



XDR 与 MR 数据关联与应用

刘韬, 吴涛, 王斌

(中国联合网络通信有限公司安徽分公司, 安徽 合肥 230069)

摘要: 结合大数据及机器学习的发展与实践, 以安徽省 O 域 XDR 与 MR 数据为基础, 详细阐述了数据采集过程及 XDR 与 MR 关联的实现技术, 并结合实践给出了关联率与准确率影响因素分析, 通过机器学习与指纹定位算法结合, 利用最小欧氏距离算法不断优化指纹定位提升定位精度, 实现了关联后数据的栅格化, 同时基于关联数据提出了面向规划、面向网络、面向客户、面向市场的创新应用场景与方向, 最后给出了建维与市场两个方向的应用案例, 验证了 XDR 与 MR 关联的应用前景。

关键词: 大数据; 数据关联; 关联成功率; 关联数据应用

中图分类号: TP319

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019057

Association and application in XDR and MR data

LIU Tao, WU Tao, WANG Bin

Anhui Branch of China United Network Communications Co., Ltd., Hefei 230069, China

Abstract: Based on the XDR and MR data records of O domain and combining with the development and practice in big data and machine learning technologies, the analysis of impact factors of correlation rate and accuracy was described in detail via practice with the help of technology of association of XDR and MR data records. By means of the integration of machine learning technology and fingerprint localization algorithm, the position accuracy was promoted constantly using the random forest algorithm, and the rasterization of associated data was realized. Simultaneously, the innovative application scenarios and directions applied to planning, network, customer and market were proposed. Finally, to verify the prospect of the association of XDR and MR data records, two examples in network construction, maintain and market were enumerated.

Key words: big data, data association, success rate of the data association, application of data association

1 引言

随着移动互联网的发展和 4G 网络的普及, O 域数据呈现爆发式增长, 而 XDR 与 MR 就是网络侧最主要的两类数据。XDR 主要采集接口包括 S11、

S1-MME、S1-U、SGS、S6a、S5/S8、Gb、IU_PS、Gn、A、IU-CS、C/D, 其中, S1-MME、S1-U 承载了用户信息、业务信息、话单信息等; MR 主要包括 MRO、MRS 和 MRE 3 种类型测试报告, 其中 MRO 代表周期性的测量报告样本数据文

件,包含了用户定位的工参信息及网络评估的覆盖信息等。就中国联合网络通信有限公司安徽分公司(以下简称安徽联通)而言,目前PS流量峰值已达到400 Gbit/s,CS流量峰值达到560 Mbit/s,某地市每小时S1-MME达到1 600万条,MRO达到1 300万条,而这个数字每年还在成倍增长。

目前各个运营商已经具备这两类数据的采集、解析与分析能力,并应用于基本的建维中。如参考文献[1]针对网络维护效率、用户体验等,提出了基于大数据的信令监测系统建设方案;参考文献[2]中通过对MR定位及数据的分析获取网络覆盖的情况,指导了宽阔无线环境和城区深度覆盖的网络规划;参考文献[3]中基于MR定位技术实现了客服支撑系统应用,使网络投诉问题自动分析和统一解释、用户投诉地点自动定位和投诉问题处理闭环管理;参考文献[4]针对MR数据采集的业务场景,提供高质量的网络覆盖,提出了基于Flume技术的数据采集

实现方案;参考文献[5]中将XDR应用于精确识别高铁用户、评估高铁用户感知、提供高铁网络优化分析等方面;参考文献[6]中提出一种基于DPI数据的家庭宽带投诉精细化分析系统,实现了家庭宽带业务全流程穿越,同时有效解决了家庭宽带投诉问题。然而在实际生产中,将XDR与MR关联分析实现用户信令轨迹回溯、KQI栅格化等方面仍然存在不足,主要表现为XDR与MR数据关联成功率与准确率低、基于MR的定位准确度不高、XDR与MR从采集到关联再到如何应用缺少整体架构、无法提供高性能的大数据平台支撑海量的XDR与MR关联分析等。

本文结合安徽联通测试实践,针对安徽联通大数据平台架构对XDR与MR关联的算法进行了研究,并对关联率与准确率的影响因素进行分析,提出将机器学习应用到指纹库定位中提高定位精度与减少盲区,同时给出了建维与市场两个方面的应用实践。实践结果表明,通过将XDR与MR

