

中文图书分类号: TP391

UDC: 004

学 校 代 码: 10005



硕 士 学 位 论 文

MASTERAL DISSERTATION

论 文 题 目: 面向轨道交通的空间密度与时空分布可视化分析研究

论 文 作 者: 孙功玉

学 科: 计算机科学与技术

指 导 教 师: 才智

论文提交日期: 2020 年 5 月

UDC: 004
中文图书分类号: TP391

学校代码: 10005
学 号: S201861205

北京工业大学工学硕士学位论文

题 目: 面向轨道交通的空间密度与时空分布可视化
分析研究

英文题目: RESEARCH ON VISUAL ANALYSIS OF
SPATIAL DENSITY AND
TEMPORAL-SPATIAL DISTRIBUTION FOR
RAIL TRANSIT

论 文 作 者: 孙功玉

学 科 专 业: 计算机科学与技术

研 究 方 向: 可视化分析

申 请 学 位: 工学硕士

指 导 教 师: 才智 副教授

所 在 单 位: 信息学部

答 辩 日 期: 2021 年 5 月

授 予 学 位 单 位: 北京工业大学

独 创 性 声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的科研成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：孙阳玉

日 期：2021 年 05 月 20 日

关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签 名：孙阳玉

日 期：2021 年 05 月 20 日

导师签名：孙

日 期：2021 年 05 月 20 日

摘要

随着中国经济以及城市化进程的快速发展,城市交通以及土地使用逐渐显露出各式问题,并且这些问题已经逐渐成为阻碍城市社会和经济可持续发展的障碍。同时,随着计算机技术以及互联网技术的快速发展,海量城市交通以及土地使用相关的数据已经被存储,如智能卡刷卡记录、兴趣点(Point of Interest, POI)数据等。可视化分析作为一种海量数据关联分析的工具,通过将抽象的数据信息转化成直观的图像信息并提供交互式操作,可以方便地向管理、规划决策者传达可理解的、有意义的信息,进而加速了信息的发现过程并为解决上述问题提供辅助支持。因此,本文基于 POI 数据和智能卡刷卡记录可视化分析城市轨道交通站点周边土地使用特征以及城市轨道交通客流量的时空分布特征,为解决上述问题提供辅助支持,主要研究内容如下:

(1) 提出了一种面向主题的非交叉层次 POI 提取方法。首先,基于 Voronoi 图确定每个城市轨道交通站点的覆盖区域并基于射线法确定每个站点覆盖区域内的 POI 集合。然后,为了提取相关主题的 POI 集合,基于相似度算法过滤站点覆盖区域的 POI 集合。然后,为了将站点覆盖区域的 POI 集合按照用户感兴趣程度进行分类,基于 Skyline 查询技术抽取层次 POI 集合。最后,为了避免不同层次 POI 构成的封闭区域在地理位置上相互交叉,基于 Graham 扫描法提取非交叉层次 POI 集合。

(2) 基于面向主题的非交叉层次 POI 提取方法以及三维等高线模型构建了一个可视化分析系统。通过该系统比较与分析不同站点周边土地使用特征的差异,为轨道交通的运营以及站点周边土地的开发与使用规划提供指导建议。同时,通过案例研究与实验分析验证系统的有效性。

(3) 设计了一个空间弦图,既可以表达 OD (Origin-Destination) 对象之间的关联关系又能表达 OD 对象蕴含的空间信息。同时,为了避免空间弦图中出现大量弦交叉的现象,提出了一种基于方向的多尺度空间聚合方法。该方法基于城市轨道交通站点的物理位置对站点进行聚合,并通过多个预设参数控制空间聚合的尺度。

(4) 基于多尺度空间聚合方法以及空间弦图构建了一个可视化分析系统。通过该系统比较与分析客流量时空分布特征,为分析土地使用的空间布局以及轨道交通系统的运营提供指导建议。同时,通过案例研究验证系统的有效性。

关键词: 可视化分析; Voronoi 图; Skyline 查询; 弦图; 空间聚合

Abstract

With rapid development of Chinese economy and urbanization, urban traffic and land use are gradually giving rise to many problems, and these problems have gradually become obstacles to the sustainable development of urban society and economy. Meanwhile, with rapid development of computer technology and Internet technology, mass data related to urban traffic and land use have been stored, such as smart card data, point of interest (POI), etc. As a tool of massive data association analysis, visual analysis can easily accelerate the process of information discovery and convey comprehensible, meaningful information to management and planning decision makers by converting abstract data information into visual image information and providing interactive operations, and provide auxiliary support to solve these problems. Therefore, based on POI data and smart card data, this paper analyze the land use characteristics around urban rail transit stations and the temporal-spatial distribution characteristics of urban rail transit passenger flow, so as to provide auxiliary support to solve these problems. The main research contents are as follows:

(1) A method of extracting uncrossed, hierarchical, topic-oriented POI set is proposed. Firstly, the coverage of each urban rail transit station is determined based on the Voronoi diagram and the POI set within the coverage of a station is determined based on the ray method. Then, in order to extract POI set of related topics, POI within a station's coverage is filtered based on similarity algorithm. Then, in order to classify the POI set within a station's coverage according to the user's interest, the hierarchical POI set is extracted based on Skyline query technology. Finally, in order to avoid the possibility that the closed region of different layers may cross each other in geographical location, the uncrossed, hierarchical POI set is extracted based on Graham scanning method.

(2) A visual analysis system is built based on the 3D contour model and method of extracting uncrossed, hierarchical, topic-oriented POI set. Through the analysis and comparison of the differences of land use characteristics around different stations, the system provides guidance and suggestions for the operation of rail transit and the development and use planning of land around the stations. The validity of the system is verified by case studies and experimental analysis.

(3) A spatial chord graph is designed, which can not only express the relationship between OD objects, but also express the spatial information contained in OD objects. Meanwhile, a multi-scale spatial aggregation method based on direction is proposed to avoid the intersection of a large number of chords in spatial chord graph. In this method, stations are aggregated based on the physical position of urban rail transit stations, and the scale of spatial aggregation is controlled by several preset

parameters.

(4) A visual analysis system is built based on the spatial chord graph and the multi-scale spatial aggregation method. Through the analysis and comparison of spatial-temporal characteristics of passenger flow distribution, the system provides guidance and suggestions for the analysis of the spatial layout of land use and the operation of rail transit system. The validity of the system is verified by case studies.

Keywords: visual analysis, Voronoi diagram, Skyline query, Chord glyph, spatial aggregation

目录

摘 要.....	I
Abstract.....	III
第 1 章 绪论.....	1
1.1 研究背景与意义.....	1
1.1.1 城市轨道交通以及土地使用的研究背景与意义.....	1
1.1.2 可视化分析的研究背景与意义.....	2
1.2 国内外研究现状.....	3
1.2.1 土地使用研究现状.....	3
1.2.2 OD 数据可视化研究现状.....	4
1.3 本文主要研究内容.....	5
1.4 本文组织结构.....	7
第 2 章 背景知识与相关技术介绍.....	9
2.1 Voronoi 图技术概述.....	9
2.2 Skyline 查询技术概述.....	10
2.3 Graham 扫描法概述.....	11
2.4 弦图概述.....	13
2.5 本章小结.....	15
第 3 章 面向主题的非交叉层次 POI 提取方法.....	17
3.1 非交叉层次 POI 提取方法.....	17
3.1.1 POI 集合划分方法.....	17
3.1.2 层次 POI 提取方法.....	18
3.1.3 非交叉层次 POI 提取方法.....	21
3.2 面向主题的非交叉层次 POI 提取.....	24
3.2.1 面向主题的非交叉层次 POI 提取方法.....	24
3.2.2 基于主题的颜色值生成方法.....	25
3.3 本章小结.....	26
第 4 章 面向轨道交通的空间密度可视化分析研究.....	27
4.1 可视化任务概述.....	27
4.2 可视化设计概述.....	28
4.2.1 POI 数据集介绍.....	28
4.2.2 可视化分析系统概述.....	28

4.2.3 三维等高线模型分析原则.....	30
4.3 案例研究与实验分析.....	30
4.3.1 案例研究.....	30
4.3.2 实验分析.....	34
4.4 本章小结.....	36
第 5 章 面向轨道交通的时空分布可视化分析研究.....	37
5.1 可视化任务概述.....	37
5.2 可视化设计概述.....	37
5.2.1 城市轨道交通刷卡记录数据集介绍.....	38
5.2.2 空间弦图设计.....	38
5.2.3 基于方向的多尺度空间聚合方法以及时间聚合方法.....	40
5.2.4 可视化分析系统概述.....	43
5.3 案例研究.....	45
5.4 本章小结.....	49
结 论.....	51
参 考 文 献.....	53
攻读硕士学位期间所发表的学术论文.....	57
致 谢.....	59

第 1 章 绪论

1.1 研究背景与意义

1.1.1 城市轨道交通以及土地使用的研究背景与意义

城市轨道交通以其速度快、污染少、安全、准时、大容量等优点，已成为大中型城市综合交通规划的重要组成部分以及公共交通发展的主要方向^{[1][2]}。21 世纪初，我国城市轨道交通建设发展迅速，截至 2020 年底，全国（不含港澳台）共计 44 个城市、233 条城市轨道交通线路投入运营，总里程 7545.5 公里。北京、上海等特大城市轨道交通运营总里程已超过 700 公里。近年来，随着客流量的增长，城市轨道交通管理面临的问题日益严峻。

(1) 高峰时段大部分城市轨道交通站点的瞬时客流量是巨大的。这会对乘客出行造成不便，一是高峰时段客流量大、候车时间较长，增加了乘客的通勤时间；二是高峰时段大部分线路的满载率都是大于 1 的，超载现象较为严重，这使得车厢变得异常拥挤，车厢内乘客的舒适度急剧下降。

(2) 部分城市轨道交通站点客流量大且辐射范围大。这会对乘客出行造成不便，一是由于客流量过大会增加乘客的候车时长；二是当乘客的目的地距离站点较远时需要中转其它交通工具。西二旗站是一个典型的客流量大且辐射范围大的站点，其辐射范围内有很多公司和住宅且它与部分公司、住宅的距离较远。因此，在西二旗站下车的大量乘客需要中转其它出行工具方可抵达目的地。

(3) 换乘站客流量较大。由于换乘站点承载着来自多条线路的客流量，所以换乘站的客流量往往要高于非换乘站。当城市轨道交通线路较多且复杂时，乘客的出行不可避免的要选择换乘站点，而换乘站的客流量又是较大的，这会对乘客的出行造成不便。

随着中国经济以及城市化进程的快速发展，城市土地使用存在的一些问题已经成为阻碍城市社会和经济可持续发展的障碍，如土地空间布局不合理、土地利用效率低下、土地需求增加与土地供给不足等问题。

(1) 城市土地使用空间布局不合理。城市土地使用类型可以分为住宅用地、工业用地、商业用地、城市绿地以及其它类型用地五种类型^[3]。目前，大部分城市这五种类型土地使用的空间分布是不合理的，比如住宅用地与商业用地相分离，即职住分离现象。职住分离使得居民通勤时间较长，截止 2020 年，北京市单程平均通勤时长可达 47 分钟。

(2) 土地需求增加与土地供给不足。随着经济的发展以及人口数量的增长，

大量的人口从农村移向城市。从 1978 到年 2010 年,城市居住人口占据总人口的比例已经从 19.7%增长到 49.7%^{[4][5]}。随着城市居住人口的大量增加,对城市土地的需求也在急剧增加。而城市土地资源是有限的,因此许多大城市正面临着土地需求的增长与土地供给不足的问题。

城市土地使用通常与社会功能是相关的^{[6][7][8]},并且人类活动与不同区域的社会功能是相互作用的^[9]。因此,城市土地使用会对人类活动产生影响,如公共交通出行^{[10][11][12][13]}。通过人类活动能够识别社会功能、推断城市土地利用,人类活动信息可以用城市轨道交通客流量数据来描述^{[14][15][16][17][18]}。因此,城市轨道交通出行与土地利用是相互影响的。Pan 等人^[19]基于上海市城市轨道交通客流量数据和手机信号数据各站点客流量研究城市轨道交通站点客流量与土地使用类型之间的相互影响,得出以下结论:(1) 客流量与就业密度、站点周边居民通勤距离呈正相关;(2) 开通时间较早且作为换乘节点的站点与客流量呈正相关;(3) 与周边商业发展结合较好的地铁站,其客流量较大。

因此,研究城市轨道交通站点周边土地使用是非常有必要的,其对轨道交通系统的运营以及站点周边土地的开发和使用规划具有指导意义。同时,研究城市轨道交通客流量的时空分布特征也是非常有必要的,其对分析土地使用的空间布局以及轨道交通系统的运营具有指导意义。

1.1.2 可视化分析的研究背景与意义

近年来,随着计算机技术以及互联网技术的快速发展,人们的日常生活已经逐渐交由各类现代信息系统进行管理,如出行、购物等。因此,各类现代信息系统每年收集的数据集越来越大且越来越复杂。这些海量且复杂的数据中蕴含着大量信息,但同时这些数据也是不易被理解的。因此,基于可视化分析探索、分析以及深入了解这些数据是非常有效的。可视化分析通过将抽象的数据信息转化成直观的图像信息并提供交互式操作,可以方便地向管理、规划决策者传达可理解的、有意义的信息,从而加速了信息的发现过程。目前对于海量且复杂数据的可视化研究主要存在以下两个挑战:

(1) 最有效的、可理解的图像被认为是相对较小、稀疏的、美观可读的图像布局,而且它们通常不会是具有数百或数千个顶点的图形。

(2) 用户解释大量视觉信息的能力是有限的,这意味着含有大量视觉信息的可视化分析的系统必须具备交互操作的能力,以允许用户对图像进行选择性的探索,保证可视化图形的复杂性对用户来说是可控的。

因此,基于可视化分析城市轨道交通站点周边土地使用现状以及城市轨道交通客流量的时空分布特征是很有价值的。通过直观的可视化图像以及丰富的可视

化交互操作,可以使得管理、规划决策者更方便地探索海量的客流量数据以及 POI (Point of Interest, 兴趣点) 数据所蕴含的丰富信息,从而为轨道交通系统的运营以及站点周边土地的开发和使用规划提供辅助支持。本文基于等高线思想提出了一种密度可视化方法,解决了传统的密度可视化方法中表达信息具有局限性的问题,如在热力图中通过颜色信息仅能够区分处于不同密度范围内的对象,而无法对同一密度范围内的对象进行区分。除此之外,本文设计了一个可视化图形,既可以呈现 OD (Origin-Destination) 对象之间的关联关系又可以呈现 OD 对象蕴含的空间信息。同时,提出了一种基于方向的多尺度空间聚合方法,用以保证可视化图形的清晰、可理解。

1.2 国内外研究现状

1.2.1 土地使用研究现状

目前,对于土地使用的研究,主要集中在城市功能区的识别。关于城市功能区识别的研究,主要是基于单一遥感数据或者社会感知数据进行的。

随着技术的飞速发展,高空间分辨率的遥感图像在城市区域功能提取和分析方面具有巨大的潜力。Bratasanu 等人^[20]在罗马尼亚进行了案例研究,首先引入了场景分类的概念,即从高分辨率遥感图像的光谱特征中提取虚拟词。Zhang 等人^[21]结合了 BoVW 模型和概率主题模型,利用低分辨率语义信息识别北京市和珠海市的城市区域功能。Li 等人^[22]基于深度学习神经网络对高空间分辨率遥感图像进行分类。遥感技术能够有效提取遥感地图中城市功能区的空间地理特征,是研究城市土地利用功能变化的关键技术。

虽然遥感技术在提取包括地表建筑和城市空间结构在内的物理特征方面具有良好的表现,但是它无法反映人类活动信息。城市区域的功能和人类活动是息息相关的。因此,仅通过遥感图像是很难准确识别城市功能区的。在一定程度上,社会感知和人类活动是城市规划中更好的动态感知和识别方法。POI 数据是一种经典的社会感知数据。Yuan 等人^[15]根据城市内的主要道路将城市划分为多个互不相连的区域,然后使用出租车轨迹数据和 POI 数据建立了一个主题推理模型来识别单个区域的功能。Hu 等人^[23]基于开放街道地图中包含的道路网络提取出地块。利用来自 POI 数据的 10 个特征和 OLI 图像的 2 个特征,将地块划分为 8 个 I 级和 16 个 II 级土地利用类别。在分类过程中,采用相似性测量和阈值法对土地利用类型进行识别。Yao 等人^[24]在交通分析区域尺度上建立了一个将 POI 纳入 World2Vec 模型的框架,并以中国广东为例进行了验证。Zhang 等人^[7]提出了一种层次语义认知结构作为识别城市功能区划的一般认知结构。利用高分辨率卫星

图像数据和 POI 数据，基于层次语义认知结构对功能区进行分类。

本文根据城市轨道交通站点将城市划分为多个互不相连的区域，基于 POI 数据研究站点周边的土地使用密度以及功能。由于功能区具有混合性，一个站点往往是多个功能的混合。因此，本文并未对功能进行分类，也没有将一个区域限定到一个功能上。而是设计了系统以支持用户交互式的探索与分析一个站点的功能特征。

1.2.2 OD 数据可视化研究现状

OD 数据通过指定出行的出发地点、出行的出发时间、出行的目的地点以及抵达目的地的时间来描述用户在空间位置上的移动，需要注意的是，OD 数据是不包括具体的出行路线的。目前常用的 OD 数据可视化方法可以概括为三种，flow map、OD 矩阵以及 OD map。下面将介绍这三种 OD 数据可视化方法以及它们的研究现状。

Flow map 通过直线或者曲线连接 OD 数据中的起点和终点，线的宽度或者颜色用以表示 OD 流量的大小^[25]。第一批著名的 flow map 可视化图形是在 20 世纪 80 年代提出的^{[26][27]}。因为这些可视化图形并未对线条进行合并，所以随着 OD 对象的增加，会出现大量线条交叉的现象，大大降低 flow map 的可读性。S.van 等人^[28]通过选择和聚合对 flow map 进行视觉化简。边缘绑定是图形学领域中的一个热门研究话题^[29]，它通过对空间位置相邻的流进行合并或分组可以实现 flow map 在视觉上的化简。Phan 等人^[30]提出了一种基于边缘绑定的层次聚类算法，并基于该算法绘制了一种单源的 flow map。由于该算法使用了一种特殊的方法来路由边缘，所以交叉现象是无法避免的。Kevin Verbeek 等人^[31]基于螺旋树提出了一种边缘绑定算法。螺旋树可以诱导目标进行聚类以及线条的平滑，大大提高了视觉的可读性。除此之外，还可以通过删除连接起点与终点的线来避免 flow map 在视觉上的混乱。I. Boyandin 等人^[32]使用一个地图中展示流的起源，另一个地图中展示流的目的地，在两个地图之间使用一个表用于展示时间序列起点与终点之间的流信息来避免 flow map 在视觉上的混乱。

在 OD 矩阵中，行和列都表示位置，每个单元格皆表示该单元格的行和列之间的流量，可以使用单元格的颜色来表示流量的大小。同时，它支持对矩阵的行和列进行重新排序，以便发现隐藏在数据内的行为。Guo 等人^[33]利用 OD 矩阵对不同地区居民的日常活动和疾病传播进行可视化，为预防传染病传播提供策略。在他们的研究中，他们用一种特殊的方式对旅行者每日移动 OD 矩阵的行和列进行重新排序，以获得对角线上的“最佳”排序区域。但是 OD 矩阵也是存在一些不足之处的，在 OD 矩阵中空间位置信息会被遗漏并且 OD 矩阵可展示的单元格

的数量是有限的。在南卡罗莱纳大学的一些学者的研究中^[34]，他们将 flow map 与重新排序矩阵交互链接，实现了地理位置信息的可解释性。

OD map 通过构建一个网格化的二维空间树状图来保存所有起点和终点的空间布局。具体来说，OD map 中的位置是按矩形布局的，每个位置由一个矩阵单元表示。任何矩阵单元都由一个与整个矩阵结构相同的小矩阵填充，它表示从(到)该位置到(从)所有其他位置的流量。Wood 等人^[35]是第一个提出 OD map 概念来解决源头与终点之间的流量交叉问题。虽然 OD map 是解决 flow map 中起点与终点线交叉、阻塞以及 OD 矩阵中缺乏空间上下文等问题的好方法，但其缺点也很明显，在使用 OD 图分析具体问题时，需要反复观察不同的矩阵。当位置增加时，视觉可读性不佳。

本文设计了一个 flow map 类型的可视化图形，既可以呈现 OD 对象之间的关联关系又可以呈现 OD 对象的空间信息。为了保证可视化图形的清晰、易理解，本文提出了一种基于方向的多尺度空间聚合方法来减少 OD 对象之间的线条数量。

1.3 本文主要研究内容

可视化分析通过将抽象的数据信息转化成直观的图像信息并提供交互式操作，可以方便地向管理、规划决策者传达可理解的、有意义的信息。本文以城市轨道交通作为研究对象，基于 POI 数据以及智能卡刷卡记录进行可视化分析研究。具体的研究内容可以划分为两个模块，面向轨道交通的空间密度可视化分析研究以及面向轨道交通的时空分布可视化分析研究，本文技术路线与研究内容如图 1-1 所示。

本文的第 3 以及第 4 章对面向轨道交通的空间密度可视化分析进行了研究。首先提出了一种面向主题的非交叉层次 POI 集合提取算法。然后基于三维等高线模型对其进行可视化并以 GIS 地图为底构建了一个可视化分析系统，系统构建过程如图 1-1 左侧所示。具体工作内容如下：

(1) 提出了一种 POI 集合划分算法。首先采用 Voronoi 图技术进行区域划分，以城市轨道交通站点为生成元绘制 VD，在绘制 VD 的过程中即可获得所有城市轨道交通站点的覆盖区域，即一组边界点。然后基于射线法确定每个站点覆盖区域内的 POI 集合。

