

基于云计算的城市轨道交通数据可视化方法及案例研究

曹惠茹¹ 成海秀² 刘永鑫³ 陈凤宜⁴ 詹锡隆⁴

¹(广州工程技术职业学院信息工程学院 广东 广州 510075)

²(华南理工大学计算机科学与工程学院 广东 广州 510006)

³(深圳大学计算机与软件学院 广东 深圳 518060)

⁴(中山大学南方学院 广东 广州 510970)

摘要 针对城市轨道交通票卡数据流无法快速处理现状,基于云计算平台提出一种城市轨道交通大数据可视化分析方法。基于数据 ID 将城市轨道交通客流数据按时间段进行分表处理;基于客流关键字段建立索引数据集并将必要客流数据字段放入对应的索引中;通过删除冗余、错误数据实现对数据筛选和清洗处理。利用 ECharts 中的 JS 插件对城市轨道交通客流数据进行可视化呈现。以 JSP 容器与 Web 服务器协同构建城市轨道交通数据可视化移动终端。以目标城市为例,对可视化系统进行验证。

关键词 城市轨道交通客流量 可视化系统 数据分析 数据清洗 数据预处理

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2021.02.006

VISUALIZATION AND CASE STUDY OF URBAN RAIL DATA BASED ON CLOUD COMPUTING

Cao Huiru¹ Cheng Haixiu² Liu Yongxin³ Chen Fengyi⁴ Zhan Xilong⁴

¹(College of Information Engineering, Guangzhou Institute of Technology, Guangzhou 510075, Guangdong, China)

²(School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, Guangdong, China)

³(College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China)

⁴(Nanfeng College of Sun Yat-Sen University, Guangzhou 510970, Guangdong, China)

Abstract As the current situation that the ticket card data flow of urban rail can't be processed rapidly, we propose a big data visualization analysis method of urban rail based on the cloud computing platform. The subway passenger flow data was divided into tables according to the time period based on the data ID; the index data set was established based on the key field of passenger flow, and the necessary passenger flow data field was put into the corresponding index; data selection and cleaning were realized by removing redundancy and error data; the JS plug-in of open source E-charts was used for the visualization design of subway passenger flow data; a metro data visualization mobile terminal was built with JSP container and Web server. Taking the target city as an example, the visualization system is validated.

Keywords Subway passenger flow Visual system Data analysis Data cleaning Data preprocessing

0 引言

随着城市规模的不断增加,城市轨道交通成为越来越多城市居民出行的首选^[1-3]。运用云计算、大数

据等信息技术对城市轨道交通客流量的数据进行分析,研究城市轨道交通客流量变化态势成为当前研究的热点。以此为依据对规划城市轨道交通进行建设与运营,不仅会降低城市轨道交通建设与运维成本,而且为城市居民的出行带来更多便利^[4-6]。城市轨道交通

客流的数据可视化,成为上述工作的关键步骤之一。因此,对城市轨道交通客流进行可视化研究是城市轨道交通数据分析的重要环节,具有重要意义。

针对城市交通数据可视化问题,国内外从不同的层面展开相关研究。文献[7]设计了一种基于 GPS 的出租车轨迹数据的三维动态可视化方法,建立了基于时空双重控制的粒子系统,构建了城市大尺度三维场景中 GPS 轨迹的模拟移动模型。文献[8]为解决公共自行车数据可视化系统的性能问题,整合 Web 服务器集群、数据库集群、缓存框架服务器的集群架构,构建基于该架构的公共自行车数据可视化系统。文献[9]先后采用数据预处理、匹配技术、聚集计算等将车辆运行状态转化为道路通行信息,实现公路交通的数据可视化。文献[10]分析了珠三角地区公交系统位置数据丢失的空间分布特征,然后采用数据可视化技术对目标城市的交通数据进行了数据可视化,证明所提方法可以有效地提高数据质量。文献[11]使用快速增量模型树漂移检测方法来分析和预测英国交通人流量,同时在所开发的预测系统的基础上,将生成的传感器点内的交通流状态预测在真实的地图仿真中可视化。上述文献对交通层面的数据可视化进行了研究,然而针对城市轨道交通票卡数据流可视化研究偏少。同时城市轨道交通数据存在数据量巨大、种类繁多、数据实时性高等特点,对数据可视化提出新的挑战与要求^[12-13]。

