



(12)发明专利申请

(10)申请公布号 CN 109871874 A

(43)申请公布日 2019.06.11

(21)申请号 201910051795.1

(22)申请日 2019.01.21

(71)申请人 华南理工大学

地址 511458 广东省广州市南沙区环市大道南路25号华工大广州产研院

(72)发明人 黄紫林 许伦辉

(74)专利代理机构 广州市华学知识产权代理有限公司 44245

代理人 李君

(51)Int.Cl.

G06K 9/62(2006.01)

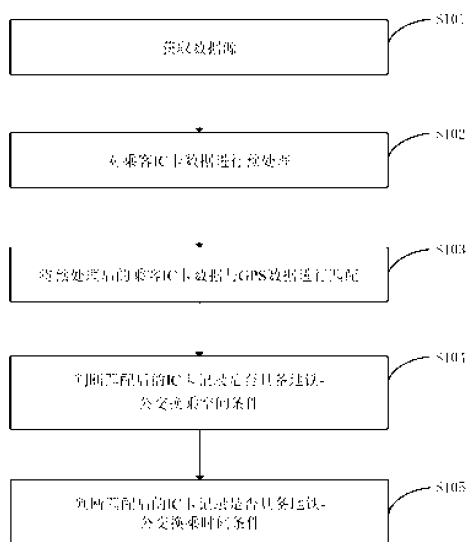
权利要求书2页 说明书9页 附图7页

(54)发明名称

地铁换乘公交行为识别方法、系统、计算机设备及存储介质

(57)摘要

本发明公开了一种地铁换乘公交行为识别方法、系统、计算机设备及存储介质,所述方法包括:获取数据源,所述数据源包括乘客IC卡数据和GPS数据;对乘客IC卡数据进行预处理;将预处理后的乘客IC卡数据与GPS数据进行匹配;判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件;当匹配后的IC卡记录具备地铁-公交换乘空间条件时,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。本发明通过结合乘客IC卡数据和GPS数据,给出乘客在地铁与公交两种交通方式之间换乘所应具备的空间和时间条件,实现了地铁换乘公交行为识别,识别精确度较高。



1. 一种地铁换乘公交行为识别方法,其特征在於,所述方法包括:
获取数据源;其中,所述数据源包括乘客IC卡数据和GPS数据;
对乘客IC卡数据进行预处理;
将预处理后的乘客IC卡数据与GPS数据进行匹配;
判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件;
当匹配后的IC卡记录具备地铁-公交换乘空间条件时,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

2. 根据权利要求1所述的地铁换乘公交行为识别方法,其特征在於,所述对乘客IC卡数据进行预处理,具体包括:

从乘客IC卡数据中选择一个地铁站点作为研究对象,筛选该地铁站点的乘客地铁IC卡数据;

对筛选后的乘客地铁IC卡数据中时间重复、有误的数据进行过滤;

以IC卡号作为标识匹配乘客地铁IC卡数据与乘客公交IC卡数据;

将匹配后的乘客IC卡数据进行清洗,确保每张IC卡对应研究时段内的一个出站刷卡时间。

3. 根据权利要求1所述的地铁换乘公交行为识别方法,其特征在於,所述将预处理后的乘客IC卡数据与GPS数据进行匹配,具体包括:

在地铁GPS数据中,以地铁站点作为标识匹配乘客地铁IC卡数据,得到乘客出站刷卡时的经纬度信息;

在公交GPS数据中,以车载机号作为标识匹配乘客公交IC卡数据,得到乘客上车刷卡时的经纬度信息。

4. 根据权利要求1所述的地铁换乘公交行为识别方法,其特征在於,所述判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件,具体包括:

将公交换乘站点与地铁换乘站点的直线距离不大于换乘空间阈值作为公交换乘空间条件;

根据公交换乘空间条件以及地铁换乘站点周围的公交换乘站点地理信息,得到地铁-公交换乘空间条件;

判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件。

5. 根据权利要求1所述的地铁换乘公交行为识别方法,其特征在於,所述判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件,具体包括:

根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,作为地铁-公交换乘时间条件;其中,所述换乘时间阈值取95%置信度水平;

判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

6. 根据权利要求5所述的地铁换乘公交行为识别方法,其特征在於,所述根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,如下式:

$$0 \leq T_{si} - T_{sj} \leq T$$

其中, T_{si} 表示公交的到站时间, T_{sj} 表示地铁到达换乘站点的时间, T 表示换乘时间阈值。

7. 根据权利要求1-6任一项所述的地铁换乘公交行为识别方法,其特征在於,所述乘客IC卡数据包括乘客地铁IC卡数据和乘客公交IC卡数据,以IC卡号作为乘客的唯一标识,匹

配乘客的地铁刷卡行为和公交刷卡行为；

所述GPS数据包括公交GPS数据和地铁GPS数据，地铁GPS数据中包括地铁线路、地铁站点和经纬度，公交GPS数据包括时间、经纬度、车载机号和车牌号。

8. 一种地铁换乘公交行为识别系统，其特征在于，所述系统包括：

获取模块，用于获取数据源；其中，所述数据源包括乘客IC卡数据和GPS数据；

预处理模块，用于对乘客IC卡数据进行预处理；

匹配模块，用于将预处理后的乘客IC卡数据与GPS数据进行匹配；

第一判断模块，用于判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件；

第二判断模块，用于当匹配后的IC卡记录具备地铁-公交换乘空间条件时，判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

9. 一种计算机设备，包括处理器以及用于存储处理器可执行程序存储器，其特征在于，所述处理器执行存储器存储的程序时，实现权利要求1-7任一项所述的地铁换乘公交行为识别方法。

10. 一种存储介质，存储有程序，其特征在于，所述程序被处理器执行时，实现权利要求1-7任一项所述的地铁换乘公交行为识别方法。

地铁换乘公交行为识别方法、系统、计算机设备及存储介质

技术领域

[0001] 本发明涉及一种地铁换乘公交行为识别方法、系统、计算机设备及存储介质,属于时空数据挖掘及交通技术领域。

背景技术

[0002] 当前,市民出行选择的公共交通工具主要包括公交车、地铁以及共享自行车等。公交车和地铁是市民优先选择的经济便捷的出行方式,这类交通工具的优点是快捷、覆盖面广且省时省力。因此,如何优化市民在这两种交通方式之间的换乘效率,提高居民公共出行效率,动态发现新的公共出行需求对交通服务者及城市规划都是非常重要的。

