



基于 XDR 大数据分析的高架用户识别技术

汪保友, 姚赛彬, 黄久成, 潘晖

(中国联合网络通信有限公司上海分公司, 上海 200080)

摘要: 城市高架路场景下, 用户行驶速度相对较快、人流密集, 客户对通信体验要求更高。首先, 介绍了网络质量管控大数据平台体系架构; 其次, 设计了基于 XDR 大数据分析的高架用户识别的算法; 最后, 通过 XDR 大数据分析处理, 与高架路段指纹库匹配, 输出高架用户运动轨迹, 提升了高架用户识别准确性, 为高架路无线网络质量监控提供辅助支撑。

关键词: 网络优化; XDR; 城市高架识别

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021214

Identification technology for city freeway users based on XDR big data analysis

WANG Baoyou, YAO Saibin, HUANG Jiucheng, PAN Hui

Shanghai Branch of China United Network Communications Co., Ltd., Shanghai 200080, China

Abstract: In the scene of city freeway, the driving speed is relatively faster, the flow of people is denser, and the customers have higher requirements for communication experience. Firstly, the architecture of big data platform for mobile network quality control was introduced. Then, the identification algorithm for city freeway users based on XDR big data analysis was proposed. Finally, through XDR big data analyzing and processing, the user track on freeway was output, which followed by matching with the fingerprints database of freeways. The identification accuracy for city freeway users was improved, and wireless network quality monitoring of city freeways was supported.

Key words: network optimization, XDR, identification of city freeway

1 引言

为有效缓解城市交通拥堵, 很多城市选择修建城市高架路, 建立城市立体交通网络。以上海为例, 上海有延安高架、南北高架、内环高架等 14 条高架线路, 日均车流量已经达到 201 万辆。

高架、高速等快速道路中, 用户行驶路线相对固定、行驶速度相对较快、人流密集, 客户对通信体验要求更高, 无线网络的移动性、保持性、接入性质量对客户感知有重要影响。如何准确识别高架用户, 利用高架用户群体的真实多维感知数据, 实现全民测试、全量测试, 快速汇聚问题道



路,更趋近用户真实感知体验,从而大幅提高网络检测效率和准确性,是迫切的实际需求,非常具有实用价值。

目前高架用户识别方法,存在如下问题。

(1) GPS 卫星定位法

由于多径效应等影响, GPS 的高度误差明显高于平面误差,仅靠 GPS 的坐标,无法区分用户处在高架桥上还是地面道路上。

(2) 位置区更新识别法

单纯位置区更新识别算法误判率高,容易把地面道路用户误判为高架用户。

(3) 基于图像检索识别法

基于图像检索识别高架用户,后台需要拍摄、存储大量高架路图片,对设备的计算能力要求高,适用于车辆导航及路线服务场景。

众所周知,运营商拥有海量的数据,其中 MR 测量报告和 XDR 信令数据是 OSS 域最主要的两类数据,MR 测量报告每日可达几 TB, XDR 原始信令每日更可达到几十 TB。面对如此巨大的数据体量,必须依托大数据支撑平台处理分析。本文提出一种基于 XDR 大数据分析的高架用户识别技术。该技术通过 XDR 大数据分析处理,输出用户完整运动轨迹,与高架路指纹库匹配,可自动输出高架用户运动轨迹,提升了高架用户识别准确性,为高架路无线网络质量监控提供了辅助支撑。

