



(12) 发明专利申请

(10) 申请公布号 CN 116452016 A

(43) 申请公布日 2023. 07. 18

(21) 申请号 202310406652.4

G06F 18/25 (2023.01)

(22) 申请日 2023.04.17

(71) 申请人 成都市规划设计研究院

地址 610000 四川省成都市高新区府城大道西段399号天府新谷9号楼1单元22层

(72) 发明人 蒋源 张晓荣 唐鹏 于儒海 周垠 曹埭 张娜 丁睿

(74) 专利代理机构 成都行之智信知识产权代理有限公司 51256

专利代理师 张宽

(51) Int. Cl.

G06Q 10/0637 (2023.01)

G06Q 50/26 (2012.01)

G06F 18/214 (2023.01)

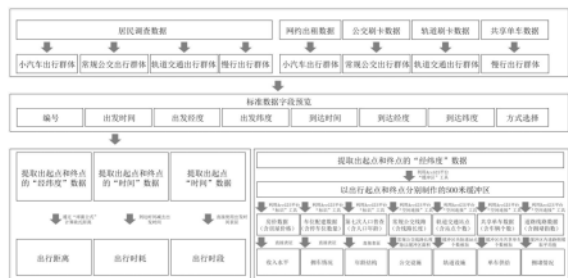
权利要求书2页 说明书6页 附图1页

(54) 发明名称

基于XGBoost模型的城市交通结构分析及系统

(57) 摘要

本发明涉及城市规划分析技术领域,公开了一种基于XGBoost模型的城市交通结构分析方法及系统,S1、通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;S2、提取所述居民出行方式数据的特征指标;S3、将所述特征指标划分为训练集和测试集,将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的的城市交通结构分析模型;S4、将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。本发明处理分析规模和空间尺度更大、更广。



1. 一种基于XGBoost模型的城市交通结构分析方法,其特征在于,包括如下步骤:

S1、通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;

S2、提取所述居民出行方式数据的特征指标;

S3、将所述特征指标划分为训练集和测试集,将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的的城市交通结构分析模型;

S4、将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

2. 根据权利要求1所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,所述居民出行方式数据包括居民入户抽样调查得到的居民出行调查数据、多元数据融合分析获取的轨道刷卡、公交刷卡、共享单车、网约出租运行数据。

3. 根据权利要求2所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,通过居民入户抽样调查得到的居民出行调查数据占所述居民出行方式数据的3.5%-4.0%。

4. 根据权利要求2所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,提取所述居民出行方式数据的特征指标的具体步骤为:

S2.1、将所述居民出行方式数据的经纬度数据在ArcGIS平台中进行空间化构建面板数据;

S2.2、以所述经纬度数据分别作500米缓冲区,将所述面板数据叠合城市用地、房屋租赁、道路拥堵、轨道站点、公交线网的数据,匹配得到每一个所述居民出行方式数据的特征指标。

5. 根据权利要求4所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,所述特征指标包括出行特征、居民属性和交通状态,其中,

所述出行特征包括居民的出行距离、出行时耗、出行时段、起点属性和终点属性;

所述居民属性包括居民的收入水平、拥车情况和年龄结构;

所述交通状态包括起点附近公交设施、轨道设施、共享单车供给情况和出行时段的道路拥堵情况。

6. 根据权利要求1所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,S3的训练过程为:

S3.1、将所述训练集的特征指标按照标准格式构建为面板数据,所述面板数据的字段内容和顺序依次为:1. 出行方式;2. 出行距离;3. 出行时段;4. 出发区域;5. 到达区域;6. 收入条件;7. 停车条件;8. 年龄情况;9. 拥堵情况;10. 轨道设施;11. 公交设施;12. 共享单车供给;