[0003] 近几年来,交通大数据技术日新月异,为交通研究者提供了大量的数据基础。其中,有效的识别换乘是进行公共交通合理规划的关键。传统的识别方法主要根据单一的换乘时间阈值或者换乘空间阈值对换乘行为进行识别,且大部分研究主要集中在公交系统内部的换乘行为。而地铁-公交系统具有其独特性,不适用于传统方法。

发明内容

[0004] 有鉴于此,本发明提供了一种地铁换乘公交行为识别方法、系统、计算机设备及存储介质,通过结合乘客IC卡数据和GPS数据,给出乘客在地铁与公交两种交通方式之间换乘所应具备的空间和时间条件,实现了地铁换乘公交行为识别,识别精确度较高。

[0005] 本发明的第一个目的在于提供一种地铁换乘公交行为识别方法。

[0006] 本发明的第二个目的在于提供一种地铁换乘公交行为识别系统。

[0007] 本发明的第三个目的在于提供一种计算机设备。

[0008] 本发明的第四个目的在于提供一种存储介质。

[0009] 本发明的第一个目的可以通过采取如下技术方案达到:

[0010] 一种地铁换乘公交行为识别方法,所述方法包括:

[0011] 获取数据源;其中,所述数据源包括乘客IC卡数据和GPS数据;

[0012] 对乘客IC卡数据进行预处理;

[0013] 将预处理后的乘客IC卡数据与GPS数据进行匹配;

[0014] 判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件;

[0015] 当匹配后的IC卡记录具备地铁-公交换乘空间条件时,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0016] 进一步的,所述对乘客IC卡数据进行预处理,具体包括:

[0017] 从乘客IC卡数据中选择一个地铁站点作为研究对象,筛选该地铁站点的乘客地铁IC卡数据;

[0018] 对筛选后的乘客地铁IC卡数据中时间重复、有误的数据进行过滤;

[0019] 以IC卡号作为标识匹配乘客地铁IC卡数据与乘客公交IC卡数据;

[0020] 将匹配后的乘客IC卡数据进行清洗,确保每张IC卡对应研究时段内的一个出站刷

卡时间。

[0021] 进一步的,所述将预处理后的乘客IC卡数据与GPS数据进行匹配,具体包括:

[0022] 在地铁GPS数据中,以地铁站点作为标识匹配乘客地铁IC卡数据,得到乘客出站刷卡时的经纬度信息;

[0023] 在公交GPS数据中,以车载机号作为标识匹配乘客公交IC卡数据,得到乘客上车刷卡时的经纬度信息。

[0024] 进一步的,所述判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件,具体包括:

[0025] 将公交换乘站点与地铁换乘站点的直线距离不大于换乘空间阈值作为公交换乘空间条件;

[0026] 根据公交换乘空间条件以及地铁换乘站点周围的公交换乘站点地理信息,得到地铁-公交换乘空间条件;

[0027] 判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件。

[0028] 进一步的,所述判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件,具体包括:

[0029] 根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,作为地铁-公交换乘时间条件;其中,所述换乘时间阈值取95%置信度水平;

[0030] 判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0031] 进一步的,所述根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,如下式:

$$[0032] \quad 0 \leq T_{si} - T_{sj} \leq T$$

[0033] 其中, T_{si} 表示公交的到站时间, T_{sj} 表示地铁到达换乘站点的时间, T 表示换乘时间阈值。

[0034] 进一步的,所述乘客IC卡数据包括乘客地铁IC卡数据和乘客公交IC卡数据,以IC卡号作为乘客的唯一标识,匹配乘客的地铁刷卡行为和公交刷卡行为;

[0035] 所述GPS数据包括公交GPS数据和地铁GPS数据,地铁GPS数据中包括地铁线路、地铁站点和经纬度,公交GPS数据包括时间、经纬度、车载机号和车牌号。

[0036] 本发明的第二个目的可以通过采取如下技术方案达到:

[0037] 一种地铁换乘公交行为识别系统,所述系统包括:

[0038] 获取模块,用于获取数据源;其中,所述数据源包括乘客IC卡数据和GPS数据;

[0039] 预处理模块,用于对乘客IC卡数据进行预处理;

[0040] 匹配模块,用于将预处理后的乘客IC卡数据与GPS数据进行匹配;

[0041] 第一判断模块,用于判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件;

[0042] 第二判断模块,用于当匹配后的IC卡记录具备地铁-公交换乘空间条件时,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0043] 本发明的第三个目的可以通过采取如下技术方案达到:

[0044] 一种计算机设备,包括处理器以及用于存储处理器可执行程序存储器,所述处理器执行存储器存储的程序时,实现上述的地铁换乘公交行为识别方法。

[0045] 本发明的第四个目的可以通过采取如下技术方案达到:

[0046] 一种存储介质,存储有程序,所述程序被处理器执行时,实现上述的地铁换乘公交行为识别方法。

[0047] 本发明相对于现有技术具有如下的有益效果:

[0048] 1、本发明先获取乘客IC卡数据和GPS数据,对乘客IC卡数据进行预处理,并将预处理后的乘客IC卡数据与GPS数据进行匹配,针对匹配后的IC卡记录,给出乘客在地铁与公交两种交通方式之间换乘所应具备的空间和时间条件,实现了地铁换乘公交行为识别,识别精确度较高。

[0049] 2、本发明改进目前普遍采用的统一阈值识别方法,将公交换乘站点与地铁换乘站点的直线距离不大于换乘空间阈值作为公交换乘空间条件,并结合地铁站点周围的公交站点地理信息,得到地铁-公交换乘空间条件,此外根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,作为地铁-公交换乘时间条件。

附图说明

[0050] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,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示出的结构获得其他的附图。