2 网络质量管控大数据平台体系架构

网络质量管控大数据平台体系架构如图 1 所示,下面 3 层是技术层面,上面两层是业务层面。

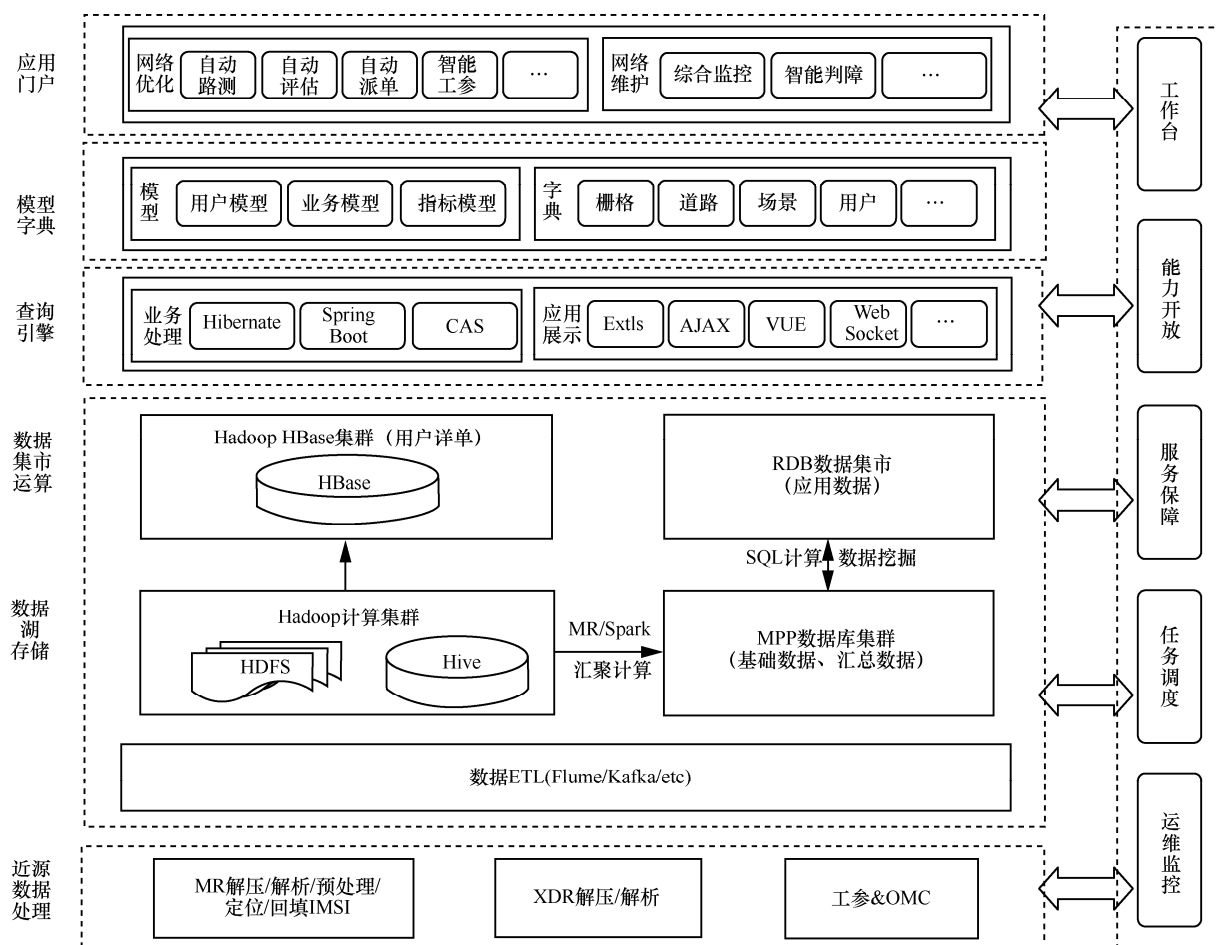


图 1 网络质量管控大数据平台体系架构

底层是对近源数据的采集处理,包括 MRO 文件解压、解析、预处理、定位、回填 IMSI 等;XDR 文件解压、解析;OMC 网管 PM 性能数据和工参数据等采集处理,采用了 Flume、Kafka 流式计算框架。往上一层,是数据湖存储和数据集市运算,数据湖存储于 Hadoop 集群的 HDFS 文件,通过 Hive 及 HBase 工具组件对外提供查询接口;通过 Spark Streaming 准实时计算工具,汇聚计算中间结果存储于 MPP 数据库(包括基础数据和汇总数据);再通过数据挖掘、SQL 计算,产生价值数据和粗粒度汇总数据,存储于 RDB 数据集市,以便外部调用。再往上,是查询引擎,通过一些框架组件,如 Hibernate、SpringBoot、CAS 框架,ExtJS、VUE、WebSocket 前端工具等,进行业务处理和可视化展现。再往上一层,是模型字典,提供计算模型(用户模型、业务模型、指标模型)以及数据字典(栅格、道路、场景、用户等)。顶层是应用门户,实现网络优化生产(自动路测、自动评估、自动派单、智能工参等)以及网络维护支撑(综合监控、智能判障等)。

3 基于 XDR 大数据分析的高架用户识别技术

3.1 相关定义

定义 1 区块路段

为有效缓解城市交通拥堵,很多城市选择修建城市高架路,建立城市立体交通网络。车辆通过上、下匝道驶入或驶离高架路。定义相邻匝道间的高架路段为一个区块路段。区块路段的集合记为 Σ , $\Sigma = \{S_1, S_2, \dots, S_k\}$, 其中, k 是区块路段的总数。根据上海高架匝道分布,将上海 14 条高架切分成 168 条具体区块路段。

