

基于大数据的城市轨道交通运营故障影响 分析系统对客流影响的分析

吴志强¹ 黄天印² 颜彦文¹ 赵时旻²

(1. 上海久誉软件系统有限公司, 201103, 上海; 2. 上海申通地铁集团有限公司, 201103, 上海//第一作者, 工程师)

摘要 运营故障影响分析系统基于大数据思想和技术, 通过机器学习技术, 可对城市轨道交通运营故障数据和客流数据实现采集、存储与处理、建模及呈现, 可实现对城市轨道交通运营事件类型的分类及影响时间的推算。介绍了该系统的工作原理, 描述了运营故障情况下的客流影响关联模型、故障影响分析方法及故障影响的可视化呈现手段。实例验证了该系统的有效性。

关键词 城市轨道交通; 运营故障; 客流分析; 大数据

中图分类号 U293.6

DOI:10.16037/j.1007-869x.2019.04.008

Impact of Urban Rail Transit Operational Fault Analysis System on Passenger Flow Based on Big Data

WU Zhiqiang, HUANG Tianyin, YAN Yanwen,
ZHAO Shimin

Abstract The working principle of operational failure impact analysis system in data collection, data processing, data modeling and visualization are introduced on the basis of big data technology and machine learning technology. It is possible to achieve the rail transit operational event classification and the influence time calculation. In this paper, the passenger flow impact association model, fault impact analysis method and the visualization method of fault impact under operational failures are introduced, the effectiveness of operational failure impact analysis system is verified by practical applications.

Key words urban rail transit; operational fault; passenger flow analysis; big data

First-author's address Shanghai Jiuyu Software Systems Co., Ltd., 201103, Shanghai, China

城市轨道交通线路连接着城市重要的客流集散点, 如商务区、铁路车站及长途客运站等。在大型活动或重要节假日期间, 由于出行客流的骤增, 一些功能性十分明显的车站, 很容易发生不同程度

的大规模客流突发聚集或消散。一旦某路段或某站点发生故障, 将对城市轨道交通网络中的乘客出行或线网运营产生直接或间接的影响^[1]。

目前, 国内外针对城市轨道交通的故障影响分析判断, 主要是根据站点工作人员的经验 and 感觉给出。在故障发生时, 由于工作人员经验不同, 其主观感觉判断存在差异, 不同的人对相同运营故障的可能持续影响时间判断会相差很大, 对客流分布估计也会与实际情况差距甚远。

本文从数据的采集、存储处理、分析、呈现等四个方面, 对城市轨道交通运营故障造成的客流影响进行体系性研究, 还对运营故障对于客流影响分析判断方法作了深入研究, 并提出了一套具有创新性的、可视化的运营故障影响分析系统。该系统能有效规避工作人员的主观差异, 客观地提供判断依据, 从而更人性化、更准确、更直观, 更智能地分析故障影响。

1 运营故障影响分析系统的工作原理

1.1 数据的采集、存储与处理

网络化运营是城市轨道交通发展成熟化的标志。网络化运营后, 故障的影响就不仅仅体现在故障发生现场了, 而是会传播到一个更大的范围。例如, 在某时刻, 网络中 A 站发生的故障将可能会使之前 B 站进站的乘客无法在 A 站出站。面对这种情况, 传统单点的客流数据采集已经无法满足分析的需求, 而是需要运用到大数据技术, 对整个网络的全量客流数据进行采集。

客流影响分析需要的多源数据有: 车站位置数据、列车行车数据、天气数据、OD 数据(站点的进站数据及出站数据), 以及故障描述数据等。有些是静态数据, 有些是有信息系统接口数据, 还有些则是需要通过互联网去收集的数据^[2]。故障对不同

车站的影响与其距离远近没有正相关性;列车行车数据能反映网络中列车的到发时间和运行控制情况,1 列列车的故障停留会影响该线路其它列车的运行,从而可以在时间和空间上推算客流分布;天气好坏在一定程度上能决定乘客乘坐城市轨道交通或者寻找其他代步工具的主观愿望;OD 数据从宏观上能反映出某个站点总的进站和出站客流量;故障描述数据记录了运营故障事件。