S3.2、运用python语言,调用XGBoost模型的python语言模块,以所述训练集的特征指标为基础进行XGBoost模型进行训练,具体包括将“1. 出行方式”字段作为因变量值,将“2. 出行距离;3. 出行时段;4. 出发区域;5. 到达区域;6. 收入条件;7. 停车条件;8. 年龄情况;9. 拥堵情况;10. 轨道设施;11. 公交设施;12. 共享单车供给”字段作为自变量值,设置所述XGBoost模型特征参数全为初始值,进行代码运行即开始XGBoost模型训练;

S3.3、等待代码运行结束即完成所述XGBoost模型训练,训练完成的所述XGBoost模型即为城市交通结构分析模型。

7.根据权利要求1所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,所述训练集和测试集的比例为4:1。

8.根据权利要求1所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,在将所述手机信令数据输入至所述城市交通结构分析模型中之前,先对手机信令数据在ArcGIS平台中进行“空间化”称为点状矢量数据。

9.根据权利要求8所述的一种基于XGBoost模型的城市交通结构分析方法,其特征在于,S4的具体步骤为:

S4.1、在ArcGIS平台中以手机信令的点状矢量数据为基础,运用“缓冲区”工具,以所述手机信令的点状矢量数据为中心构建半径为500米的圆形缓冲区;

S4.2、利用“空间连接”工具,统计以所述手机信令的点状矢量数据为中心的空间位置在圆形缓冲区中的特征指标;

S4.3、将关于所述手机信令的特征指标输入至所述城市交通结构分析模型,得到手机信令所在位置的交通结构分析结果;

S4.4、循环S4.1-S4.3,直到将需要调查的地区的所有居民的手机信令数据全部输入至所述城市交通结构分析模型中,最终得到该地区的城市交通结构分析结果。

10.一种基于XGBoost模型的城市交通结构分析系统,其特征在于,包括:

采集模块,用于通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;

提取模块,用于提取所述居民出行方式数据的特征指标,将所述特征指标划分为训练集和测试集;

训练模块,用于将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的的城市交通结构分析模型;

输入模块,用于将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

## 基于XGBoost模型的城市交通结构分析方法及系统

### 技术领域

[0001] 本发明涉及城市交通规划分析技术领域,具体涉及一种基于XGBoost模型的城市交通结构分析方法及系统。

### 背景技术

[0002] 城市交通结构是指在城市居民出行总量中,小汽车、轨道交通、常规公交及慢行交通(含非机动车和步行)等出行方式的占比情况。城市交通结构是评估分析城市交通发展状况、制定城市交通发展战略目标的基本指标,其在城市交通领域的重要性类似于“用地结构”在城市规划领域中的重要性。在“低碳绿色、集约高效”的发展导向下,以发展绿色交通(轨道交通、常规公交、慢行交通等)为目标、优化调整城市交通结构是北京、上海、广州、深圳等超大城市共同的战略选择,如北京提出力争中心城区绿色出行比例达到76.5%;上海将“中心城绿色交通(轨道交通、公共汽电车、轮渡、自行车和步行)出行比例不低于75%”作为未来10年城市交通结构转型的发展目标;成都提出“2025年,中心城区(12+2城区范围)绿色出行比例达到70%”。因此,为实现全国城市交通结构优化调整转型,准确掌握城市交通结构现状情况是国内城市普遍存在的现实需求。

[0003] 目前针对城市居民的出行方式的研究数据主要来自两个方面:居民入户抽样调查和多元数据融合分析,但是,这两个数据具有一定的缺陷。

[0004] 居民入户抽样调查是最直接的获取城市居民出行方式的方法,该方法以直接抽样询问居民的出行方式为主,一定时段内是最主要的获取方法,许多城市为此建立了“5年大规模调查3年小规模调查”的机制,从而得到城市居民出行结构的变化情况。该方法的好处是相对客观、直接,但受制于抽样率和覆盖面的限制,难以反应城市居民普遍情况,同时,调查周期较长、人力成本较高;