定义 2 区块路段指纹

高架路上每一个区块路段,分别均由若干个主控小区覆盖,称为该区块路段指纹,记为: $S_i = \{C_1, C_2, \dots, C_n\}$, 其中, S_i 为第 i 个区块路段,

n 代表该区块路段有 n 个主控小区覆盖。

定义 3 高架路指纹库

将城市里所有高架切分为若干个区块路段,每个区块路段包含若干个指纹小区,由此构成 Σ^2 邻接小区的高架路指纹库。

定义 4 运动特征用户

基于移动性原理,用户在行驶过程中,会从一个位置区移动到另一个位置区,如果用户在短时间内发生多个位置区更新,则将其定义为运动特征用户。

定义 5 高架用户

某时间段内在高架路段行驶的运动特征用户,定义为高架用户。

3.2 技术思路

基于 XDR 大数据分析的高架用户识别,首先,通过 XDR 大数据关键信息,生成用户运动轨迹;其次,把每个用户的运动轨迹与高架路指纹库进行比对匹配,判定用户是否经过某个区块路段、经过多少条区块路段,进而判定是否为高架用户、运动场景经过哪些高架;最后,通过阈值和条件选择实现高架用户自动识别,为无线网质量感知问题提供用户聚焦手段。数据处理流程如图 2 所示。

3.3 用户运动行程生成算法

XDR 数据包括基础信息、业务信息、控制面信息、用户面信息等,选择关键时域位置信息,构成五元组 ($Event_{type}, Time_{begin}, UID, Cell_{uk}, Net_{type}, \dots$) 输入信息。UE 在移动过程中占用的基站小区会发生改变,如与上次变换时间间隔在 t_{span} 之内(设定 $t_{span} = 15 \text{ min}$),则运动继续,否则代表此次运动结束,据此生成用户行程 $UTrips$ 输出。用户运动行程生成算法流程如图 3 所示。