针对城市轨道交通数据的特点和城市轨道交通客流量新挑战,本文提出基于云计算的城市轨道交通客流量数据可视化解决方案。将城市轨道交通客流量数据按时间段进行分表处理、建立索引数据集合;删除冗余、错误数据实现对数据筛选和清洗处理;对城市轨道交通客流量数据进行了可视化设计;以目标城市为例,对可视化系统进行了验证,并对客流规律进行了分析。

1 基于云计算的城市轨道交通数据系统框架

云计算采用简化接口,将海量计算任务通过通信网络分发到多个并行计算机上进行处理,以提高数据处理的实时性^[3,14-15]。用户访问分布式服务器可以减小延时,增强访问的实时性。为此,本文设计了如图 1 所示的基于云计算与数据存储的城市轨道交通客流数据系统框架。该框架主要包括用户端、云服务器、管理中心三个模块。用户端是指访问城市轨道交通数据库

移动端,主要以出行的客流人群为主体。云服务器用于处理、存储城市轨道交通数据,具体分为客户、中心数据存储和计算三种类型的服务器。管理中心主要负责对整个系统的管理和运维。首先,为降低网络安全风险,保护核心数据,采用中心服务器和客户服务器模式完成用户端的数据访问。其次,以用户提供高效的数据访问服务为目标,在城市不同位置部署客户服务器,采用高速网络与中心数据服务器建立高效连接,进而构建高效的分布式客户服务器集群。

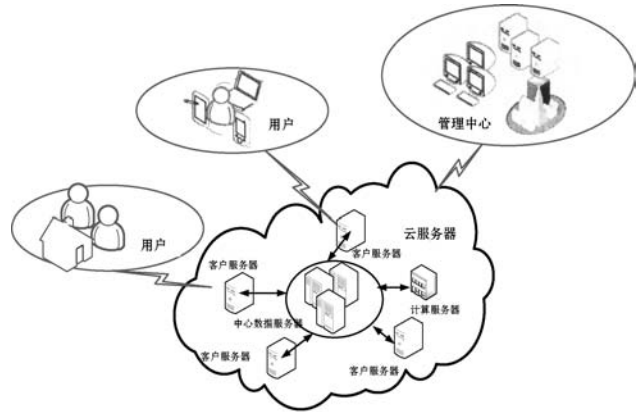


图 1 基于云计算服务器的城市轨道交通数据系统框架

不失一般性,客户服务器访问中心服务器可以进行如下描述。记 $S^U = \{s_1^u, s_2^u, \dots, s_k^u\}$ 为客户服务器; $S^C = \{s_1^c, s_2^c, \dots, s_l^c\}$ 为中心数据存储服务器; U 为用户库。当某个客户服务器 s_k^u 向数据服务器 S^C 请求数据访问时,自动在用户库 U 进行注册;注册的信息集合 u 包括:客户端服务器 IP 地址,要调用数据的入口地址 (Port)。数学描述如下:

$$U = U + u(s_k^u, IP, Port) \quad (1)$$

完成上述步骤之后, S^C 发送可以访问的中心数据存储服务器 $S_{available}^C$ 与 s_k^u 。最后, s_k^u 从 $S_{available}^C$ 随机选取某个服务器下载数据,同时在客户端服务器保留一份数据备份以供其他客户端服务器下载。从上述过程可以得出,客户访问服务器越多则效率越高,作为服务器提供者只需要一台被调用服务器即可。

从数据流的角度,基于云服务器城市轨道交通数据系统总的工作过程,可以从数据上传和下发进行描述。首先,城市轨道交通系统通过相应的装置,收集用户出行信息,并通过通信网络将数据上传至本地客户服务器。本地服务器周期性地数据上传给中心数据存储服务器,进而完成数据上传。其次,当用户需要访问相关交通数据时,客户服务器从中心数据服务器访问相关数据,并在本地进行数据备份存储,然后将用户需要数据下发至用户端,进而完成数据的数据下发。