(2) 提出了一种非交叉层次 POI 集合提取算法。首先基于 Skyline 查询技术抽取层次 POI 集合。因为基于 Skyline 查询技术提取的层次 POI 集合位于不同层次 POI 构成的封闭区域在地理位置上可能是相互交叉的，所以在此基础上基于

Graham 扫描法提取非交叉层次 POI 集合。

(3) 提出了一种面向主题的非交叉层次 POI 集合提取算法。面向主题的非交叉层次 POI 提取方法是在非交叉层次 POI 提取方法的基础上实现的，在提取层次 POI 前使用相似度算法对 POI 进行过滤。

(4) 基于三维等高线模型可视化面向主题的非交叉层次 POI 集合，并构建了一个可视化分析系统，用于分析与比较不同站点周边土地使用特征的差异。同时，设计了两个案例研究以及一个实验分析验证系统的有效性。

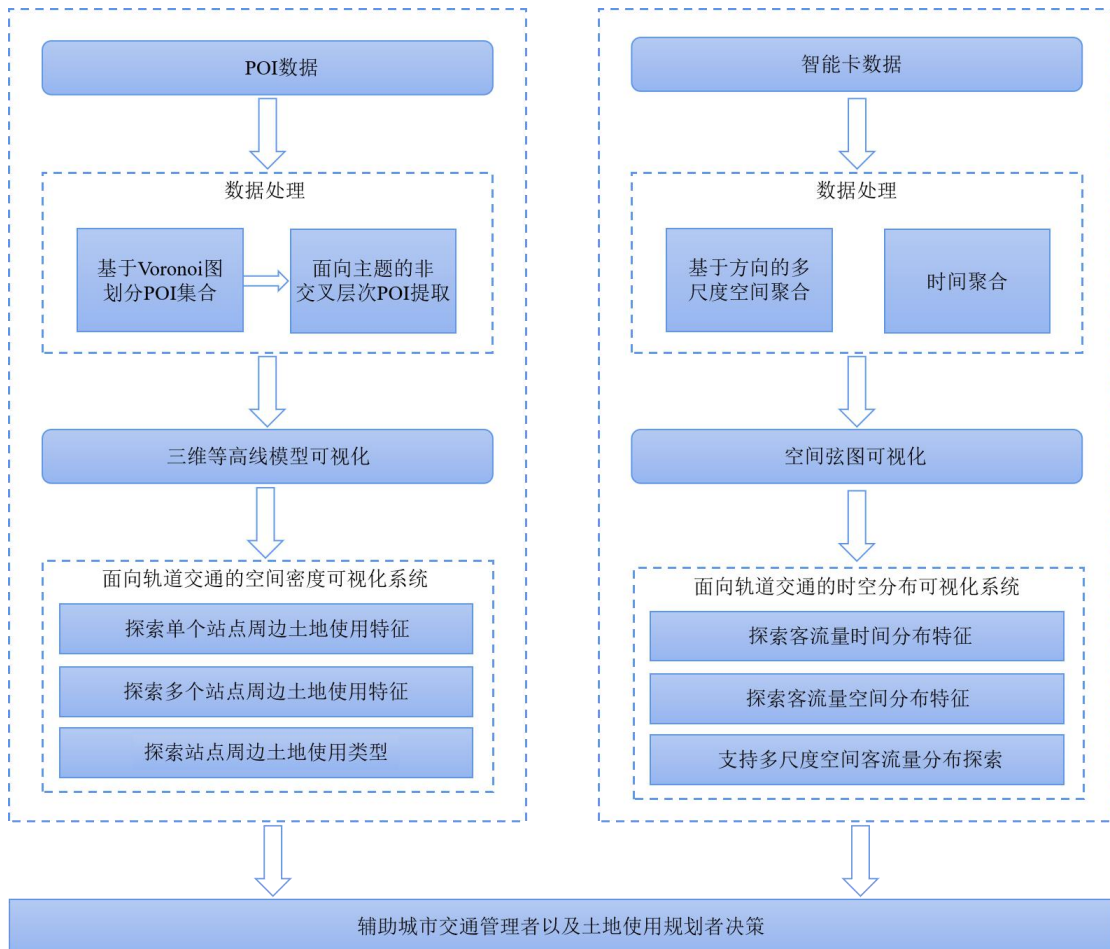


图 1-1 本文技术路线与研究内容

Fig. 1-1 The technical route and research content

本文的第 5 章对面向轨道交通的时空分布可视化分析进行了研究。首先提出了一种基于方向的多尺度空间聚合方法，然后设计了一个空间弦图对时空聚合后的数据进行可视化并构建了一个可视化分析系统，系统构建过程如图 1-1 右侧所示。具体工作内容如下。

(1) 设计了一个空间弦图，既可以表达 OD 对象之间的关联关系又能表达 OD 对象蕴含的空间信息。同时，为了避免空间弦图中出现大量弦交叉的现象，提出了一种基于方向的多尺度空间聚合方法。该方法基于城市轨道交通站点的物理位

置对站点进行聚合，并通过多个预设参数控制空间聚合的尺度。

(2) 基于多尺度空间聚合方法以及空间弦图构建了一个可视化分析系统。通过该系统比较与分析客流量时空分布特征，为分析土地使用的空间布局以及轨道交通系统的运营提供指导建议。同时，通过案例研究验证系统的有效性。

1.4 本文组织结构

本文一共含有 5 个章节，各个章节的具体安排如下：

第 1 章：绪论。本章首先介绍了本文的研究背景与意义。然后介绍了土地使用以及 OD 数据可视化在国内外的研究现状。最后介绍了本文的主要研究内容以及本文的组织结构。

第 2 章：背景知识与相关技术介绍。本章首先介绍了本文第 3 章中实现的非交叉层次 POI 提取方法中所涉及到的一些关键技术，其中包括 Voronoi 图技术、Skyline 查询技术以及 Graham 扫描法。然后对本文第 5 章中提出的空间弦图所参考的弦图进行了介绍。

第 3 章：面向主题的非交叉层次 POI 提取方法。首先，基于 Voronoi 图确定每个城市轨道交通站点的覆盖区域并基于射线法确定每个站点覆盖区域内的 POI 集合。然后，为了提取相关主题的 POI 集合，基于相似度算法过滤站点覆盖区域的 POI 集合。然后，为了将站点覆盖区域的 POI 集合按照用户感兴趣程度进行分类，基于 Skyline 查询技术抽取层次 POI 集合。最后，为了避免不同层次 POI 构成的封闭区域在地理位置上相互交叉，基于 Graham 扫描法提取非交叉层次 POI 集合。

第 4 章：面向轨道交通的空间密度可视化分析研究。本章首先介绍了面向轨道交通的空间密度可视化分析系统需要完成的任务。然后介绍了 POI 数据集以及可视化分析系统的详细设计。最后，通过两个案例研究以及实验分析验证系统的有效性。

第 5 章：面向轨道交通的时空分布可视化分析研究。本章首先介绍了面向轨道交通的时空分布可视化分析系统需要完成的任务。然后依次介绍了系统构建过程中用到的基础数据集、可视化图形以及数据处理方法。即城市轨道交通刷卡记录数据集、空间弦图的设计以及多尺度空间聚合方法。然后介绍了可视化分析系统的详细设计。最后通过案例研究验证系统的有效性。

第2章 背景知识与相关技术介绍

本章首先介绍了本文第3章中介绍的面向主题的非交叉层次POI提取方法中涉及到的一些关键技术,其中包括 Voronoi 图技术、Skyline 查询技术以及 Graham 扫描法。然后对本文第5章中提出的空间弦图所参考的弦图进行了介绍。

2.1 Voronoi 图技术概述

Voronoi 图, 又称 Dirichlet 或者泰森多边形, 是一种用于描述空间邻近关系的数据结构^[36]。其最初是由俄罗斯数学家 M.G.Voronoi 提出的, 而后被广泛应用于计算机科学、数学、物理、化学等多个领域^[37]。本节主要介绍 Voronoi 图的相关概念及其在区域划分方面的应用, 图 2-1 为 Voronoi 图的示意。

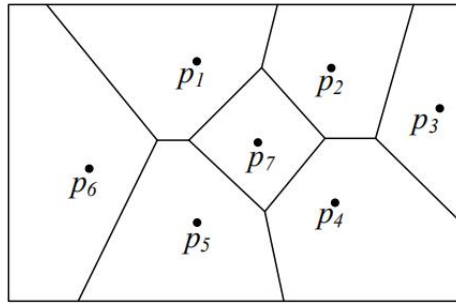


图 2-1 Voronoi 图示意

Fig. 2-1 Demo of Voronoi diagram

定义 2-1 (VD 与 VP) Voronoi 图 (Voronoi Diagram, VD) 基于目标点的最邻近原则将整个空间划分为多个区域^[38]。假设二维平面上存在一个点集 $P = \{p_1, p_2, \dots, p_i, \dots, p_n\}$, 点 p_i 所处的 Voronoi 多边形 (Voronoi Polygon, VP) 为到点 p_i 欧式距离最小的点的集合, 如公式 2-1 所示。

$$V(p_i) = \{x \in V(p_i) \mid \text{dist}(x, p_i) \leq \text{dist}(x, p_j), j \neq i, j = 1, 2, \dots, n\} \quad (2-1)$$

P 的 VD 为该点集构成的所有 VP 的集合, 如公式 2-2 所示。

$$V(D) = \{V(p_1), V(p_2), \dots, V(p_n)\} \quad (2-2)$$

由上述关于 VD 和 VP 的定义可知, VD 具有如下性质:

- (1) 每个 VP 内有且仅有一个生成元, 点 p_i 即为生成元。
- (2) 不同 VP 之间的交集为空集; 相邻的 VP 之间除边界外, 交集为空集。
- (3) 对于任意的 VP 来讲, 该多边形内任意一点到该多边形的生成元的距离是最小的。

(4) 位于 VP 边上的点到两侧生成元的距离是相等的。

通过上述 VD 和 VP 相关概念可知 VD 可以被用来进行区域划分。近年来,越来越多的来自各个领域的学者使用 VD 进行区域划分。马少蒙^[39]以北京市铁路内环线上的站点作为 VD 的生成元划分北京市的物流配送区,进而确认每个站点的覆盖面积。斐秋艳^[40]基于 VD 进行区域划分,进而确定受灾区域人员最合适的集中撤离点。任静^[41]以临界点作为生成元对海洋流场进行区域划分,进而得到每个临界点的影响范围,这将会有助于更好的理解复杂的海洋流场现象。

2.2 Skyline 查询技术概述

Skyline 查询技术最初是由 M.Shimrat 等人提出的^[42],是一种基于多维度决策从数据库中筛选出用户感兴趣的数据集合的数据库查询技术,被广泛应用于多目标优化问题^[43]。本节主要介绍 Skyline 命名的来源、Skyline 查询技术涉及的关键定义以及 Skyline 查询技术的经典实现算法。

当我们站在大海的一岸望向另一岸时,我们视野范围内所能看见的建筑物要么是靠近海岸线的建筑物要么是高耸的建筑物,即我们可以看见的只是那些没有被其它建筑物遮挡的建筑物。这个思想与 Skyline 查询技术不谋而合,即 Skyline 查询技术会从数据库中筛选出不被其它数据元组支配的数据集合。这些我们所能看到的建筑物构成了对岸风景的天际线, Skyline 即取名于此。

定义 2-2(支配)已知两个数据元组 p_1, p_2 均有 n 个属性,其中 $p_1 = (a_1, a_2, \dots, a_m, \dots, a_n)$ 且 $p_2 = (b_1, b_2, \dots, a_m, \dots, b_n)$, 那么 p_1 支配 p_2 当且仅当满足 $\forall i \in \{1, 2, \dots, m, \dots, n\}$ 使得 $a_i \leq b_i$ 成立且 $\exists j \in \{1, 2, \dots, m, \dots, n\}$ 使得 $a_j < b_j$ 成立。

由上述定义可知,支配满足如下性质:

- (1) 支配具有传递性,已知 p_1 支配 p_2 且 p_2 支配 p_3 , 那么 p_1 是支配 p_3 的。
- (2) 已知 p_1 无法支配 p_2 , 那么 p_2 亦无法支配 p_1 。

定义 2-3 (Skyline 查询) Skyline 查询是指是从给定的数据集合中选择出那些未能被任何其它数据元组支配的数据元组的集合。该集合被称为 Skyline 数据集, Skyline 数据集中的每一个数据元组皆被称为 Skyline 点。

下面以一个经典的实例对 Skyline 查询进行阐述。假设某个用户想去 Nassau 度假,他希望可以找到一家酒店,其距离海滩是更近的且消费价格是更便宜的。事实上,往往一个酒店距离海滩越近它的消费价格越昂贵,为了能够筛选出“最佳”酒店给用户,可以应用 Skyline 查询技术。假设待选择的酒店集合 $hotels = \{a, b, c, d, e, f, g, h, i, k, l, m, n\}$, 且酒店与海滩之间的距离 $distance$ 以及酒店的消费价格 $price$ 如表 2-1 所示。

表 2-1 Skyline 查询示例数据
Tab. 2-1 Sample data of Skyline query

	a	b	c	d	e	f	g	h	i	k	l	m
distance	1	2	4	6	8	5	4	3	9	10	6	8
price	9	10	8	7	5	6	3	2	1	4	2	3

通过计算得, a 、 i 、 h 三个酒店是未被任何其它酒店支配的 Skyline 点, 即 a 、 i 、 h 三个 Skyline 点从距离以及价格两个维度是不输于其它酒店的。因此, 用户只需从这三个 Skyline 点中选出其最感兴趣的酒店即可。图 2-2 即为该实例的 Skyline 查询示意图, 其中纵轴 $price$ 表示酒店的消费价格, 横轴 $distance$ 表示酒店与海滩之间的距离, 圆点表示酒店, 折线上的点表示 Skyline 点。

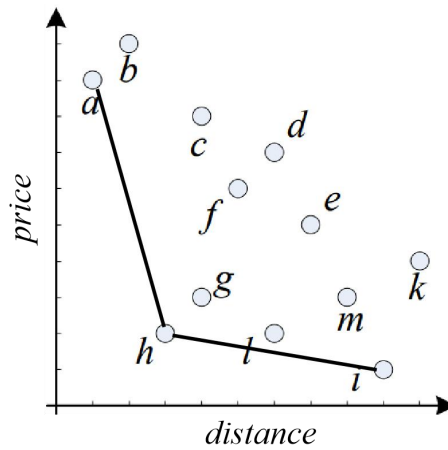


图 2-2 Skyline 查询示意
Fig. 2-2 Demo of Skyline query

目前, 已经有大量学者参与了 Skyline 算法的研究。以检索过程中索引是否被使用作为划分标准, Skyline 算法被划分为两类, 一类是基于索引的 Skyline 算法, 另一类是非索引的 Skyline 算法。关于基于索引的 Skyline 算法, 文献[44][45]提出了基于 R-树的提高查询效率的 Skyline 算法, 文献[46]提出了基于 B-树的提高查询效率的 Skyline 算法。关于非索引 Skyline 算法, 文献[47][48][49][50][51]提出了基于排序框架的 Skyline 算法。

2.3 Graham 扫描法概述

凸包是计算机图形学中的一个经典概念, 其被广泛用于解决各种实际问题[52][53][54]。Graham 扫描法是一种经典的二维平面上的凸包生成算法, 该算法是美国数学家 Ronald Graham 在 1972 年提出的。本节主要介绍凸包的概念以及基于

Graham 扫描法生成凸包的详细过程。

定义 2-4（凸包）在实数向量空间 V 中，假设存在点集 $P \subset V$ ，凸包就是点集 P 的所有凸组合构成的集合，如公式 2-3 所示。

$$H(P) = \sum_{i=1}^m \lambda_i x_i \mid x_i \in P, \lambda_i \geq 0, i=1,2,\dots,m, \sum_{i=1}^m \lambda_i = 1, m \in N \quad (2-3)$$

通俗来讲假设二维平面上存在一个点集，凸包就是包含点集的最小凸多边形，如图 2-3 所示。

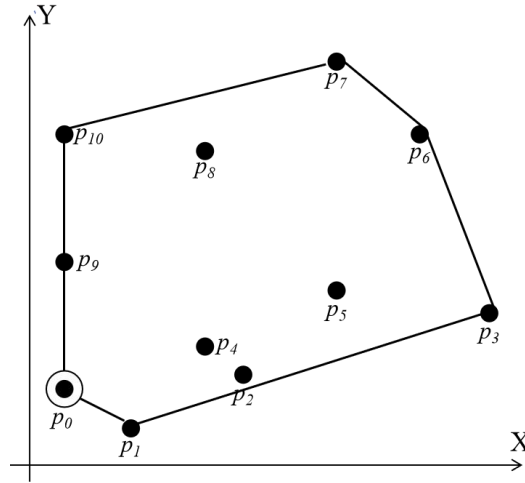


图 2-3 凸包示意

Fig. 2-3 Demo of convex hull

假设二维平面点集合 $P=\{p_0, p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}\}$ ，Graham 扫描法生成凸包的过程描述如下：

步骤 1：从点集 P 中筛选出横坐标 X 最小的点 p_0 （当横坐标 X 相同时选择纵坐标 Y 最小的点）， p_0 一定是在凸包上的，如图 2-4(a)所示。

步骤 2：将点 p_0 视作极点，对点集 P 中剩余的点依据极角进行升序排序，当极角相同时依据点与极点 p_0 的距离进行升序排序，如图 2-4(b)所示。

步骤 3：创建一个栈 S 用以存储构成凸包的点，首先将极点 p_0 压入栈中，然后将步骤 2 中排序最小的两个点 p_1 、 p_2 压入栈中。

步骤 4：判断栈顶元素 p_2 与剩余的点中排序最小的点 p_3 的凹凸性。若线段 p_2p_3 的方向相对于 p_1p_2 是向左旋转的，则说明点 p_3 是凸点，将点 p_3 入栈；若线段 p_2p_3 的方向相对于 p_1p_2 是向右旋转的，则说明点 p_2 不是凸点，将栈顶元素 p_2 出栈，如图 2-4(c)所示。

步骤 5：继续按照步骤 4 遍历剩余点直至算法结束，此时，栈中所有元素形成的集合即为凸包，如图 2-4(d)所示。

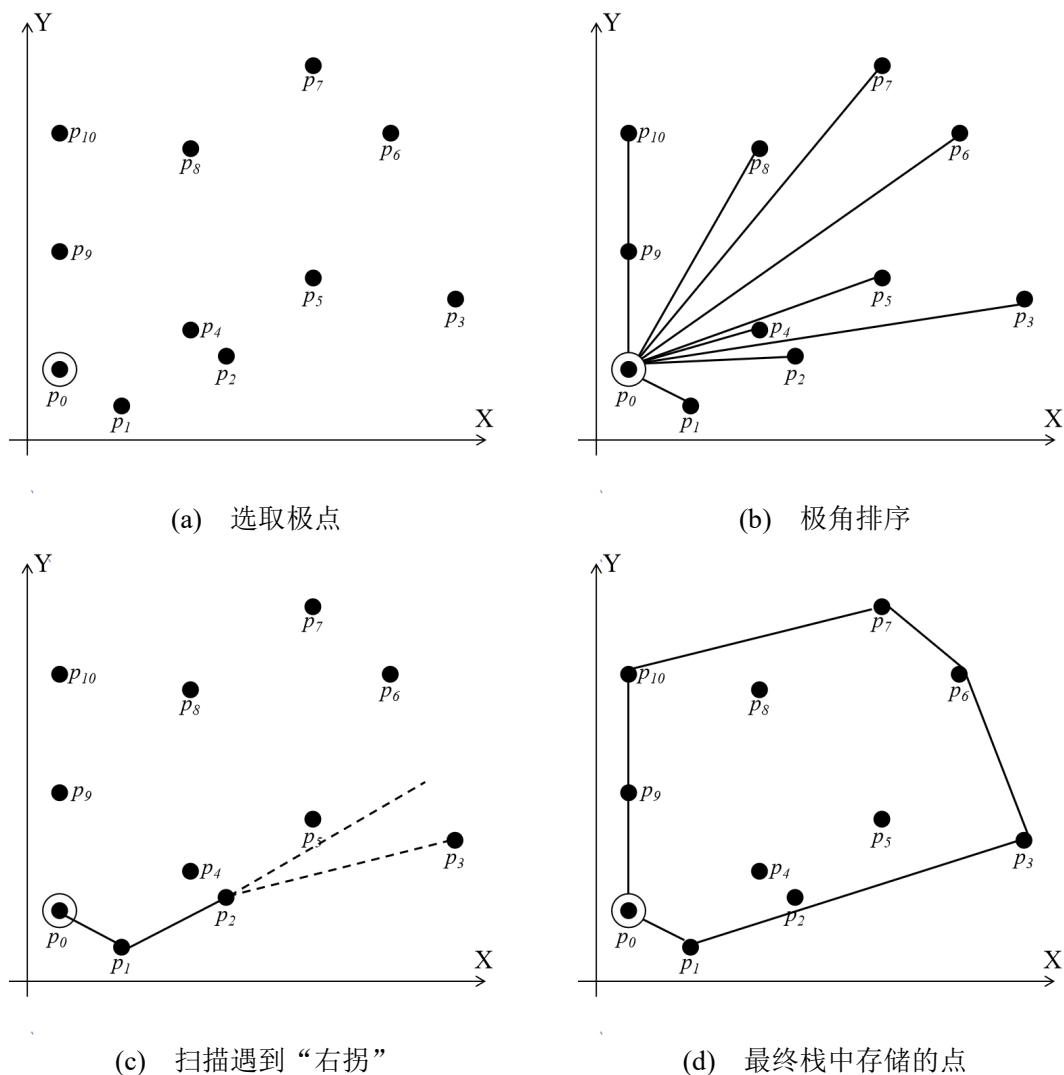


图 2-4 Graham 扫描法生成凸包过程示意

Fig. 2-4 Demo of generating convex hull through Graham scan algorithm

2.4 弦图概述

弦图是一种常见的数据可视化图形，被广泛地用于表示对象之间的关联关系^[55]。本节主要介绍弦图及其在 OD 数据可视化方面的应用。

假设我们想要将 5 个对象(A, B, C, D, E)之间的关系展示在弦图上面，我们需要输入一个 5×5 的矩阵 M 。需要注意的是，虽然 M_{AB} 和 M_{BA} 都表示了对象 A 和对象 B 之间的关系，但它们具有的含义是不同的。当起点为 A 且终点为 B 时，我们使用 M_{AB} 的值来表示它们之间的关系；当起点为 B 且终点为 A 时，我们使用 M_{BA} 的值来表示它们之间的关系。图 2-5(a)是一个标准的弦图，它表现了所有对象之间的关系。当鼠标悬浮在图 2-5(a)中的节点 C 上，我们可以得到图 2-5(b)，图 2-5(b)突出强调了对象 C 和其它对象之间关系。