3.4 用户运动行程合并算法

通过算法 1 生成的用户运动行程 $UTrips$,可能存在行程碎片,即用户单次行程被分割记录成多条行程,为此需要进行用户行程合并。选择用户

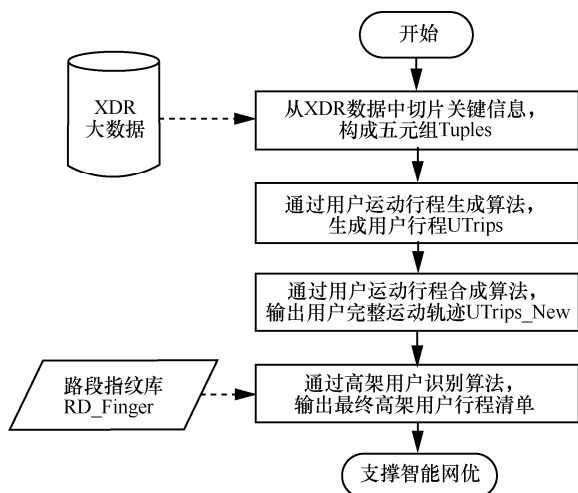


图2 高架用户识别数据处理流程

运动行程 $UTrips$ 作为输入信息, 首先将用户运动行程 $UTrips$ 与网络工参信息匹配, 增添用户起始占用基站小区经纬度数据和结束占用基站小区经纬度数据; 然后对用户行程碎片进行时序排列, 比较同一用户相邻两条行程记录的尾头信息, 即比较当前行程的起始时间与上次行程的结束时间, 以及当前行程的起始占用基站小区经纬度与上次行程的结束占用基站小区经纬度, 通过经纬度计算站点距, 并与相邻行程首尾时间差相除, 估算相邻行程碎片间用户运动速度 V , 如果运动速度 V 大于判定阈值 $V_{threshold}$ (如 $V_{threshold} = 30 \text{ km/h}$), 则认为用户在

