

# 城市轨道交通客流预测的历史相似日选择研究<sup>\*</sup>

光志瑞<sup>1,2</sup> 魏 运<sup>1,2\*\*</sup> 薛云雷<sup>3</sup> 谢莎婷<sup>1,2</sup> 吴雁军<sup>1,2</sup>

(1. 北京市地铁运营有限公司技术创新研究院, 100039, 北京; 2. 地铁运营安全保障技术北京市重点实验室, 100039, 北京;  
3. 京投新岸线技术有限公司, 100089, 北京//第一作者, 高级工程师)

**摘 要** 选择合理的历史相似日(以下简称“相似日”),可有效提高城市轨道交通客流预测的精度。提出了适用于不同客流预测需求的相似日选择方法:首先将相似日选择过程转化为客流模式识别过程;其次提出了前向逐日检索和双向定位检索两种相似日选择算法策略;然后采用定量转化、量纲一化、构造函数法等方法分别计算了星期类型、与预测日的间隔时长、工作日属性、日最高温度等影响因素的相似度,构建了含各影响因素的累乘函数;最后,结合北京城市轨道交通客流数据对相似日选择方法进行拟合及验证。结果表明:该方法的预测结果准确、可靠,可为不同的城市轨道交通客流预测场景提供参考。

**关键词** 城市轨道交通;客流预测;历史相似日;客流模式;构造函数法

**中图分类号** U293.13

**DOI:**10.16037/j.1007-869x.2022.07.011

## Study on Historical Similarity Day Selection in Urban Rail Transit Passenger Flow Prediction

GUANG Zhirui, WEI Yun, XUE Yunlei, XIE Shating, WU Yanjun

**Abstract** Choosing reasonable historical similar days is an effective way to improve the prediction accuracy of urban rail transit passenger flow. A method of choosing similar days to meet different demands of passenger flow prediction is proposed. First, similar day selection process is converted to passenger flow pattern recognition process. Secondly, two strategies of similar day selection are proposed, which are 'forward day by day retrieval' and 'bidirectional positioning retrieval'. Then, methods of quantitative transformation, normalization, and function construction are adopted to calculate the similarity degree of each influencing factor, such as week attribute, length of time between forecast days, date attribute, daily maximum temperature, a multiplicative function including all

factors is constructed. Finally, Beijing urban rail transit passenger flow data is adopted for calibration and validation with similar day selection method. Results show that the prediction result of this method is accurate and reliable, providing reference for passenger flow prediction in various urban rail transit scenarios.

**Key words** urban rail transit; passenger flow prediction; historical similarity day; passenger flow mode; function construction method

**First-author's address** Technical Innovation Research Institute of Beijing Mass Transit Railway Operation Co., Ltd., 100039, Beijing, China

选择合理的历史相似日(以下简称“相似日”),是城市轨道交通客流预测工作的重要组成部分。电力系统在负荷预测中对相似日进行了广泛研究,如文献[1-3]将温度、湿度、日类型及天气状况等影响电力负荷的因素量化,给出了各因素相似度映射表计算日期的相似度。在交通领域,文献[4]通过对OD(起讫点)概率矩阵进行聚类分析,将运营日分为5类;文献[5]在对节假日客流进行预测时,提出了构建日期特征向量、选取合理评估函数的相似日选择思路。但以上研究均缺乏对城市轨道交通客流影响因素的量化分析,以及日期相似度的系统化模型构建。