[0005] 多元数据融合分析是随着数据科学的兴起出现的获取城市居民出行方式的方法,该方法是可以获取大体量、广覆盖、低成本的交通大数据,这已逐渐成为获取居民出行方式的主要来源,浮动车、网约车、公交刷卡、轨道刷卡、共享单车等数据为研究者直接识别单一交通方式提供了极大的便利性,但是,在众多交通数据中,轨道刷卡、公交刷卡、共享单车、网约车运行数据仅能标定特定群体的出行方式,如轨道刷卡数据仅标定乘坐轨道出行的居民群体,无法分析得到小汽车、轨道、常规公交、慢行等方式出行的所有群体;

[0006] 因此,上述两种城市居民的出行方式的数据获取均不能很精确的体现一个地区的城市交通结构。

### 发明内容

[0007] 本发明提供一种基于XGBoost模型的城市交通结构分析方法及系统,能够精确的体现一个地区的城市交通结构。

[0008] 本发明通过下述技术方案实现:

[0009] 一种基于XGBoost模型的城市交通结构分析方法,包括如下步骤:

[0010] S1、通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;

[0011] S2、提取所述居民出行方式数据的特征指标;

[0012] S3、将所述特征指标划分为训练集和测试集,将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的城市交通结构分析模型;

[0013] S4、将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

[0014] 作为优化,所述居民出行方式数据包括居民入户抽样调查得到的居民出行调查数据、多元数据融合分析获取的轨道刷卡、公交刷卡、共享单车、网约出租运行数据。

[0015] 作为优化,通过居民入户抽样调查得到的居民出行调查数据占所述居民出行方式数据的3.5%-4.0%。

[0016] 作为优化,提取所述居民出行方式数据的特征指标的具体步骤为:

[0017] S2.1、将所述居民出行方式数据的经纬度数据在ArcGIS平台中进行空间化构建面板数据;

[0018] S2.2、以所述经纬度数据分别作500米缓冲区,将所述面板数据叠合城市用地、房屋租赁、道路拥堵、轨道站点、公交线网的数据,匹配得到每一个所述居民出行方式数据的特征指标。

[0019] 作为优化,所述特征指标包括出行特征、居民属性和交通状态,其中,

[0020] 所述出行特征包括居民的出行距离、出行时耗、出行时段、起点属性和终点属性;

[0021] 所述居民属性包括居民的收入水平、拥车情况和年龄结构;

[0022] 所述交通状态包括起点附近公交设施、轨道设施、共享单车供给情况和出行时段的道路拥堵情况。

[0023] 作为优化,S3的训练过程为:

[0024] S3.1、将所述训练集的特征指标按照标准格式构建为面板数据,所述面板数据的字段内容和顺序依次为:1.出行方式;2.出行距离;3.出行时段;4.出发区域;5.到达区域;6.收入条件;7.停车条件;8.年龄情况;9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给;

[0025] S3.2、运用python语言,调用XGBoost模型的python语言模块,以所述训练集的特征指标为基础进行XGBoost模型进行训练,具体包括将“1.出行方式”字段作为因变量值,将“2.出行距离;3.出行时段;4.出发区域;5.到达区域;6.收入条件;7.停车条件;8.年龄情况;9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给”字段作为自变量值,设置所述XGBoost模型特征参数全为初始值,进行代码运行即开始XGBoost模型训练;

[0026] S3.3、等待代码运行结束即完成所述XGBoost模型训练,训练完成的所述XGBoost模型即为城市交通结构分析模型。

[0027] 作为优化,所述训练集和测试集的比例为4:1。

[0028] 作为优化,在将所述手机信令数据输入至所述城市交通结构分析模型中之前,先对手机信令数据在ArcGIS平台中进行“空间化”称为点状矢量数据。

[0029] 作为优化,S4的具体步骤为:

[0030] S4.1、在ArcGIS平台中以手机信令的点状矢量数据为基础,运用“缓冲区”工具,以所述手机信令的点状矢量数据为中心构建半径为500米的圆形缓冲区;