[0051] 图1为本发明实施例1的地铁换乘公交行为识别方法的流程图。

[0052] 图2为本发明实施例1的乘客IC卡数据预处理的流程图。

[0053] 图3为本发明实施例1的地铁-公交换乘空间条件判断的流程图。

[0054] 图4为本发明实施例1的地铁-公交换乘时间条件判断的流程图。

[0055] 图5为本发明实施例2的地铁换乘公交行为识别系统的结构框图。

[0056] 图6为本发明实施例2的预处理模块的结构框图。

[0057] 图7为本发明实施例2的第一判断模块的结构框图。

[0058] 图8为本发明实施例2的第二判断模块的结构框图。

[0059] 图9为本发明实施例3的计算机设备的结构框图。

具体实施方式

[0060] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明的一部分实施例,而不是全部的实施例,基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0061] 实施例1:

[0062] 如图1所示,本实施例提供了一种地铁换乘公交行为识别方法,该方法包括以下步骤:

[0063] S101、获取数据源。

[0064] 本实施例的数据源包括乘客IC卡数据和GPS数据,该数据源可以从数据库查找获取,例如预先在数据库内存储乘客IC卡数据和GPS数据,从数据库中搜索乘客IC卡数据和GPS数据即可得到。其中,所述乘客IC卡数据包括乘客地铁IC卡数据和乘客公交IC卡数据,

以IC卡号作为乘客的唯一标识,匹配乘客的地铁刷卡行为和公交刷卡行为;所述GPS数据包括公交GPS数据和地铁GPS数据,地铁GPS数据中包括地铁线路、地铁站点、经纬度等信息,公交GPS数据包括时间、经纬度、车载机号、车牌号等信息。

[0065] 本实施例选取深圳市地铁三号线(龙岗线)作为研究对象,它连接了福田、罗湖和龙岗三个行政区,由双龙站至益田站,总长41.7公里,共有30个地铁站点,在进行地铁-公交换乘行为识别时,首先需要定义查询规则,以便正确地将一个地铁换乘与其随后的公交换乘有着相同的IC卡编号。

[0066] 本实施例的数据源主要由四张表组成,分别是:深圳市地铁三号线地铁GPS数据表、公交GPS数据表、2018年5月23日乘客地铁IC卡数据表和乘客公交IC卡数据表。

[0067] 乘客地铁IC卡数据表主要包含五个字段,即:CardID、StationID、TrnsctTime、TrnsctyType、Statiom,其中CardID是智能交通卡的唯一标识;StationID是地铁站的唯一标识;TrnsctTime是交易(刷卡)时间,TrnsctTime有两种类型,进站和出站,分别用21,22标识,Statiom是交易(刷卡)站点,详情如下表1所示。

[0068] 表1乘客地铁IC卡数据表

[0069]

卡编号	交易类型	交易时间	地铁线路	站点	车载机号
361955655	22	2016/5/23 0:01	地铁三号线	爱联	AGM-105
660896555	22	2016/5/23 0:01	地铁三号线	爱联	AGM-113
667356942	22	2016/5/23 0:01	地铁三号线	双龙	AGM-122
.....
684144991	22	2016/5/23 0:01	地铁三号线	爱联	AGM-119

[0070] 乘客公交IC卡数据表记录着每一位乘客乘坐公交的出行信息,其中卡编号作为是被公交卡的唯一识别码,每一张IC卡对应一位持卡乘客,CardID代表卡编号,TradeTime 代表交易时间,LineID代表线路号,BusID代表车牌号,详情如下表2所示。

[0071] 表2乘客公交IC卡数据表

卡编号	交易类型	交易时间	公司名称	线路号	车牌号
326091092	31	2016/5/23 1:05	巴士集团	216 路	DN2471
20733829	31	2016/5/23 0:00	巴士集团	63 路	DCM425
80464253	31	2016/5/23 0:09	巴士集团	10 路	DCJ411
.....
296270453	31	2016/5/23 0:22	巴士集团	10 路	DCJ411

[0073] 通过查询地铁站点地理信息系统,可以得知深圳市地铁系统每个站点的经纬度地理信息与站点编号,通过匹配地铁站点地理信息,可以获得乘客出站地理位置,地铁GPS数据表详情如下表3所示。

[0074] 表3地铁GPS数据表

[0075]

站点	X 轴坐标	Y 轴坐标	线路	编号
益田	114.0579	22.52203	三号线	301
石厦	114.0602	22.52945	三号线	302
购物公园	114.0612	22.54034	三号线	303
.....
华新	114.0933	22.55487	三号线	307

[0076] 公交GPS数据表记录了车辆实时的运行信息,主要记录着线路号、车载机号、站点编号、业务时间、经纬度、线路编号等信息,公交GPS数据表详情如下表4所示。

[0077] 表4公交GPS数据表

[0078]

线路号	车载机号	站点编号	业务时间	经度	纬度	GPS 速度	车牌号	线路编号
817	31299	2	2016-05-23 00:00:17.000	114.0881	22.54337	7.83	粤 BC0965	00N41
62	29120	1	2016-05-23 00:00:57.000	114.104	22.54412	8.61	粤 B83470	30
.....
101	1171	2	2016-05-23 00:01:40.000	113.7861	22.73817	0	粤 B87332	3380

[0079] S102、对乘客IC卡数据进行预处理。

[0080] 该步骤S102如图2所示,具体包括:

[0081] S1021、地铁站点M选择:从乘客IC卡数据中选择一个地铁站点M作为研究对象,筛选该地铁站点M的乘客地铁IC卡数据。

[0082] S1022、第一次数据清洗:对筛选后的乘客地铁IC卡数据中时间重复、有误的数据进行过滤。

[0083] S1023、IC卡数据匹配:以IC卡号作为标识匹配乘客地铁IC卡数据与乘客公交 IC卡数据。

[0084] S1024、第二次数据清洗:将匹配后的乘客IC卡数据进行清洗,确保每张IC卡对应研究时段内的一个出站刷卡时间。

[0085] S103、将预处理后的乘客IC卡数据与GPS数据进行匹配。

[0086] 具体地,在地铁GPS数据中,以地铁站点作为标识匹配乘客地铁IC卡数据,得到乘客出站刷卡时的经纬度信息;在公交GPS数据中,以车载机号作为标识匹配乘客公交IC卡数据,得到乘客上车刷卡时的经纬度信息。

[0087] 在本实施例中,当IC卡记录中含车载机编号时,可以通过IC卡数据与GPS数据进行匹配,得到公交线路GPS数据,识别规则为车载机号后5位与公交车辆车牌号匹配(巴士集团也存在部分公交匹配不上GPS信息,应删除该部分数据;部分公交车载号是A结尾,识别算法应为删除字符“A”后5个字符)。以地铁站点的GPS数据直线1Km作为换乘空间阈值,以30min作为换乘时间阈值,匹配结果如表5所示。