本文从日期相似度算法策略设计、各影响因素相似度计算、日期相似度模型构建等方面,提出了适合城市轨道交通客流预测的相似日选择方法。

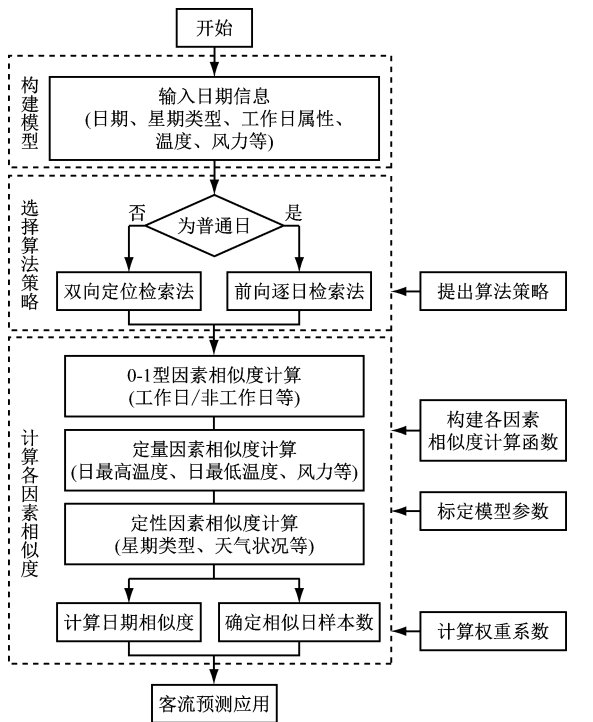
## 1 客流预测相似日选择算法设计

### 1.1 相似日选择方法的技术路线

城市轨道交通客流预测相似日选择的技术路线如图1所示。

<sup>\*</sup> 国家重点研发计划项目(2020YFB1600705)

<sup>\*\*</sup> 通信作者



注:普通日是指正常工作日与双休日;与之对应的是特殊日,特殊日含节假日、节假日的前一日、重大政治文化活动日等。

图1 城市轨道交通客流预测相似日选择方法的技术路线  
Fig. 1 Technical route of similar day selection method for urban rail transit passenger flow forecast

## 1.2 客流模式

客流模式是指当日城市轨道交通线路/线网所对应的客流影响因素合集。客流模式越接近,客流的影响因素越相似,预测日与相似日在客流总量、客流分布情况或客流变化趋势等方面越相似,因此,可将相似日的选择过程转化为客流模式的识别过程。定义第日客流模式的特征向量  $M(i)$  为:

$$M(i) = [D_{i,we}, D_{i,lt}, D_{i,wo}, T_{i,h}, T_{i,l}, W_{i,we}, W_{i,wi}] \quad (1)$$

式中:

$D_{i,we}$ ——星期类型,如星期一、星期二等;

$D_{i,lt}$ ——与预测日的间隔时长;

$D_{i,wo}$ ——工作日属性;

$T_{i,h}$ ——日最高温度;

$T_{i,l}$ ——日最低温度;

$W_{i,we}$ ——晴、大雨等天气状况;

$W_{i,wi}$ ——风力。

## 1.3 算法策略

普通日的客流量与日期距离密切相关。一般地,  $D_{i,lt}$  越小,客流总量越接近,可称为“客流总量相

当日”。节假日、节假日的前一日(以下简称“节前一日”)等特殊日与邻近普通日的客流规律差别较大,但和往年相同时期或相似节假日的变化趋势相似,可称为“客流变化相似日”。

在进行客流模式识别时,为节约运算成本、提高算法效率,可针对客流总量相当日和客流变化相似日,分别采用不同的相似日选择算法策略。

策略一:选择客流总量相当的相似日时,采用前向逐日检索法,从最临近预测日的历史日开始,逆时序逐日计算每个历史日与预测日的相似度,选取最近  $N$  天中相似度最高的  $n$  天作为与之客流总量相当的历史相似日,或选取相似度大于某个阈值的  $n$  天作为历史相似日。其中,  $n$  为确定选用的历史相似日的总天数。

策略二:选择客流变化相似的相似日时,采用双向定位检索法,将往年(如最近3年)同期作为定位点,如同一特殊阶段、假期天数相同的法定节假日中与预测日时序相同的某天,从定位点分别正向、逆向逐日计算每个历史日与预测日的相似度,确定  $n$  个历史相似日。