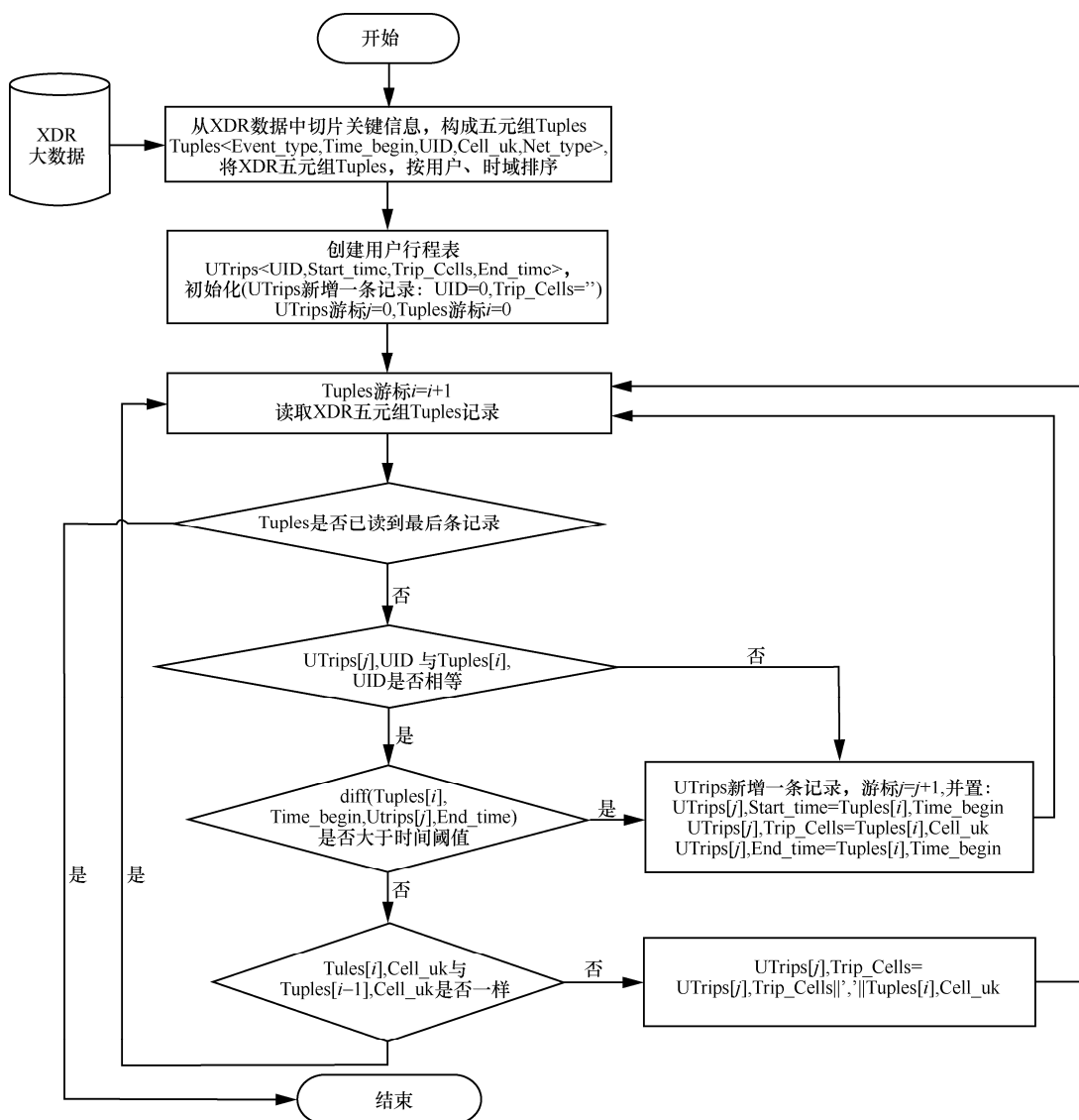


图3 用户运动行程生成算法流程

2 个运动碎片之间依然保持运动状态, 用户相邻运动碎片可以合并, 遍历操作后, 最终输出用户完整运动轨迹 $UTrips_New$ 。用户运动行程

合并算法流程如图4所示。

3.5 高架用户识别算法

将合并后的用户运动轨迹 $UTrips_New$, 与高

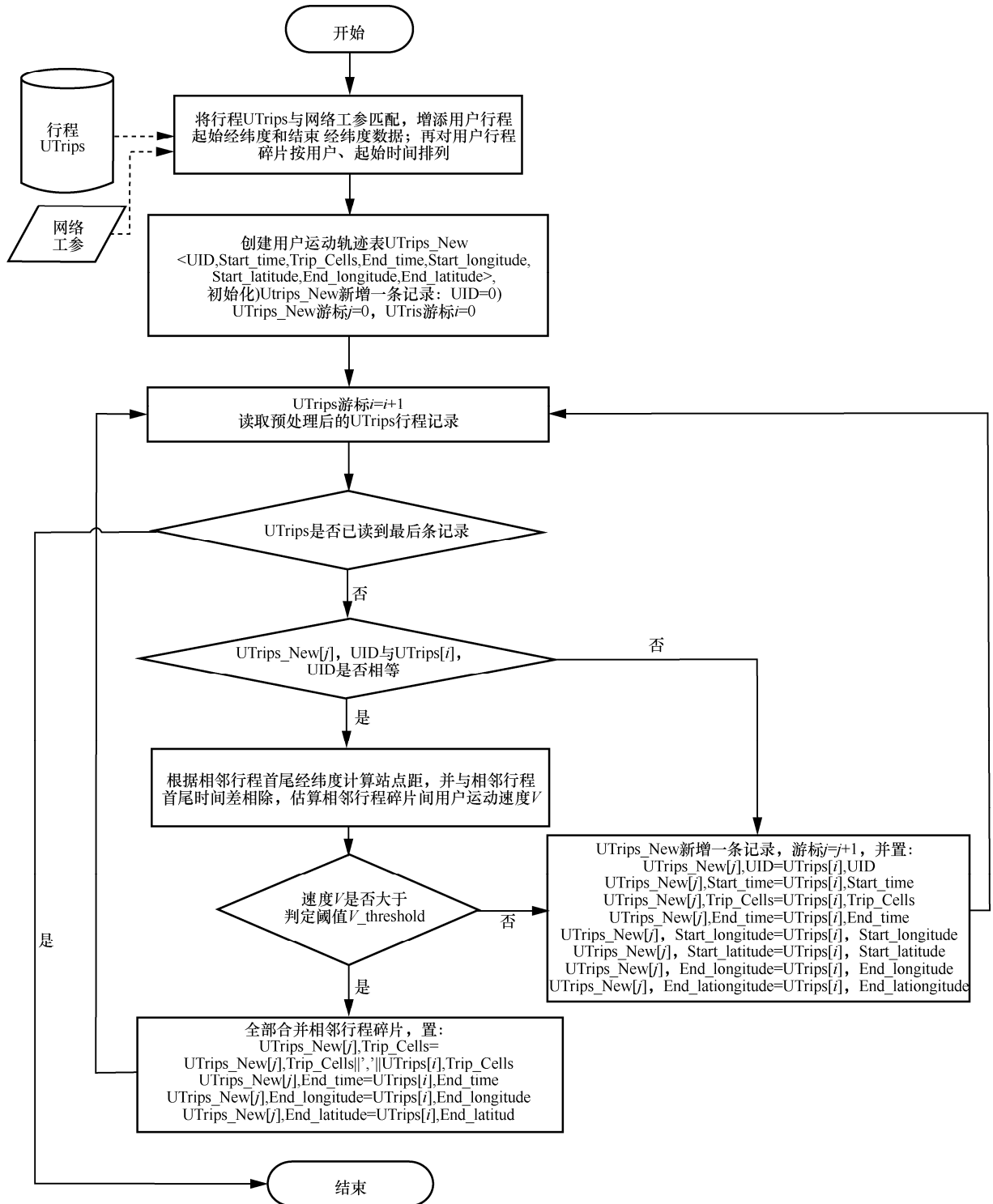


图4 用户运动行程合并算法流程



高架路指纹库进行匹配,判定用户是否经过某区块路段、经过多少条区块路段,进而判定是否高架

用户、运动场景经过哪些高架路。高架用户识别算法流程如图5所示。

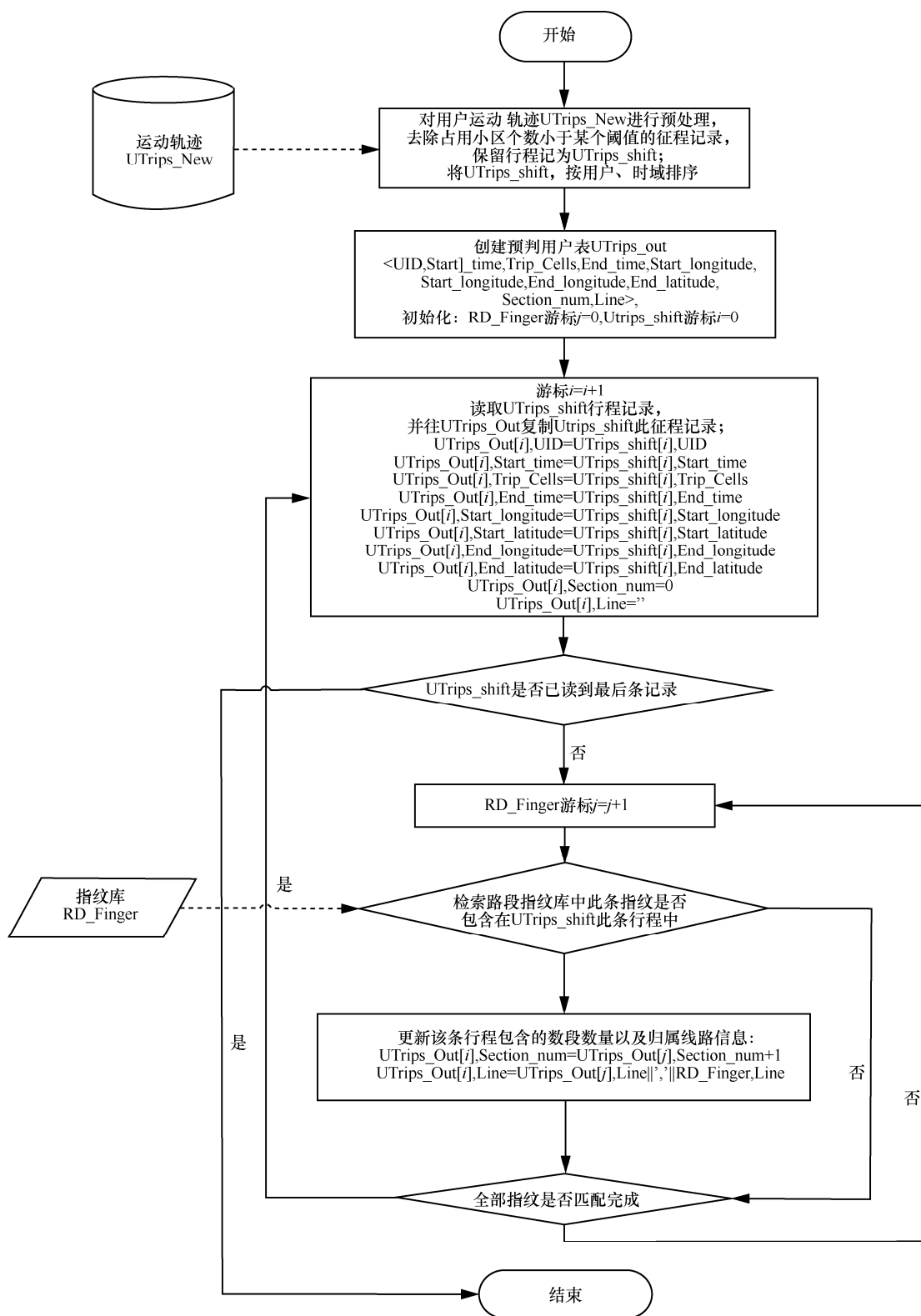


图5 高架用户识别算法流程

此算法基本思路分为以下4步。