[0088] 表5 IC卡数据与GPS数据的匹配表

[0089]

IC卡编号	离开地铁时间	站点	经度	纬度	上车时间	运营公司	公交线路	车牌号	时间间隔
293189842	2016-05-23 21:03:09.000	双龙	114.2841	22.73479	2016-05-23 21:05:58.000	巴士集团	351路	粤BCJ350	169
294087379	2016-05-23 11:48:47.000	双龙	114.2841	22.73479	2016-05-23 11:58:01.000	巴士集团	351路	粤BCJ350	554
329234753	2016-05-23 16:22:36.000	双龙	114.2841	22.73479	2016-05-23 16:32:37.000	巴士集团	351路	粤BCJ350	601
.....
329438378	2016-05-23 20:13:14.000	双龙	114.2841	22.73479	2016-05-23 20:15:59.000	巴士集团	351路	粤 BCK910	165

[0090] S104、判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件。

[0091] 该步骤S104如图3所示,具体包括:

[0092] 结合深圳市道路交通规划设计规范中有关路口设置公交站点的要求,初步设定公交换乘站点与地铁换乘站点至少具备直线距离不大于换乘空间阈值(即1Km),作为公交换乘空间条件,结合地铁换乘站点周围的公交换乘站点地理信息,得到地铁-公交换乘空间条件,再判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件,若是,则进入步骤S105。

[0093] 公交换乘站点与地铁换乘站点至少具备直线距离不大于换乘空间阈值,如下式:

[0094]
$$d_{R,B} \leq D$$

[0095] 其中,D表示换乘空间阈值; $d_{R,B}$ 表示地铁站点与公交站点的直线距离,其可通过站点地理信息数据表中的站点经纬度计算得到。

[0096] 地铁站点与公交站点的直线距离的计算如下式:

$$d_{R,B} = R \cdot \arccos(\sin(s.lat) \cdot \sin(b.lat) + \cos(s.lat) \cdot \cos(b.lat) \cdot \cos(s.lon - b.lon)) * \pi / 180$$

[0098] 其中,R表示地球半径,一般取6378.137km;s.lat表示地铁换乘站点的纬度;s.lon表示地铁换乘站点的经度;b.lat表示公交换乘站点的纬度;b.lon表示为公交换乘站点的经度。

[0099] S105、判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0100] 根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,作为地铁-公交换乘时间条件,换乘时间阈值取95%置信度水平,即30min;判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0101] 该步骤S105如图4所示,具体包括:

[0102] 根据居民的出行特点,当明确目的地后,在选择出行路径时,乘客一般选择直达、换乘步行时间最少的站点进行换乘,因此地铁-公交的刷卡时间差不应太大。设定乘客换乘时间间隔不应大于换乘时间阈值,得到换乘时间条件如下式:

[0103]
$$0 \leq T_{si} - T_{sj} \leq T$$

[0104] 其中, T_{si} 表示公交的到站时间; T_{sj} 表示地铁达到换乘站点的时间;T表示换乘时间阈值。

[0105] 由上式,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0106] 至此,完成了地铁换乘公交行为识别方法的构建及实现,识别结果如下表6所示。

[0107] 表6地铁换乘公交行为识别结果表

[0108]

站点编号	换乘人次	出行人次	换乘系数	站点编号	换乘人次	出行人次	换乘系数
301	3835	19780	1.194	316	1553	19585	1.079
302	570	16582	1.034	317	4990	17242	1.289
303	83	6040	1.014	318	3618	21707	1.167
304	654	19313	1.034	319	1734	23586	1.074
305	468	3821	1.122	320	1000	7991	1.125
306	400	6607	1.061	321	1064	10956	1.097
307	315	17046	1.018	322	2003	16185	1.124
308	250	10043	1.025	323	1467	4932	1.297
309	194	10157	1.019	324	731	6361	1.115
310	551	42617	1.013	325	3073	8395	1.366
311	571	13980	1.041	326	989	8909	1.111
312	1575	17107	1.092	327	612	7712	1.079
313	529	12175	1.043	328	1492	10002	1.149
314	172	6998	1.025	329	984	9796	1.1
315	2123	17843	1.119	330	5445	15053	1.362

[0109] 从上述统计发现,研究时段内地铁三号线的出行总量为408521人次,其中换乘公交人数为43045人次,换乘公交人数占总出行总量的10.54%。由匹配结果可知,本实施例的方法能够准确高效地将乘客公交IC卡数据与乘客地铁IC卡数据进行匹配,方法适于推广。

[0110] 本领域技术人员可以理解,实现上述实施例的方法中的全部或部分步骤可以通过程序来指令相关的硬件来完成,相应的程序可以存储于计算机可读存储介质中。

[0111] 应当注意,尽管在附图中以特定顺序描述了上述实施例的方法操作,但是这并非要求或者暗示必须按照该特定顺序来执行这些操作,或是必须执行全部所示的操作才能实现期望的结果。相反,描绘的步骤可以改变执行顺序。附加地或备选地,可以省略某些步骤,将多个步骤合并为一个步骤执行,和/或将一个步骤分解为多个步骤执行。

[0112] 实施例2:

[0113] 如图5所示,本实施例提供了一种地铁换乘公交行为识别系统,该系统包括获取模块501、预处理模块502、匹配模块503、第一判断模块504和第二判断模块505,各个模块的具体功能如下:

[0114] 所述获取模块501,用于获取数据源;其中,所述数据源包括乘客IC卡数据和GPS数据。

[0115] 所述预处理模块502,用于对乘客IC卡数据进行预处理;该预处理模块502如图6所示,具体包括:

[0116] 地铁站点选择单元5021,用于从乘客IC卡数据中选择一个地铁站点M作为研究对象,筛选该地铁站点M的乘客地铁IC卡数据。

[0117] 第一数据清洗单元5022,用于对筛选后的乘客地铁IC卡数据中时间重复、有误的数据进行过滤。

[0118] IC卡数据匹配单元5023,用于以IC卡号作为标识匹配乘客地铁IC卡数据与乘客公交IC卡数据。

[0119] 第二数据清洗单元5024,用于将匹配后的乘客IC卡数据进行清洗,确保每张IC卡对应研究时段内的一个出站刷卡时间。

[0120] 所述匹配模块503,用于将预处理后的乘客IC卡数据与GPS数据进行匹配。

[0121] 所述第一判断模块504,用于判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件;该第一判断模块如图7所示,具体包括:

[0122] 公交换乘空间条件获取单元5041,用于将公交换乘站点与地铁换乘站点的直线距离不大于换乘空间阈值作为公交换乘空间条件。

[0123] 地铁-公交换乘空间条件获取单元5042,用于根据公交换乘空间条件以及地铁换乘站点周围的公交换乘站点地理信息,得到地铁-公交换乘空间条件。

[0124] 第一判断单元5043,用于判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件。

[0125] 所述第二判断模块505,用于当匹配后的IC卡记录具备地铁-公交换乘空间条件时,判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件;该第二判断模块505如图8所示,具体包括:

[0126] 地铁-公交换乘时间条件获取单元5051,用于根据不同地铁站点与公交线路之间的换乘,基于实际数据采用不同的换乘时间阈值,作为地铁-公交换乘时间条件;其中,所述换乘时间阈值取95%置信度水平。

[0127] 第二判断单元5052,用于判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。

[0128] 本实施例中各个模块的具体实现可以参见上述实施例1,在此不再一一赘述;需要说明的是,本实施例提供的系统仅以上述各功能模块的划分进行举例说明,在实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。

[0129] 可以理解,上述实施例的系统所使用的术语“第一”、“第二”等可用于描述各种模块,但这些模块不受这些术语限制。这些术语仅用于将第一个模块与另一个模块区分。举例来说,在不脱离本发明的范围的情况下,可以将第一判断模块称为第二判断模块,且类似地,可将第二判断模块称为第一判断模块,第一判断模块和第二判断模块两者都是判断模块,但不是同一判断模块。

[0130] 实施例3:

[0131] 本实施例提供了一种计算机设备,该计算机设备可以是计算机,如图9所示,其包括通过系统总线901连接的处理器902、存储器、输入装置903、显示器904和网络接口905,该处理器用于提供计算和控制能力,该存储器包括非易失性存储介质906 和内存存储器907,该非易失性存储介质906存储有操作系统、计算机程序和数据库,该内存存储器907为非易失性存储介质中的操作系统和计算机程序的运行提供环境,处理器902执行存储器存储的计算机程序时,实现上述实施例1的地铁换乘公交行为识别方法,如下:

[0132] 获取数据源;其中,所述数据源包括乘客IC卡数据和GPS数据;

[0133] 对乘客IC卡数据进行预处理;

- [0134] 将预处理后的乘客IC卡数据与GPS数据进行匹配；
- [0135] 判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件；
- [0136] 当匹配后的IC卡记录具备地铁-公交换乘空间条件时，判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。
- [0137] 实施例4：
- [0138] 本实施例提供了一种存储介质，该存储介质为计算机可读存储介质，其存储有计算机程序，所述程序被处理器执行时，处理器执行存储器存储的计算机程序时，实现上述实施例1的地铁换乘公交行为识别方法，如下：
- [0139] 获取数据源；其中，所述数据源包括乘客IC卡数据和GPS数据；
- [0140] 对乘客IC卡数据进行预处理；
- [0141] 将预处理后的乘客IC卡数据与GPS数据进行匹配；
- [0142] 判断匹配后的IC卡记录是否具备地铁-公交换乘空间条件；
- [0143] 当匹配后的IC卡记录具备地铁-公交换乘空间条件时，判断匹配后的IC卡记录是否具备地铁-公交换乘时间条件。
- [0144] 本实施例中所述的存储介质可以是磁盘、光盘、计算机存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、U盘、移动硬盘等介质。
- [0145] 综上所述，本发明先获取乘客IC卡数据和GPS数据，对乘客IC卡数据进行预处理，并将预处理后的乘客IC卡数据与GPS数据进行匹配，针对匹配后的IC卡记录，给出乘客在地铁与公交两种交通方式之间换乘所应具备的空间和时间条件，实现了地铁换乘公交行为识别，识别精确度较高。
- [0146] 以上所述，仅为本发明专利较佳的实施例，但本发明专利的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明专利所公开的范围内，根据本发明专利的技术方案及其发明构思加以等同替换或改变，都属于本发明专利的保护范围。