## 1.4 相似日选择算法描述

### 1.4.1 日期相似度的计算

设  $R_i$  为历史第  $i$  日与预测日的日期相似度,  $r_{i,m}$  为历史第  $i$  日与预测日在第  $m$  种客流影响因素上的相似度,  $k_m$  为第  $m$  种客流影响因素的权重系数(通过自适应过程标定)。采用各客流影响因素相似度累乘的方法计算  $R_i$ ,通过  $k_m$  确定各因素的影响指数。  $k_m$  标定的目标是使  $R_i$  的排序与对应客流相似度排序尽可能一致。  $R_i$  的计算式为:

$$R_i = \prod_{m=1}^n (r_{i,m})^{k_m} \quad (2)$$

### 1.4.2 相似日样本数 $n$ 的确定

可借鉴聚类分析中确定类别数的方法确定相似日的样本数:

1) 结合实际法。根据实际应用经验,进行客流预测时,普通日可选择4~6个相似日,特殊日可选择1~2个相似日。

2) 肘方法。确定与预测日相似度最高的1~ $n$ 个相似日,参考相似日的客流情况对预测日的客流进行预测,绘制样本数与预测误差的折线图,折线的拐点处即为合适的样本数。

2 各客流影响因素的相似度计算

客流的影响因素主要包括星期类型、与相似日的间隔时长、工作日属性、日最高温度、日最低温度、天气状况、风力等。这些因素可分为 0-1 型指标、定性指标、定量指标 3 种类型。应按照影响因素不同的类型分别进行相似度计算。

2.1 0-1 型指标的相似度计算

工作日与非工作日、法定节假日与非法定节假日、重大政治文化活动日与非重大政治文化活动日间的客流特点明显不同,两两绝非相似日。若历史日与预测日同属同一个类型,如均为工作日,则相似度取 1,否则取 0。式(1)中的  $D_{i,wo}$  属 0-1 型指标。

2.2 定性指标的相似度计算

式(1)中的  $D_{i,we}$ 、 $W_{i,we}$  属定性指标。计算定性指标的相似度时,可直接匹配已标定的相似度映射表。映射表标定流程如下:① 通过给定性指标赋值或借助其他定量指标,对定性指标进行定量转化;② 对定量转化后的指标进行量纲一化处理;③ 计算相似度,形成定性指标相似度映射表。

本文以星期类型相似度为例,介绍定性指标的相似度计算方法。星期类型对客流量有较大影响,构建星期类型相似度函数时,需遵循以下基本原则:① 星期类型的相似度函数分布在(0,1]区间;② 若星期类型相同,则相似度为 1;工作日与双休日相似度最小,最小相似度应大于 0;③ 通过构造函数法构建相似度计算函数,使不同星期类型的相似度能够较均匀地分布在(0,1]内。

根据以上原则,构建出星期类型的两个相似度函数,其计算式分别为:

$$r_{Dt,1} = 1 - |x_p - x_q| \tag{3}$$

$$r_{Dt,2} = \frac{1}{e^{\sum_k |x_p - x_q|}} \tag{4}$$

式中:

- $r_{Dt,1}$ ——星期类型的相似度函数 1;
- $r_{Dt,2}$ ——星期类型的相似度函数 2;
- $x_p$ ——星期类型为  $p$  的日期所对应客流量的量纲一化值;
- $x_q$ ——星期类型为  $q$  的日期所对应客流量的量纲一化值;
- $p、q$ ——均为星期类型,星期一取 1,星期二取 2,⋯,星期日取 7。

结合北京城市轨道交通线网 2017 年 7 月份和 8 月份的客流数据,采用上述两个函数得到星期类型相似度映射表,结果显示:这两个函数所得到的工作日(星期一至星期五)的相似度均在 0.9 以上,双休日(星期六、星期日)的相似度均在 0.8 以上。如表 1 所示,工作日与双休日的相似度均低于 0.500,其中: $r_{Dt,1}$ 均低于 0.250, $r_{Dt,2}$ 大多接近 0.500。

表 1 基于两个相似度函数得到的各星期类型间客流相似度结果

Tab. 1 Results of passenger flow similarity between week types based on two similarity functions

