

基于信令大数据的国际漫入用户通信质量分析体系研究

Research on Communication Quality Analysis System of International Roaming Users Based on Signaling Big Data

王雪莹(中国联通网络运营事业部,北京 100037)

Wang Xueying(China Unicom Network Operations Division, Beijing 100037, China)

摘要:

通过算法研究、模型开发等工作,研发了基于信令大数据的国际漫入用户通信质量分析体系,该体系专门针对国际漫入用户建立大数据分析模型,有效分析国际漫入用户信息变化趋势,解决国际漫入用户网络质量无法集中采集的问题。模型的测试结果表明,该模型在国际漫入用户信息数据分析、通信服务质量监控等方面取得了良好的效果。

关键词:

手机信令;国际漫入;通信服务优化

doi:10.12045/j.issn.1007-3043.2025.04.012

文章编号:1007-3043(2025)04-0062-04

中图分类号:TN915

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Through algorithm research and model development, a communication quality analysis system for international roaming users based on signaling big data is developed. The system establishes a big data analysis model for international roaming users, effectively analyzes the change trend of international roaming user information, and solves the problem that the network quality of international roaming users cannot be collected centrally. The model test results show that the model has achieved good results in the analysis of international roaming user information data and the monitoring of communication service quality.

Keywords:

Mobile phone signaling; International roaming; Communication services optimization

引用格式:王雪莹. 基于信令大数据的国际漫入用户通信质量分析体系研究[J]. 邮电设计技术, 2025(4): 62-65.

0 前言

手机信令数据是一种新型的大数据源,与其他类型的数据相比,其具有实时性、完整性、出行时空全覆盖性等其他数据源所不拥有的优势^[1-2]。大数据是运用所有相关数据来进行挖掘分析的技术^[3],人们通过大数据技术获取海量数据,并对数据进行处理和分析,得到许多对未来具有研究和改进意义的结果^[4]。目前基于手机信令分析用户行为的研究已扩展到居

民动态、职住空间、通勤出行等方面^[5-7]。通过对海量的数据进行深入分析和利用,移动通信运营商能够更准确地把握用户需求,优化网络结构和服务策略,提升用户体验和满意度^[8]。

1 国际漫入用户基于信令大数据分析的现状

当前国际漫入用户基于网络信令大数据分析的过程中存在如下问题:第一,国际漫入用户数据在全网中相对分散,用户数据在全国每个省份都占有一定份额,并且网络制式包含5G、4G、2G/3G;第二,用户数据量较少,对比国内在网用户,数据量占比不高;第

收稿日期:2025-03-03

三,目前信令大数据分析没有专门针对国际漫入用户的模型,不利于集中分析,在重大会议、国际赛事等国际漫入用户集中的时刻,无法集中掌握用户服务质量,影响用户对运营商服务质量评价。

现有国际漫入用户通信质量分析流程一般应用hive SQL技术。分析流程如下:首先,基于所有用户信令数据,确定其中的国际漫入用户,根据号码归属国家、归属运营商和IMSI前3位识别出国际漫入用户;然后,根据IMSI关联语音业务信令数据、数据业务信令数据,得到待分析用户通信服务质量数据集。最后,分析国际漫入用户通信服务质量。

对比现有流程,本文提出的国际漫入用户通信质量分析体系具有如下优势。

a) 数据分省分市。本文提出的国际漫入用户通信质量分析体系得到分省分地(市)的数据结果集,分析人员可以根据自身需求筛选出相应省分地(市)的国际漫入用户。

b) 减少大量计算。现有的国际漫入用户质量分析流程需要先在全量数据中识别用户,再关联用户语音和数据的信令数据,最后分析通信质量,需要进行3次计算,应用本文的模型只需要进行1次分析计算即可。