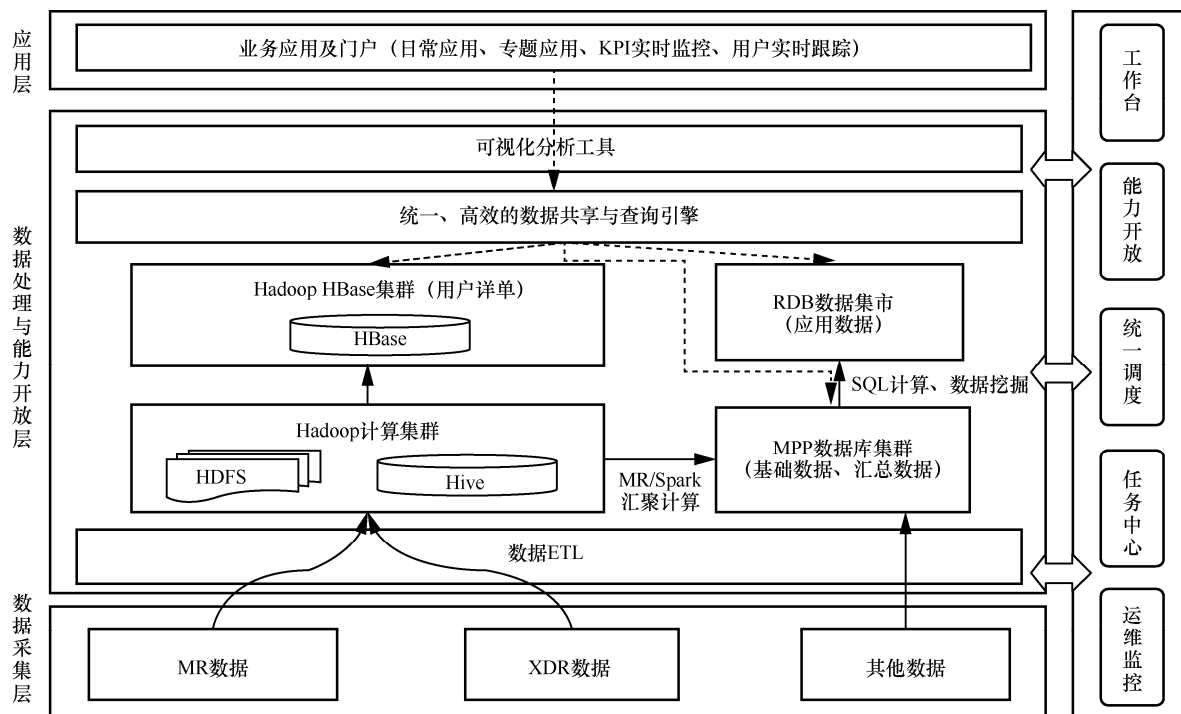


图1 XDR与MR关联的融合架构



关联,可以有效实现用户信令轨迹回溯、KQI 栅格化等,从而助力运营商提高数据使用价值,拓展数据变现业务。

2 XDR 与 MR 数据关联技术

2.1 XDR 与 MR 关联的融合架构

安徽联通 XDR 与 MR 关联的融合平台基于 x86 架构,以主流开源通用框架技术为主,结合特定专业场景定制开发,采用架构分层、功能模块化的设计思路进行构建。如图 1 所示,支持三层架构,具有良好的灵活性和扩展性以支撑上层应用;支持 Hadoop、虚拟机(分布式处理等)、Docker 等大数据模式。在实际工作中,利用三层架构模式的大数据平台可以充分发挥数据挖掘的能力,把数据挖掘和大数据平台有效结合,对于数据分析处理、用户感知提升起到巨大作用。

2.2 XDR 数据与 MR 数据关联的分析技术

针对 XDR 数据,安徽联通通过部署华为 SEQ 平台,利用探针的深度解析能力,将原始码流解析成中国联通集团 DPI 3.0 规范的 XDR 数据,其中包括详细的用户事件等重要信息。而 MR 数据则是安徽联通通过部署诺基亚的 FAST 平台来采集、解析和定位的。

XDR 数据与 MR 数据的关联技术,主要通过关联字段 MME_UE_S1AP_ID 将同一时间、同一地点和同一用户的 XDR 数据和 MR 数据联系起来,并回填所要分析的信令字段和业务字段,并在此基础上进行相关指标的统计及分析。其中,XDR 与 MR 关联的具体字段如图 2 所示。

XDR 与 MR 数据的关联有 3 种,分别为 MME_MR 关联、HTTP_MR 关联和 COMMON_MR 关联。每种关联的区别在于 XDR 数据不同,分别是 MME 数据、HTTP 数据和 S1_COMMON 数据。在每种关联中均有 3 个过程:预处理—关联—回填。现将关联的项目过程描述如下。