星期类型	$r_{Dt,1}$		$r_{Dt,2}$	
	星期六	星期日	星期六	星期日
星期一	0.215	0.080	0.456	0.399
星期二	0.228	0.093	0.462	0.404
星期三	0.232	0.098	0.464	0.406
星期四	0.239	0.104	0.467	0.408
星期五	0.135	0.001	0.421	0.368
星期六	1.000	0.866	1.000	0.874
星期日	0.866	1.000	0.874	1.000

在特殊情况下(如星期六调休加班下的客流与工作日客流间差别较小),星期类型不应作为客流预测的主导因素。为避免星期类型在任何条件下均为主导因素,应使星期类型相似度较均匀地分布在(0,1]内,因此,选用式(4)作为星期类型的相似度函数,其计算结果如表 2 所示。

表 2 基于式(4)得到的各星期类型间客流相似度结果

Tab. 2 Results of passenger flow similarity between different week types based on formula (4)

星期类型	星期一	星期二	星期三	星期四	星期五	星期六	星期日
星期一	1.000	0.987	0.982	0.976	0.924	0.456	0.399
星期二		1.000	0.995	0.989	0.912	0.462	0.404
星期三			1.000	0.994	0.907	0.464	0.406
星期四				1.000	0.902	0.467	0.408
星期五					1.000	0.421	0.368
星期六						1.000	0.874
星期日							1.000

2.3 定量指标的相似度计算

式(1)中的  $D_{i,lt}$ 、 $T_{i,h}$ 、 $T_{i,l}$ 、 $W_{i,wi}$  属定量指标。本文以  $D_{i,lt}$ 、 $T_{i,h}$  为例,构造定量指标的相似度计算函数。

### 2.3.1 $D_{i,lt}$ 的相似度计算

$D_{i,lt}$  对客流的影响具有“近大远小”的特点,即:当预测日为普通日时, $D_{i,lt}$  和  $D_{i,we}$  为客流预测的主导因素。离待预测日越近,其客流特征越相似,且随着与预测日距离的增加相似度缓慢衰减。当预测日为节假日及节前一日时, $D_{i,lt}$  为客流预测的主导因素,离定位点的距离越近其客流特征越相似,且随着与定位点距离的增加相似度骤减。因此,引入特殊日因子  $\lambda$ ,用以解决不同情况下主导因素变换的问题。同时,考虑到普通日距离衰减的缓慢性,分别引入了周衰减比例和日衰减比例,用以表述该指数函数的衰减程度。 $D_{i,lt}$  相似度的计算式为:

$$r_{Dd}(d_1, d_2) = \omega_1^{(1-\lambda_u)\text{int}(d_1/7)} \omega_2^{(1-\lambda_u)\text{mod}(d_1/7)} \omega_3^{\lambda d_2} \quad (5)$$

式中:

- $r_{Dd}(d_1, d_2)$ —— $D_{i,lt}$  相似度;
- $d_1$ ——第  $i$  个历史日距离预测日的天数;
- $d_2$ ——第  $i$  个历史日距离定位点的天数;
- $\omega_1$ ——普通日的周相似衰减比例;
- $\omega_2$ ——普通日的日相似衰减比例;
- $\omega_3$ ——特殊日的日期距离衰减系数。

式(5)中: $\lambda$  为 0-1 变量,当预测日为特殊日时, $\lambda$  取 1,否则取 0;int 为取整运算,mod 为取余运算。

### 2.3.2 $T_{i,h}$ 相似度计算

$T_{i,h}$ 、 $T_{i,l}$ 、 $W_{i,wi}$  等定量指标计算方法类似。以  $T_{i,h}$  为例,不同范围内  $T_{i,h}$  变化对客流的影响程度不同:当  $T_{i,h}$  低于某阈值时,客流对最高温度的反应不敏感,如  $T_{i,h}$  为 23℃ 与 28℃ 时(相差 5℃)的客流量相当;但是一旦超过某阈值,客流对温度的反应将变得敏感,如  $T_{i,h}$  为 36℃ 与 31℃ 时(相差 5℃)的客流量相差较大。因此,通过构建分段函数来计算  $T_{i,h}$  的相似度,分段函数的影响系数用  $\alpha_k$  表示, $T_{i,h}$  低于阈值时  $k$  取 1,高于阈值时  $k$  取 2:

$$r_{Th} = 1 - \alpha_k |T_{u,h} - T_{v,h}| \quad (6)$$

式中:

- $r_{Th}$ —— $T_{i,h}$  相似度;
- $T_{u,h}$ ——第  $u$  日  $T_{i,h}$  量纲一化值;
- $T_{v,h}$ ——第  $v$  日  $T_{i,h}$  量纲一化值;
- $\alpha_k$ ——影响系数( $k$  取 1,2)。

当  $T_{i,h}$  低于某一阈值(如 34℃),此时  $T_{i,h}$  不是主导因素, $k$  取 1,  $\alpha_k$  可取 0.001(结合星期类型相

似度赋值);当  $T_{i,h}$  高于某一阈值(如 34℃),此时  $T_{i,h}$  为主导因素, $k$  取 2,  $\alpha_k$  可取 0.020(结合星期类型相似度赋值)。

## 3 实例分析

### 3.1 相似日选择的评价标准

相似日选择是否合理,主要通过评估预测精度来判断。设城市轨道交通客流预测的平均绝对误差为  $E_{MAE}$ ,平均绝对百分比误差为  $E_{MAPE}$ ,其计算式分别为:

$$E_{MAE} = \frac{1}{t} \sum |\hat{y}_s - y_s| \quad (7)$$

$$E_{MAPE} = \frac{1}{t} \sum \left| \frac{\hat{y}_s - y_s}{y_s} \right| \times 100\% \quad (8)$$

式中:

- $\hat{y}_s$ ——预测客流;
- $y_s$ ——实际客流;
- $t$ ——预测样本数。

### 3.2 普通日相似日选择

本文基于经验法与相似日选择法,对 2017 年 11 月 23 日(星期四)进行相似日选择,对该日的北京城市轨道交通全路网客流进站量进行预测,并对传统经验法的预测结果与本文提出的相似日选择法的预测结果进行对比。

在选择普通日相似日时,通常采用无异常情况的 4 个与预测日星期类型相同的历史日期作为相似日。采用经验法选择的相似日为 2017-11-16(星期四)、2017-11-09(星期四)、2017-11-02(星期四)、2017-10-26(星期四)。采用本文提出的相似日选择法,式(5)的  $\omega_1$  取 0.98,  $\omega_2$  取 0.99,得到的计算结果如下:2017-11-22(星期三)的相似度为 0.984、2017-11-21(星期二)的相似度为 0.969、2017-11-16(星期四)的相似度为 0.980、2017-11-15(星期三)的相似度为 0.964。

在此基础上,采用平均值与二次指数平滑法(平滑系数  $l$  分别取 0.1、0.2、0.3)计算  $E_{MAE}$  和  $E_{MAPE}$ 。经验法和相似日选择法得到的  $E_{MAE}$  和  $E_{MAPE}$  如表 3 所示。由表 3 可看出,对 4 种客流预测结果的误差进行平均,可得到采用相似日选择法得到的  $E_{MAE}$  为 5 080 人次,  $E_{MAPE}$  为 0.08%,其预测精度高于经验法的预测精度。

表 3 基于两种相似日选择法的普通日相似日客流预测误差对比

Tab. 3 Comparison of normal day and similar day passenger flow prediction based on similar day selection methods

项目	经验法的 预测误差		相似日选择法的 预测误差	
	$E_{MAE}/\text{人次}$	$E_{MAPE}/\%$	$E_{MAE}/\text{人次}$	$E_{MAPE}/\%$
平均值法	25 225	0.40	6 805	0.11
指数平滑法( $l=0.1$ )	25 693	0.41	8 110	0.13
指数平滑法( $l=0.2$ )	8 107	0.13	4 021	0.06
指数平滑法( $l=0.3$ )	8 082	0.13	1 383	0.02
各方法的平均值	16 777	0.27	5 080	0.08