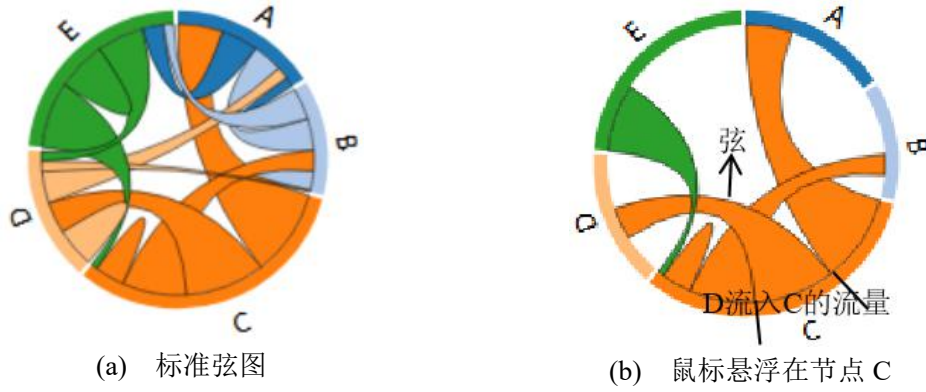


图 2-5 弦图示意

Fig. 2-5 Demo of chord graph

弦图主要由两部分组成：节点和弦。在图 2-5 所示的弦图中，总计 5 个节点，其中每一个节点的弧长皆是矩阵中该对象所在行的数值总和的比例映射，节点 C 的弧长如公式 2-4 所示：

$$ArcLengthC = \frac{M_{CA} + M_{CB} + M_{CC} + M_{CD} + M_{CE}}{\sum_{i \in \{A, B, C, D, E\}} (M_{iA} + M_{iB} + M_{iC} + M_{iD} + M_{iE})} \times 6.28 \quad (2-4)$$

由节点弧长的计算公式可知，节点的弧长可以用来呈现该对象的总入流量。弦用来表示两个对象之间的关系，如节点 C 和节点 D 之间的弦表示对象 C 和对象 D 之间的关系。其中，该弦与节点 C 相切的弧长表示从对象 C 流入对象 D 的总流量，该部分弧长如公式 2-5 所示；同样的，该弦与节点 D 相切的弧长表示从对象 D 流入对象 C 的总流量，该部分弧长如公式 2-6 所示。

$$ArcDC = \frac{M_{DC}}{ArcLengthC} \times 6.28 \quad (2-5)$$

$$ArcCD = \frac{M_{DC}}{ArcLengthC} \times 6.28 \quad (2-6)$$

基于上述有关弦图概念的叙述可知弦图是一种非常适合呈现 OD 数据的可视化图形。近年来，越来越多的来自各个领域的学者使用弦图作为 OD 数据可视化的工具。冯涛等人^[56]基于弦图对出租车行动轨迹 OD 流数据进行时空多尺度可视化分析，来研究居民的出行模式。鲁亚楠等人^[57]基于弦图对土地使用格局的变化情况进行可视化分析，来研究土地使用格局变化对生态服务价值的影响。王倩^[58]基于弦图对上海市公共交通枢纽的 OD 流数据进行可视化分析，来研究数据可视化对信息化城市建设价值。

2.5 本章小结

本章首先介绍了本文第 3 章中实现的非交叉层次 POI 提取方法中所应用到的三种技术，Voronoi 图技术、Skyline 查询技术以及 Graham 扫描法。本文使用 Voronoi 图技术进行区域划分，使用 Skyline 查询技术获取支配点集合，使用 Graham 扫描法获取凸包。然后对本文第 5 章中提出的空间弦图所参考的弦图进行了介绍。

第3章 面向主题的非交叉层次 POI 提取方法

POI 描述了那些和人们生产、生活息息相关的地理实体（如酒店、银行、住宅等）的空间信息以及语义信息（如地理实体的名称、地址、坐标等）^{[59][60]}。因此，POI 具有描述地理实体以及反映城市活动和土地使用信息的能力。为了了解城市轨道交通站点覆盖范围内的土地使用特征，本节介绍了一种面向主题的非交叉层次 POI 提取方法。

3.1 非交叉层次 POI 提取方法

非交叉层次 POI 提取方法步骤如下：首先按照城市轨道交通站点划分 POI 集合，然后基于 Skyline 查询技术提取层次 POI 集合，最后基于 Graham 扫描法提取非交叉层次 POI 集合。

3.1.1 POI 集合划分方法

按照城市轨道交通站点划分 POI 集合的详细流程如算法 3-1 所示。该算法的输入为城市轨道交通站点地理坐标的集合 $Stations = \{S_1, S_2, \dots, S_i, \dots, S_n\}$ 以及 POI 集合 $POIs = \{P_1, P_2, \dots, P_i, \dots, P_m\}$ ，其中 $S_i = (lng_i, lat_i)$ 且 m 是远大于 n 的。该算法的输出为集合 $Stations$ 中所有城市轨道交通站点覆盖范围内的 POI 集合 $PA = \{P_1, P_2, \dots, P_i, \dots, P_k\}$ ，其中城市轨道交通站点 S_i 覆盖范围内的 POI 集合 $P_i = \{p_i^1, p_i^2, \dots, p_i^l, \dots, p_i^l\}$ ， $p_i^l \in POIs$ ， $\forall p_i^l \in P_i$ 且 $i \neq j$ 皆满足 $dist(p_i^l, S_i) < dist(p_i^l, S_j)$ 。

在按照城市轨道交通站点进行 POI 集合划分的过程中，首先确定所有城市轨道交通站点的覆盖区域 $coverages = \{Cov(S_1), Cov(S_2), \dots, Cov(S_i), \dots, Cov(S_n)\}$ （line1），其中城市轨道交通站点 S_i 的覆盖区域 $Cov(S_i) = \{(lng_i^1, lat_i^1), (lng_i^2, lat_i^2), \dots, (lng_i^l, lat_i^l), \dots, (lng_i^k, lat_i^k)\}$ 。本文采用 Voronoi 图技术进行区域划分，以城市轨道交通站点为生成元绘制 VD，在绘制 VD 的过程中即可获得所有城市轨道交通站点的覆盖区域 $coverages$ 。然后，获取城市轨道交通站点 S_i 覆盖区域内的 POI 集合 P_i 。遍历集合 $POIs$ ，判断点 p_j 是否位于 S_i 的覆盖区域 $Cov(S_i)$ 内，若 p_j 是位于 S_i 的覆盖区域内的，则将点 p_j 加入集合 P_i 中（line 7 至 line 10），直至集合 $POIs$ 遍历结束便可获取集合 P_i 。最后，再将集合 P_i 加入集合 PA 中（line11），并将集合 P_i 中的元素从集合 $POIs$ 中移除（line12）。其余城市轨道交通站点覆盖区域内的 POI 集合均可由上述方法获得，因此，遍历集合 $Stations$ 即可获得集合 PA 。

算法 3-1: POI 集合划分算法

输入: *Stations*: 城市轨道交通站点地理坐标的集合;

POIs: POI 集合;

输出: *PA*: 所有的城市轨道交通站点覆盖区域内 POI 集合的集合;

```

1.  coverages  $\leftarrow$  Voronoi(Stations);
2.  PA  $\leftarrow$  {};
3.  for each  $s_i \in$  Stations do
4.       $P_i \leftarrow$  {};
5.      for each  $p_j \in$  POIs do
6.          boolean tag == line( $p_j$ , Cov( $s_i$ ))
7.          if tag == true then
8.               $P_i.add(p_j)$ ;
9.      PA.add( $P_i$ );
10.     POIs.remove( $P_i$ );
11.  return PA;

```

本文判断点 p_j 是否在 S_i 的覆盖区域 $Cov(S_i)$ 内的方法如下。首先从点 p_j 出发绘制一条射线。然后, 计算该射线与 $Cov(S_i)$ 构成的凸多边形交点的个数。若点 p_j 与凸多边形交点的个数为奇数, 则点 p_j 是在城市轨道交通站点 S_i 的覆盖区域内的, 如图 3-1(a) 所示; 若点 p_j 与凸多边形交点的个数为偶数, 则点 p_j 是不在城市轨道交通站点 S_i 的覆盖区域内的, 如图 3-1(b) 所示^{[61][62][63]}。

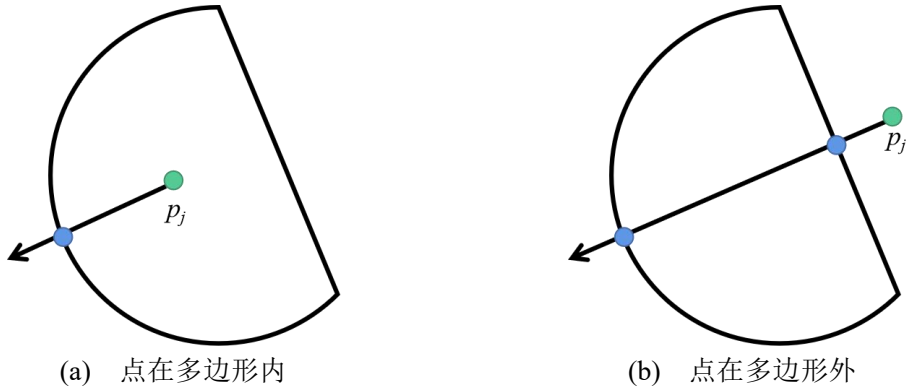


图 3-1 射线法判断点是否在凸多边形内

Fig. 3-1 Ray method to determine whether a point is within a convex polygon

3.1.2 层次 POI 提取方法

基于 3.1.1 节介绍的 POI 集合划分方法, 可以得到集合 *Stations* 中所有城市轨道交通站点覆盖区域内 POI 集合的集合 *PA*。本章主要的目的是提取城市轨道交通站点 S_i 覆盖区域内的土地使用特征以供可视化分析使用。如果直接使用城市轨道交通站点 S_i 覆盖区域内 POI 集合 P_i 作为土地使用特征进行可视化分析, 用

户能够观察到的仅仅是城市轨道交通站点覆盖区域的 POI 数量特征,而无法区分 POI 之间的差异。因此,基于等高线思想,本节提出了一种按照层次进行 POI 集合提取的方法,基于该方法可以根据用户感兴趣程度对 POI 集合进行分层。在得到城市轨道交通站点 S_i 的层次 POI 集合后,本文将基于提取的层次 POI 集合进行可视化分析。

为了基于用户对 POI 的感兴趣程度按照层次进行 POI 集合的划分,首先重要的是如何确定用户对 POI 的感兴趣程度。本文主要关注两个维度,一个是 POI 的评分,另一个是 POI 与城市轨道交通站点之间的距离。本文主要关注这两个维度信息的原因是,这两个维度是在城市轨道交通站点下车的乘客是否选择某个 POI 点的重要衡量标准。一个 POI 点的评分越高,用户选择该 POI 的可能性越大;一个 POI 点与城市轨道交通站点之间的距离越小,用户选择该 POI 的可能性越大。基于 POI 评分以及 POI 与城市轨道交通站点之间的距离获取用户感兴趣的 POI 集合是一个典型的多目标优化问题,在本文我们将采用 2.2 节介绍的 Skyline 查询技术作为该问题的解决方案。因此,本文将在 Skyline 查询技术的基础上,将城市轨道交通站点覆盖区域的 POI 集合按照用户的感兴趣程度进行分层。下面将具体介绍基于 Skyline 查询技术的层次 POI 集合提取方法的实现过程。

算法 3-2: 层次 POI 提取算法

输入: S_i : 城市轨道交通站点地理坐标;
 P_i : 城市轨道交通站点 S_i 覆盖区域内的 POI 集合;
 输出: H_i : 城市轨道交通站点 S_i 覆盖区域内的层次 POI 集合;

1. while $P_i \neq \{\}$ do
2. $l \leftarrow 1$;
3. $h_i^l \leftarrow \{\}$;
4. $skylinePoints \leftarrow \text{findSkylinePoint}(P_i, S_i)$;
5. for each $p_j \in skylinePoints$ do
6. $h_i^l.add(p_j)$;
7. $P_i.delete(p_j)$;
8. $H_i.add(h_i^l)$;
9. $l++$;
10. return H_i ;

基于 Skyline 查询技术提取城市轨道交通站点 S_i 的层次 POI 集合的详细流程如算法 3-2 所示。该算法的输入为第 i 个城市轨道交通站点的地理坐标 S_i 以及该城市轨道交通站点覆盖区域内的 POI 集合 P_i , 该算法的输出为该城市轨道交通站点 S_i 覆盖区域内的层次 POI 集合 $H_i = \{h_i^1, h_i^2, \dots, h_i^l, \dots, h_i^l\}$, 其中 $h_i^l = \{p_1, p_2, \dots,$

$p_j, \dots\}$ 表示城市轨道交通站点第 t 层的 POI 集合, $p_j \in P_i \subset POIs$ 。在算法实现过程中, 首先, 基于 Skyline 查询技术提取出集合 P_i 的支配点集合 $skylinePoints$ (line4)。然后, 将集合 $skylinePoints$ 作为层次 POI 集合的第一层 (line6)。最后, 再从集合 P_i 中删除集合 $skylinePoints$ 中所有元素 (line7)。重复上述过程, 直至集合 P_i 为空, 即可获得城市轨道交通站点 S_i 覆盖区域内的层次 POI 集合 H_i 。

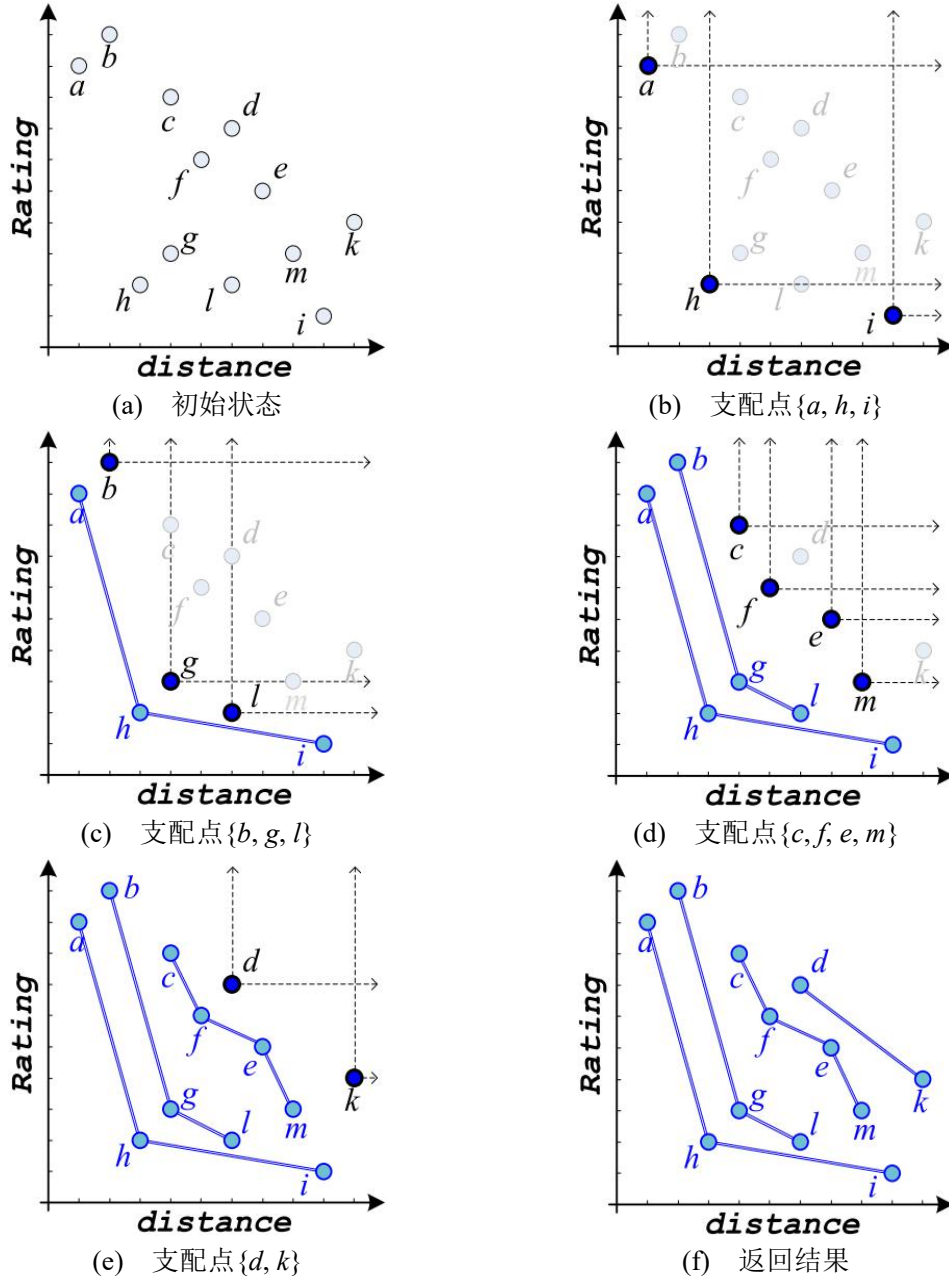


图 3-2 基于 Skyline 查询的层次 POI 抽取实例

Fig. 3-2 Example of hierarchic extraction based on Skyline query

为了更好的理解基于 Skyline 查询技术实现的层次 POI 集合提取算法, 下面以一个具体实例对上述方法进行阐述。该实例的层次 POI 集合提取过程如图 3-2

所示, 其中 *Rating* 表示数以百万计的用户对 POI 的评分, *distance* 表示 POI 点距离城市轨道交通站点的距离。假设该站点的覆盖区域内的 POI 集合 $P_i = \{a, b, c, d, e, f, g, h, i, k, l, m\}$, 如图 3-2(a)所示, 该站点的层次 POI 集合提取过程描述如下。首先, 如图 3-2(b)所示, 基于 Skyline 查询技术提取出集合 P_i 的所有支配点集合 $h_i^1 = \{a, h, i\}$ 作为该站点的层次 POI 集合的第一层。然后, 如图 3-2(c)所示, 从集合 P_i 中移除集合 h_i^1 中所有元素后 $P_i = \{b, c, d, e, f, g, k, l, m\}$, 基于 Skyline 查询技术提取出集合 P_i 的所有支配点集合 $h_i^2 = \{b, g, l\}$ 作为该站点的层次 POI 集合的第二层。然后, 如图 3-2(d)所示, 从集合 P_i 中移除集合 h_i^2 中所有元素后 $P_i = \{c, d, e, f, k, m\}$, 基于 Skyline 查询技术提取出集合 P_i 的所有支配点集合 $h_i^3 = \{c, e, f, m\}$ 作为该站点的层次 POI 集合的第三层。最后, 如图 3-2(e)所示, 从集合 P_i 中移除集合 h_i^3 中所有元素后 $P_i = \{d, k\}$, 基于 Skyline 查询技术提取出集合 P_i 的所有支配点集合 $h_i^4 = \{d, k\}$ 作为该站点的层次 POI 集合的第四层。最后, 再从集合 P_i 中移除集合 h_i^4 中所有元素后集合 P_i 为空, 停止遍历。因此, 该站点的层次 POI 集合 $H_i = \{h_i^1, h_i^2, h_i^3, h_i^4\}$ 。

3.1.3 非交叉层次 POI 提取方法

基于 3.1.2 节介绍的方法, 给定第 i 个城市轨道交通站点的地理坐标 S_i 以及该站点覆盖区域内的 POI 集合 P_i , 可以获取该站点覆盖区域内的层次 POI 集合 H_i 。但是直接使用层次 POI 集合 H_i 进行可视化分析是不可行的, 这是因为位于不同层次 POI 构成的封闭区域在地理位置上可能是相互交叉的, 这是会影响可视化效果的。本节将介绍这个问题出现的原因并提出了一种非交叉层次 POI 提取方法解决该问题。

在本文中, 基于 Skyline 查询技术提取的层次 POI 集合数据是由 POI 评分以及 POI 与城市轨道交通站点之间的距离两个维度信息决定的。但是当我们对抽取到的层次 POI 集合数据进行可视化时, 只有位置信息会体现在可视化图形上。更准确地说, 基于 Skyline 查询技术提取的层次 POI 集合中上一层 POI 是支配下一层 POI 的, 而 POI 的支配是由 POI 的评分以及 POI 与城市轨道交通站点之间的距离两个维度信息决定的。所以, 在可视化图形中, 不同层次 POI 构成的封闭区域在地理位置上可能是相互交叉的。

继续以 3.1.2 节介绍的实例为例, 对于给定的城市轨道交通站点地理坐标 S_i 及该站点覆盖区域内的 POI 集合 P_i , 该站点的层次 POI 集合 $H_i = \{h_i^1, h_i^2, h_i^3, h_i^4\}$, 其中 $h_i^1 = \{a, h, i\}$, $h_i^2 = \{b, g, l\}$, $h_i^3 = \{c, e, f, m\}$, $h_i^4 = \{d, k\}$ 。该站点覆盖区域内 POI 集合 P_i 的空间分布如图 3-3(a)所示, h_i^1 、 h_i^2 、 h_i^3 各自构成的封闭区域在二维平面上的空间位置分别如图 3-3(b)、图 3-3(c)、图 3-3(d)所示。由于 h_i^4 只包