c) 模型长期固定。本文的模型可以每小时定期执行,结果保存在hive库当中,现有分析流程通常只能完成一次分析需求,无法长期固定。

2 基于信令大数据的国际漫入用户通信质量分析技术方案

2.1 整体思路

基于信令大数据的国际漫入用户基本信息模型技术方案如图1所示。

本文将现有的O域海量信令大数据用户进行分析,用国际漫入用户基本模型替代实现,以加快计算速度,获得国际漫入用户分析结果。

a) 在用户小区漫游数据、用户小区数据业务数据及用户小区语音数据各自的数据集中筛选出国际漫入用户。

b) 建立国际漫入用户基本模型,分析用户语音数据、网络数据、归属国家、归属运营商,确定网络服务质量。

2.2 整体技术方案

基于信令大数据的国际漫入用户信息模型体系

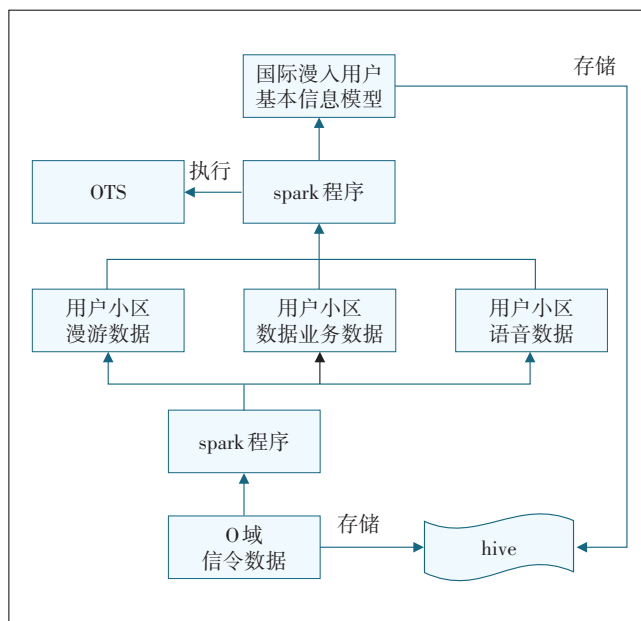


图1 基于信令大数据的国际漫入用户基本信息模型技术方案

方案采用spark+OTS的架构,实现对国际漫入用户基本信息模型构建。具体包括以下2个部分内容。

a) 利用spark框架筛选获取用户小区漫游数据、用户小区数据业务数据、用户小区语音数据3个数据集,经二次加工获得国际漫入用户基本模型。

b) OTS每小时调度1次spark加工大数据,完成每个小时的数据加工,获取分省、分天、分小时的模型数据,同时监督异常数据,补充数据执行,停止延迟的数据执行。

2.3 关键技术原理与实现

2.3.1 数据预处理

原始手机信令数据需要经过一系列算法处理才能用来提炼、分析相应指标^[9]。本文选取原始手机信令数据加工生成的用户小区漫游数据、用户小区数据业务数据及用户小区语音数据进行进一步的加工分析。其中,用户小区漫游数据、用户小区数据业务数据及用户小区语音数据都已经经过省分、分天、分小时的筛选。

2.3.2 国际漫入用户基本信息模型算法

2.3.2.1 用户筛选

建立国际漫入用户基本信息模型时,首先分别在用户小区漫游数据、用户小区数据业务数据及用户小区语音数据各自的数据集中进行国际漫入用户的筛选。筛选条件是数据产生时间为当前时间的前一天(即 $T-1$ 天),并且国际移动用户识别码为非空,前3位

不是460且大于200,长度为15,是否有数字匹配。具体代码如图2所示。

```
WHERE day_id=' $ {dayId} '
AND imsi is not null
AND substr(imsi,1,3)! ='460'
AND substr(imsi,1,3)>'200'
AND length(imsi)=15
AND imsi rlike '^\\d+ $ '
```