结构化的数据和非结构化的数据(如故障描述)需要分别采用不同的存储方式。而数据存储方式不是本文所要重点研究的内容,故不作更具体的阐述。另外,客流分析对实时性要求较高,传统的数据统计处理方式已无法满足^[4];合理的数据清洗及预处理也将直接影响客流分析的效率和效果。因此,从准确性的角度出发,在预处理阶段应主要对数据进行去重和去异常操作。

1.2 分析模型的选择

在完成数据采集、数据存储与处理后,如何挖掘运营故障事件对城市轨道交通客流的影响是十分重要的问题。文献[5]分析了客流的时空特性后,设计了乘客需求模型和乘客行为模型用来预测事故发生后的客流变化情况,并以东京地铁为例,利用客流数据及地铁事故数据,探索分析每天的客流变化模式、受事故影响的乘客行为、事故后乘客流动预测等问题。文献[6]发现社交媒体中的事件报道与站点出站客流间存在一定联系,结合线性回归与 SARIMA 模型,提出了一种基于凸优化的方法。

本文采用文献[7-8]中的 LDA(Latent Dirichlet Allocation)模型对客流影响进行分析。LDA 是一种非监督机器学习技术,是一种主题生成模型,通过挖掘故障描述中“每个词以一定概率选择了某个主题,并从这个主题以一定概率选择某个词”这样一个过程来实现。

基于 LDA 模型的特点,运营故障影响分析系统选用 LDA 模型作为故障影响分析模型,根据历史运营故障的描述,结合实时运营故障的描述,能够准确地将该次运营故障划分到特定的故障主题中,预测得到该次故障的持续时间,进而计算在此时间段所影响到的整个网络客流分布情况。这样,能有效地避免工作人员由于经验不同给出的持续时间判

断差异大的主观影响。

1.3 数据呈现

数据呈现是数据分析效果的窗口,可直观、有效地对分析结果进行数据可视化,可为运营者提供参考并支持做出决策^[9-10]。

2 客流影响分析

2.1 构建 LDA 模型

基于 LDA 算法的思想,建立一个城市轨道交通故障信息主题模型的具体过程为:城市轨道交通运营事件的故障描述是由不同故障主题组成的;通过分析各类故障主题和事故信息中出现的词,可得各种故障主题的关键词;通过挖掘主题下关键词中的“故障持续时间”,可计算该类故障主题的持续时间。

为了使得预测的持续时间更准确,还可将发生运营故障时的天气数据添加到故障描述中。以第 m 个事故描述在主题 i 上的分布作为权重 $w_{m,i}$,结合对应主题的持续时间 $x_{m,i}$ 可以得到第 m 个事故的预测时间 t_m ,即:

$$t_m = w_{m,1} x_{m,1} + w_{m,2} x_{m,2} + \cdots + w_{m,i} x_{m,i} + \cdots + w_{m,k} x_{m,k} \quad (1)$$

式中:

k ——选定的故障主题个数。

采用基于主题模型的故障持续时间预测方法,可在故障发生时预测得到故障持续时间。这一预测结果将用在后续的故障影响分析当中。

2.2 故障影响分析

故障影响分析模型的输入参数包括故障发生地、发生时间及预测持续时间。按故障点对整个网络的影响,客流分布分为直接影响的客流分布和间接影响的客流分布。

分析直接影响客流分布时,首先,根据网络历史断面客流数据,计算在故障发生时刻时,网络中经过故障点的客流量 D_1 ,并对 D_1 进行数据处理,进而得到故障直接影响人数 R_1 ;然后,计算在故障持续时间内将要进入故障车站的客流量。取历史同时间段进入故障车站的客流数据集,得到客流数据集 D_2 ;统计 D_2 的目的分布情况得到故障直接影响客流分布情况的输出值 R_2 :

$$R_2 = \{x_1, x_2\} \quad (2)$$

式中:

x_1 ——目的车站的名称代码;

x_2 ——受影响客流至该目的车站的人数。

分析间接影响的客流分布时,首先,根据故障预测持续时间,结合历史断面客流量,计算该时间段内将要经过故障点的客流量,形成客流数据集 D_3 ,进而得到间接受影响的人数 R_3 。由于 D_3 中包含了 OD(出发地和目的地)属性,故能统计出客流的来源车站分布和终点车站分布;结合 D_3 ,即可得到输出间接受影响 OD 对的客流分布 R_4 ,即