含两个 POI 点, 无法构成封闭区域, 故此处未进行展示。由图 3-3 显然可知站点 S_i 的不同层次的 POI 构成的封闭区域在地理位置上是相互交叉的。

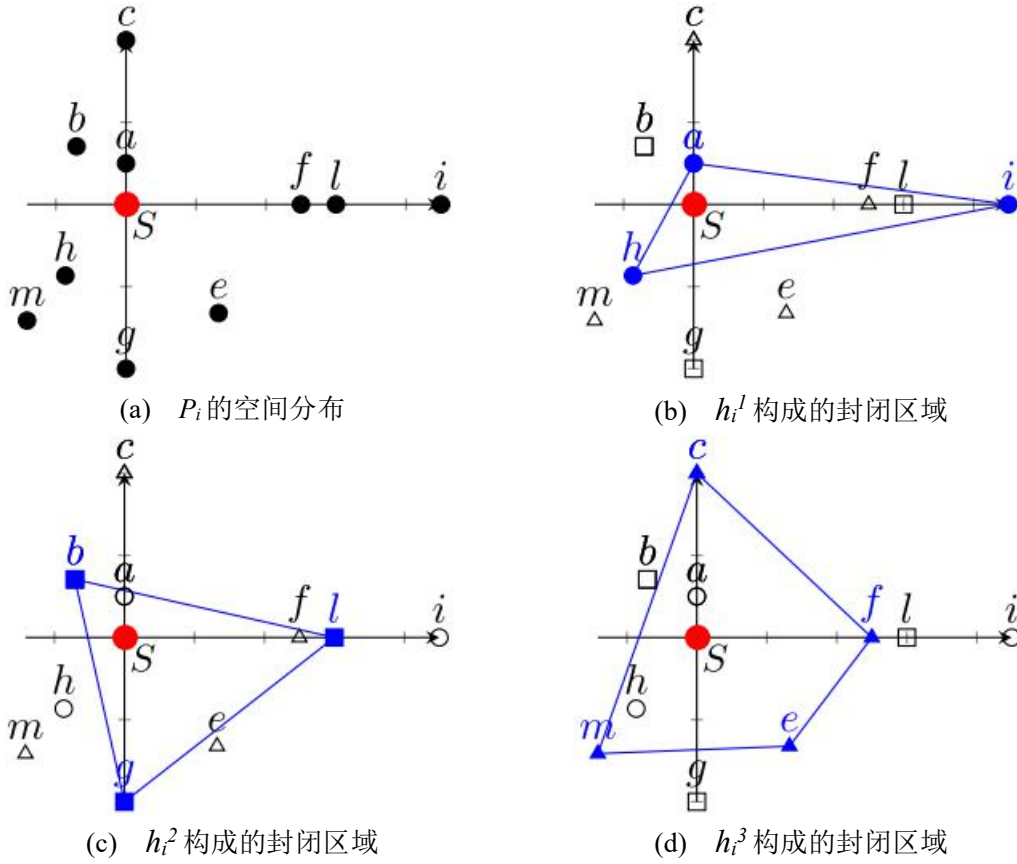


图 3-3 层次 POI 的空间信息

Fig. 3-3 Spatial information of hierarchical POI

为了解决上述问题, 本节提出了一种非交叉层次 POI 提取算法, 如算法 3-3 所示。该算法的输入为第 i 个城市轨道交通站点的地理位置坐标 S_i 以及算法 3-2 的输出, 即第 i 个城市轨道交通站点覆盖区域内的层次 POI 集合 H_i 。该算法的输出为第 i 个城市轨道交通站点的非交叉层次 POI 集合 IH_i , 即该站点不同层次 POI 构成的封闭区域在地理位置上是互不交叉的。在算法的实现过程中, 首先将城市轨道交通站点的地理位置坐标 S_i 添加到层次 POI 集合 H_i 的第一层 h_i^1 (line1)。然后, 获取每一层的非交叉层次 POI 集合 (line 3 至 line 12)。对于 H_i 的第 l 层非交叉 POI 集合的获取, 计算第 l 层的 POI 集合和第 $l-1$ 层的非交叉层次 POI 集合的并集 (line 4 至 line 10), 并将其作为 Graham 扫描法的输入, Graham 扫描法输出该集合的凸包 (line11), 即为第 l 层的非交叉层次 POI 集合。同理可得其它层的非交叉 POI 集合。

算法 3-3: 非交叉层次 POI 提取算法

输入: S_i : 第 i 个城市轨道交通站点的地理坐标;

H_i : 第 i 个城市轨道交通站点覆盖区域内的层次 POI 集合;

输出: IH_i : 第 i 个城市轨道交通站点覆盖区域内的非交叉层次 POI 集合;

```

1.   $H_i.get(0).add(S_i)$ ;
2.   $IH_i \leftarrow \{\}$ ;
3.  for each  $h_i^l \in H_i$  do
4.       $points \leftarrow \{\}$ ;
5.       $ih_i^l \leftarrow \{\}$ ;
6.      for each  $p_j \in h_i^l$  do
7.           $points.add(p_j)$ ;
8.      if  $l \geq 1$  then
9.          for each  $p_j \in ih_i^{l-1} \subset IH_i$  do
10.              $Points.add(p_j)$ ;
11.          $ih_i^l = \text{grahamScan}(points)$ ;
12.          $IH_i.add(ih_i^l)$ ;
13. return  $IH_i$ ;
```

为了更好的理解算法 3-3 介绍的基于 Graham 扫描法实现的非交叉层次 POI 集合提取算法, 下面将继续以 3.1.2 节使用的实例为例对其进行阐述。首先, 将该城市轨道交通站点的地理坐标 $S_i = (0, 0)$ 添加到 h_i^l 中。然后, 提取非交叉层次 POI 集合的第一层, 以 $h_i^l = \{a, h, i, S_i\}$ 作为输入, 基于 Graham 扫描法得到凸包 $ih_i^l = \{a, h, i\}$, 即非交叉层次 POI 的第一层, 图 3-4(a) 中浅色阴影区域即为 ih_i^l 构成的封闭区域。然后, 提取非交叉层次 POI 集合的第二层, 以 $ih_i^l = \{a, h, i\}$ 和 $h_i^2 = \{a, h, i, b, g, l\}$ 的并集 $\{a, h, i, b, g, l\}$ 作为输入, 基于 Graham 扫描法得到凸包 $ih_i^2 = \{g, h, i, b\}$, 即非交叉层次 POI 的第二层, 图 3-4(b) 中浅色阴影区域即为 ih_i^2 构成的封闭区域。然后, 提取非交叉层次 POI 集合的第三层, 以 $ih_i^2 = \{g, h, i, b\}$ 和 $h_i^3 = \{c, e, f, m\}$ 的并集 $\{g, h, i, b, c, e, f, m\}$ 作为输入, 基于 Graham 扫描法得到凸包 $ih_i^3 = \{g, i, b, c, f, m\}$, 即非交叉层次 POI 的第三层, 图 3-4(c) 中浅色阴影区域即为 ih_i^3 构成的封闭区域。直至遍历完所有层的 POI 集合, 停止轮询。综上所述, 该站点的非交叉层次 POI 集合 $IH_i = \{ih_i^l, ih_i^2, ih_i^3, ih_i^4\}$ 。 ih_i^l 、 ih_i^2 、 ih_i^3 的非交叉层次 POI 集合构成的封闭区域在空间上的相对位置如图 3-4(d) 所示。由图 3-4(d) 显然可知站点 S_i 的不同层次的非交叉 POI 集合构成的封闭区域在地理位置上是互不交叉的。

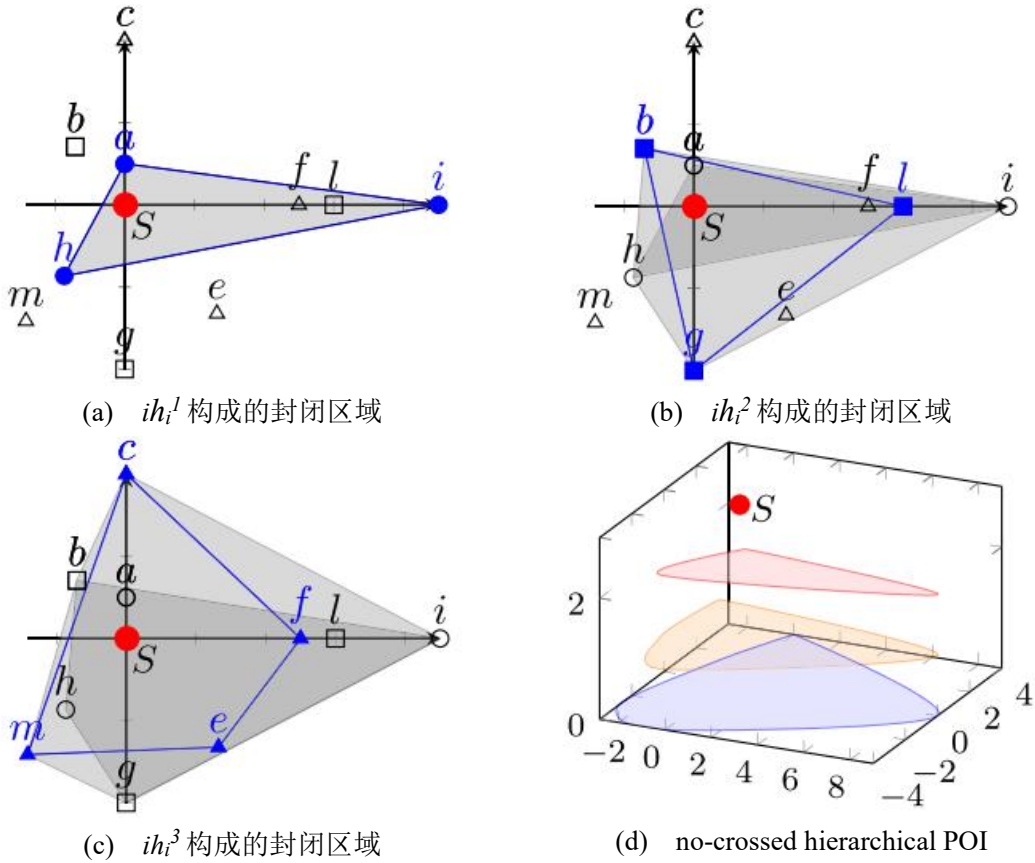


图 3-4 非交叉层次 POI 的空间信息

Fig. 3-4 Spatial information of uncrossed hierarchical POI

3.2 面向主题的非交叉层次 POI 提取

基于 3.1 节介绍的非交叉层次 POI 提取方法, 本节提出了一种面向主题的非交叉层次 POI 提取方法。基于该方法可以提取指定主题下的城市轨道交通站点覆盖区域内的用地特征, 进而更好的了解不同城市轨道交通站点覆盖范围内的土地使用类型。研究城市轨道交通站点覆盖范围内的土地使用类型, 对轨道交通系统的运营以及站点周边土地的开发和使用规划具有指导意义。

3.2.1 面向主题的非交叉层次 POI 提取方法

面向主题的非交叉层次 POI 提取方法是在非交叉层次 POI 提取方法的基础上实现的, 其实现步骤与非交叉层次 POI 提取方法的实现步骤 (见 3.1 节) 基本是一致的。不同之处是在提取面向主题的非交叉层次 POI 集合时, 算法 3-2 的输入是城市轨道交通站点 S_i 覆盖区域内相关主题的 POI 集合 TP_i , 而非城市轨道交通站点 S_i 覆盖区域内的 POI 集合 P_i 。

为了提取城市轨道交通站点 S_i 覆盖区域内相关主题的 POI 集合 TP_i , 需要遍历 P_i 中的任意元素 p , 并计算其与指定主题词 $topic$ 的相似度。若 p 与 $topic$ 的相

似度是大于 0 的, 则将其加入集合 TP_i 中; 若 p 与 $topic$ 的相似度为 0, 则舍弃 p 。受^{[64][65]}的启发, 本文使用 Jaccard 系数来计算 POI 点 p 与指定的主题词 $topic$ 之间的相似度, 如公式 3-1 所示。

$$sim(topic, p) = \frac{topic \cap p}{topic \cup p} \quad (3-1)$$

假设城市轨道交通站点 S_i 覆盖区域内的 POI 集合 $P_i = \{Beijing Sports University, College student apartment, Yuanmingyuan Ruins Park\}$ 并且指定的主题词为 *university*。为了提取相关主题的 POI 集合 TP_i , 基于公式 3-2 计算 P_i 中所有元素与 *university* 的相似度, 其结果如表 3-1 所示。*Beijing Sports University*、*College student apartment* 以及 *Yuanmingyuan Ruins Park* 三个 POI 点中只有 *Beijing Sports University* 与 *university* 的相似度是大于 0 的, 因此城市轨道交通站点 S_i 覆盖区域内 *university* 主题词相关的 POI 集合 $TP_i = \{Beijing Sports University\}$ 。

表 3-1 P_i 中的元素与主题词 *university* 的相似度

Tab. 3-1 Similarity between p and *university*

	Beijing Sports University	College student apartment	Yuanmingyuan Ruins Park
<i>sim</i>	0.33	0	0

3.2.2 基于主题的颜色值生成方法

无论是基于 3.1 节提取的非交叉层次 POI 集合, 还是基于 3.2.1 节提取的面向主题的非交叉层次 POI 集合, 在第 4 章都需要将其作为数据进行可视化。为了能够区分不同主题的可视化模型, 本小节设计了一种基于主题的颜色值生成方法。将基于该方法获得的颜色值用于填充可视化模型, 即可区分不同主题的可视化模型。

基于主题生成颜色值的具体过程如算法 3-4 所示。该算法的输入为指定的主题词 *topic*, 该算法的输出为 RGB 颜色值 *RGB*。RGB 色彩模式是一种经典的色彩模型, 它有三个通道 R、G、B。因此, 为了生成一个颜色值, 将主题词 *topic* 均分为三个部分 (line 2 至 line 4)。本文使用每部分所有字母 ASCII 值的均值分别作为 R、G、B 的值。然而, 由于不同字母的 ASCII 值差异较小, 通过这种方法生成的颜色值差异较小。因此, 本文在上述计算得到的 R、G 和 B 值的基础上分别添加一个 0 到 100 之间的随机数 (line 5 至 line 13)。考虑到 RGB 色彩模式的各个通道取值范围为 0-255, 若计算得到的数值大于 255, 需要对其取余 (line 14 至 line 15)。基于算法 3-4 生成的颜色值, 当主题词为 *university* 时, 国贸站填充的颜色如图 4-5(a) 所示。

 算法 3-4: 基于主题的颜色值生成算法

 输入: *topic*: 给定的主题词;

 输出: *RGB*: RGB 颜色值;

```

1.  RGB  $\leftarrow$  {};
2.  lengthR  $\leftarrow$   $\text{len}(\text{topic}) / 3$ ;
3.  lengthG  $\leftarrow$   $\text{len}(\text{topic}) / 3$ ;
4.  lengthB  $\leftarrow$   $\text{len}(\text{topic}) / 3$ ;
5.  for i  $\leftarrow$  0; i < lengthR; i++ do
6.      r  $\leftarrow$  r +  $\text{ord}(\text{lower}(\text{topic}[i]))$ ;
7.  RGB[0]  $\leftarrow$  r / lengthR +  $\text{random.random}() \times 100$ ;
8.  for i  $\leftarrow$  0; i < lengthG; i++ do
9.      g  $\leftarrow$  g +  $\text{ord}(\text{lower}(\text{topic}[i]))$ ;
10. RGB[1]  $\leftarrow$  g / lengthG +  $\text{random.random}() \times 100$ ;
11. for i  $\leftarrow$  0; i < lengthB; i++ do
12.     b  $\leftarrow$  b +  $\text{ord}(\text{lower}(\text{topic}[i]))$ ;
13. RGB[2]  $\leftarrow$  b / lengthB +  $\text{random.random}() \times 100$ ;
14. for i  $\leftarrow$  0; i < 3; i++ do
15.     R[i]  $\leftarrow$  R[i] % 255;
16. return RGB;

```

3.3 本章小结

本章介绍了一种非交叉层次 POI 提取方法。首先, 基于 Voronoi 图按照城市轨道交通站点对 POI 集合进行划分。然后, 基于 Skyline 查询技术提取层次 POI 集合。最后, 基于 Graham 扫描法提取非交叉层次 POI 集合。同时, 本章还在上述非交叉层次 POI 提取方法的基础上提出了一种面向主题的非交叉层次 POI 集合提取方法。本章介绍的方法为第 4 章中面向轨道交通的空间密度可视化分析研究提供数据支持。

第4章 面向轨道交通的空间密度可视化分析研究

为了帮助土地使用以及城市轨道交通管理、规划决策者更好地比较不同城市轨道交通站点周边用地特征的差异,本章基于第3章介绍的面向主题的非交叉层次POI提取方法以及三维等高线模型构建了一个面向轨道交通的空间密度可视化系统。

4.1 可视化任务概述

本系统是专门为城市轨道交通以及土地使用管理、规划决策者设计的,因此本文通过采访相关领域的工作人员了解用户的真实需求。他们期望面向轨道交通的空间密度可视化分析系统能够完成T1、T2、T3可视化分析任务。

T1. 单个城市轨道交通站点覆盖区域内土地使用密度特征的可视化。站点周边土地使用强度与客流量成正比,因此通过比较不同城市轨道交通站点覆盖区域内的土地使用特征的差异,可以发现不同站点客流量特征差异以及周边土地使用特征差异,进而为土地使用以及城市轨道交通管理、规划决策者提供指导建议。比如,通过比较不同站点覆盖区域内土地使用特征,可以挖掘出覆盖范围大且客流量大的站点,为新的城市轨道交通站点建址提供指导建议。

T2. 多个城市轨道交通站点覆盖区域内土地使用密度特征的可视化。探索单个城市轨道交通站点覆盖区域内的土地使用密度特征,只能了解单一站点的土地使用特征,无法了解站点群的土地使用特征。通过探索多个城市轨道交通站点覆盖区域内的土地使用密度特征,可以从更宏观地角度观察由几个城市轨道交通站点构成的站点群的土地使用密度特征。

T3. 城市轨道交通站点覆盖区域内土地使用功能特征的可视化。城市轨道交通站点覆盖区域内的土地使用功能特征可以影响站点的客流量,比如早晨以居住用地较为集中的站点其客流量以进站为主,以工作用地较为集中的站点其客流量以出站为主。通过探索一个站点覆盖区域在不同主题下的土地使用特征差异,可以更好地了解该站点覆盖区域内的土地使用功能特征,进而可以分析城市轨道交通站点的客流量随时间变化情况。通过探索不同站点覆盖区域在同一主题下的土地使用特征差异,可以更好地了解一个功能在城市的空间分布信息。

4.2 可视化设计概述

本节主要介绍面向轨道交通的空间密度可视化分析系统使用到的 POI 数据集、可视化分析系统的构建以及三维等高线模型的分析原则。

4.2.1 POI 数据集介绍

POI 描述了那些和人们生产、生活息息相关的地理实体（如酒店、银行、住宅等）的空间信息以及语义信息（如地理实体的名称、地址、坐标等）^{[59][60]}。本文基于 JavaScript 语言以及 Python 语言从高德地图中爬取了北京市 POI 数据约 520,000 条^[65]。每条 POI 数据包含 6 个属性，它们分别是 ID、名称、经度、纬度、评分以及类别，如表 4-1 所示。

表 4-1 POI 属性示意
Tab. 4-1 Example of attribute of POI

ID	名称	经度	纬度	评分	类别
23714	Century Hualian Supermarket	116.769049	39.602736	3.9	shopping
211316	Map of Coffee	116.761949	39.607981	3.0	cater

POI 的经度和纬度能够确定地理实体的空间位置。POI 的评分能够确定用户对地理实体的喜爱程度，取值范围为[0, 5]。POI 的类别属性共计 20 种，其中包含 shopping、cater 等。

4.2.2 可视化分析系统概述

为了实现 4.1 节描述的可视化分析任务，本节基于 Python 以及高德 API 构建了以 GIS 地图为底绘制的面向轨道交通的空间密度可视化系统，如图 4-1 所示。本系统共包含两个视图。其中视图 A 展示了国贸站、立水桥站、北京大学东门站、东单站等共计 227 个城市轨道交通站点（截止 2012 年，北京市共计 227 个运营的城市轨道交通站点）未指定主题词的俯视三维等高线模型并为系统提供了交互操作，如图 4-1 中区域 A 所示。通过该视图，用户可以对北京市所有城市轨道交通站点的用地特征有一个总体的概览。视图 B 展示了一个或多个用户感兴趣的城市轨道交通站点的三维等高线可视化模型，如图 4-1 中区域 B、C、D、E 所示。通过该视图，用户可以比较不同城市轨道交通站点之间用地特征的差异。用户可以通过视图 A 中设计的交互操作得到视图 B。