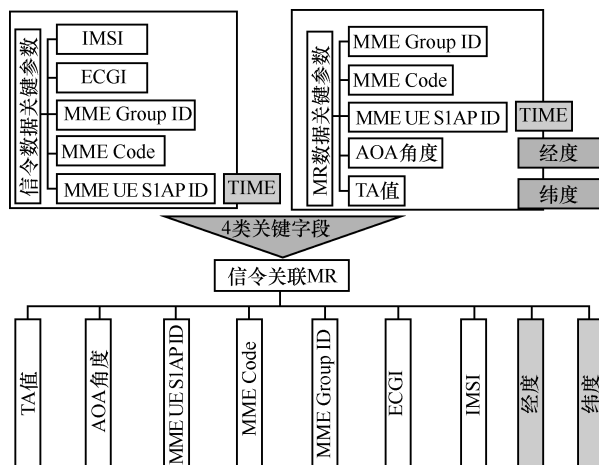


图2 XDR 与 MR 关联的具体字段

(1) 预处理

预处理部分包括数据清洗和数据处理。

数据关联之前,先将小区号为空、开始时间不合理、结束时间不合理等无效数据进行清洗,只保留有效的数据进行关联,以此提高数据关联的效率,保证关联所得到数据的有效性。

(2) 关联

关联部分包括两个过程:第一个过程是使用关联字段 MME_UE_S1AP_ID 对 XDR 数据和 MR 数据进行关联;第二个过程是针对第一个过程中没有关联的记录进行第二次关联。

数据关联的条件是站号相等、长码 MME_UE_S1AP_ID 相同,同时挑选出开始时间在 MR 的开始时间和结束时间之间的 XDR 信令数据,计算 XDR 与 MR 的时间差,然后按照时间差进行排序,最后挑选出时间差最小的 XDR 数据作为可信的关联记录,至此第一个关联过程结束。

对于没有关联上的部分,将会对 MR 数据的开始、结束时间前后的 20 min 分别进行滑窗搜索,向前滑窗搜索结果和向后滑窗搜索结果进行合并、去重,这就是第二次的关联过程。最后再将两次关联过程的结果进行去重操作,得到 XDR 数据与 MR 数据的关联记录。

(3) 回填

回填部分是根据关联部分得到的关联记录,

将一一对应的 MR 数据中的测量信息、位置信息填写到存储 XDR 感知数据的表中,全部回填完成后,就完成了 XDR 数据与 MR 数据的关联工作。

关联得到的所有字段中,只有位置信息是不存在于 XDR 数据和 MR 数据中的,所以想要得到位置信息,需要对 MR 数据进行定位处理。安徽联通采用的是基于指纹库的 MR 定位算法。指纹库定位算法不过分依赖工参数据,定位精度高,完全满足精确规划、精细规划的精度要求。

安徽联通通过数据建模和大量的 MR、XDR 数据关联运算,把数据查看视角进一步放大,实现从小区级别的定位扩展到用户具体的业务位置定位,并可以借此带来一系列的应用。

2.3 XDR 与 MR 数据的关联率及准确率研究

MR 与 XDR 的关联率,一般都不可能达到 100%,根据实际的关联统计,关联率可达到 70%左右,关联后的数据准确率可达到 85%左右。经分析,影响关联率和准确率的因素主要有数据质量问题 and 基站系统工作原理问题两个方面。

数据质量方面的问题可分为如下两类。

- 由于某种原因导致 MR 与 XDR 包含的基站不能完全一样,各自丢掉一定基站、小区的报告现象比较常见。
- 由于某种原因导致 MR 与 XDR 相同基站包含的时间段不能完全一样,各自丢掉一定时间段的报告的现象比较常见。

基站系统工作原理方面的原因可分为如下两个原因。

- 在通常考查的时段内,比如 1 h 内,总是存在有些终端没有上报 MR 可能性,而却在 S1MME 里留有记录的情况,比如一个用户在 1 h 内既没有业务,又没有移动,这时候可能仅有周期性位置更新的信令记录,而没有上报 MR 的情况,事实上根据对原始 XDR 和 MR 报告的统计,1 h 内 MR 报告中的用户数总是仅有 XDR 用户数据的

70%左右。

- 重要关联字段 MME_UES1APID 有时候存在短码的现象,导致不能关联。

2.4 关联后 XDR 与 MR 数据的栅格化

XDR 数据和 MR 数据完成关联之后,就可以将 MR 数据中的位置信息和 XDR 数据中的感知信息结合起来,根据关联数据中的位置信息进行 KQI/KPI 的栅格化。栅格化之后,就可以对 KQI/KPI 等指标进行统计分析。