图2 国际漫入用户的筛选条件

2.3.2.2 分组聚合

经过筛选获得用户小区漫游数据、用户小区数据业务数据及用户小区语音数据各自的数据集,对每个数据集进行分组聚合。

a) 用户小区漫游数据。将用户小区漫游数据以天、省分、数据时间、用户号码、国际移动用户识别码、国际移动设备标识、终端类型分配码、拜访省分、拜访地(市)、归属运营商国家码、归属运营商移动网号、小区编码、接入网络类型等进行分组,将归属运营商国家码、归属运营商移动网号最大值聚合,获得用户小区漫游数据集。

b) 用户小区数据业务数据。将用户小区数据业务数据以天、省分、数据时间、用户号码、国际移动用户识别码、国际移动设备标识、终端类型分配码、拜访省分、拜访地(市)、归属运营商国家码、归属运营商移动网号、小区编码、接入网络类型等进行分组,将归属运营商国家码、归属运营商移动网号最大值,总流量加和聚合,获得用户小区数据业务数据集。

c) 用户小区语音数据。将用户小区语音数据以天、省分、数据时间、用户号码、国际移动用户识别码、国际移动设备标识、终端类型分配码、拜访省分、拜访地(市)、归属运营商国家码、归属运营商移动网号、小区编码、接入网络类型等进行分组,将归属运营商国家码、归属运营商移动网号最大值,通话时长加和,短信条数加和聚合,获得用户小区语音数据集。

2.3.3 形成数据集

最终,将来自用户小区漫游数据、用户小区数据业务数据及用户小区语音数据的数据集并集到一起,形成一个数据集。具体代码如图3所示。

2.3.4 模型生成

国际漫入用户基本信息模型以天、省分、数据时

```
user_profile_cell_roam
union all
user_serviceaware_cell_exp
union all
user_serviceaware_cell_voice
```

图3 并集形成数据集

间、用户号码、国际移动用户识别码、国际移动设备标识、终端类型分配码、拜访省分、拜访地(市)、归属运营商国家码、归属运营商移动网号、小区编码、接入网络类型分组,将总流量加和,通话时长加和,短信条数加和聚合,生成模型。

2.4 测试结论

本文以亚运会期间A省的数据来验证国际漫入用户基本信息模型。首先,国际漫入用户情况分析随时间变化而变化,重点以亚运会赛事期间(2023-09-23—2023-10-08)为分析基准,为测算同比、环比及观察下一周期情况,分别选择2023-08-23—2023-09-08、2023-09-07—2023-09-22、2023-10-09—2023-10-24为比对参照时间范围。其次,A省国际漫入用户情况需要从人数、通话时长、数据流量3个方面分析用户信息,人数计算方法为:取对应时间拜访省份为A省的用户号码去重取总数量;通话时长计算方法为:取对应时间拜访省份为A省的所有用户主被叫通话时间之和;数据流量计算方法为:取对应时间拜访省份为A省的所有用户总流量之和。再次,分析数据过程中涉及到如下2个比率的计算:

同比增长率=(当月的指标值-上月同期的值)÷上月同期的值×100%

环比增长率=(本期数-上期数)÷上期数×100%

最终,得到A省国际漫入用户情况分析结果如表1所示。

表1 A省国际漫入用户情况分析结果

分析项目	2023-08-23—2023-09-08(同比)	2023-09-07—2023-09-22(环比)	2023-09-23—2023-10-08(亚运会赛事期间)	2023-10-09—2023-10-24(观察期)
人数/人	253 552	602 915	604 900	553 288
通话时长/h	19 609	22 600	21 679	21 063
数据流量/PB	94 027	113 502	143 233	127 076

2.4.1 国际漫入用户数分析

从表1可知,亚运会赛事期间,国际漫入用户数同

比增长 351 348 人,同比增长率为 138.57%,国际漫入用户数环比增长 1 985 人,环比增长率为 0.33%。

国际漫入用户数在亚运会赛事期间之前已经达到和赛事期间一样的量级。结合观察期数据分析,国际漫入用户数在亚运会赛事前期开始增长,亚运会期间数量最高,赛事结束后观察期内,略有减少,依旧处于比较高的水平。

2.4.2 国际漫入用户通话时长分析

从表 1 可知,亚运会赛事期间,国际漫入用户通话时长同比增长 2 070 h,同比增长率为 10.56%,国际漫入用户通话时长减少 921 h,环比下降率为 4.1%。