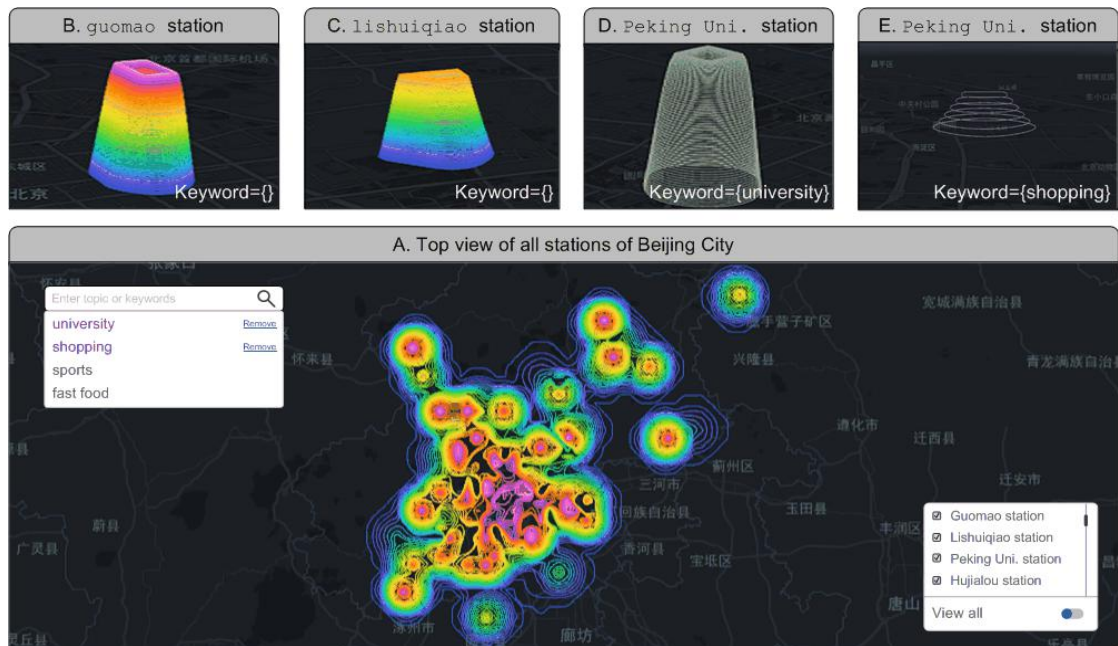


图 4-1 空间密度可视化系统概览图

Fig. 4-1 Overview of spatial density visualization system

基于 POI 数据绘制三维等高线模型的过程如图 4-2 所示。在视图 A 中，参数 *stations* 为北京市 227 个城市轨道交通站点的集合，参数 *keyword* 为空集。在视图 B 中，参数 *stations* 和 *keyword* 是由视图 A 提供的交互文本框提供。

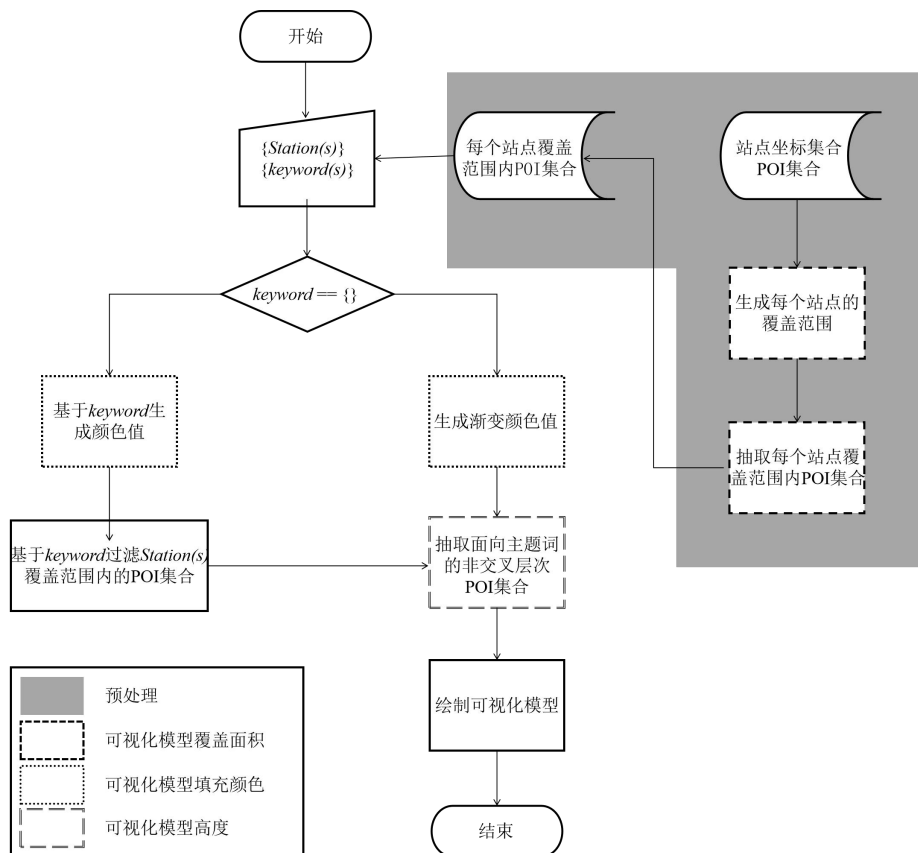


图 4-2 三维等高线模型绘制流程图

Fig. 4-2 Flow chart of drawing 3D contour model

该系统提供的用户交互操作如下：

(1) 指定城市轨道交通站点。当用户在视图 A 中的 *stations* 复选框选中一个或者多个城市轨道交通站点时，可以在视图 B 中得到相应的三维等高线模型。当用户在视图 A 中的 *stations* 复选框选中 *guomao station* 时，其三维等高线可视化模型如图 4-1 中区域 B 所示。具体实现过程如图 4-2 所示，其中参数 *stations* 为 *guomao station*，参数 *keyword* 为空。

(2) 指定主题词。仅指定 *keyword* 而不指定 *stations* 是没有意义的，在视图 A 中的 *stations* 复选框中选中一个或者多个城市轨道交通站点的基础上，用户通过在视图 A 中的 *keyword* 搜索框输入主题词，可以再视图 B 中得到特定主题的三维等高线模型。当用户在 *stations* 复选框中选中 *Peking Uni station* 且在 *keyword* 搜索框中输入主题词 *university* 时，其三维等高线模型如图 4-1 中区域 D 所示。具体实现过程如图 4-2 所示，其中参数 *stations* 为 *Peking Uni station*，参数 *keyword* 为 *university*。

4.2.3 三维等高线模型分析原则

基于三维等高线模型分析城市轨道交通站点覆盖区域内的土地使用特征，需要遵从下列原则。

(1) 三维等高线模型的高度：在其他条件相同的情况下，三维等高线模型越高，城市轨道交通站点覆盖区域内的土地使用密度越大。

(2) 三维等高线模型的覆盖面积：在其他条件相同的情况下，三维等高线模型的覆盖面积越小，城市轨道交通站点覆盖区域内的土地使用密度越大。

(3) 三维等高线模型的颜色跨度：在其他条件相同的情况下，三维等高线模型的颜色跨度越大，城市轨道交通站点覆盖区域内的土地使用密度越大。特别地，该规则只能用于分析未指定主题词的三维等高线模型，这是因为面向主题词的三维等高线模型的颜色不是默认的渐变色，而是由主题词决定的单一颜色值。

4.3 案例研究与实验分析

本节基于两个案例研究以及一个实验分析来验证本系统生成的三维等高线模型的有效性。

4.3.1 案例研究

(1) 基于本系统比较不同城市轨道交通站点或者区域的土地使用强度(T1、T2)。依据 4.2.3 节描述的三维等高线模型分析原则对三维等高线模型进行分析。

首先，基于三维等高线模型分析不同城市轨道交通站点土地使用强度差异。下面以国贸站三维等高线模型和立水桥站三维等高线模型为例进行分析，其三维等高线模型分别如图 4-3(a)、4-3(b)所示。由图 4-3(a)、4-3(b)可知，国贸站三维等高线模型和立水桥站三维等高线模型的覆盖面积几乎是相同的，但是国贸站三维等高线模型的高度要高于立水桥站三维等高线模型并且国贸站三维等高线模型的颜色跨度比立水桥站三维等高线模型的颜色跨度大。因此，通过上述两个三维等高线模型的对比，可以得出国贸站土地使用强度大于立水桥站土地使用强度的结论。

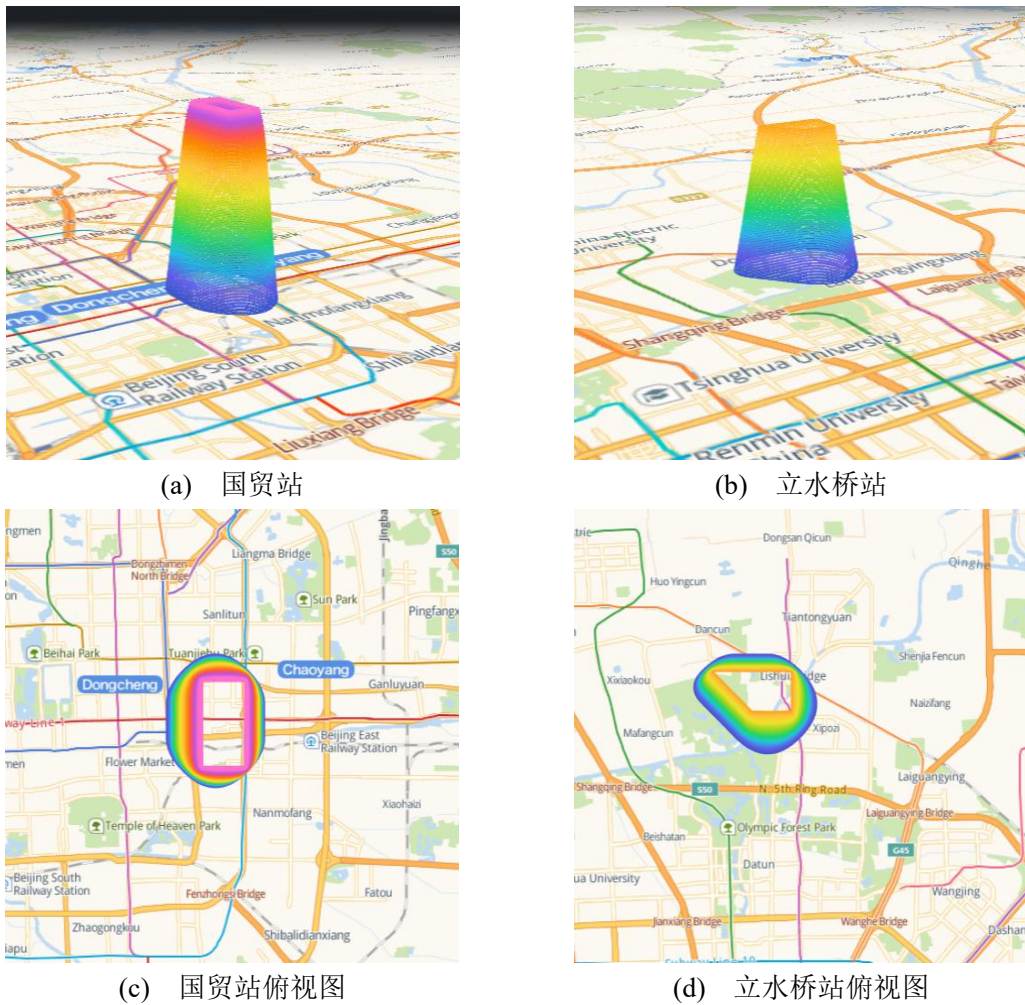


图 4-3 国贸站和立水桥站三维等高线模型

Fig. 4-3 3D contour model of guomao station v.s. lishuiqiao station

然后，基于三维等高线模型分析不同站点群土地使用强度差异，一个站点群由几个城市轨道交通站点组成。下面以两个站点群的三维等高线模型为例进行分析，站点群 A 包括国贸站和朝阳门站，站点群 B 包括立水桥站和北苑北路站，其三维等高线模型分别如图 4-4(a)、4-4(b)所示。由图 4-4(a)、4-4(b)可知，站点群 A 三维等高线模型和站点群 B 三维等高线模型的覆盖面积几乎是相同的，但

站点群 A 三维等高线模型的高度要高于站点群 B 三维等高线模型并且站点群 A 三维等高线模型的颜色跨度大于站点群 B 三维等高线模型。因此，通过上述两个三维等高线模型的对比，可以得出站点群 A 土地使用强度是大于站点群 B 土地使用强度的结论。

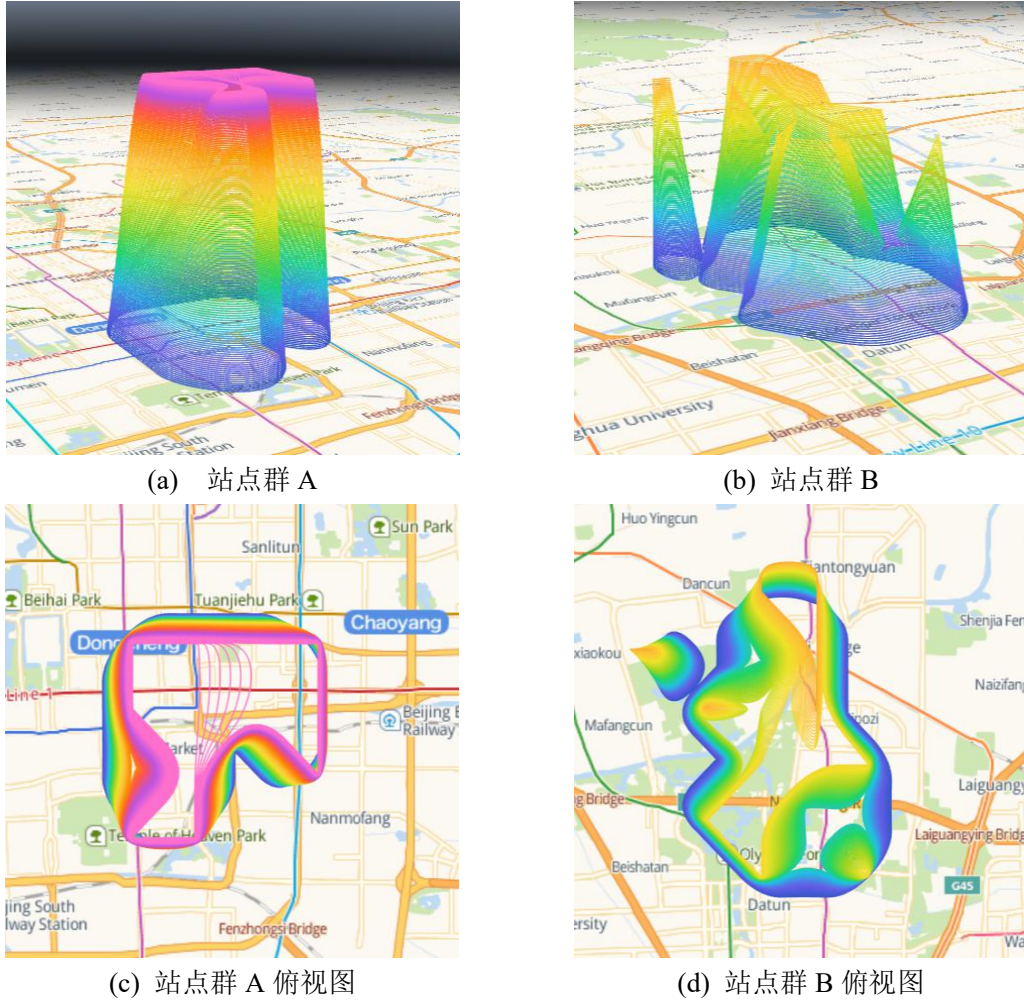


图 4-4 站点群 A 和站点群 B 的三维等高线模型

Fig. 4-4 3D contour model of group A v.s. group B

(2) 基于本系统比较不同城市轨道交通站点土地使用主题特征(T3)。具体来说，以主题词 *university* 和 *shopping* 下的国贸站以及北京大学东门站作为本案例的研究对象。主题 *university* 下的国贸站三维等高线模型以及北京大学东门站的三维等高线模型分别如图 4-5(a)、4-5(b)所示。主题 *shopping* 下的国贸站三维等高线模型以及北京大学东门站的三维等高线模型分别如图 4-6(a)、4-6(b)所示。依据 4.2.3 节描述的三维等高线模型分析原则对三维等高线模型进行分析。

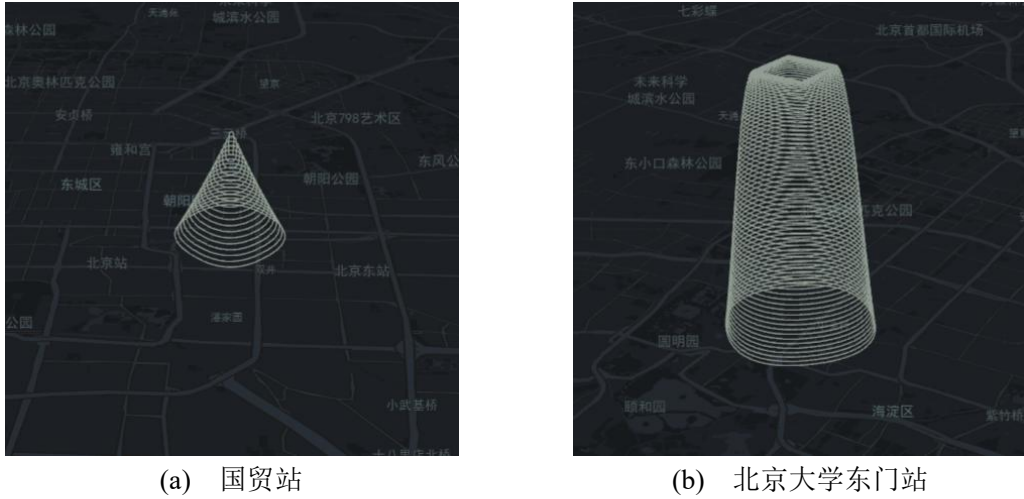


图 4-5 主题词 *university* 下的国贸站和北京大学东门站的三维等高线模型

Fig. 4-5 *guomao station v.s. Peking Uni. station (kw={university})*

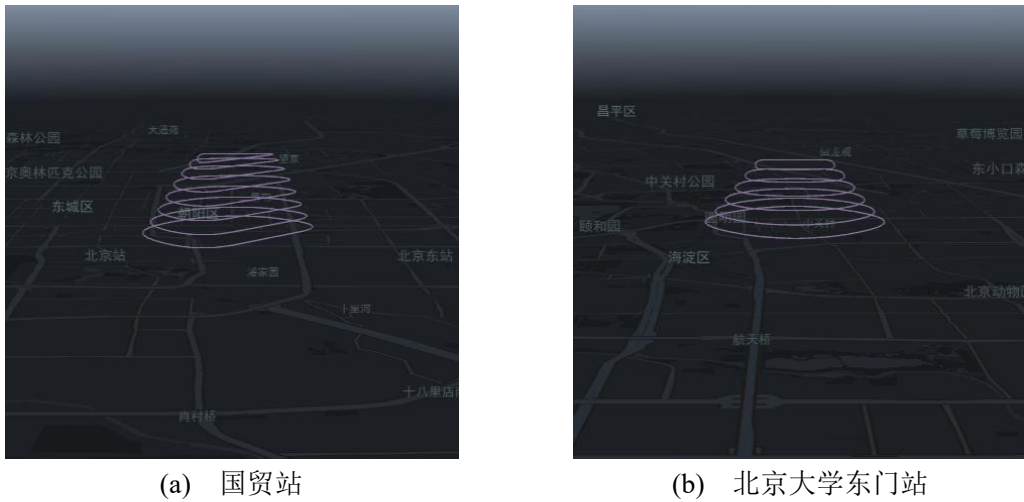


图 4-6 主题词 *shopping* 下的国贸站和北京大学东门站的三维等高线模型

Fig. 4-6 *guomao station v.s. Peking Uni. Station(kw={shopping})*

由图 4-5(a)、4-5(b)可知,在 *university* 主题词下国贸站的覆盖面积与 *university* 主题词下北京大学东门站的覆盖面积几乎是相同的,但是 *university* 主题词下北京大学东门站的三维等高线模型的高度远比 *university* 主题词下国贸站的三维等高线模型要高。由此可以得出,北京大学东门站覆盖范围内 *university* 相关的土地使用强度比国贸站覆盖范围内 *university* 相关的土地使用强度要大很多。由图 4-6(a)、4-6(b)可知,在 *shopping* 主题词下国贸站的覆盖面积与 *shopping* 主题词下北京大学东门站的覆盖面积几乎是相同的,并且在 *shopping* 主题词下国贸站的三维等高线化模型的高度比在 *shopping* 主题词下北京大学东门站的三维等高线模型要高一点。由此可以得出,国贸站点覆盖范围内 *shopping* 相关的土地使用强度比北京大学东门站覆盖范围内 *shopping* 相关的土地使用强度要大一些。

由图 4-5(a)、4-6(a)可知,在 *shopping* 主题词下国贸站的覆盖面积与在

university 主题词下国贸站的覆盖面积几乎是相同的, 但是 *shopping* 主题词下国贸站的三维等高线模型的高度比 *university* 主题词下国贸站的三维等高线模型要高一点。由此可以得出, 国贸站覆盖范围内的土地使用以 *shopping* 为主。由图 4-5(b)、4-6(b)可知, 在 *shopping* 主题词下北京大学东门站的覆盖面积与 *university* 主题词下北京大学东门站的覆盖面积几乎是相同的, 但是 *university* 主题词下北京大学东门站的三维等高线模型的高度比 *shopping* 主题词下北京大学东门站的三维等高线模型要高很多。由此可以得出, 北京大学东门站覆盖范围内的土地使用以 *university* 为主。

北京大学东门站位于北京大学校园附近, 该站点与北京大学东门距离不足 100 米, 该站点主要为北京大学的学生、北京大学的教师以及参观访问北京大学校园的游客等提供交通服务。国贸站毗邻国际贸易服务公司以及 CBD。因此, 国贸站覆盖范围内的土地在商业方面利用更高, 北京大学东门站覆盖范围内的土地在教育方面的利用更高。这与通过上述两个场景分析得到的结论是一致的。

4.3.2 实验分析

为了验证本文提出的未指定主题词的三维等高线模型的正确性, 本节基于城市轨道交通站点每周出站客流量进行交叉验证。城市轨道交通站点覆盖范围内的土地使用强度对城市轨道交通站点的客流量是有显著影响的, 城市轨道交通站点覆盖区域内的土地使用强度越高, 其容积率和单位面积客流量越大。以案例(1)中分析的国贸站和立水桥站为例, 这两个站点的周客流量如图 4-7 所示。如图 4-7 可知国贸站每日客流总量远高于立水桥站的每日客流总量。因此, 国贸站的土地使用强度是大于立水桥站的, 这与案例(1)的结论是一致的。