KQI/KPI 栅格化,可以实现相关指标的栅格化定位,进而实现相关指标的地理化分析、GIS 图形呈现等。在实际应用中,KQI/KPI 栅格化可用于高流量区域、高价值区域、流量受压抑区域的分析以及精准的基站价值投放规划、精细化的网络优化维护和 VIP 栅格保障等方面。

由于关联后的数据量过大,传统的数据库、运算模式将无法支撑其运算,所以关联数据的栅格化过程是在 Hadoop 环境下进行的。借助 Hadoop 的 MapReduce 编程模型以及分布式文件存储系统,可以将关联后的数据分别存储在不同的子节点服务器上,并通过主节点进行文件的调度和数据的计算,以完成大规模数据的栅格化运算。

栅格化后的字段包括 TOTAL_TRAFFIC、USER_COUNT、DL_RATE、TCP_LETANCY、AVGRSRP、SAMPLES 等,分别对应着流量、用户数、下载速率、TCP 时延等感知类指标以及平均电平、采样数、弱覆盖比例等覆盖类指标。栅格化之后的关联数据,能够以图形化的形式进行呈现,实现多种指标的联动分析,清晰地显示出异常小区的分布和质差区域的分布,图形化呈现如图 3 所示。

2.5 XDR 与 MR 数据的关联率及准确率优化

第 2.2 节分析了影响 XDR 与 MR 数据关联准确性的因素,想要提升数据的关联率和准确率,就需要从关联准确率的影响因素下手,尽量减少这些影响因素对关联准确率的影响。接下来将从

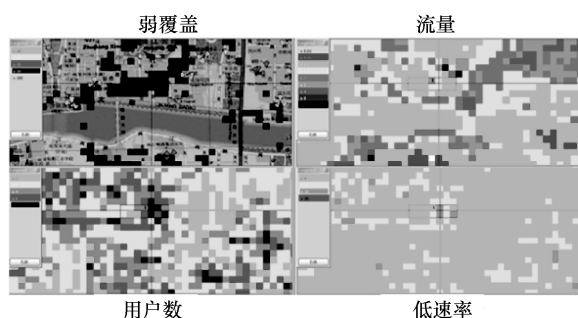


图3 XDR与MR数据栅格化后的图形化呈现

数据质量和基站系统原理两方面进行分析。

在数据质量方面，需要对采集到的 XDR 数据和 MR 数据的质量进行把控，尽量提升数据的质量。只有提高了数据的采全率和质量，将无效数据的占比降到最低，才能提升数据的关联率和准确率。通过对数据质量进行监管，比如按时监控采集服务器的状态，保证服务器能够正常运行，并有足够的空间来存储上报的文件，保证 XDR 和 MR 数据在采集时不会发生文件的丢失；此外还可以分别对 XDR 数据中的小区数和 MR 数据中的小区数进行统计，统计的小区个数可以和现网的小区数进行对比，如果发现统计得到的小区数小于现网的小区数，则可以查找采集侧的原因，查看基站的配置、查看数据传输过程是否有文件丢失等，解决发现的问题，提升数据的采全率。XDR 数据统计得到的小区数和 MR 数据统计得到的小区数也可以进行对比，若 MR 数据统计得到的小区数小于 XDR 数据统计得到的小区数，则可以排查 MR 数据采集侧的原因，解决后也可以提升 XDR 数据和 MR 数据的关联准确率。

在基站系统方面，需要注意系统上报的 MME_UES1APID 不能为短码，因为关联时默认短码的 MME_UES1APID 无效，这些短码的数据将会被清洗掉。如果发现了短码的存在，可以统计短码数据的占比，并查看采集侧的配置，查找原因并进行问题的解决。

XDR 数据和 MR 数据关联之前，会先进行数

据清洗工作，清洗掉无效数据，比如小区号为空的数据、开始时间/结束时间不合理的数据以及 MME_UES1APID 短码的数据等，所以加强数据采集传输过程的监管，保证采集配置无误，数据不丢失，从而提升数据质量，是提升数据关联成功率的主要方法。