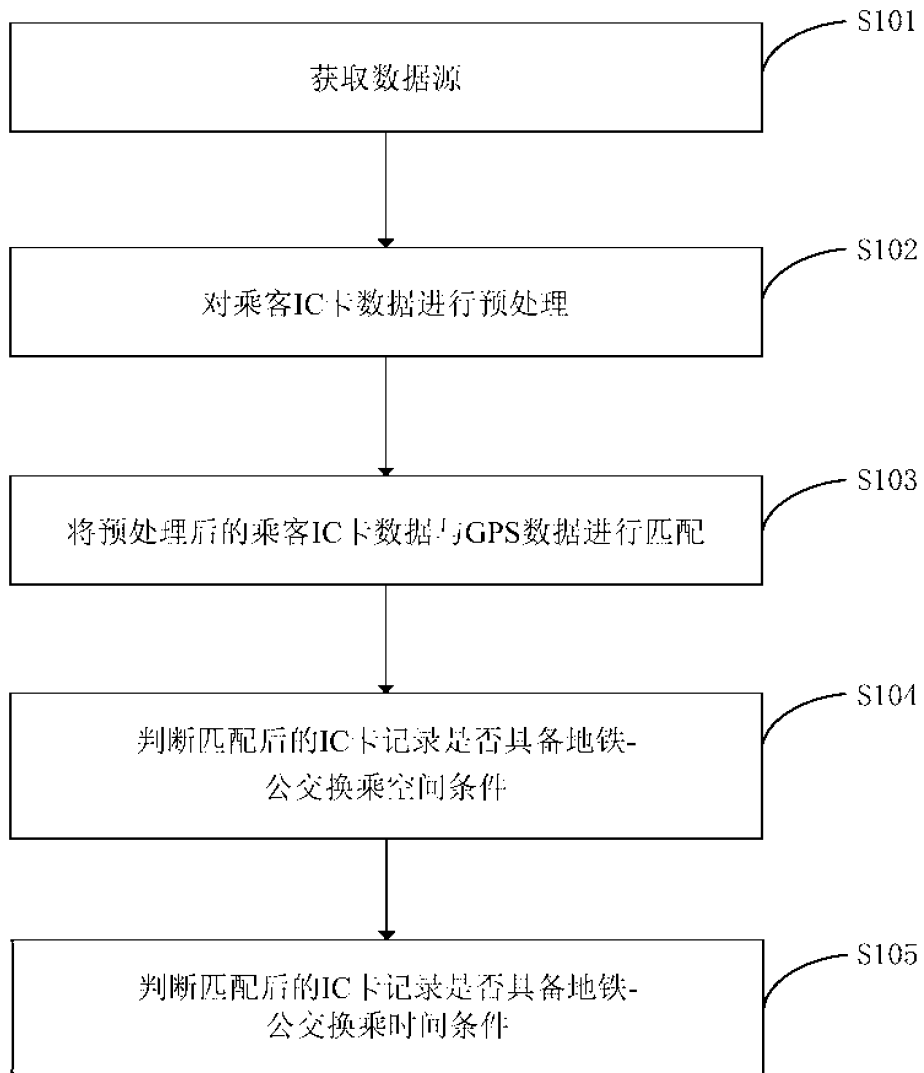


图1

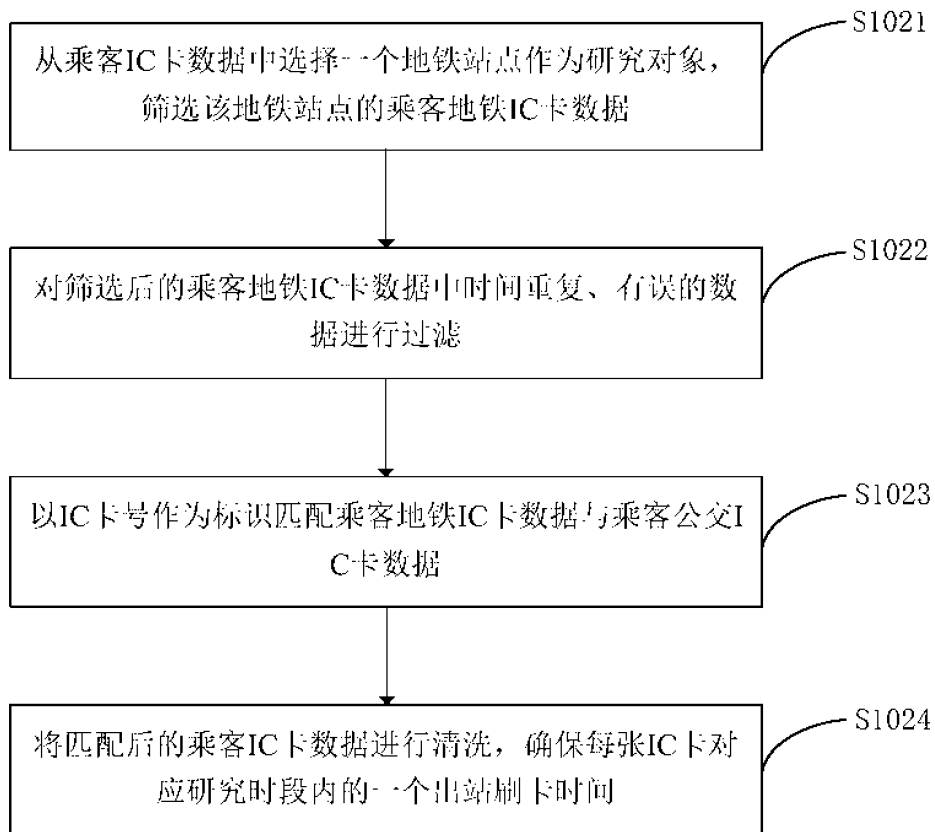


图2

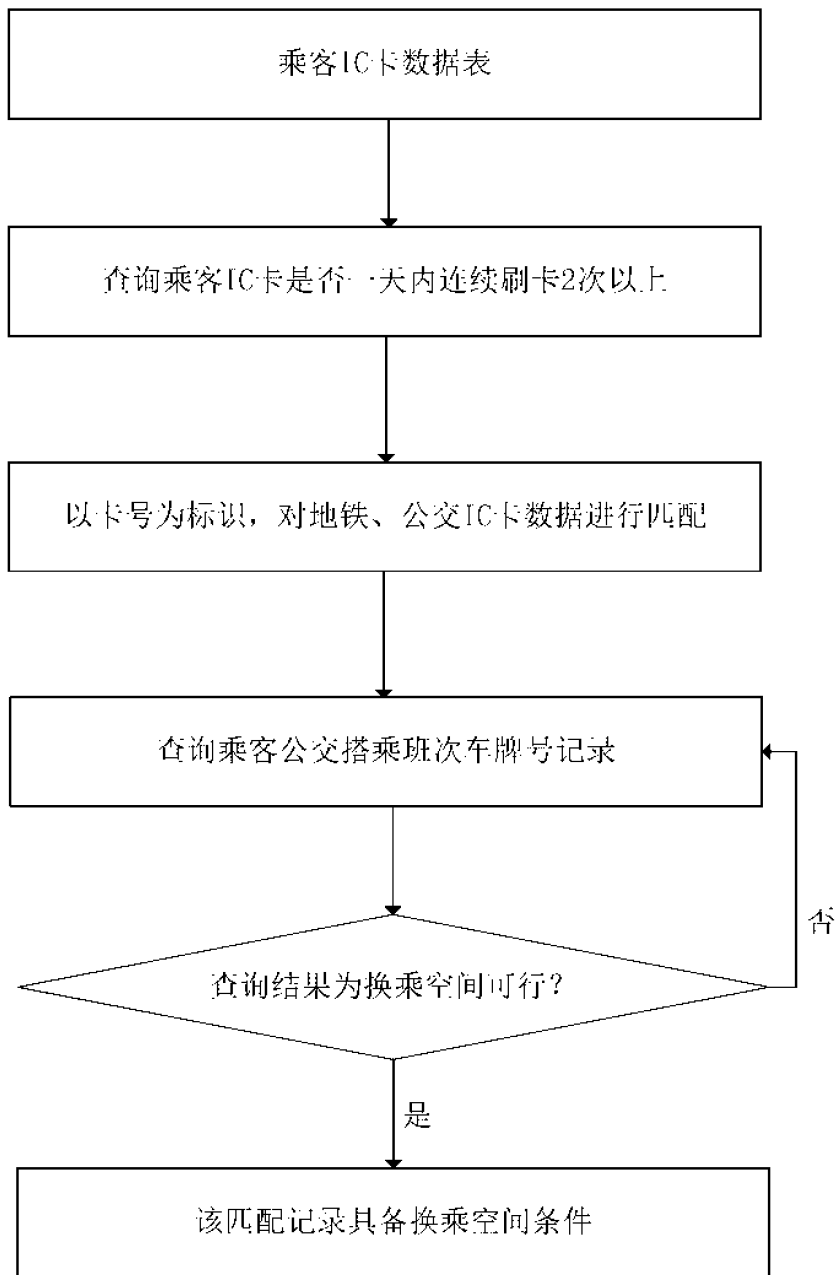


图3