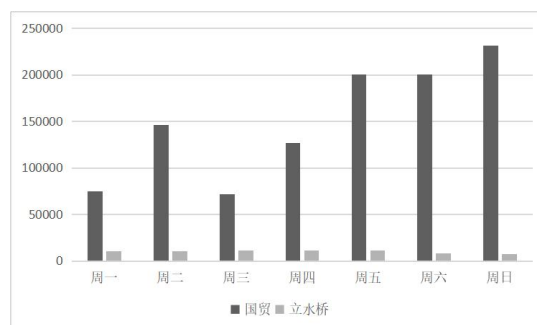


图 4-7 每日出站客流量

Fig. 4-7 Daily exiting passenger volume

为了验证指定主题词下三维等高线模型的正确性, 本节基于城市轨道交通站点一天内不同时间的入站客流量进行交叉验证。城市轨道交通站点覆盖范围内的土地使用类型直接影响着城市轨道交通车站的客流情况, 上午, 居住用地较多的

城市轨道交通站点的客流量以入站为主,就业用地较多的城市轨道交通站点的客流量以出站为主。此处以呼家楼站和光熙门站作为研究对象,主题词 *resident* 下呼家楼站和光熙门站的三维等高线模型分别如图 4-8(a)、4-8(b)所示。由图 4-8(a)、4-8(b)可知,在 *resident* 主题词下呼家楼站的覆盖面积与在 *resident* 主题词下光熙门站的覆盖面积几乎是相同的,但是 *resident* 主题词下光熙门站的三维等高线模型的高度是远高于 *resident* 主题词下呼家楼站的三维等高线模型,即光熙门站覆盖范围内住宅建筑数量多于呼家楼站覆盖范围内住宅建筑数量。通过上述分析可以推断出,呼家楼站上午入站客流量较小,而光熙门站上午入站客流量较大。图 4-9(a)以及图 4-9(b)分别展示了呼家楼站和光熙门站一天内不同时间入站客流量详情,由图 4-9(a)、4-9(b)可知呼家楼站上午入站客流量较小,而光熙门站上午入站客流量较大。这与依据三维等高线模型分析得到的结论是一致的。

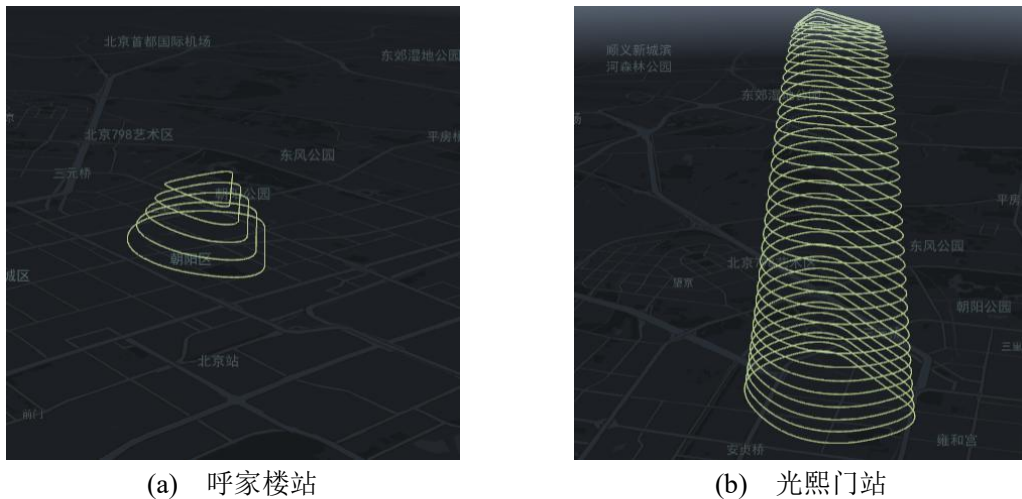


图 4-8 主题词 *resident* 下的呼家楼站和光熙门站的三维等高线模型

Fig. 4-8 *hujialou station v.s. guangximen station (kw={resident})*

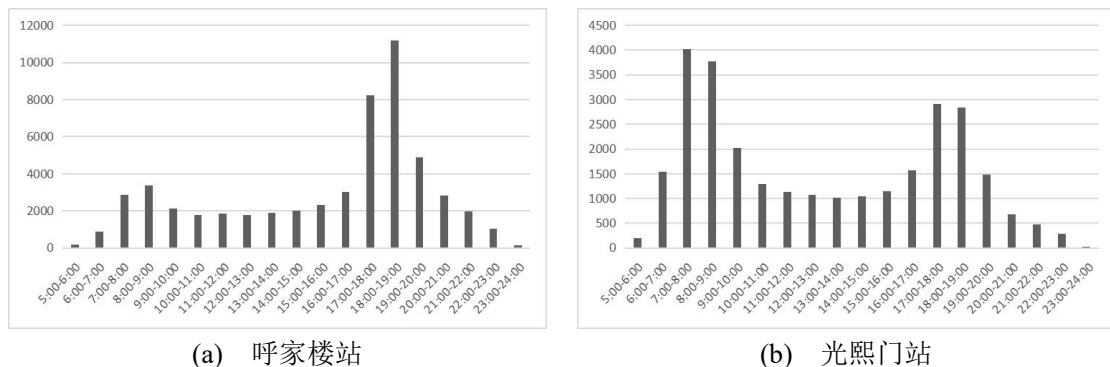


图 4-9 每时入站客流量

Fig. 4-9 Hourly entering passenger volume

4.4 本章小结

本章首先介绍了面向轨道交通的空间密度可视化分析系统需要完成的任务。然后介绍了面向轨道交通的空间密度可视化分析研究中使用到的 POI 数据集及系统设计。最后，通过两个案例研究以及实验分析来验证本文设计的面向轨道交通的空间密度可视化分析系统的有效性。

第5章 面向轨道交通的时空分布可视化分析研究

为了帮助城市轨道交通以及土地使用管理、规划决策者更好地分析城市轨道交通客流量时空分布特征，本章首先设计了一个空间弦图，既可以表达 OD 对象之间的关联关系又能表达 OD 对象蕴含的空间信息。同时，为了避免空间弦图中出现大量弦交叉的现象，提出了一种基于方向的多尺度空间聚合方法。然后基于一组空间弦图构建了一个面向轨道交通的时空分布可视化系统。最后，通过案例研究验证系统的有效性。

5.1 可视化任务概述

本系统是专门为城市轨道交通以及土地使用管理、规划决策者设计的，因此本文通过采访相关领域的工作人员了解用户的真实需求。他们期望面向轨道交通的时空分布可视化分析系统能够完成 T1、T2、T3 可视化分析任务。

T1. 探索客流量时间分布特征。客流量的时间分布特征可以反映大规模出行规律，如客流量的高峰时段。因此，通过探索客流量的时间分布特征可以为城市轨道交通以及土地使用管理、规划决策者提供指导建议。

T2. 探索客流量空间分布特征。客流量的空间分布特征可以反映出不同区域的土地使用强度以及土地使用类型。例如，如果一个区域在早高峰时段的流入客流量大且流出客流量小，可知该区域土地使用类型以住宅为主。类似的，如果一个区域在早高峰时段的流出客流量大且流入客流量小，可知该区域土地使用类型以办公场所为主。因此，通过探索客流量的空间分布特征可以为城市轨道交通以及土地使用管理、规划决策者提供指导建议。

T3. 支持多种空间分布的探索。为了更好的探索人员流动的时空分布特征，希望系统可以支持多种空间分布的探索，即用户可以通过系统自定义探索不同空间划分下客流量时空分布特征。例如，用户可以探索南北两个区域的客流量分布特征，也可以探索东西两个区域的客流量分布特征。

5.2 可视化设计概述

本节将主要介绍面向轨道交通的时空分布可视化系统的设计。系统的实现过程如下：首先基于时空聚合方法处理智能卡刷卡记录，然后将聚合后的数据呈现在空间弦图上，最后以一组空间弦图为基础进行可视化分析系统的搭建。

5.2.1 城市轨道交通刷卡记录数据集介绍

城市轨道交通自动售检票系统(Automatic Fair Collection System, AFC 系统)是一种由计算机集中控制的网络系统,其主要包含了自动售票、自动检票、自动计费、自动收费以及信息收集与统计等功能。旅客在车站终端设备刷城市交通卡入站、出站的同时,AFC 系统会把旅客的乘车信息(如入站站点名称、出站站点名称、入站时间、出站时间等)记录并存储起来,即智能卡刷卡记录。智能卡刷卡记录是一种用于表达以及分析城市轨道交通系统运行状况的重要数据源[66][67]。越来越多的研究人员基于智能卡刷卡记录分析城市轨道交通系统的运行。

面向轨道交通的时空分布可视化分析研究采用北京市 2013 年 6 月和 7 月的智能卡刷卡记录数据集。该数据集记录了由北京市的 17 个城市轨道交通线路、227 个城市轨道交通站点、150,081,258 张市政交通卡、483,614,919 条进出站记录产生的进出站客流量数据。由于 AFC 系统的复杂性以及重要性,每条智能卡刷卡记录都包含了多个属性信息。本文选择了 5 个属性作为研究对象,如表 5-1 所示。

表 5-1 城市轨道交通刷卡记录属性示意

Tab. 5-1 Demo of attribute information of urban rail transit card swipe record

入站时间	入站站点编号	入站线路编号	出站站点编号	出站线路编号
2013/6/3 5:37	29	9	33	13
2013/6/3 5:40	10	2	33	13
2013/6/3 6:03	25	5	47	13
2013/6/3 6:09	39	13	47	13

5.2.2 空间弦图设计

为了可视化分析客流量的时空分布特征,本文将基于智能卡刷卡记录构建面向轨道交通的时空分布可视化系统。智能卡刷卡记录一种典型的 OD 数据,其可以使用弦图进行可视化。但是由 2.4 节可知弦图仅能用于表达 OD 对象之间的关联关系,而无法表达 OD 对象本身蕴含的空间信息。因此,本节基于弦图设计了一个空间弦图,其既能表达 OD 对象之间的关联关系又能表达 OD 对象本身蕴含的空间信息。下面介绍本文设计的空间弦图,如图 5-1 所示。

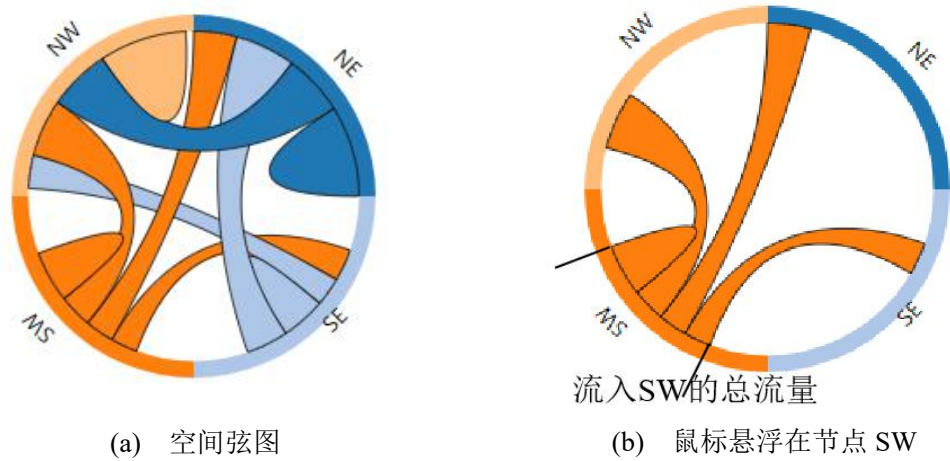


图 5-1 空间弦图示意

Fig. 5-1 Demo of spatial chord graph

与传统弦图一致的是，空间弦图也是由节点和弦两部分组成的。但是，两种弦图节点和弦的弧长计算公式以及表达的信息是不一致的。在传统弦图中，节点的弧长是由其相应 OD 对象的总入流量来决定的；在空间弦图中，节点的弧长是由其相应 OD 对象的空间位置信息决定的。因为节点相应 OD 对象空间位置信息映射的弧长与 OD 对象总入流量映射的弧长不一定是相等的，所以会出现弦溢出节点的现象，即属于某个节点的弦不与该节点相切而是与其它节点相切，如图 5-2 所示。因此，传统弦图弦的弧长计算公式在空间弦图中是不适用的。下面介绍空间弦图的弧长计算公式。

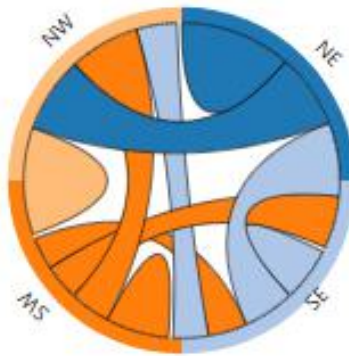


图 5-2 弦溢出其所属节点

Fig. 5-2 Chord spill node it belongs to

假设用户想要将 4 个对象 NE（地理位置为正东北）、SE（地理位置为正东南）、SW（地理位置为正西南）、NW（地理位置为正西北）之间的关联关系呈现在空间弦图上，则在空间弦图中，节点 NE 的弧长 $ArcLength_{NE} = 6.28/4$ 。

在计算弦的弧长过程中,为了保证属于任意节点的弦一定与该节点相切,首先需要对单位流量所占的弧长进行归一化处理,如公式 5-1 所示。

$$UnitFlowArcLength = \frac{6.28/4}{\max(flow(NE), flow(SE), flow(SW), flow(NW))} \quad (5-1)$$

其中 $flow(NE)$ 表示对象 NE 的总入流量。然后,即可计算弦的弧长,其中节点 NW 与节点 NE 相连的弦与节点 NE 相切的弧长如公式 5-2 所示。

$$ArcLengthOfChrod_{NW_NE} = UnitFlowArcLength \times ChrodFlow_{NW_NE} \quad (5-2)$$

其中 $ChrodFlow_{NW_NE}$ 表示对象 NW 流入对象 NE 的总流量。

5.2.3 基于方向的多尺度空间聚合方法以及时间聚合方法

因为只要不同的 OD 对象之间存在关联关系就可以相互连接,这是不受地理位置约束的,所以随着 OD 对象的增加,这很难以一种可理解的方式进行可视化。因此,本节提出了一种基于方向的多尺度空间聚合方法对 OD 对象进行聚合。同时,为了探索客流量的时间分布特征,需要对其进行时间聚合。

(1) 基于方向的多尺度空间聚合方法

算法 5-1: 基于方向的多尺度空间聚合算法

输入: $Stations$: 城市轨道交通站点地理坐标的集合;

n : 聚合后的对象个数;

O : 中心点的物理位置;

$initial_angle$: 起始方向与中心点 O 的正上方的夹角;

输出: R_s : 聚合后的 n 个对象集合;

1. for $i \leftarrow 0; i < n; i++$ do
 2. $R_s[i] \leftarrow \{\}$;
 3. for $S_i \in Stations$ do
 4. $angle \leftarrow \text{computeAngle}(O, initial_angle, S_i)$;
 5. boolean $isAboveLine \leftarrow \text{aboveLine}(O, initial_angle, S_i)$;
 6. if $isAboveLine == \text{true}$ do
 7. $angle \leftarrow 360 - angle$;
 8. $j \leftarrow (n \times angle) / 360$;
 9. $R_s[j].\text{add}(S_i)$;
 10. return R_s ;
-

基于方向的多尺度空间聚合方法可以基于站点所处的物理位置对其进行聚合,如算法 5-1 所示。该算法的输入为城市轨道交通站点地理坐标的集合 $Stations = \{S_1, S_2, \dots, S_i, \dots, S_k\}$, 其中 $S_i = (lng_i, lat_i)$; 空间聚合后对象的个数 n ; 集合 $Stations$ 中所有城市轨道交通站点形成的凸包中心点的物理位置 $O = (lng_o, lat_o)$,

O 的物理位置可以是基于用户经验或者用户兴趣自行设置的,也可以是基于公式 5-3 和公式 5-4 获取的;

$$\ln g_o = \frac{1}{n} \times \sum_{i=1}^{i \leq n} \ln g_i \quad (5-3)$$

$$lat_o = \frac{1}{n} \times \sum_{i=1}^{i \leq n} lat_i \quad (5-4)$$

起始方向与中心点 O 正上方的夹角 $initial_angle$, 其取值范围为 $[0^\circ, 180^\circ]$ 。该算法的输出为聚合后 n 个对象的集合 $R_s = \{R_1, R_2, \dots, R_i, \dots, R_n\}$, 其中 $R_i = \{S_1, S_2, \dots, S_i, \dots, S_m\}$ 且 $S_i \in Stations$ 。通过动态调节参数 O 、 $initial_angle$ 以及 n , 可以对 OD 数据进行多尺度的空间聚合。

在算法的实现过程中, 需要计算集合 $Stations$ 中每个城市轨道交通站点被聚合到哪个对象上 (line3 至 line10)。计算城市轨道交通站点 S_i 被聚合到哪个对象的步骤如下: 首先, 计算沿起始方向出发的射线 OA 与中心点 O 至 S_i 的线段 OS_i 之间的夹角 (line4), 其取值范围为 $[0^\circ, 180^\circ]$ 。然后, 将射线 OA 与线段 OS_i 之间的夹角从 $[0^\circ, 180^\circ]$ 映射到 $[0^\circ, 360^\circ]$ 。在图 5-3 中, 射线 OA 与线段 OS_1 之间夹角为 90° , 射线 OA 与线段 OS_2 之间夹角也为 90° 。但是, S_1 与 S_2 的空间位置显然是不相同的。因此当射线 OA 与线段 OS_i 之间的夹角在顺时针方向是大于 180° 时 (S_2), 需要将其取值范围设置成 $(180^\circ, 360^\circ)$ (line 5 至 line 7)。判断射线 OA 与线段 OS_i 之间的夹角在顺时针方向是否大于 180° 的方法是判断 S_i 是否在射线 OA 上方。若 S_i 是在射线 OA 的上方, 则射线 OA 与线段 OS_i 之间的夹角在顺时针方向是大于 180° 的; 否则, 射线 OA 与线段 OS_i 之间的夹角在顺时针方向是小于 180° 的 (line 5)。最后, 基于公式 5-5 即可确定城市轨道交通站点 S_i 被聚合到哪个对象上 (line 8)。

$$j = \frac{1}{360} \times (n \times angle) \quad (5-5)$$

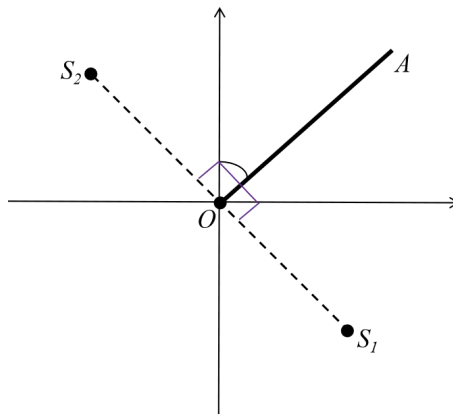


图 5-3 射线 OA 与线段 OS_i 夹角示意图

Fig. 5-3 Demo of angle between ray OA and line segment OS_i

基于算法 5-1 介绍的基于方向的多尺度空间聚合方法,可以得到 n 个聚合对象 R_s 。此时,从聚合后的 OD 对象 R_o 流入 R_d 的客流量为 R_o 内所有的城市轨道交通站点到 R_d 内所有的城市轨道交通站点的客流量总和,如公式 5-6 所示。

$$flow(R_o, R_d) = \sum_{S_i \in R_o} (\sum_{S_j \in R_d} flow(S_i, S_j)) \quad (5-6)$$

假设给定的 $Stations = \{\text{雍和宫站, 呼家楼站, 十里河站, 北京南站, 人民大学站}\}$ 、中心点物理位置为天安门、起始角度为 0° 以及聚合后对象的个数为 4, 则获取聚合后的对象集合以及计算其客流量的步骤如下: 基于站点的空间分布情况由算法 5-1 可确定每个站点被聚合到的对象, 这些站点的空间分布如图 5-4 所示。雍和宫站和呼家楼站被聚合到对象 R_0 上, 十里河站被聚合到对象 R_1 上, 北京南站被聚合到对象 R_2 上, 人民大学站被聚合到对象 R_3 上。因此聚合后的对象集合 $R_s = \{R_0, R_1, R_2, R_3\}$, 其中 $R_0 = \{\text{雍和宫站, 呼家楼站}\}$ 、 $R_1 = \{\text{十里河站}\}$ 、 $R_2 = \{\text{北京南站}\}$ 、 $R_3 = \{\text{人民大学站}\}$ 。