2 轨道交通数据处理与可视化设计

2.1 系统概述

本系统主要用于对城市轨道交通客流量的可视化,方便对城市轨道交通客流量的各类数据进行分析,直观地分析出数据蕴含的规律。图2为城市轨道交通数据可视化总体设计过程。总体上来讲,该过程分为原始数据读取、数据预处理、数据可视化三个主要步骤。

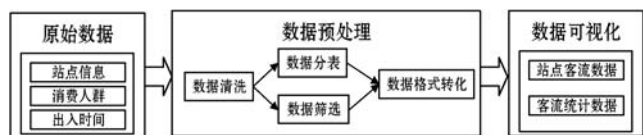


图2 系统设计过程

2.2 数据预处理

数据预处理与数据可视化是该系统重要模块。从数据存储服务器中读取的原始数据,为不同人群(ID)的刷卡时间(h)和进出站点记录($data$),其数据字段包括站点信息、消费人群、出入站点时间、乘车次数等关键性数据信息。在数据预处理阶段要对原始数据进行数据清洗,并且转换成JSON格式进而实现数据的可视化。根据数据特性,从多层次、多角度分析设计最终以Web页面形式进行效果展示。

在数据预处理层面,针对城市交通原始数据会存在缺失值、重复值等问题,在使用之前需要进行数据预处理。数据预处理没有标准的流程,通常针对任务和数据集属性的不同而不同。数据预处理的常用流程为:去除唯一属性,处理缺失值,属性编码,数据标准化正则化,特征选择,主成分分析。如图3所示,针对城市城轨交通数据的特点,采用如下所述的步骤进行数据预处理。

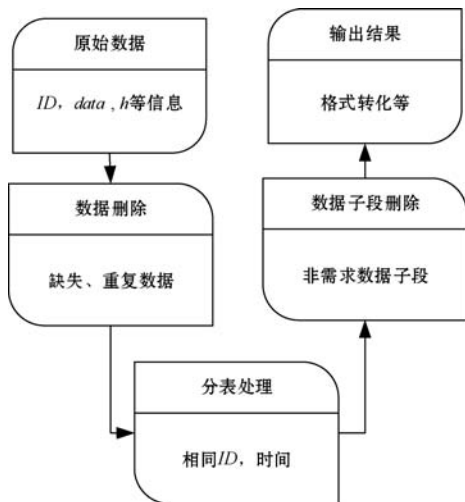


图3 数据预处理过程

不失一般性,记数据集为 $D = \{d_1, d_2, \dots, d_n\}$, 数据 $d_i = \{ID, data_i, h_i\}$ 。首先删除缺失值、重复值的数据,数学描述如下:

$$D = \begin{cases} D - d_j & \exists d_i = d_j \\ D - d_i & data_i = \text{NULL} \end{cases} \quad (2)$$

式中: $i \neq j \in n$ 。然后,按照读取数据ID,按时间为单位对数据进行分表处理,并建立索引。记 $data_i = \{ds_1, ds_2, \dots, ds_m\}$ 即数据 d_i 信息 $data_i$ 包含 m 个字段, DS 为所需信息字段。将需要用到 DS 包含在索引中,查询时只返回所需的字段,将不需要字段进行查询并且删除,从而实现数据预处理。其数学描述如下:

$$data_i = data_i - ds_j \quad ds_j \in DS, i \in m \quad (3)$$

2.3 数据可视化设计

城市轨道交通数据可视化是该系统的最后环节,直接面对用户端。通过对数据不同角度的分析和展示,以数据可视化形式展示了各种客流信息,进而为用户出行、城市轨道交通管理等提供参考依据。

图4为城市轨道交通数据可视化过程示意图。总体来讲,该过程可以分为数据检索、数据统计、数据图形内核调用、数据展示等步骤。具体来说,其关键步骤描述如下:

Step1 计算服务器根据可视化目标的具体要求,从数据存储服务器中读取与该目标相关,并通过预处理后的完整数据集。

Step2 以可视化目标为出发点,采用数据 ID 为关键字对统计数据关键信息进行检索。

Step3 完成上述步骤,建立检索结果表,并进行结果存储。

Step4 基于数学统计方法,调用ECharts统计模块内核,完成对客流关键信息的数据统计。

Step5 根据Step4的统计结果,建立统计客流数据与可视化图形之间的映射关系。

Step6 调用ECharts中相应图形的内核模块,按照不同统计结果,最终完成城市轨道交通数据可视化。

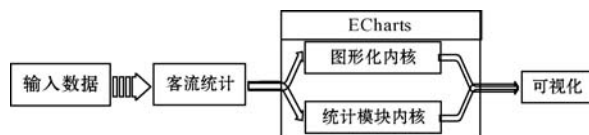


图4 数据可视化过程

3 应用案例与分析

3.1 城市轨道交通数据案例实现

以南方某城市轨道交通数据为案例,采用第2节所述方案进行了数据可视化实现。数据集为该城市

2015 年 1 月至 2 月的 9 条地铁线路乘客上下站点共计 750 000 条数据。对该数据集合进行清洗后以小时为单位对数据进行分表处理,进而建立客流小时统计表。然后对处理后的表进行格式转化,形成适合 ECharts 的 JSON 数据格式。最后,以客流站点热力图和统计客流量为目标进行可视化。

该案例的城市轨道交通数据可视化系统提供总览、线路对比和单线统计三种模式。总览模式采用地理数据可视化的热力图形式呈现所有站点的客流信息。首先以时间单位统计各个线路、站点客流量,利用不同的明亮程度表示各个站点不同路段拥堵程度(稀疏/正常/拥挤),如图 5 所示。在多线路对比模式下,在单位时间内对多条城市轨道交通线路的数据进行统计,然后采用柱状图的形式呈现不同线路的客流量。单线统计模式下,分别呈现选中线路各种客流量信息,其中包括该线路客流类型、各换乘站点日客流量、早晚高峰 Top5 出站点、一周内容客流量变化等方面的信息。

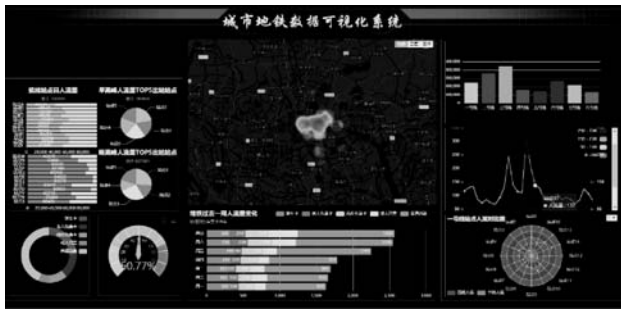


图 5 城市轨道交通客流量可视化系统界面

3.2 基于数据可视化的城市轨道交通客流分析

图 6 展示了城市轨道交通客流量在各个站点分布情况。用热力图来展示不同站点在一段时间内的总体客流量,从暗到亮来表示客流量密集程度,灰色表示客流量较小,亮的区域表示客流量较大。同时可将地图进行放大和缩小,从整体到局部了解不同站点的客流量的分布情况。可以看出,分布在商业密集区或居民区站点的客流较大。

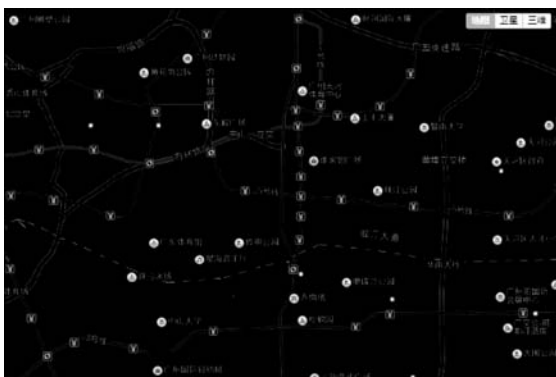


图 6 城市轨道交通客流量热力图

客流量高峰期的计算定义:早高峰时间为上午 7

点至 9 点,晚高峰时间为下午 5 点至 7 点半。因此将客流量数据按小时进行分类,提取各换乘站点与早、晚高峰热门站点数据并转换成 JSON 格式,进而以图形化形式展示出来,如图 7 所示。从早、晚高峰客流量饼形图可以看出,早高峰站点 1、3、4 为前三,晚高峰时站点 1、站点 3 客流量依然较大,而第三热门站点则由站点 5 替代早高峰的站点 4;而从换乘站点日客流量情况可以看出,左上方柱状表示日客流量在 60 000 以下的站点,而左下方柱状则为日客流量 60 000 以上的站点,且站点 R 和站点 S 为日流量最大的两个站点。因此,根据早、晚高峰以及各换乘站点日客流量情况分析,可合理调配相关工作人员以应对突发事件的处理。