$$R_4 = \{y_1, y_2\} \tag{3}$$

式中:

y_1 ——途经车站名称代码;
 y_2 ——间接受影响客流途经车站的人数。

结合故障直接和间接影响的客流分布,可计算出运营故障对于客流的影响情况,包括故障直接影响的客流量、直接影响客流的目的地分布、间接影响的客流量及间接影响客流车站分布。故障影响分析流程如图 1 所示。值得注意的是,历史同时刻指在相同网络情况下(网络结构相同)的相同时刻(即相同月份、相同周、相同星期、相同时刻或相同时间段)。

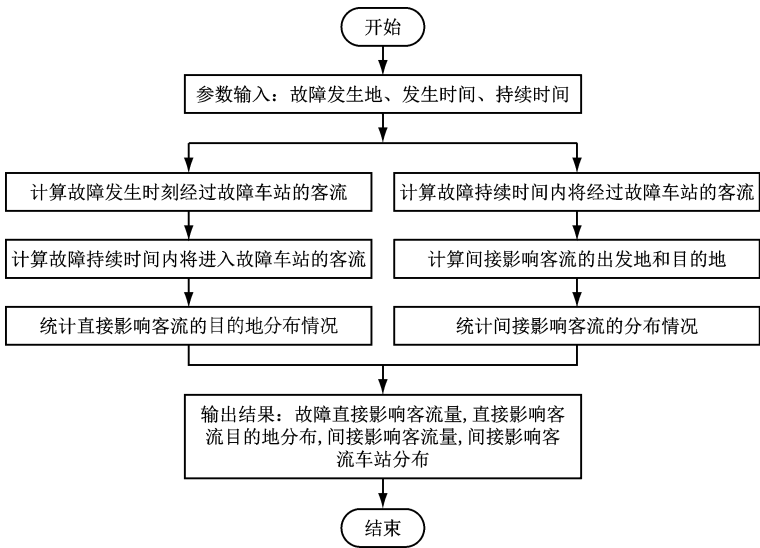


图 1 故障影响分析流程示意图

2.3 故障影响呈现

操作人员输入场景相关维度参数(包括故障车站、故障类型、发生时间等参数)后,数据呈现层根据特征数据在界面中呈现受影响车站和路径的客流情况。呈现方式包括时间推演、客流动态变化、受影响车站的闪烁提示,以及车站或路径的颜色及宽度变化等,并在辅助信息栏中显示故障车站、故障持续时间、受影响人数及受影响人群的线路分布等相关信息。

3 实例验证

该运营故障影响分析系统在上海轨道交通进行了实例验证。在运营故障影响分析系统主界面网络图中输入故障车站、故障类型和故障发生后,即可在城市轨道交通网络地图中突出显示故障发生车站,并根据客流影响大小进行了影响区域的颜色区分;在该系统主界面右侧信息栏显示故障车

站、预测的故障持续时间、受影响人数及受影响的站点分布等信息。

首先,在主界面的右侧上方位置输入故障信息。例如:发生时间为 08:57,故障类型为轨旁设备,发生位置为中山公园 04 号线,处理状态为已处理。然后,根据输入信息,主界面右侧下方位置会显示主要受到影响站点的分布情况。其中,上海火车站、金远路、莘庄等站的客流情况都受到了不同程度的影响。在主界面右侧中间位置,显示了当前事故导致的不同线路上受到影响客流量情况。系统显示,7 号线与 1 号线受到的影响更为明显。为了让使用人员更直观地分析故障造成的影响,线网图中还标出了事故发生地点和受影响的路段及站点,并利用颜色来表示受影响的程度,越趋向于红色说明影响越大,越趋向于绿色说明影响越小。

实例的运用展示说明:本系统有效地实现了故障对客流影响的提前预估,并展现了影响情况。

4 结语

基于大数据技术的运营故障影响分析系统,能够对城市轨道交通运营故障时的客流影响情况进行可视化分析。本文从数据采集、数据存储与处理、数据分析、数据呈现等 4 个方面进行了论述。运营故障影响分析系统以 LDA 模型作为故障影响分析模型,能有效预测运营故障及天气条件所造成的对客流的持续影响情况,可避免人为主观因素造成的“感觉误差”。根据 LDA 模型的故障持续时间预测,运营故障影响分析系统可计算该故障持续时间内整个网络受影响的客流分布情况,并通过友好的界面将故障车站、故障持续时间、影响人数、受影响的线路分布等直观呈现出来,为相关工作提供参考。