3 指纹库定位算法

手机终端在每一个位置都会观测到一组小区的信号强度样本值，并且在不同位置所接收到的这组小区的 RSRP 样本值又不相同，这种信息的集合，定义为指纹库。指纹库中的数据来源包括 CQT 拨打测试数据、DT 路测数据、MDT 数据以及 OTT 数据等。由于采集的局限性，指纹库中的数据无法覆盖所有的区域，所以需要利用这些准确的数据来训练传播模型，获得最逼近真实无线环境的传播模型，利用该模型计算所有栅格里面各小区的强度，然后补充到指纹库中，使得指纹库中的数据可以覆盖全网区域。利用 k NN (k -nearest neighbor) 算法，根据 MR 报告中的主服务小区 RSRP 和邻区 RSRP 匹配指纹库中的数据，即可得到 MR 上报时的位置信息，实现 MR 数据的定位。

指纹库定位算法的详细过程描述如下。

(1) 将原有 MRO 数据中的无效内容去掉，只提取关键的站址 ID、MME UE S1AP ID、TA、频点、PCI、RSRP 等信息，同时考虑结合 MME UE S1AP ID 和 TA 数据，通过对多条 MR 数据进行统一分析、滤波，排除不符合要求的错误上报数据。

(2) 通过站址 ID 和 PCI 信息，对比工参，将复用的 PCI 转为全网唯一的小区 ID。

(3) 通过最小化欧氏距离算法将 MR 定位到指纹库对应栅格。对于一个待定位的 MR 记录，首先确定它的本小区和邻小区的信号强度，然后在该 MR 归属主服务小区的指纹栅格集合中，查找特征信息与当前 MR 中包含的特征信息最接近

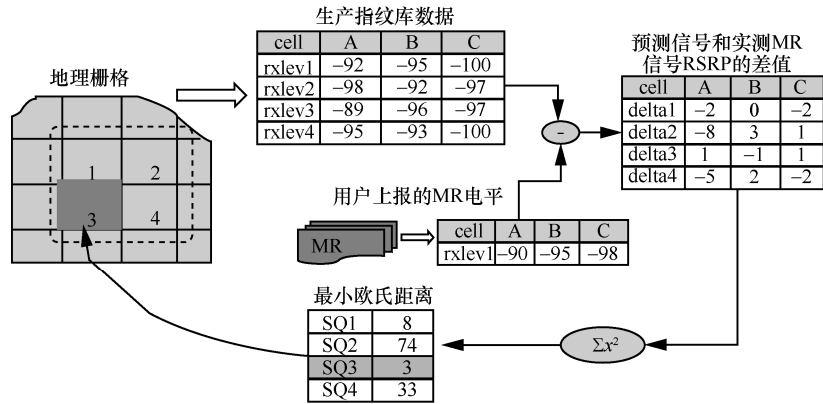


图4 指纹库定位算法流程

的栅格，最终以此栅格的位置作为该MR的位置。即：以主服务小区为键key值搜索，仅提取满足主服务小区匹配的指纹库；假设对于第*i*条MRO数据，包含*K*个主邻小区ID，对应的参考信号功率记为 $RSRP_{i,k}$ ，选取第*j*条指纹库信息中的*K*个主邻小区进行比较，对应的参考信号功率记为 $RSRP_{j,k}$ ，则第*i*条MRO数据与第*j*条指纹库信息的归一化欧氏距离匹配可以表示为：

$$d_{i,j} = \sqrt{\sum_{k=1}^K (RSRP_{i,k} - RSRP_{j,k})^2} \quad (1)$$

那么，将指纹库序号的全集记为*J*，则第*i*条MRO数据匹配到第*s*条指纹库的最小化归一化欧氏距离匹配算法可以表示为：

$$s = \arg \min_{j \in J} d_{ij} \quad (2)$$

可以认为第*i*条MRO数据匹配到第*s*条指纹库对应的栅格，将数据整理之后加上经纬度信息输出，具体流程如图4所示。

对于室内室外用户可以通过用户的位置变化、移动速率、主服务小区变化率、RSRP水平等多个维度进行判决，赋予每条MR室内外属性，基本算法流程如图5所示。

(1) 对上报的MR数据以MME UES1APID为单位进行统计，对同一MME UES1APID的MR数据，若主服务小区切换符合主小区切换特征（依据该用户前后若干时刻所在小区是否变化以及变