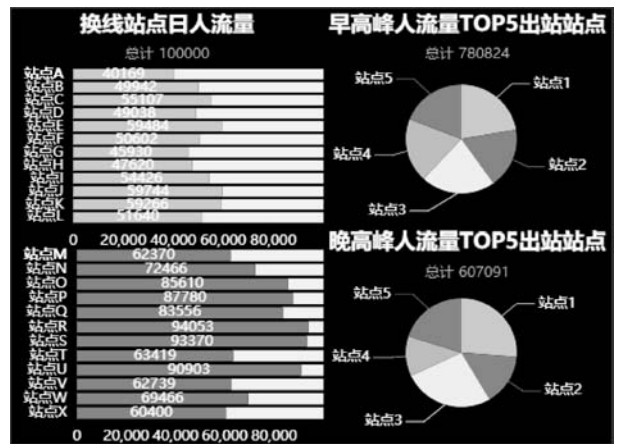


图 7 城市轨道交通热门换乘站点和早晚高峰热门出站点

图 8 为一号线各站点平时客流与高峰期客流的对比,可以看出,同一线路时高峰期的客流量明显高于平时的客流量,尤其突出表现在站点 9、站点 14、站点 15、站点 16。通过图中右上角下拉三角位置还可以选择查看不同线路的平时客流量与高峰期客流量的对比图。而上述站点均为换乘站点,说明换乘站点的客流量较大。

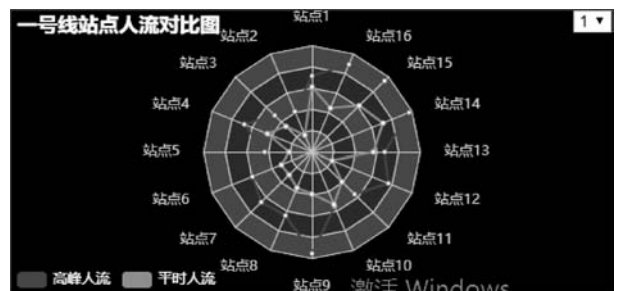


图 8 城市轨道交通同一线路不同站点对比图

4 结 语

本文以大数据环境下的城市轨道交通票卡数据流实时客流态势图的可视化为研究对象,基于云计算平台提供了一种城市轨道交通大数据可视化分析方法,

(下转第 49 页)

sponse for combined manufacturing system and heating, venting and air-conditioning (HVAC) system [J]. *Journal of Cleaner Production*, 2016, 135: 1650 – 1657.

- [8] 钟潇. 基于热舒适度的建筑空调用电优化研究[D]. 北京: 华北电力大学, 2016.
- [9] House J M, Smith T F. System approach to optimal control for HVAC and building systems [C]//Proceedings of 1995 American Control Conference-ACC' 95. IEEE, 1995, 6: 4326 – 4330.
- [10] 张青. 中央空调系统节能运行控制方法研究[D]. 南京: 东南大学, 2016.
- [11] Spitler J D, Fisher D E, Pedersen C O. The radiant time series cooling load calculation procedure [J]. *ASHRAE Transactions*, 1997, 103(3): 503 – 518.
- [12] Hosni M H, Wilkins C. Heat gain from office equipment [J]. *ASHRAE Transactions*, 2000, 6(42): 31 – 39.
- [13] 练斯甄. 冷水机组能效预测及节能优化运行研究[D]. 广州: 华南理工大学, 2014.
- [14] Swide D J. A comparison of empirically based steady-state models for vapor-compression liquid chillers [J]. *Applied Thermal Engineering*, 2003, 23(5): 539 – 556.
- [15] Brundage M P, Chang Q, Li Y, et al. Energy efficiency management of an integrated serial production line and HVAC system [J]. *IEEE Transactions on Automation Science and Engineering*, 2014, 11(3): 789 – 797.
- [16] Data-driven modelling and real time optimization for diversified chiller plants [EB/OL]. Advanced Singapore: Digital Sciences Center. [2019 – 03 – 06]. <https://publish.illinois.edu/chiller-modelling-and-optimization/home/>.
- [17] 许珊珊. 中央空调系统的控制策略研究 [J]. *内江科技*, 2018, 10(06): 42 – 44.
- [18] 刘志斌. 大型商场建筑中央空调冷冻水系统运行能效及参数优化研究[D]. 广州: 华南理工大学, 2016.