[0031] S4.2、利用“空间连接”工具,统计以所述手机信令的点状矢量数据为中心的空间位置在圆形缓冲区中的特征指标;

[0032] S4.3、将关于所述手机信令的特征指标输入至所述城市交通结构分析模型,得到手机信令所在位置的交通结构分析结果;

[0033] S4.4、循环S4.1-S4.3,直到将需要调查的地区的所有居民的手机信令数据全部输入至所述城市交通结构分析模型中,最终得到该地区的城市交通结构分析结果。

[0034] 本发明还公开了一种基于XGBoost模型的城市交通结构分析系统,包括:

[0035] 采集模块,用于通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;

[0036] 提取模块,用于提取所述居民出行方式数据的特征指标,将所述特征指标划分为训练集和测试集;

[0037] 训练模块,用于将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的的城市交通结构分析模型;

[0038] 输入模块,用于将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

[0039] 本发明与现有技术相比,具有如下的优点和有益效果:

[0040] 本发明利用XGBoost模型,并构建对应的特征指标库,来进行城市交通结构的识别,较现有技术具有(1)处理分析规模和空间尺度更大、更广;(2)处理分析速度和精度更快、更优;(3)处理涵盖方式更全,包含小汽车、轨道交通、常规公交和慢行交通,该发明充分满足城乡规划领域“量大、高效、精准”的分析需求,可持续应对相关领域的技术分析需求。

## 附图说明

[0041] 为了更清楚地说明本发明示例性实施方式的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本发明的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。在附图中:

[0042] 图1为特征指标的匹配过程图。

## 具体实施方式

[0043] 为使本发明的目的、技术方案和优点更加清楚明白,下面结合实施例和附图,对本发明作进一步的详细说明,本发明的示意性实施方式及其说明仅用于解释本发明,并不作为对本发明的限定。

[0044] 实施例1的一种基于XGBoost模型的城市交通结构分析方法,如图1所示,包括如下步骤:

[0045] S1、通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据。具体的,居民出行方式数据包括居民入户抽样调查得到的居民出行调查数据(小汽车出行群体、常规公交出行群体、轨道交通出行群体和慢行出行群体)、多元数

据融合分析获取的轨道刷卡、公交刷卡、共享单车、网约出租运行数据。

[0046] 因为训练XGBoost模型需要部分已经明确了出行方式的样本数据。目前能掌握到的能够100%确定出行方式数据有(1) 50余万份的居民入户抽样调查;(2) 千万级体量的轨道刷卡、公交刷卡、共享单车和网约出租车(可代表小汽车)数据。其中,居民入户抽样调查数据每条数据代表一种出行方式,也就是说居民入户抽样调查数据可以确切地划分出所有出行方式。但轨道刷卡、公交刷卡、共享单车和网约出租车(可代表小汽车)数据只能代表各自数据对应的交通方式,例如轨道刷卡数据只能代表选择轨道出行乘客,为了使得XGBoost模型训练结果更加贴近现实,本发明将50余万份的居民入户抽样调查和千万级体量的轨道刷卡、公交刷卡、共享单车和网约出租车(可代表小汽车)数据共同作为训练数据,通过居民入户抽样调查得到的居民出行调查数据大致占所述居民出行方式数据的3.5%-4.0%,本发明的比例为居民入户抽样调查仅占总数据体量(居民出行调查数据+多元数据融合分析数据)的3.8%。

[0047] S2、提取所述居民出行方式数据的特征指标,具体步骤为:

[0048] S2.1、将所述居民出行方式数据的经纬度数据在ArcGIS平台中进行空间化构建面板数据;

[0049] 经纬度坐标是表征空间位置的数值型数据,上述数据(居民出行方式数据相关的城市用地、房屋租赁、道路拥堵、轨道站点、公交线网的数据)都具有且形式一样。在城市规划领域,数据的“空间化”是一切分析的基础,而数据的经纬度坐标是其“空间化”的重要依据。