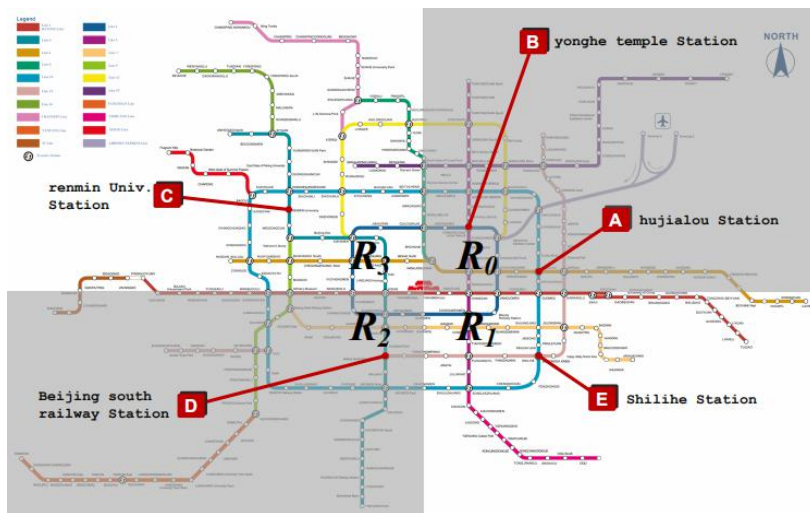


图 5-4 城市轨道交通站点的空间分布

Fig. 5-4 Spatial distribution of Stations

假设聚合后的对象 R_0 和 R_1 所包含的城市轨道交通站点间的客流量如表 5-2 所示。

表 5-2 R_0 和 R_1 包含的站点之间的客流量

Tab. 5-2 OD flow between stations in R_0 and R_1

	雍和宫站	呼家楼站	十里河站
雍和宫站	0	1000	1500
呼家楼站	2000	0	2500
十里河站	3000	3500	0

基于公式 5-4, 可以得到对象 R_0 和 R_1 之间的客流量如表 5-3 所示。

表 5-3 R_0 和 R_1 之间的客流量
Tab. 5-3 OD flow between R_0 and R_1

	R_0	R_1
R_0	3000	4000
R_1	6500	0

(2) 时间聚合方法

对于时间的聚合，需要基于数据所处的时间范围将其划分为不同的时间间隔。时间间隔的划分取决于所研究数据的时间周期（如日、周、月）、特定时间周期的相关性以及所研究数据的时空密度等^{错误!未找到引用源。}。本文以每日的周期作为研究对象且一个时间间隔是由一个空间弦图呈现的，当空间弦图的个数过多时无法清晰的观察客流量的时间分布信息。因此，为了能够捕捉到一天中人群流动行为的自然差异且控制划分的时间间隔个数，本文将每天作为一个周期按照如下标准进行间隔的划分：早高峰(6:00, 10:00)、营业时间(10:00, 16:00)、晚高峰(16:00, 20:00)和深夜(20:00, 6:00)。

通过比较空间弦图中弦的弧长，可以确定各个区域间客流量分布情况。然而，基于 5.2.3 节介绍的方法，弦的弧长大小只在当前空间弦图内是有比较意义的，在不同的空间弦图中比较弦的弧长大小是没有意义的。为了能够分析客流量的时空分布特征，不仅要比较单个空间弦图内弦的弧长大小，还需要比较不同空间弦图弦的弧长大小。因此，本文采用如下方法对不同空间弦图内单位流量所占的弧长进行归一化处理，如公式 5-7 所示。

$$ArcLengthOfUnitFlow = \frac{6.28}{n \times \max(flow(t_1), flow(t_2), flow(t_3), flow(t_4))} \quad (5-7)$$

其中 n 表示划分的区域个数， $flow(t_1)$ 表示 t_1 时间间隔内 n 个对象中的最大总入流量。然后，基于公式 5-2 即可计算弦的弧长。

5.2.4 可视化分析系统概述

为了了解客流量时空分布特征，本文基于 Python 以及 D3 构建了一个时空分布可视化系统，如图 5-5 所示。该系统共包含两个视图，其中视图 A 主要为该系统提供交互式操作，如图 5-5 中区域 A 所示。在视图 A 中，用户通过自定义参数 *division scale*、*initial angle*、*center coordinates*、*data selection* 可以探索多尺度的空间客流量分布特征（T3）。视图 B 基于一组空间弦图展示了用户自定义的时空客流量特征，如图 5-5 中区域 B 所示。空间弦图虽然可以展示空间信息，但不能展示时间信息，有关空间弦图的介绍见 5.2.3 节。如果想用空间弦图同时显

示空间和时间信息，最直接的方法是使用多个空间弦图，这也是比较合适的，因为若是在空间弦图上继续过度设计会降低其可读性。因此，在该系统中，使用一组空间弦图展示时空客流量分布特征。通过视图 B，用户可以分析时空客流量情况。为城市轨道交通以及土地使用管理、规划决策者提供指导建议。

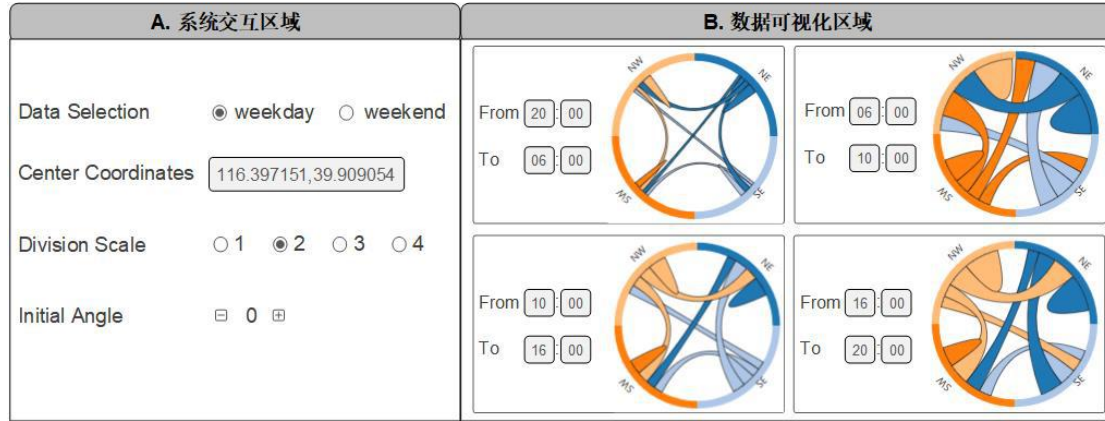


图 5-5 时空分布可视化系统概览图

Fig. 5-5 Overview of temporal and spatial distribution visualization system

该系统提供的用户交互操作如下：

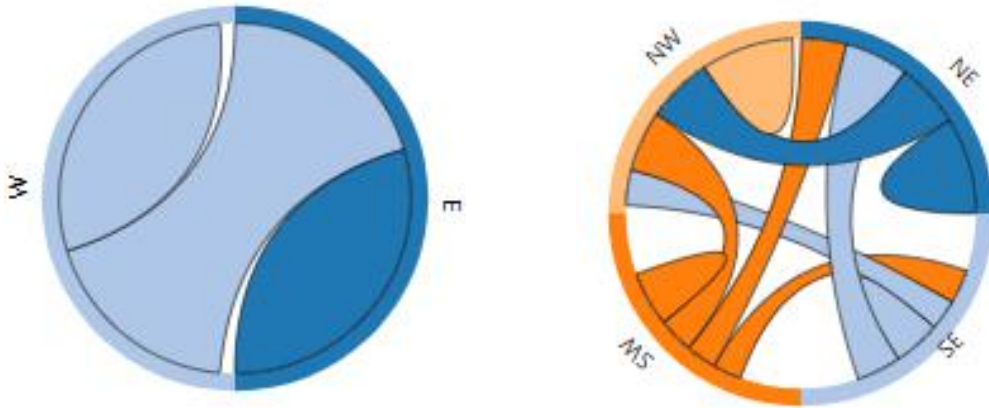
(1) 数据选择。通过在视图 A 中设置参数 *data selection*，用户可以选择不同日期的数据。该参数共有两个可选项 *workday* 和 *weekend*。若选择 *workday* 则视图 B 中呈现的是工作日的客流量数据；若选择 *weekend* 则视图 B 中呈现的的工作日的客流量数据。

(2) 设置空间划分的中心坐标。通过在视图 A 中设置参数 *center coordinates*，用户可以将指定空间划分的中心点坐标。参数 *center coordinates* 的默认值如公式 5-3 以及公式 5-4 所示。

(3) 设置空间划分尺度。通过在视图 A 中设置参数 *division scale*，用户可以将整体空间划分成指定个数的区域。若将视图 A 中的参数 *division scale* 设置成 1，则视图 B 中的空间弦图是被划分成 2 个区域的，如图 5-6(a)所示；若将视图 A 中的参数 *division scale* 设置成 2，则视图 B 中的空间弦图是被划分成 4 个区域的，如图 5-6(b)所示。因为空间聚合粒度过细是没有比要的，所以视图 A 中参数 *division scale* 取值范围为[1, 4]。

(4) 设置空间划分方向。通过在视图 A 中设置参数 *initial angle*，用户可以将整体空间按照指定方向进行划分。参数 *initial angle* 表示的是沿中心坐标的正上方向顺时针方向旋转的角度，取值范围为(0, 360 / *division scale*)。假设参数 *division scale* 设置为 2，若参数 *initial angle* 设置为 0，则整体空间被划分为东、西两个区域，如图 5-7(a)所示；若参数 *initial angle* 设置为 90，则整体空间被划

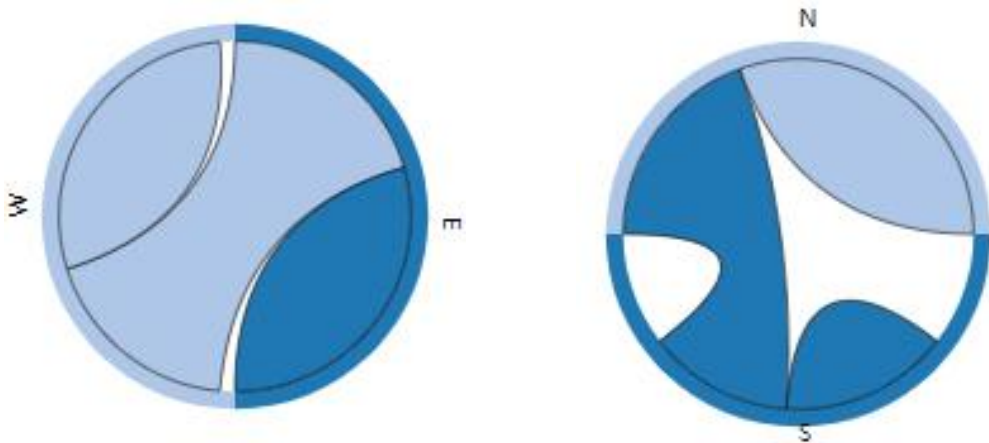
分为南、北两个区域，如图 5-7(b)所示。



(c) 2 个区域的空间弦图 ($division\ scale = 1$) (d) 4 个区域的空间弦图 ($division\ scale = 2$)

图 5-6 $division\ scale$ 影响空间弦图的绘制

Fig. 5-6 Influence of $division\ scale$ on drawing spatial chord graph



(a) 南北区域的空间弦图 ($initial\ angle = 90$) (b) 东西区域的空间弦图 ($initial\ angle = 0$)

图 5-7 $initial\ angle$ 影响空间弦图的绘制

Fig. 5-7 Influence of $initial\ angle$ on drawing spatial chord graph

5.3 案例研究

本节基于两个案例研究来验证面向轨道交通的时空分布可视化系统的有效性。首先，探索工作日与非工作日客流量时间分布特征差异 (T1)。然后，探索南、北与东、西客流量分布特征差异 (T2)。

(1) 讨论如何基于本系统探索工作日与非工作日客流量时间分布特征差异。

首先，探索工作日客流量时间分布特征。在视图 1 中进行参数设置，*data selection* 选中 *weekday*，参数 *center coordinates* 设置为(116.397151, 39.909054)，参数 *initial angle* 设置为 0，参数 *division scale* 设置为 4。视图 A 按照上述参数设

置后，视图 B 中生成的一组空间弦图如图 5-8 所示。由图 5-8 可知四个时间段的客流量存在显著差异，四个时间段的总客流量大小关系如下： $(6:00, 10:00) > (16:00, 20:00) > (10:00, 16:00) > (20:00, 6:00)$ ，并且客流量主要集中在 $(6:00, 10:00)$ 和 $(16:00, 20:00)$ 两个时段。因此，可以得出工作日客流量高峰时段为 $(6:00, 10:00)$ 和 $(16:00, 20:00)$ 。发现客流量高峰时段后，通过比较高峰时段内客流量的空间分布特征，可以为城市轨道交通的运营提供指导建议。

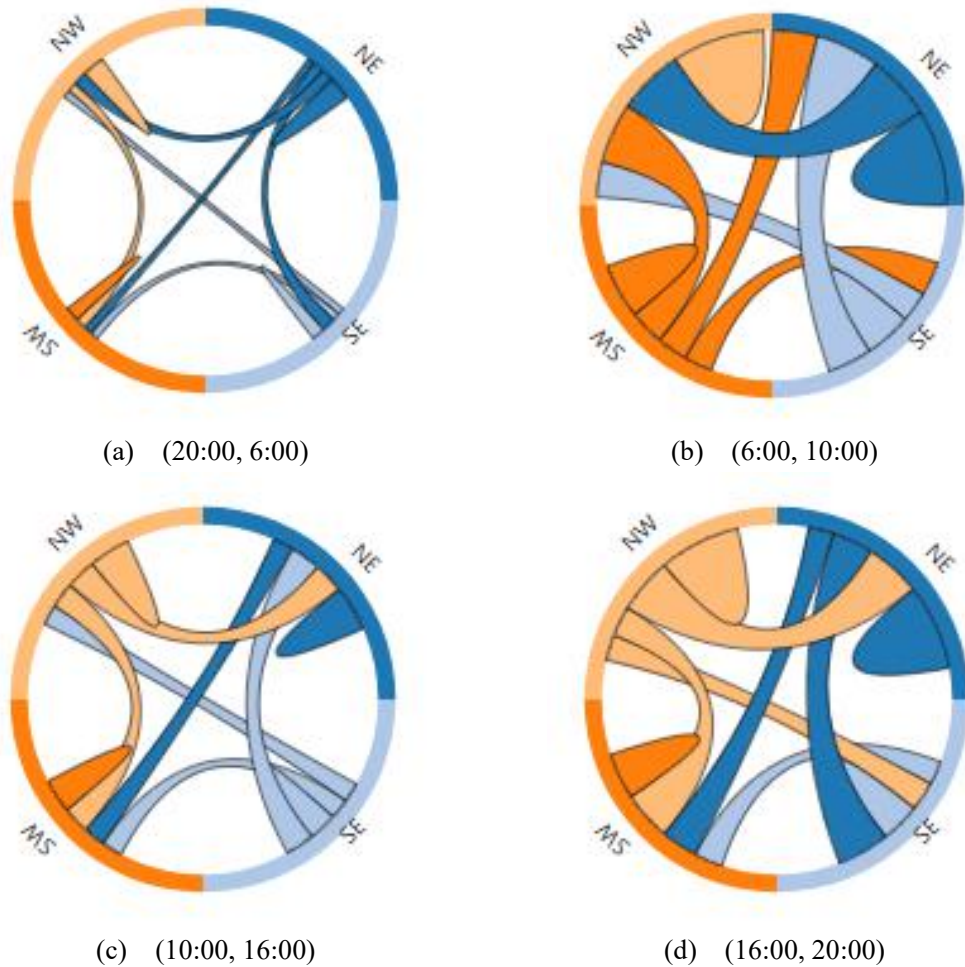


图 5-8 工作日客流量时空分布特征

Fig. 5-8 Temporal and spatial distribution characteristic of passenger flow on weekday

保持其它参数不变，*data selection* 选中 *weekend*，非工作日的客流量时空分布特征如图 5-9 所示。由图 5-9 可知非工作日四个时间段的总客流量大小关系如下： $(10:00, 16:00) > (16:00, 20:00) > (6:00, 10:00) > (20:00, 6:00)$ ，并且除 $(20:00, 6:00)$ 时段外，其它三个时段的客流量分布较均匀。因此，可以得出工作日的客流量分布具有比较强的时间聚集性，非工作日客流量时间分布是更均匀的结论。

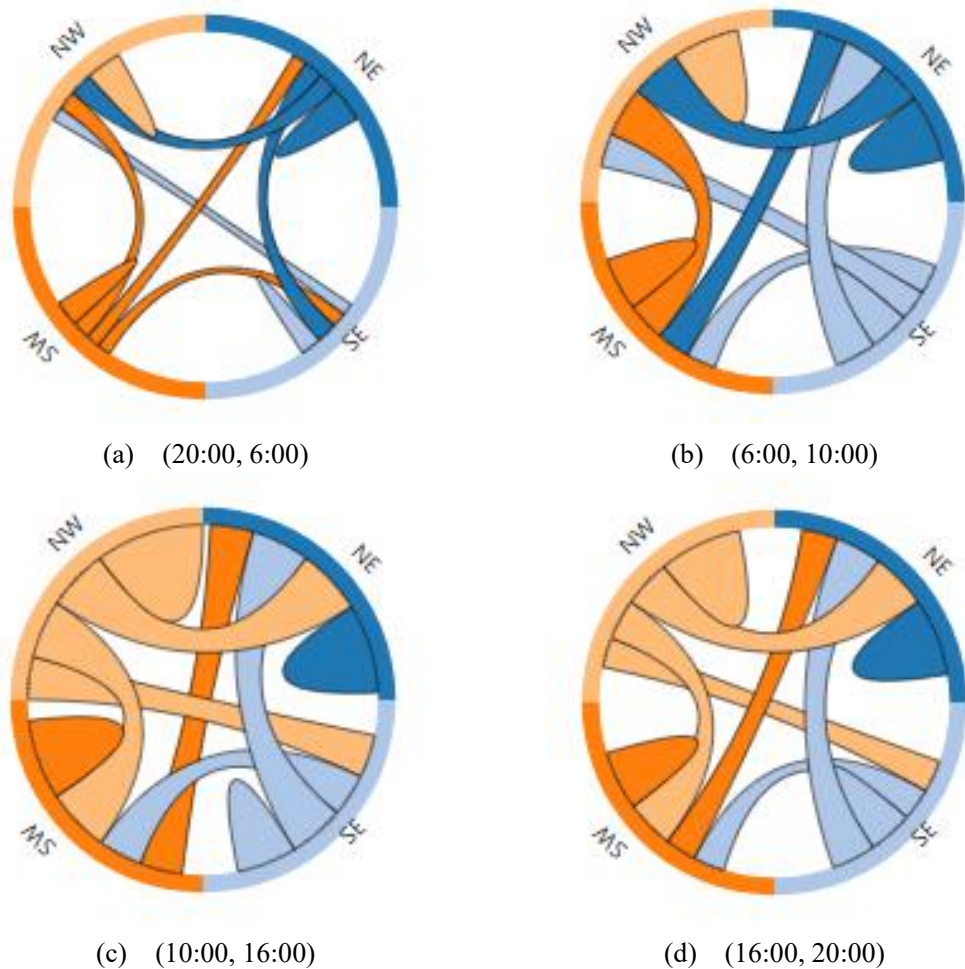


图 5-9 非工作日客流量时空分布特征

Fig. 5-9 Temporal and spatial distribution characteristic of passenger flow on weekend

(2) 讨论如何基于本系统探索南、北与东、西客流量分布特征差异。

首先，探索南、北两个方向空间客流量分布特征。在视图 A 中进行参数设置，*data selection* 选中 *weekday*，参数 *center coordinates* 设置为(116.397151, 39.909054)，参数 *initial angle* 设置为 90，参数 *division scale* 设置为 2。视图 A 按照上述参数设置后，视图 B 中生成的一组空间弦图如图 5-10 所示。由图 5-10(b) 可知区域 N 流入区域 S 弦的弧长远大于区域 S 流入区域 N 弦的弧长，故北京早高峰阶段人员流动方向为由南到北。由图 5-10(d) 可知区域 S 流入区域 N 弦的弧长远大于区域 N 流入区域 S 弦的弧长，故北京晚高峰阶段人员流动方向为由北向南。由图 5-10(a) 以及 5-10(c) 可知，其它两个时段区域 N 流入区域 S 弦的弧长基本等于区域 S 流入区域 N 弦的弧长，无明显的流动方向。因为早高峰时段的出行目的以上班为主，乘客的出发地是家，目的地是公司。晚高峰时段的出行目的以回家为主，出发地是公司，目的地是家。由此可知北京市大部分的公司位于其北部。这将导致城市居民生活和工作地点分离，使得居民出行通勤时间较长。因此，提议为土地使用管理、规划决策者提供指导建议。

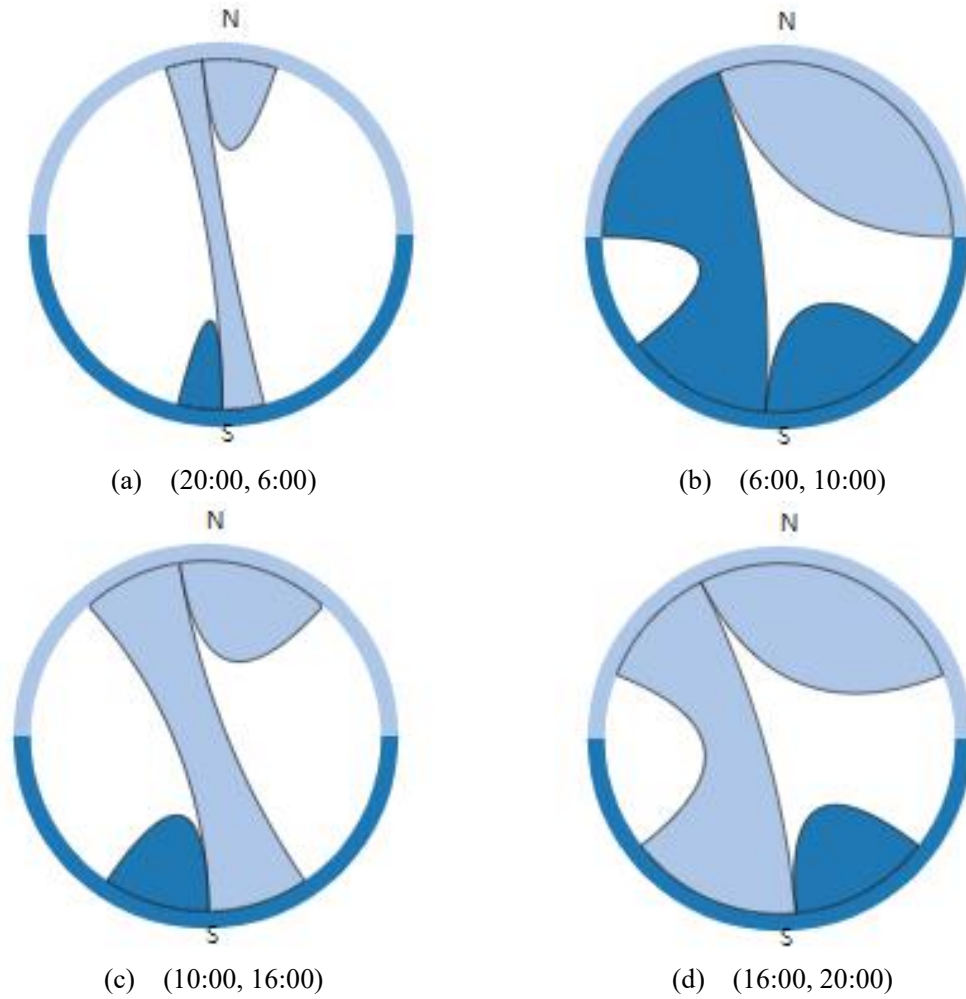


图 5-10 南北区域客流量时空分布特征

Fig. 5-10 Temporal and spatial distribution characteristic of passenger flow(N & S)

