434 条公交卡数据为算例, 测试和验证提出的公交换乘总量模型和算法, 得到南宁市公交卡乘客换乘总量模型, 见式(13), 相关计算结果, 见表 3。

$$P_t = \sum_{i=2}^{120} [p_{di} - (1.115 \times p_{si} + 14.01)] \quad (13)$$

表 3 算法试验结果

Tab. 3 Algorithm experiment results

项目	计算结果	备注
公交卡乘客总乘车次数/人次	362 434	公交 IC 卡系统 24 h 的总刷卡人次
重复刷卡次数/人次	31 401	刷卡时间间隔小于 2 min
用于计算的公交卡乘客总乘车次数/人次	331 033	扣除重复刷卡数据
计算得到的公交卡乘客总换乘量/人次	33 047	
计算得到的公交卡乘客出行总量/人次	297 986	
换乘系数	1.11	
公交卡乘客换乘总量/人次	36 181	将重复刷卡数据按比例计入
公交卡乘客出行总量/人次	326 253	
全部公交乘客换乘总量/人次	144 724	根据刷卡率(25%)扩样得到
全部公交乘客出行总量/人次	1 305 012	

5.2 与统一阈值判断方法比较

统一阈值方法采用一个固定的乘车时间间隔来判断是否为换乘关系, 当同一卡号, 连续两次刷卡的时间差小于该换乘阈值, 且两次乘车线路不同, 将两次乘车判断为换乘关系。本文假定

55 min 为统一换乘判断阈值, 比较换乘总量模型与统一阈值方法的关系如图 3 所示。

图 3 中细实线表示不同时间间隔选择不同线路的人次分布, 粗实线表示换乘总量模型计算得到的换乘人次分布。统一阈值方法计算的换乘量为

图3中A区与B区面积之和, 换乘总量模型计算的换乘量为图3中B区与C区面积之和. 统一阈值方法计算换乘量是将某一固定的时间间隔作为所有换乘判断的阈值, 漏计了图3中C区换乘量, 而多计A区换乘量.

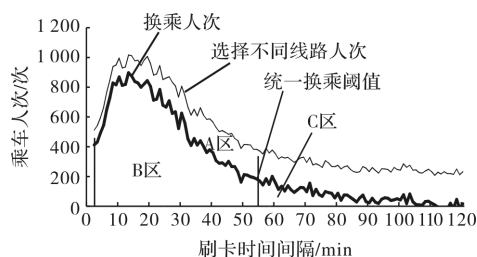


图3 换乘总量模型与统一阈值方法比较

Fig. 3 Comparison of total transit transfer volume model with unified threshold method

5.3 与独立阈值判断方法比较

将提出的换乘总量模型与现有研究基于独立阈值的换乘判断模型进行比较. 独立阈值判断模型的基本思路是: 逐个卡号计算连续两次刷卡之间的时间间隔是否超过合理的换乘时间阈值, 如果实际刷卡的时间间隔小于或等于换乘时间阈值则判断为换乘^[3-4]. 独立阈值换乘判断方法需要的计算数据有公交IC卡数据、公交GPS数据、运行记录数据和公交站点坐标数据(或者公交IC卡数据和公交站点时刻表数据), 本文提出的换乘总量模型仅需要公交IC卡数据.

将2008年12月3日(星期三)南宁市57条智能调度线路的192 971条公交IC卡数据作为验证数据, 采用独立阈值判断模型得到换乘总人次为8 442次, 计算过程详见文献[4], 采用本文提出的换乘总量模型得到换乘总人次为9 532次. 2种方法得到的不同时间间隔的换乘人次分布, 如图4所示.

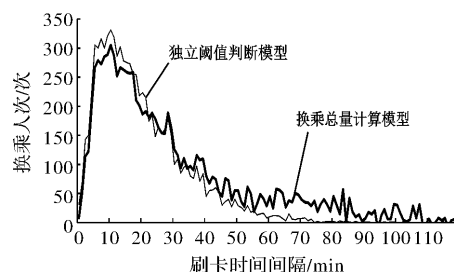


图4 换乘总量模型与独立阈值方法比较

Fig. 4 Comparison of total transit transfer volume model with independent threshold method

从图4中可以看出, 2种方法计算结果的曲线分布基本一致. 两者存在的偏差, 经过分析其主要原因是: 验证算例受数据条件限制, 仅采用了

57条智能调度线路的数据进行计算, 建立换乘总量模型时 $p_{ai} = f(p_{si})$ 的相关性受到了影响, 57条智能调度线路数据的相关系数为0.90, 而111条公交IC卡线路的相关系数为0.95, 南宁市实际公交线路为133条. 当采用全部公交线路数据计算时, 2种方法的计算结果应更加接近.