国际漫入用户通话时长在亚运会赛事前处于最高,赛事期及赛事结束后观察期内,逐步下降,下降较为缓慢,总体量级不变。

2.4.3 国际漫入用户流量分析

从表 1 可知,亚运会赛事期间,国际漫入用户流量同比增长 49 206 PB,同比增长率为 52.33%,国际漫入用户流量环比增长 29 731 PB,环比增长率为 26.19%。

亚运会赛事期间,国际漫入用户流量有所增长,增长比例高于国际漫入用户通话时长,少于国际漫入用户数。国际漫入用户流量在亚运会赛事期间为最高,变化趋势与国际漫入用户数一致,在亚运会赛事前期开始增长,亚运会期间数量最高,赛事结束后观察期内,国际漫入用户数略有减少,依旧处于比较高的水平。

根据数据分析结果,接下来分析在通信服务优化方面可以提升的地方。第一,用户数、通话时长、数据流量 3 项指标在亚运会赛事之前就开始增长,因此通信服务支撑需要在赛事开始之前进行,以保证用户业务体验,而赛事结束后的观察期,用户数、通话时长、数据流量数据量级依旧庞大,通信服务支撑仍然需要稳定保障。第二,用户数、通话时长、数据流量 3 项指标在亚运会前后赛事数据量级一直居高不下,可以根据用户通话、流量特点,为用户推出赛事期间的 16 天专属语音、流量包,月度优惠语音、流量包,以满足用户不同需求。大数据时代,市场营销的主要任务就是要为企业提供有效信息,而在这个过程中,市场营销也需要以客户需求作为导向^[10]。第三,从国际漫入用户数、通话时长、数据流量整体分析中可以发现,通话时长的变化规律不正相关于用户数,赛事期间保障流量使用的质量更为重要,并且短期流量包相较于短期语音包更易于销售。

3 基于信令大数据的国际漫入用户通信质量分析应用前景

本文提出一种基于信令大数据的国际漫入用户通信质量分析体系,采用国际漫入用户基本信息模型,在国际漫入用户场景集中的情况下,可以得到用户数、通话时长、数量流量的具体数据和趋势,为通信服务优化提供明确的数据支撑和依据。该模型可以应用于正常时段和国际漫入用户集中时段,将极大改善现有国际漫入用户监测情况,提升运营商的品牌效应,提高通信服务质量。同时,该模型也已应用于各省分公司、国际公司日常报表的用户分析当中。

参考文献:

- [1] 刘李佳,白山,邹鑑珉. 信令数据让新闻推送更精准[EB/OL]. [2024-10-08]. doi: dx. doi. org/10.19483/j. cnki. 11-4653/n.2017.04.040.
- [2] 卢梦丽,邹思强,陈长超,等. 基于手机信令数据的交通时空大数据分析挖掘系统[J]. 计算机科学与应用,2020,10(8): 1471-1479.
- [3] 郭鹏飞,李刚. 大数据方法及其应用[J]. 计算机科学与应用,2019,9(9):1724-1731.
- [4] 刘凯悦. 大数据综述[J]. 计算机科学与应用,2018,8(10): 1503-1509.
- [5] 田钊,张乾钟,赵轩,等. 基于手机信令数据的城市居民动态 OD 矩阵提取方法[J]. 郑州大学学报(工学版),2024,45(3): 46-54.
- [6] 高瑜堃,赵鹏军. 手机信令数据在职住空间、出行行为和交通碳排放研究中的应用进展与前景[J]. 热带地理,2024,44(5): 877-890.
- [7] 葛浩菁,吕远,焦朋朋. 基于信令数据的中型城市通勤公交站点优化方法[J]. 交通信息与安全,2024,42(1): 142-149.
- [8] 任飞云. 基于大数据下移动通信网络优化的思考[J]. 智能城市应用,2024,7(6): 19-21.
- [9] 李秋曼. 基于手机信令的交通出行特征分析[J]. 现代交通技术,2023,20(5): 84-88.
- [10] 吴伍思梦. 大数据背景下市场营销面临的机遇与挑战分析[J]. 可持续发展,2022,12(6): 1789-1794.

作者简介:

王雪莹,工程师,学士,主要从事网络大数据分析、模型开发工作。