步骤1 对合并后的用户运动轨迹 $UTrips_New$ 进行预处理, 去除占用小区个数小于某个阈值 $C_{threshold}$ (如 $C_{threshold} = 3$) 的行程记录, 保留行程记为 $UTrips_shift$ 。

步骤2 将 $UTrips_shift$ 按用户、时域排序。

步骤3 行程记为 $UTrips_shift$ 与高架路指纹库 RD_Finger 匹配, 判断该条行程是否在指纹中。如是, 则判断属于哪条线路, 并输出到 $UTrips_Out$ 数据集中。

步骤4 从 $UTrips_Out$ 中, 筛选区块路段数 (Section_num) 大于某个阈值 $S_{threshold}$ (如 $S_{threshold} = 3$) 的记录, 得出最终的高架用户行程清单。

4 高架用户识别的应用

通过以上算法的大数据分析处理, 可实现高架用户自动识别, 为无线网质量感知提供用户聚焦手段。准确识别高架用户后, 再利用高架用户群体的真实多维感知数据, 实现全民测试、全量测试, 快速汇聚道路问题, 更趋近用户真实感知体验, 从而大幅提高网络检测效率和准确性, 达到“节能增效”的目的。通过可视化大屏可实现区块路段 KPI 监控、异常路段告警提

示, 也可提供单用户、单区块路段的感知指标查询分析, 辅助投诉和问题定位。可视化监控系统门户如图6所示。

5 结束语

基于 XDR 大数据分析的高架用户识别技术, 为无线网质量感知提供用户聚焦手段, 有效提升了全网高架路评估能力和频次。传统 DT 路测法成本高、数据样本量小、测试终端数极少、随机性大, 无法模拟真实用户感知; 传统指标监控法, 通过筛选出高架路沿线主控小区进行 KPI 性能统计, 会引入大量非高架路用户, 准确性偏低。本文所提方法克服了传统 DT 路测法及指标监控法的问题, 从实际应用效果看, 该方法已推广使用到高架、高速、高铁等“三高”定轨道路场景, 质差和高掉话等隐性问题的发现及时率从 30% 提高到 95%, 日常优化问题点覆盖率超过 90%, 路测工作自动完成率超过 70%, 节约了运维成本, 提高了优化效率。