5.4 与公交出行调查数据比较

换乘系数是指公交乘车人次与公交出行人次的比值, 用来衡量公交出行的直达程度. 将本文算法试验得到的换乘系数 β 与2009年南宁市公共交通规划的公交出行调查数据^[9]进行比较. 2009年南宁市采用问卷方式对5 992名公交乘客的出行进行抽样调查, 得到南宁市公交换乘系数为1.1. 本文算例 β 计算结果为1.11与南宁市公交出行调查数据一致, 证明了公交换乘总量模型的精确性.

6 结论

(1) 本文应用公交IC卡数据对公交乘车时间间隔特征和规律进行分析, 发现“同一乘客连续独立的两次公交出行(非换乘关系), 在不同乘车时间间隔, 后一次出行选择乘坐不同线路的人次和选择乘坐相同线路的人次之间存在强相关性”. 通过对该规律内在机理的理论分析, 证明了其具有普适性.

(2) 根据以上规律, 提出公交换乘总量模型和算法. 该方法仅需要公交IC卡数据, 能够计算公交换乘量、公交出行总量和公交换乘系数3个指标, 为公交系统的规划和管理提供了决策数据.

(3) 该方法与出行调查方法相比, 费用低、数据更新快, 与独立阈值方法相比, 数据条件低、计算简便、适用范围广, 并能够克服统一阈值方法存在的固有缺陷.

参考文献 References

- [1] GUO Zhan. Transfers and path choice in urban public transport systems[D]. MIT PHD; Boston; Massachusetts Institute of Technology. Thesis, 2008.
- [2] BAGCHI M, WHITE P R. The potential of public transport smart card data[J]. Transport Policy, 2005, (12):464-474.
- [3] 陈学武, 戴霄, 陈茜. 公交IC卡信息采集、分析与应用研究[J]. 土木工程学报, 2004, 37(2): 105-110.
CHEN Xuewu, DAI Xiao, CHEN Qian. Approach on the information collection, analysis and application of bus intelligent Card[J]. China Civil Engineering Journal, 2004, 37(2): 105-110.

- [4] APTS Data Archiving and Mining System (ADAMS) [R]. Florida Department of Transportation, 2007.
- [5] HOFMANN M, O'MAHONY M. Transfer journey identification and analyses from electronic fare collection data [C]//Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria;2005.
- [6] SEABORN C W. Application of smart card fare payment data to bus network planning in london [D]. Cambridge: Massachusetts Institute of Technology, 2008.
- [7] CHU K K A, CHAPLEAU R. Enriching archived smart card transaction data for transit demand modeling [J]. Transportation Research Record, 2008, (2063): 63-72.
- [8] 陈君,杨东援. 融合智能调度数据的公交 IC 卡乘客换乘站点判断方法[J],长安大学学报(自然科学版). 2013,33(4):92-98.
CHEN Jun, Yang Dongyuan. Method of identifying transfer bus stops of smart card passengers integrating bus intelligent dispatching data[J]. Journal of Chang'an University(Natural Science Edition), 2013, 33(4): 92-98.
- [9] 南宁市规划管理局. 南宁市公共交通运输规划(2008-2020)[R]. 上海:同济大学,2010.
Nanning City Planning Bureau. Public transportation planning of Nanning City(2008-2020)[R]. Shanghai: Tongji University, 2010.
- (编辑 吴海西)

(上接第 819 页)

- [14] 潘振华. 具有自复位能力的钢框架体系研究[D]. 北京:清华大学,2010: 13-78.
PAN Zhenhua. Study of self-centering steel moment frames [D]. Beijing: Tsinghua University, 2010: 13-78.
- [15] 魏宇翔. 自复位结构体系柱脚节点的抗震性能研究 [D]. 苏州:苏州科技学院,2012:20-27.
WEI Yuxiang. Research on seismic performance of post-tensioned column base connection for self-centering seismic resistant steel frame[D]. Suzhou: Suzhou University of Science and Technology, 2012: 20-27.
- [16] Dassault Systemes Simulia Corp. ABAQUS Analysis user's manual version 6.10 [M]. Providence, RI: Dassault Systemes Simulia Corp., 2010.
- (编辑 吴海西)