[0050] 数据根据经纬度进行“空间化”已经是行业共同认定的基本操作,基于常见的软件平台就可实现。本发明采用的是市场占有率最高的ArcGIS平台,空间化的目的是为了构建面板数据。初始状态下,样本数据(包括居民入户抽样调查和多元数据)只有“编号”、“出行经度”、“出行纬度”、“出行方式”、“出行时间”、“达到时间”、“达到经度”、“达到纬度”,仅能测算得到1.出行方式;2.出行距离;3.出行时段;对于4.出发区域(起点属性);5.到达区域(终点属性);6.收入条件;7.停车条件(拥车情况);8.年龄情况(年龄结构);9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给则需要通过“空间化”进行拾取得到,最终形成面板数据。

[0051] S2.2、以所述经纬度数据分别作500米缓冲区,将所述面板数据叠合城市用地、房屋租赁、道路拥堵、轨道站点、公交线网的数据,匹配得到每一个所述居民出行方式数据的特征指标。

[0052] 如果只是用“点状数据”进行拾取,会造成拾取成功率极低,因为,在ArcGIS平台中,数据之间的相互拾取的前提是数据之间要相互重合,如果使用“点状数据”进行相互叠合,非常难以实现“叠合”,通过使用“缓冲区”将“点”变“面”提升数据间相互叠合的几率。

[0053] 所述特征指标包括出行特征、居民属性和交通状态,其中,

[0054] 所述出行特征包括居民的出行距离、出行时耗、出行时段、起点属性和终点属性;

[0055] 所述居民属性包括居民的收入水平、拥车情况和年龄结构;

[0056] 所述交通状态包括起点附近公交设施、轨道设施、共享单车供给情况和出行时段的道路拥堵情况。

[0057] S3、将所述特征指标划分为训练集和测试集,将训练集输入至XGBoost模型中对所

述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的城市交通结构分析模型;

[0058] 训练过程可以为:

[0059] (1)将所有已知出行方式的特征指标按照标准格式构建为“面板数据”,如表1所示,“面板数据”字段内容和顺序依次为:1.出行方式;2.出行距离;3.出行时段;4.出发区域;5.到达区域;6.收入条件;7.停车条件;8.年龄情况;9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给;

[0060] (2)将“面板数据”按照行业共识比例划分为“训练集”和“测试集”,其中“训练集”占比80%、“测试集”占比20%;

[0061] (3)运用python语言,调用XGBoost模型的python语言模块,以“训练集”数据为基础进行XGBoost模型,具体包括将“1.出行方式”字段作为因变量值,将“2.出行距离;3.出行时段;4.出发区域;5.到达区域;6.收入条件;7.停车条件;8.年龄情况;9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给。”字段作为自变量值,设置XGBoost模型特征参数全为初始值,进行代码运行即开始XGBoost模型训练;

[0062] (4)等待运行结束即完成XGBoost模型训练,得到合格的城市交通结构分析模型。

[0063] 表1面板数据集的样例数据

数据编号	出行方式*	出行距离	出行时段	起点属性**	终点属性**	收入水平	拥车情况	年龄结构***	拥堵状况	轨道设施	公交设施	单车供给
[0064] 1	1	14.2	8:45	1	3	2.5	1140	2	2.81	0	0.95	11
2	1	13.8	11:32	3	4	1.6	568	2	1.32	0	1.12	23
3	2	10.6	12:45	3	5	1.3	1240	2	1.26	2	0.96	15
4	4	2.1	17:55	3	1	2.8	816	3	1.89	1	1.36	18
5	3	4.6	16:41	3	1	3.0	1510	2	1.67	2	1.41	22

[0065] \*:1=“小汽车”;2=“轨道交通”;3=“常规公交”;4=“慢行交通”。\*\*:1=“住区”;2=“学校”;3=“商务区”;4=“商圈”;5=“公园”。\*\*\*:1=“12岁及以下”;2=“12-26岁”;3=“26岁-55岁”;4=“55-65岁”;5=“65岁及以上”。