参 考 文 献

- [1] 李祖文, 何增镇, 苏红帆, 等. 基于自匹配模块的城市交通数据动态融合处理系统 [J]. *计算机应用与软件*, 2019, 36(6): 88 – 92.
- [2] Zhang T, Li D, Qiao Y. Comprehensive optimization of urban rail transit timetable by minimizing total travel times under time-dependent passenger demand and congested conditions [J]. *Applied Mathematical Modelling*, 2018, 58: 421 – 446.
- [3] Sun H, Wu J, Ma H, et al. A Bi-Objective timetable optimization model for urban rail transit based on the time-dependent passenger volume [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(2): 604 – 615.
- [4] Kyriakidis M, Hirsch R, Majumdar A. Metro railway safety: An analysis of accident precursors [J]. *Safety Science*, 2012, 50(7): 1535 – 1548.
- [5] Lee K, Jung W S, Park J S, et al. Statistical analysis of the metropolitan seoul subway system: Network structure and passenger flows [J]. *Physica A: Statistical Mechanics and Its Applications*, 2008, 387(24): 6231 – 6234.
- [6] Vijayan A, Islam S, Jones M, et al. Degradation of fluorescent high-visibility colors used in safety garments for the Australian railway industry [J]. *Journal of Safety Research*, 2016, 56: 1 – 7.
- [7] 王静雯, 艾廷华. 双比例尺交互模式的粒子系统动态可视化——以出租车数据为例 [J]. *测绘科学*, 2020, 260(2): 166 – 171.
- [8] 刘汪洋, 林菲, 马晓婷. 基于集群架构的公共自行车数据可视化系统 [J]. *传感器与微系统*, 2019, 38(2): 80 – 82.
- [9] 赵庶旭, 董亮. 城市交通 GPS 数据可视化分析 [J]. *计算机应用与软件*, 2016, 33(10): 279 – 283.
- [10] Liu Y, Weng X, Wan J, et al. Exploring data validity in transportation systems for smart cities [J]. *IEEE Communications Magazine*, 2017, 55(5): 26 – 33.
- [11] Wibisono A, Jatmiko W, Wisesa H A, et al. Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD) [J]. *Knowledge-Based Systems*, 2016, 93(1): 33 – 46.
- [12] Cao R, Cheng M, Mi C, et al. Modeling of a complementary and modular linear flux-switching permanent magnet motor for urban rail transit applications [J]. *IEEE Transactions on Energy Conversion*, 2012, 27(2): 489 – 497.
- [13] Yang X, Li X, Ning B, et al. A survey on energy-efficient train operation for urban rail transit [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(1): 2 – 13.
- [14] Al-Dhuraibi Y, Paraiso F, Djarallah N, et al. Elasticity in cloud computing: State of the art and research challenges [J]. *IEEE Transactions on Services Computing*, 2018, 11(2): 430 – 447.
- [15] Yang C, Huang Q, Li Z, et al. Big data and cloud computing: Innovation opportunities and challenges [J]. *International Journal of Digital Earth*, 2017, 10(1): 1 – 41.

(上接第 36 页)

搭建了实用化的应用平台并提供南方某城市应用案例。首先, 基于云计算构建了城市轨道交通数据系统框架。其次, 将城市轨道交通客流数据按时间段进行分表处理、建立索引数据集, 删除冗余、错误数据, 实现对数据筛选和清洗处理。最后, 以目标城市为案例构建了城市轨道交通数据可视化终端, 同时分析了城市轨道交通客流规律。实验结果表明, 该数据可视化方法的可行性高, 同时表明分布在商业密集区或居民区站点、换乘站点、早晚高峰的客流较大。该研究为城市轨道交通的规划和列车运营提供了可靠的依据, 通过交通状态的判别, 减缓交通压力, 均衡交通量分布, 从而提高城市轨道交通路网的利用效率和安全性。