参考文献:

- [1] 刘韬, 吴涛, 王斌. XDR 与 MR 数据关联与应用[J]. 电信科学, 2019, 35(4): 120-130.
LIU T, WU T, WANG B. Association and application in XDR



图6 可视化监控系统门户



- and MR data[J]. Telecommunications Science, 2019, 35(4): 120-130.
- [2] YAO S B, HUANG J C, WANG B Y, et al. Design and application of network optimizing integrated platform based on rasterized big data[M]. Signal and Information Processing, Networking and Computers. Singapore: Springer, 2020: 789-799.
- [3] 汪保友, 姚健, 张正卿. 基于 FKS 的信令采集与监控技术[J]. 电信科学, 2018, 34(3): 145-155.
WANG B Y, YAO J, ZHANG Z Q. Signaling collection and monitoring technology based on FKS[J]. Telecommunications Science, 2018, 34(3): 145-155.
- [4] 伏玉笋, 杨根科. 人工智能在移动通信中的应用: 挑战与实践[J]. 通信学报, 2020, 41(9): 190-201.
FU Y S, YANG G K. Application of artificial intelligence in mobile communication: challenge and practice[J]. Journal on Communications, 2020, 41(9): 190-201.
- [5] 王莹, 苏壮. 无线网络中的移动预测综述[J]. 通信学报, 2019, 40(8): 157-168.
WANG Y, SU Z. Survey of mobility prediction in wireless network[J]. Journal on Communications, 2019, 40(8): 157-168.
- [6] 许佳捷, 郑凯, 池明旻, 等. 轨迹大数据: 数据、应用与技术现状[J]. 通信学报, 2015, 36(12): 97-105.
XU J J, ZHENG K, CHI M M, et al. Trajectory big data: data, applications and techniques[J]. Journal on Communications, 2015, 36(12): 97-105.
- [7] 朱近康, 柴名扬, 周武旻. 面向 B5G/6G 的三三三网络体系架构和优化学习机制[J]. 通信学报, 2021, 42(4): 62-75.
ZHU J K, CHAI M Y, ZHOU W Y. Three-three-three network architecture and learning optimization mechanism for B5G/6G[J]. Journal on Communications, 2021, 42(4): 62-75.
- [8] 汤雅妃, 魏进武, 张云勇. 基于大数据的信令监测系统研究[J]. 邮电设计技术, 2014(7): 47-52.

TANG Y F, WEI J W, ZHANG Y Y. Research on big data based signaling monitoring platform[J]. Designing Techniques of Posts and Telecommunications, 2014(7): 47-52.

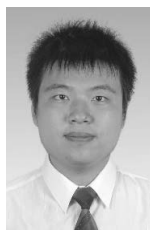
[作者简介]



汪保友 (1968—), 男, 博士, 中国联合网络通信有限公司上海分公司高级工程师, 主要研究方向为商业智能与大数据分析、无线网优化、网络与数据安全。



姚赛彬 (1982—), 男, 中国联合网络通信有限公司上海分公司无线网优化中心总经理、高级工程师, 主要研究方向为无线网优化、通信网络规划、智能网络。



黄久成 (1987—), 男, 中国联合网络通信有限公司上海分公司无线网优化中心专项优化室经理、工程师, 主要研究方向为无线网优化、通信网络规划。

潘晖 (1992—), 男, 中国联合网络通信有限公司上海市分公司工程师, 主要研究方向为无线网优化、5G 技术。