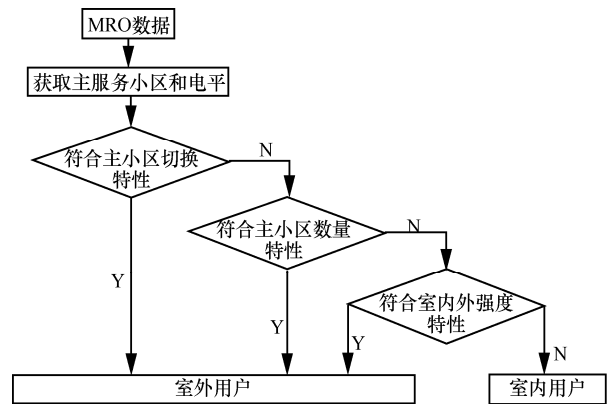


图5 室内外用户判决

化快慢），则定位为室外用户，否则进入下一步。

(2) 对同一MME UE S1AP ID的MR数据，若符合总小区数量特征（依据室内外交收到的邻区数量的差别），则分类为室外用户，否则进入下一步。

(3) 对同一MME UE S1AP ID的MR数据，若主服务小区电平符合室内外电平强度特征（依据室内外交收到的服务小区及邻区的电平强度的差别），则分类为室外用户，否则为室内用户。

4 XDR与MR关联数据的应用

XDR数据与MR数据关联后，可基于专家及实践的优化分析经验建立用户位置库，进行用户特征及覆盖场景的多维融合分析，实现从覆盖、感知、价值、竞对等多方面构建大数据多维分析体系，基于大数据及机器学习打造MR和XDR关



联的智能分析手段，支撑优化、规划、投诉、维护等多维生产。

4.1 XDR 与 MR 关联的室内应用

如图 6 所示，利用 MR 数据特征，与 XDR 多种定位技术结合，实现室内外精确定位特征指纹库、针对需要定位的 MR 数据采用模式匹配定位法实现精准定位，为优化分析、规划建设、市场营销、客户感知分析等提供基础支撑。

通过基于数据挖掘的数据清洗、特征识别的室内外用户区分、多种精确定位技术、模型校正过程实现特征指纹库，如图 7 所示。

如图 8 所示，展示了从室内用户分类、室内用户业务感知得分、精准规划建站、深度覆盖不足问题点、家庭宽带需求客户 5 个方面，以图表形式直观呈现现网状况，层层下钻，快速发现问

题，提出解决建议。

4.2 XDR 与 MR 数据关联后在建维的应用

XDR 数据与 MR 数据关联之后，可以将 XDR 的感知信息与 MR 数据中的位置信息相关联，从而可以实现端到端的应用分析，实现全面的智能诊断。

以投诉分析为例，结合关联后的数据，可以实现面向全网任意用户准实时地进行分析诊断，提供用户级、小区级的端到端感知指标和无线性能指标，判断用户感知问题点，提供诊断建议，用于用户投诉处理和感知保障。

在大数据融合平台可视化界面部分的投诉分析模块，可以通过输入投诉用户的 MSISDN，查询到与该投诉用户相关的 XDR 记录、XDR 详细信息、MR 记录、MR 详细信息以及用户的运动轨迹和感知等信息。图 9 展示了用户的运动轨迹和