保持其它参数不变, *initial angle* 设置为 90, 东、西客流量空间分布如图 5-11 所示。不同于南、北两个方向在早、晚高峰阶段有明显的大规模人员流动方向, 东、西两个方向四个时间阶段皆是区域 N 流入区域 S 弦的弧长基本等于区域 S 流入区域 N 弦的弧长, 无明显流动方向。因为早高峰时段的出行目的以上班为主, 乘客的出发地是家, 目的地是公司。晚高峰时段的出行目的以回家为主, 出发地是公司, 目的地是家。由此可知住宅以及公司的空间分布在东、西两个方向上分布是较均匀的。

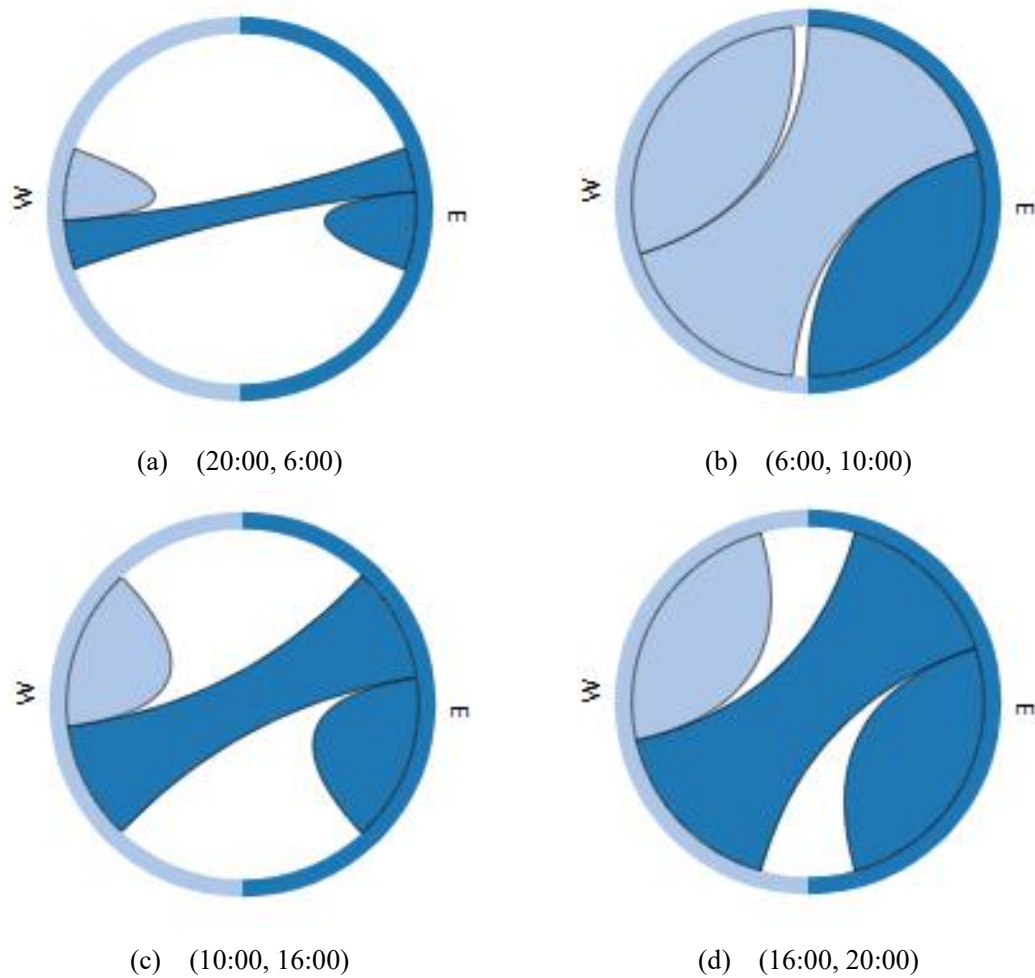


图 5-11 东西区域客流量时空分布特征

Fig. 5-11 Temporal and spatial distribution characteristic of passenger flow between(E & W)

5.4 本章小结

本章首先设计了一个空间弦图,既可以表达 OD 对象之间的关联关系又能表达 OD 对象蕴含的空间信息。同时,为了避免空间弦图中出现大量弦交叉的现象,提出了一种基于方向的多尺度空间聚合方法。然后基于一组空间弦图构建了一个面向轨道交通的时空分布可视化系统。最后,通过案例研究验证系统的有效性。

结 论

本文旨在研究可视化分析及其在轨道交通以及土地使用上面的应用，主要包括两个方面。一是基于密度可视化分析站点周边土地使用特征，为轨道交通系统的运营以及站点周边土地的开发和使用规划提供指导意义。二是基于时空分布可视化分析城市轨道交通客流量的时空分布特征，为分析土地使用的空间布局以及轨道交通系统的运营提供指导意义。本文的主要研究内容如下：

(1) 提出了一种面向主题的非交叉层次 POI 提取方法。首先，基于 Voronoi 图确定每个城市轨道交通站点的覆盖区域并基于射线法确定每个站点覆盖区域内的 POI 集合。然后，为了提取相关主题的 POI 集合，基于相似度算法过滤站点覆盖区域的 POI 集合。然后，为了将站点覆盖区域的 POI 集合按照用户感兴趣程度进行分类，基于 Skyline 查询技术抽取层次 POI 集合。最后，为了避免不同层次 POI 构成的封闭区域在地理位置上相互交叉，基于 Graham 扫描法提取非交叉层次 POI 集合。

(2) 基于面向主题的非交叉层次 POI 提取方法以及三维等高线模型构建了一个可视化分析系统。通过该系统比较与分析不同站点周边土地使用特征的差异，为轨道交通的运营以及站点周边土地的开发与使用规划提供指导建议。同时，通过案例研究与实验分析验证系统的有效性。

(3) 设计了一个空间弦图，既可以表达 OD (Origin-Destination) 对象之间的关联关系又能表达 OD 对象蕴含的空间信息。同时，为了避免空间弦图中出现大量弦交叉的现象，提出了一种基于方向的多尺度空间聚合方法。该方法基于城市轨道交通站点的物理位置对站点进行聚合，并通过多个预设参数控制空间聚合的尺度。

(4) 基于多尺度空间聚合方法以及空间弦图构建了一个可视化分析系统。通过该系统比较与分析客流量时空分布特征，为分析土地使用的空间布局以及轨道交通系统的运营提供指导建议。同时，通过案例研究验证系统的有效性。

虽然本文在基于密度可视化分析站点周边土地使用特征以及基于时空分布可视化分析城市轨道交通客流量的时空分布特征上取得了一些阶段性成果，但是仍然存在一些工作需要继续深入研究。

(1) 在获取面向主题词的非交叉层次 POI 集合的过程中，本文使用 Jaccard 系数过滤相关主题的 POI 集合。由于 Jaccard 系数在计算文本相似度上是有一定局限性的，所以后续将研究更合适的过滤相关主题 POI 集合的方法。

(2) 方向不是唯一影响客流量的因素，距离也是一个影响客流量的关键因素，所以后续将研究基于方向以及距离的空间聚合方法。

参考文献

- [1] Lu K, Han B, Baoming Lu, et al. Urban Rail Transit in China: Progress Report and Analysis (2008-2015)[J]. Urban Rail Transit, 2016, 2(3):93-105.
- [2] Yang X, Yin H, Wu J, et al. Recognizing the Critical Stations in Urban Rail Networks: An Analysis Method Based on the Smart-Card Data[J]. IEEE Intelligent Transportation Systems Magazine, 2018, 11(1):29-35.
- [3] Pojani, Dorina, Stead, et al. Sustainable Urban Transport in the Developing World: Beyond Megacities[J]. Sustainability, 2015, 7(6):7784-7805.
- [4] Nation Bureau of Statistics, PRC. China's Main Demographic Indicators from the 3rd National Population Census, 1982.
- [5] Nation Bureau of Statistics, PRC. China's Main Demographic Indicators from the 6rd National Population Census, 2011.
- [6] Vanderhaegen S, Canters F. Mapping Urban Form and Function at City Block Level Using Spatial Metrics[J]. Landscape and Urban Planning, 2017, 167:399-409.
- [7] Zhang X, Du S, Wang Q. Hierarchical Semantic Cognition for Urban Functional Zones with VHR Satellite Images and POI Data[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2017, 132:170-184.
- [8] Xing H. Integrating Landscape Metrics and Socioeconomic Features for Urban Functional Region Classification[J]. Computers, Environment and Urban Systems, 2018, 72:134-145.
- [9] Crane B R. The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies[J]. Transportation Research Part A: Policy and Practice, 2001, 35(9):823-845.
- [10] Small K A, Song S. "Wasteful" Commuting: A Resolution[J]. Journal of Political Economy, 1992, 100.
- [11] Giuliano G, Small K. Is the Journey to Work Explained by Urban Structure?[J]. Urban Studies, 1993, 30(9):1485-1500.
- [12] Wang F. Modeling Commuting Patterns in Chicago in a GIS Environment: A Job Accessibility Perspective[J]. Professional Geographer, 2000, 52(1):120-133.
- [13] Black W R. Transportation: A Geographical Analysis[J]. Geographic Information Systems, 2003, 49(2):221-223.
- [14] Yu L, Yu X, et al. Urban Land Uses and Traffic 'source-sink areas': Evidence from GPS-enabled Taxi Data in Shanghai[J]. Landscape & Urban Planning, 2012, 106(1):73-87.
- [15] Yuan J, Zheng Y, Xie X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs[C]// ACM SIGKDD international conference on Knowledge discovery and data mining, 2012:186-194.
- [16] Crooks A, Jenkins A, et al. Crowd Sourcing Urban Form and Function[J]. International Journal of Geographical Information Science, 2015.

- [17] Shan J A, Aa B, Fr C, et al. Mining point-of-interest Data from Social Networks for Urban Land Use Classification and Disaggregation - ScienceDirect[J]. Computers, Environment and Urban Systems, 2015, 53:36-46.
- [18] Jenkins A, Croitoru A, Crooks A T, et al. Crowd Sourcing a Collective Sense of Place[J]. PLoS ONE, 2016, 11(4):e0152932.
- [19] Pan H, Li J, Shen Q, et al. What Determines Rail Transit Passenger Volume? Implications for Transit Oriented Development Planning[J]. Transportation Research Part D Transport & Environment, 2017, 57:52-63.
- [20] Bratanu D, Nedelcu I, Datcu M. Bridging the Semantic Gap for Satellite Image Annotation and Automatic Mapping Applications[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2011, 4(1):193-204.
- [21] Zhang X, Du S. A Linear Dirichlet Mixture Model for Decomposing Scenes: Application to Analyzing Urban Functional Zonings[J]. Remote Sensing of Environment, 2015, 169:37-49.
- [22] Li W, Liu H, Wang Y, et al. Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas[J]. IEEE Access, 2019, 7:36274-36284.
- [23] Hu T, Yang J, Li X, Gong P. Mapping Urban Land Use by Using Landsat Images and Open Social Data[J]. Remote Sensing, 2016, 8(2):151.
- [24] Yao Y, Xia L, Liu X, et al. Sensing Spatial Distribution of Urban Land Use by Integrating Points-of-Interest and Google Word2Vec Model[J]. International Journal of Geographical Information Systems, 2016, 31(4):825-848.
- [25] Wei Z, Fu C W, Arisona S M, et al. Visualizing Waypoints-Constrained Origin-Destination Patterns for Massive Transportation Data[J]. Computer Graphics Forum, 2015, 35(8):732-733.
- [26] Francis A M, Schneider J B. Using Computer Graphics to Map Origin-Destination Data Describing Health Care Delivery Systems[J]. Social Science & Medicine, 1984, 18(5):405-420.
- [27] Waldo, R, Tobler. Experiments In Migration Mapping By Computer[J]. American Cartographer, 2013.
- [28] Van D, Wijk J V. Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations[J]. IEEE Transactions on Visualization & Computer Graphics, 2014, 20(12):2310-2319.
- [29] Landesberger T V, Kuijper A, Schreck T, et al. Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges[J]. Computer Graphics Forum, 2011, 30(6):1719-1749.
- [30] Phan D, Ling X, Yeh R, et al. Flow Map Layout[C]// IEEE Symposium on Information Visualization. IEEE Computer Society, 2005.
- [31] Buchin K, Speckmann B, Verbeek K. Flow Map Layout via Spiral Trees[J]. IEEE Transactions on Visualization & Computer Graphics, 2011, 17(12):2536.

- [32] Boyandin I, Bertini E, Bak P, et al. Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data[C]// Computer Graphics Forum. John Wiley & Sons, Ltd. 2011:971-980.
- [33] Guo D. Visual Analytics of Spatial Interaction Patterns for Pandemic Decision Support[J]. International Journal of Geographical Information Science, 2007, 21(8):859-877.
- [34] Andrienko G, Andrienko N, Fuchs G, et al. Revealing Patterns and Trends of Mass Mobility Through Spatial and Temporal Abstraction of Origin-Destination Movement Data[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(9):2120-2136.
- [35] Wood J, Dykes J, Slingsby A. Visualisation of Origins, Destinations and Flows with OD Maps[J]. Cartographic Journal, 2010, 47(2):117-129.
- [36] Okabe A, Boots B, Sugihara K, et al. Kendall.Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, Second Edition[M]. Wiley Series in Probability and Mathematical Statistics. Wiley, 2020.
- [37] 刘金义, 刘爽. Voronoi图应用综述[J]. 图学学报, 2004(02):131-138.
- [38] Qiao C, Zhao R, Chen J, and Chen Y. Contour Line Generation Method Based on Proximity within Voronoi[J]. Geomatics Inf. Sci. Wuhan Univ., vol. 30, no. 9, pp. 801-804, 2005.
- [39] 马少蒙. 站点能力约束下的城区铁路站点配送区域规划[D]. 北京交通大学, 2018.
- [40] 裴秋艳. 核事故应急响应撤离关键技术研究及系统实现[D]. 中国科学技术大学, 2020.
- [41] 任静, 季民, 陈兆宁. 一种海洋流场拓扑区域划分方法[J]. 测绘科学技术学报, 2020,37(05):545-550.
- [42] Shimrat, M. Algorithm 112: Position of Point Relative to Polygon[J]. Communications of the Acm, 1962, 5(8):434.
- [43] 肖晓伟, 肖迪, 林锦国,等. 多目标优化问题的研究概述[J]. 计算机应用研究, 2011, 28(003):805-808,827.
- [44] D Kossmann, Ramsak F, Rost S. Shooting Stars in the Sky[C]// VLDB '02: Proceedings of the 28th International Conference on Very Large Databases, 2002:275-286.
- [45] Papadias D, Tao Y, Fu G, et al. Progressive Skyline Computation in Database Systems[J]. ACM Transactions on Database Systems, 2005, 30(1):41-82.
- [46] Tao Y, Xiao X, Pei J. Efficient Skyline and Top-k Retrieval in Subspaces[J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(8):1072-1088.
- [47] Chomicki J, Godfrey P, Gryz J, et al. Skyline with Presorting[C]// IEEE Computer Society. IEEE Computer Society, 2003:717-719.
- [48] Godfrey P, Shipley R, Gryz J. Maximal Vector Computation in Large Data Sets[C]// VLDB '02: Proceedings of the 31st International Conference on Very Large Data Bases, 2005:229-240.
- [49] Bartolini I, Ciaccia P, Patella M. SaLSa: Computing the Skyline without Scanning the Whole Sky[C]// ACM International Conference on Information & Knowledge Management. ACM, 2006:405-414.

- [50] Bartolini I, Ciaccia P, Patella M. Efficient Sort-Based Skyline Evaluation[J]. *Acm Transactions on Database Systems*, 2008, 33(4):1-49.
- [51] Zhang S, Mamoulis N, Cheung D W. Scalable Skyline Computation Using Object-Based Space Partitioning[C]// *SIGMOD international conference on Management of data*, 2009:483-494.
- [52] 王伟璐. 基于快速凸包算法的4D航迹规划研究[D]. 中国民航大学, 2020.
- [53] 周先林. GNSS/INS组合导航性能改善技术研究[D]. 中国科学院大学(中国科学院国家授时中心), 2020.
- [54] 李国涛. 可展开桁架式抓取机械手设计与控制研究[D]. 哈尔滨工业大学, 2020.
- [55] Scholz S S, Dillmann M, Flohr A, et al. Contemporary Scientometric Analyses Using a Novel Web Application: the Science Performance Evaluation (SciPE) Approach[J]. *Clinical Research in Cardiology*, 2020, 109(7):810-818.
- [56] 冯涛, 艾延华, 杨伟, 等. 基于Circos弦图的居民出行模式可视分析[J]. *华中师范大学学报(自科版)*, 2016, 50(003):471-480.
- [57] 鲁亚楠, 姚顺波, 邓元杰, 等. 陕北地区土地利用及景观格局变化对生态服务价值的影响——基于退耕还林(草)背景[J]. *中国农业资源与区划*, 2019, 40(11):180-192.
- [58] 王倩. 信息可视化设计在上海市公共交通信息发布中的应用研究[D]. 哈尔滨工业大学, 2015.
- [59] 池娇, 焦利民, 董婷, 等. 基于POI数据的城市功能区定量识别及其可视化[J]. *测绘地理信息*, 2016, 41(02):68-73.
- [60] 陈香, 李晓明, 詹然, 徐卫民. 从城市兴趣点中提取多层地标方法探究[J]. *测绘与空间地理信息*, 2015, 38(10):129-136.
- [61] Sutherland I E S R F. A Characterization of Ten Hidden-Surface Algorithms[J]. *ACM Computing Surveys (CSUR)*, 1974.
- [62] Shimrat, M. Algorithm 112: Position of Point Relative to Polygon[J]. *Communications of the ACM*, 1962, 5(8):434.
- [63] Ferguson, Helaman Rolfe Pratt Holmes Stewart, Linda. Point in Polygon Algorithms. , 1973.
- [64] Fakas G J, Cai Z, Mamoulis N. Diverse and Proportional Size-l object Summaries Using Pairwise Relevance[J]. *The VLDB journal*, 2016, 25(6):791-816.
- [65] Cai Z, Li T, Su X, et al. Research on Analysis Method of Characteristics Generation of Urban Rail Transit[J]. *IEEE Transaction on Intelligent Transportation Systems*, 2020, 21(9):3608-3620.
- [66] Pelletiera M P. Smart Card Data Use in Public Transtic: A literature Review[J]. *Transportation Research Part C*, 2011, 10.
- [67] Zheng Y, et al. Urban Computing: Concepts, Methodologies, and Applications[J]. *ACM Transactions on Intelligent Systems and Technology (TIST) - Special Section on Urban Computing*, 2014.

攻读硕士学位期间所发表的学术论文

- [1] Z.Cai, G.Sun, X.Su, T.Li, L.Guo, "Visual Analysis of Land Use Characteristics Around Urban Rail Transit Stations" in IEEE Transactions on Intelligent Transportation Systems(Online Publish), doi: 10.1109/TITS.2020.2989811.
- [2] 才智, 孙功玉, 苏醒. 一种面向主题的城市轨道交通站点层次 POI 提取方法, 发明, ZL201910681682.X, 授权, 2021.02.04.
- [3] 才智, 孙功玉, 米庆, 苏醒, 刘芷凝, 朗垠. 一种城市轨道交通站点的出行模式可视化分析方法, 发明, 202010590251.5, 实审, 2020.11.20.
- [4] Z.Cai, G.Sun, Q.Mi, X.Su, L.Ju, L.Guo, "Pattern Analysis of Travel Behavior for Urban Rail Transit" in IEEE Transactions on Intelligent Transportation Systems, 审稿.

致 谢

时光荏苒，百感交集，目之所及，皆是回忆。在这三年的研究生学习和生活中，我的导师、同学和家人都给予了我很多指导、帮助和支持，在此我要衷心的感谢大家。

首先，我要感谢我的研究生导师才智副教授在研究生期间对我学业上的指导和生活上的照顾。在科研工作上，才老师兢兢业业、勤勤恳恳，他的钻研态度深深地影响了我。同时，在毕业课题的完成过程中，他给予了我很多指导，尤其是在论文选题、实验设计以及论文撰写等方面。在此，我深深地感谢才老师对我的细心栽培，并一直铭记于心。

此外，我还要感谢苏醒、米庆老师。在我完成小论文初稿时，苏老师和米老师在百忙之中抽出时间来对我的论文认真批改，严格把关论文内容，提出许多中肯的指导意见，使我顺利完成论文，万分感谢苏老师和米老师的帮助。

其次，我要感谢实验室的全体师生以及我的室友，感谢师兄、师姐、室友们的帮助与陪伴。再次，我要感谢我的朋友和家人。感谢家人们对我的鼓励和支持，陪我走过这近二十年的求学之路。

最后，感谢评阅、评议和答辩委员会的各位老师们在百忙之中给予我的指导。