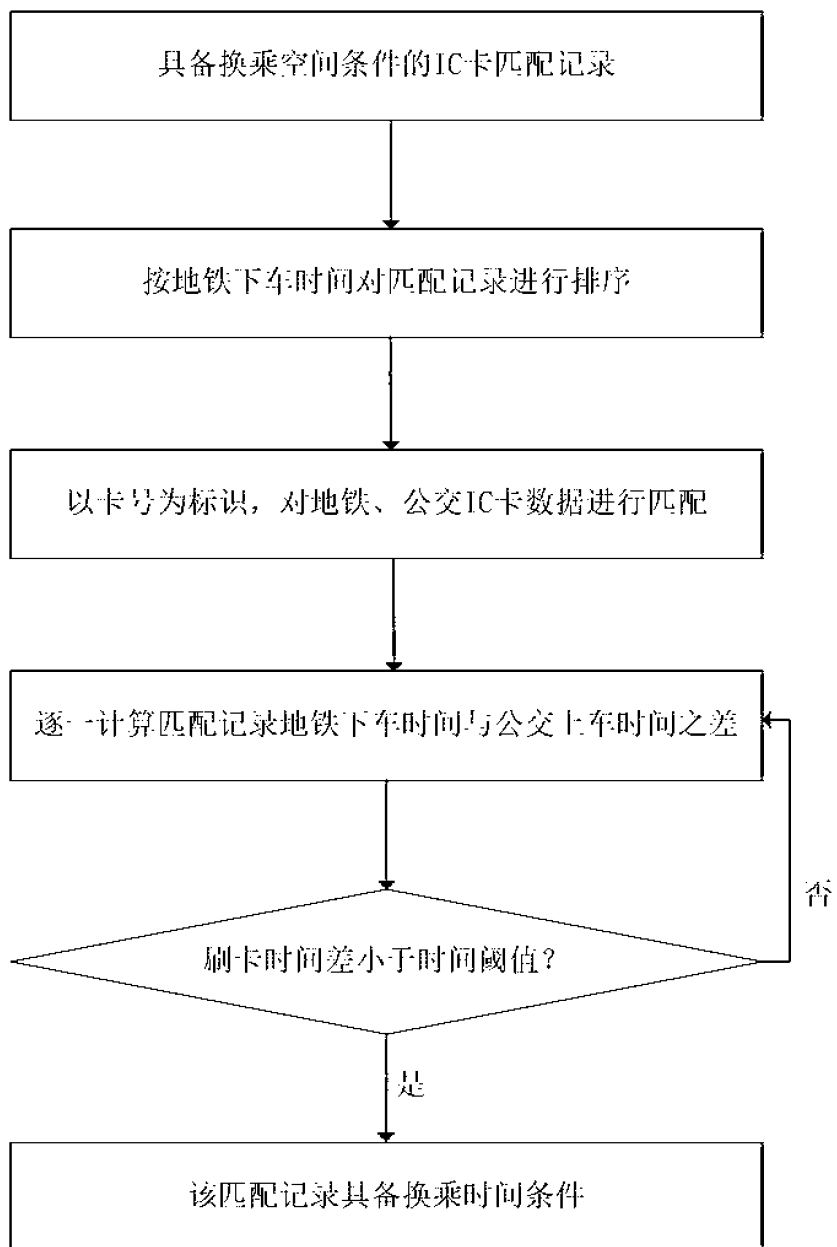


图4

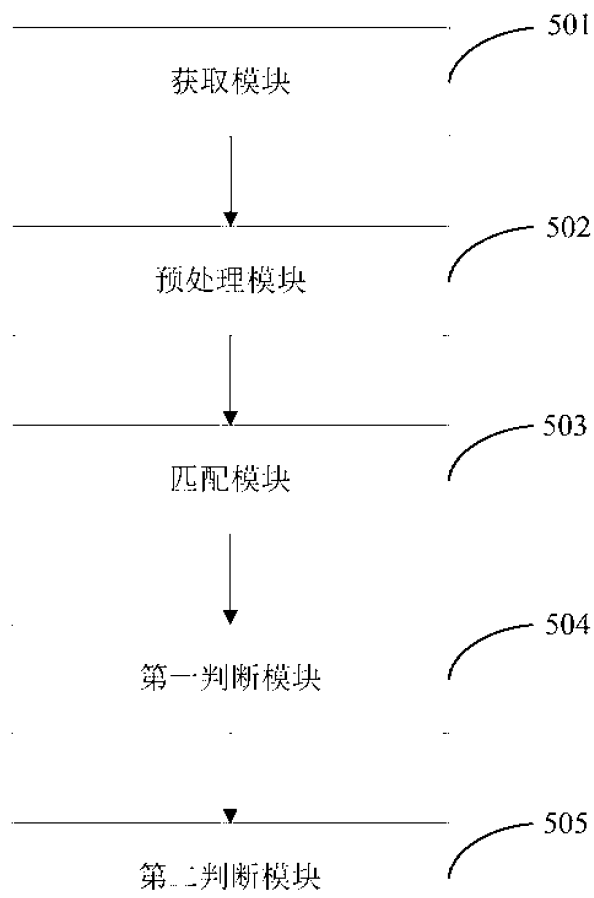


图5

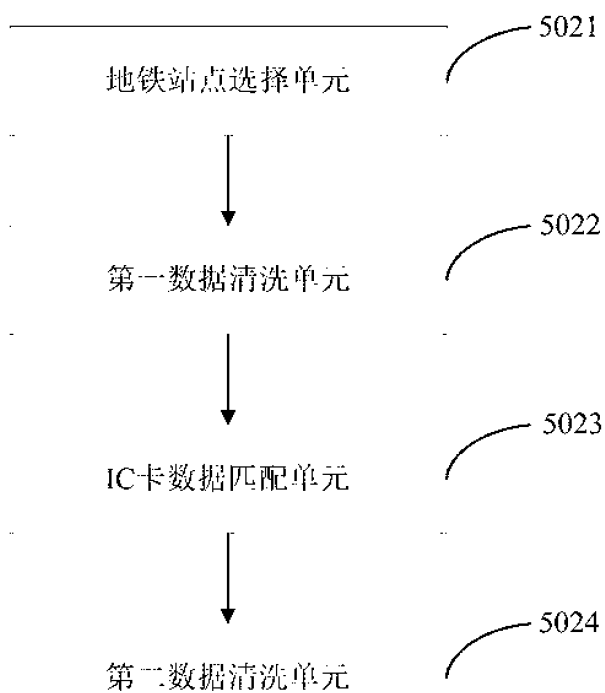


图6