3.3 节前一日相似日选择

本文基于经验法与相似日选择法,对 2017 年国庆节的前一日(2017-09-30,星期六)进行相似日选择,并预测该日北京城市轨道交通全路网的客运量。

经验法选用 2016-09-30(星期五)作为相似日,由于星期五与星期六相似度较低(仅为 0.421),不能作为本次客流预测的参考。采用相似日选择法得到的相似日为 2017 年端午节的前一日(2017-05-27,星期六),该日与预测日均为调休加班日,二者的相似度为 0.786。

如表 4 所示,采用增长系数法进行预测得到相似日选择法的  $E_{MAPE}$  为 1.27%,预测误差小于经验法的预测误差(6.95%)。

表 4 基于两种相似日选择法对节前一日客流预测的误差对比

Tab.4 Comparison of two similar day selection methods on passenger flow prediction of the previous day of holiday

方法	$E_{MAE}/\text{人次}$	$E_{MAPE}/\%$
经验法	816 933	6.95
相似日选择法	149 390	1.27

4 结语

选择合理的相似日,是提高城市轨道交通客流预测精度的有效途径。本文提出的用于城市轨道交通客流预测中的相似日选择法,综合考虑了星期类型、与预测日的间隔时长、工作日属性、日最高温度等各类影响城市轨道交通客流的因素,并根据各影响因素的特征构建了不同影响因素的相似度量方法,可为不同的城市轨道交通客流预测场景提供参考。

参考文献

[1] 莫维仁,张伯明,孙宏斌,等. 短期负荷预测中选择相似日的探讨[J]. 清华大学学报(自然科学版),2004(1):106.  
MO Weiren, ZHANG Boming, SUN Hongbin, et al. Method to select similar days for short-term load forecasting[J]. Journal of Tsinghua University (Science and Technology), 2004(1):106.

[2] 蔡佳宏,刘俊勇. 超短期负荷预测中相似日的选择方法[J]. 华北电力大学学报(自然科学版),2006(1):38.  
CAI Jiahong, LIU Junyong. Selecting method for similar day in super short-term load forecasting[J]. Journal of North China Electric Power University (Natural Science Edition), 2006(1):38.

[3] 康重庆,周安石,王鹏,等. 短期负荷预测中实时气象因素的影响分析及其处理策略[J]. 电网技术,2006(7):5.  
KANG Chongqing, ZHOU Anshi, WANG Peng, et al. Impact analysis of hourly weather factors in short-term load forecasting and its processing strategy [J]. Power System Technology, 2006(7):5.

[4] 栾文波,汪林,张宁,等. 基于数据聚类技术的城市轨道交通运营特征日分类研究[J]. 城市轨道交通研究,2018(4):14.  
LUAN Wenbo, WANG Lin, ZHANG Ning, et al. Classification of operational characteristic days based on data clustering technology for urban rail transit[J]. Urban Mass Transit, 2018(4):14.

[5] 许俊,周峰,徐瑞华. 基于数据分析的城市交通节假日客流预测[J]. 华东交通大学学报,2016(2):72.  
XU Jun, ZHOU Feng, XU Ruihua. Forecast of urban rail transit passenger flow in holidays based on data analysis[J]. Journal of East China Jiaotong University, 2016(2):72.

(收稿日期:2020-05-15)

(上接第 50 页)

[6] 刘世平. 测速误差分析与计算[J]. 兵工学报,1988(4):41.  
LIU Shiping. Error analysis and calculation in velocity measurements[J]. Acta Armamentarii, 1988(4):41.

[7] 孙晓光. 一种适用于中低速磁浮的测速测距系统研究[J]. 铁路通信信号工程技术,2018(9):67.  
SUN Xiaoguang. Research on speed and distance measurement system of medium-low-speed maglev train[J]. Railway Signalling

& Communication Engineering, 2018(9):67.

[8] 崔亦博,焦怡博,孙旺,等. 基于多传感器综合的中低速磁浮列车测速定位系统[J]. 铁道通信信号,2017(3):74.  
CUI Yibo, JIAO Yibo, SUN Wang, et al. Location and velocity detecting system of medium and low speed maglev train based on multi-sensor synthesis[J]. Railway Signalling & Communication, 2017(3):74.

(收稿日期:2020-05-18)