[0066] S4、将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

[0067] 手机信令数据含有“经度”和“纬度”信息,可以据此在ArcGIS平台中进行“空间化”,“空间化”后为点状矢量数据。同理,轨道站点数据、公交设施数据、轨道设施数据等均可转为点状矢量数据,然后,在ArcGIS平台中以手机信令的点状数据为基础,运用“缓冲区”工具,以其为中心构建半径为500米的圆形缓冲区,再利用“空间连接”工具,统计空间位置在圆形缓冲区中的轨道站点总数、公交设施总数、共享单车车辆总数、道路平均拥堵值等。通过上述操作,手机信令数据也可以具备2.出行距离;3.出行时段;4.出发区域;5.到达区域;6.收入条件;7.停车条件;8.年龄情况;9.拥堵情况;10.轨道设施;11.公交设施;12.共享单车供给的全部字段内容(即可以通过手机信令数据得到对应的特征指标),最后统计以所述手机信令的点状矢量数据为中心的空间位置在圆形缓冲区中的特征指标,最后将手机信令的特征指标输入至城市交通结构分析模型中,最终得到该地区的城市交通结构分析结果。



[0068] 实施例2的一种基于XGBoost模型的城市交通结构分析系统,包括:

[0069] 采集模块,用于通过居民入户抽样调查和多元数据融合分析对居民出行方式进行采集,获取居民出行方式数据;

[0070] 提取模块,用于提取所述居民出行方式数据的特征指标,将所述特征指标划分为训练集和测试集;

[0071] 训练模块,用于将训练集输入至XGBoost模型中对所述XGBoost模型进行训练,并对训练后的XGBoost模型进行测试,得到合格的城市交通结构分析模型;

[0072] 输入模块,用于将需要调查的地区的所有居民的手机信令数据输入至所述城市交通结构分析模型中,得到该地区的城市交通结构分析结果。

[0073] 以上所述的具体实施方式,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

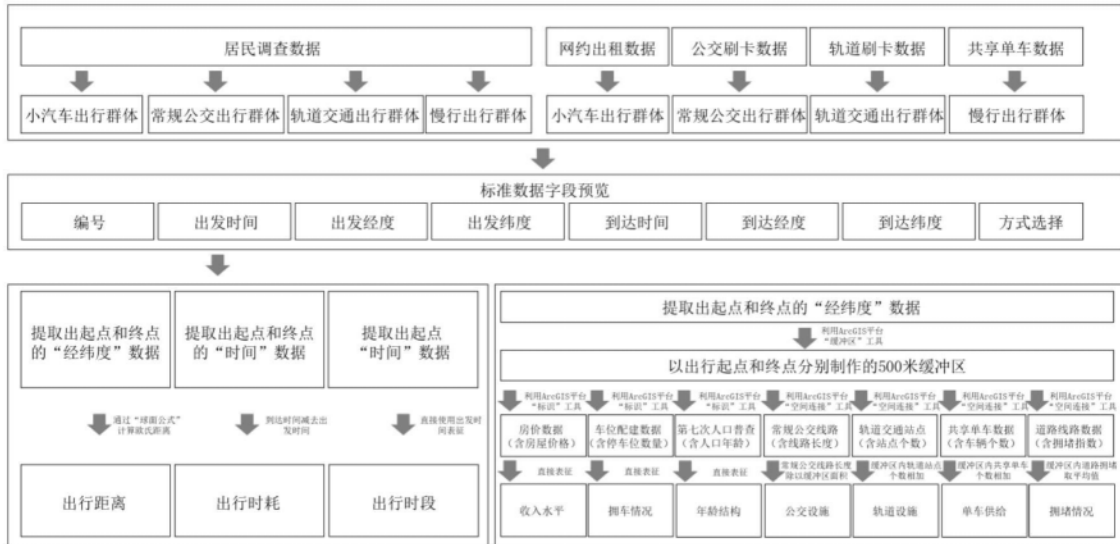


图1