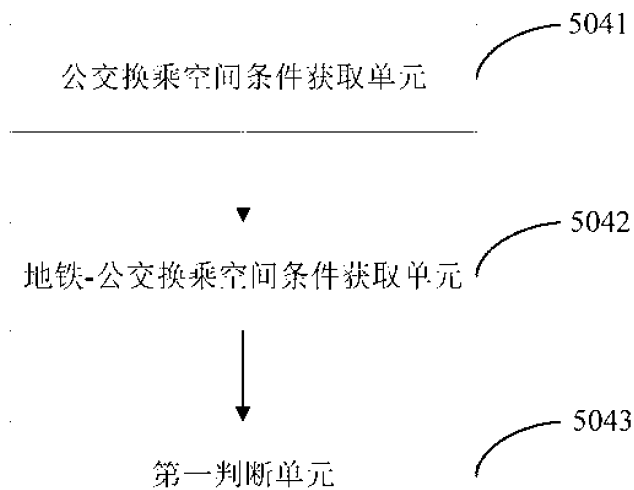


图7

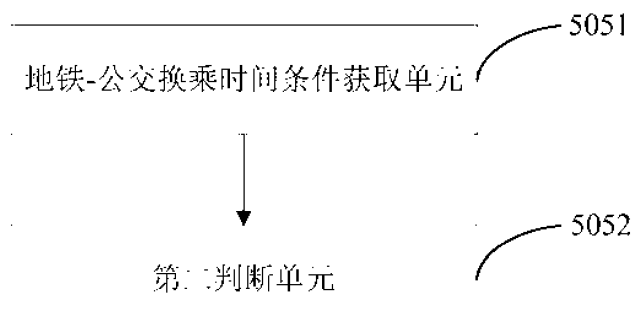


图8

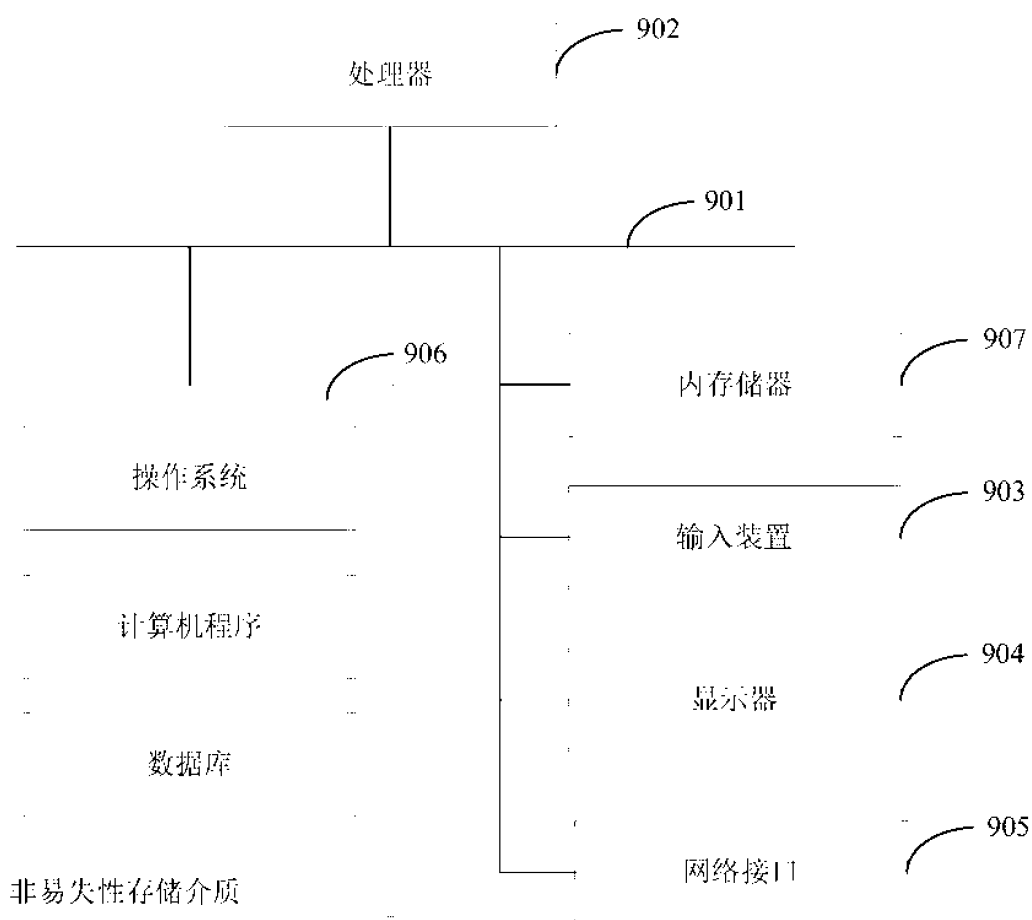


图9