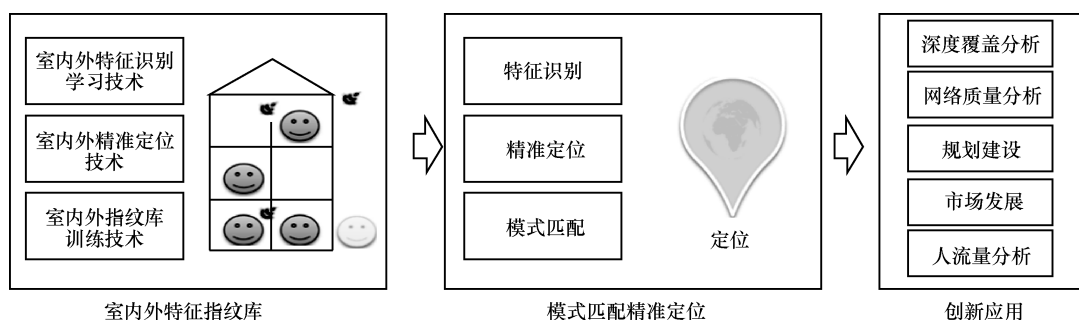


图 6 室内定位应用方向

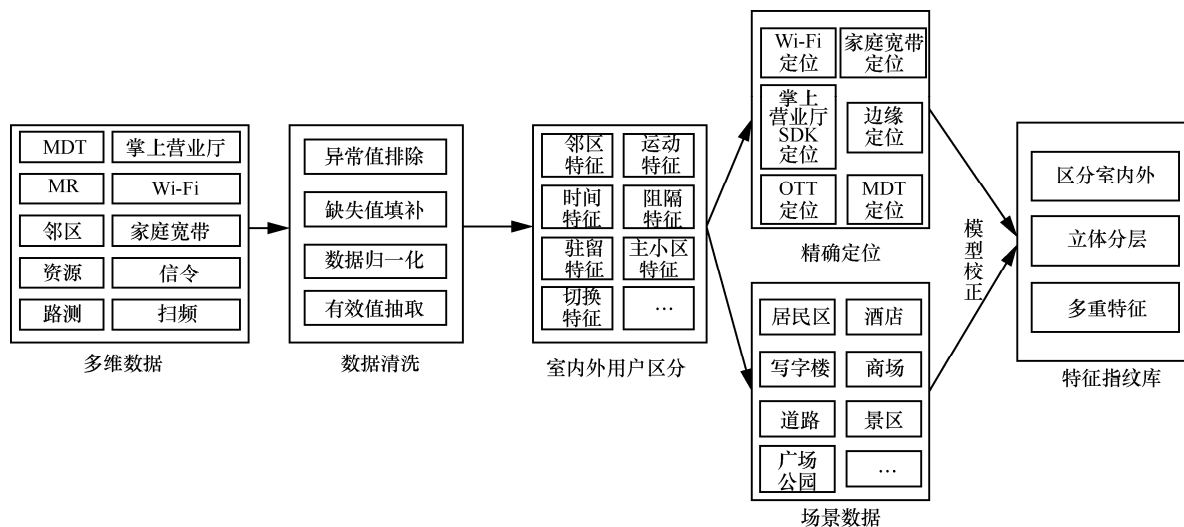


图 7 特征指纹实现过程



图8 室内应用案例

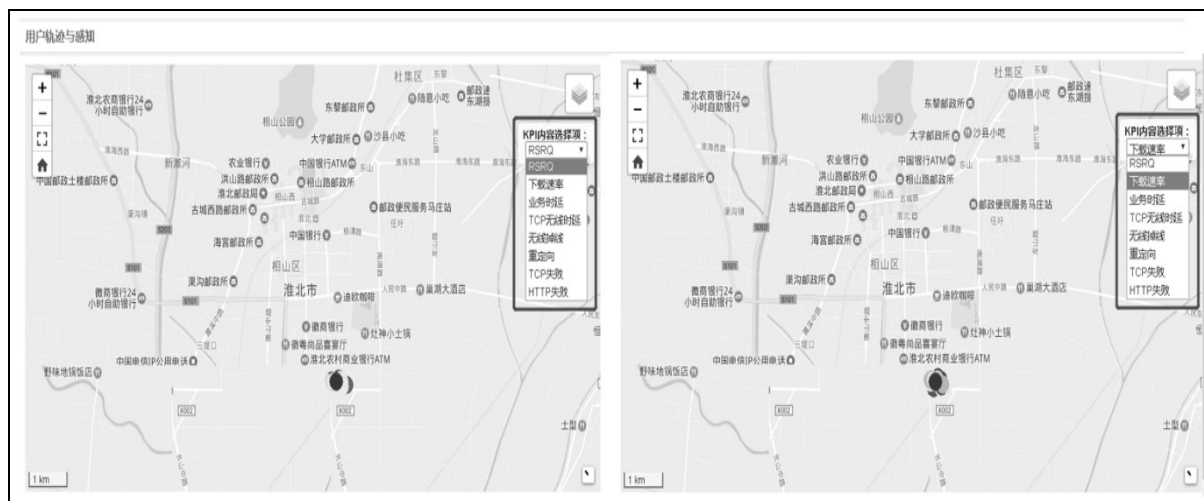


图9 用户轨迹与感知信息

感知, 结合这些信息, 可以回溯到用户投诉时正在进行的业务, 并界面化呈现用户的移动轨迹, 直观地呈现用户当时的感知信息、覆盖信息, 从而综合分析用户的感知和覆盖情况, 实现快速的问题定位, 达到快速响应用户投诉、解决网络问题、提升服务质量的效果。

4.3 XDR 与 MR 数据关联的市场线应用

XDR 数据与 MR 数据关联之后, 可以将 XDR

的用户内容信息与 MR 数据中的位置信息相关联, 从而实现用户群体的行为分析, 支撑市场活动。

4.3.1 应用一

根据用户的网络行为及内容信息, 实现用户的精准化细分; 根据 MR 的位置信息, 针对不同场景下的重点用户, 实现针对内容加区域位置的自动聚合, 为市场定制化业务和针对性推广套餐等业务提供大数据支撑, 应用的步骤如下。