参考文献

- [1] BROWN D E. Text mining the contributors to rail accidents [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(2): 346.
- [2] 李巍. 半结构化数据挖掘若干问题研究[D]. 长春: 吉林大学, 2013.
- [3] CEAPA I, SMITH C, CAPRA L. Avoiding the crowds: under-

standing tube station congestion patterns from trip data [C]// Proceedings of the ACM SIGKDD international workshop on urban computing. ACM, 2012: 134.

- [4] 陈争航. 分布式数据库系统研究与应用[D]. 成都: 西南交通大学, 2002.
- [5] NI M, HE Q, GAO J. Forecasting the subway passenger flow under event occurrences with social media [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(6): 1623.
- [6] YOKOYAMA D, ITOH M, TOYODA M, et al. A framework for large-scale train trip record analysis and its application to passengers' flow prediction after train accidents [C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Berlin: Springer, Cham, 2014: 533-544.
- [7] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993.
- [8] BLEI D M. Probabilistic topic models [J]. Communications of the ACM, 2012, 55(4): 77.
- [9] 曾悠. 大数据时代背景下的数据可视化概念研究[D]. 杭州: 浙江大学, 2014.
- [10] 张鲁鲁. 多维数据可视化方法研究[D]. 北京: 北京交通大学, 2016.

(收稿日期: 2018-07-09)

(上接第 30 页)

型及编组、车站分布、运行组织模式等关键技术指标进行研究,得出以下结论:

(1) 结合功能定位及客流特征, 4 号线推荐采用 120 km/h 作为列车运行速度目标值, B 型车 6 辆编组, 车厢内站立密度为 5 人/m²。

(2) 车站分布应根据本线在线网中的功能定位, 考虑与其它线路换乘需求及大型枢纽集散客流需求, 合理选择站间距, 并充分结合用地规划, 引导城市土地开发, 发挥 TOD 效应。

(3) 应结合客流特征、快慢车开行比例及越行站的设置等因素, 综合考虑时间与经济效益, 确定合理的运行组织方案。4 号线经过综合分析, 推荐采用站站停运行组织模式, 并组织大小交路运行。

参考文献

- [1] 王灏. 关于城市轨道交通快线发展的研究 [J]. 都市快轨交通, 2006, 19(3): 4.
- [2] 万学红, 李忍相, 冯爱军. 市域快速轨道交通的技术特征与标准研究 [J]. 城市轨道交通研究, 2016(6): 10.

- [3] 丁树奎, 姜传治. 北京市域快轨新机场线关键技术及建设标准研究 [J]. 都市快轨交通, 2016, 29(4): 12.
- [4] 胡志晖, 饶雪平. 关于市域轨道交通快速功能设计的探讨-以上海市轨道交通 11 号线南段为例 [J]. 交通与运输 (学术版), 2009(1): 22.
- [5] 李鸿战. 天津市域快速轨道交通建设模式的探讨 [J]. 铁道标准设计, 2015, 59(1): 47.
- [6] 荆敏. 市域轨道交通规划设计关键指标和主要运输组织模式研究 [D]. 北京: 北京交通大学, 2016.
- [7] 周庆源, 陈春娇. 东京京王线及京王电铁株式会社简介 [J]. 城市轨道交通研究, 2008(2): 15.
- [8] 潘寒川, 杨涛. 市域轨道交通快慢车组合运营的通行能力研究 [J]. 城市轨道交通研究, 2009(10): 48.
- [9] 中铁第四勘察设计院集团有限公司. 厦门市轨道交通 4 号线工程 (后溪至翔安机场段) 可行性研究报告 [R]. 武汉: 中铁第四勘察设计院集团有限公司, 2016.
- [10] 中国城市规划设计研究院. 厦门市轨道交通 4 号线工程可行性研究阶段客流预测 [R]. 北京: 中国城市规划设计研究院, 2016.

(收稿日期: 2017-07-12)