步骤 1 对 LTE 典型高带宽需求业务（视频、音频、下载类业务）的分组大小流量进行量化分析，保证用户和业务的统计渗透率，确定大小分组业务过滤阈值，为大小分组特征业务分钟并发用户数与视频业务速率关联提供有效依据。

步骤 2 根据用户的网络行为及内容，区分视频用户、直播用户、网页用户等，根据 MR 的位置区域，利用聚合半径，实现细分用户区域的自动聚合。

步骤 3 利用 XDR 数据与 MR 数据，实现用户消费及套餐能力的综合评估，给出市场套餐及定制化的业务建议。

手机直播用户，根据 XDR 数据发现抖音及斗鱼等直播软件每天在线时间大于 3 h 的用户，利用 MR 的数据实现该用户的常驻区域，业务量统计抖音日均上行 3.25 GB，斗鱼日均上行流量 1.89 GB，该类用户主要使用抖音与斗鱼直播上行业务需求高，发现聚集在某大型居民区，针对这部分用户，引导市场电话和现场给客户推荐高流量的套餐。

4.3.2 应用二

根据用户的网络行为及内容信息，叠加互联网的营业厅的信息，建立营业厅覆盖范围智能评

估模型，根据覆盖的范围，集合客户的常驻地址，给出营业厅选址的精准规划，应用架构如图 10 所示。

满足安徽联通市场营销、网络运营等大数据分析需求，建立基于网络行为数据、位置定位数据的大数据分析手段，以 App 呈现各种信息，并推荐营业厅选址建议，以大数据提升营销规划、市场营销决策效率及精准性。

步骤如下。

步骤 1 构建以区域、栅格、用户为主的数据模型，其中栅格维度以栅格特征识别为主，用户维度如图 11 所示。

步骤 2 利用 XDR+MR 实现业务+经纬度信息回填，经过预处理、服务封装、业务封装等操作，对外发布脱敏后的位置洞察类数据服务，支撑业务营销、网络规划。

步骤 3 从常驻区类型扩展、识别精度提升两个方面入手，提升用户常驻区识别能力，为新业务产品、数据变现类产品的精准推送提供基础支撑。

步骤 4 系统按照营业厅选址的模型要求，对照区域、客户分析的数据特征，基于已有营业厅，输出选址建议列表，应用系统如图 12 所示。

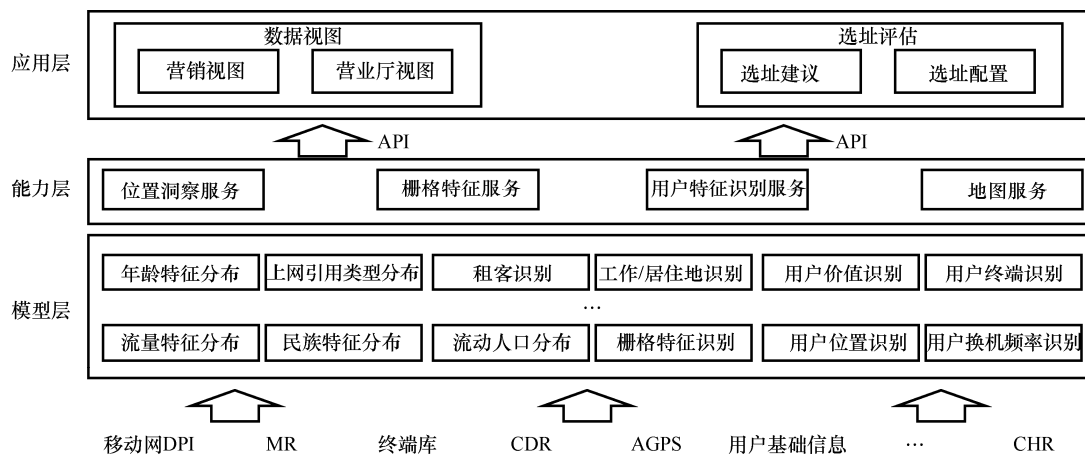


图 10 选址规划应用架构



图 11 用户与栅格维度数据

4.3.3 应用三

基于 XDR+MR 数据对接经营分析数据、资源管理数据、网络优化及投诉等数据，从居民区及用户维度开展分析，为 4G 深度覆盖规划、家庭宽带/4G 套餐精确营销、异网用户策反、带宽提升或光改提供支撑建议，如图 13 所示。

5 结束语

大数据在运营商的运营体系中已经起到了关键性作用，基于数据的融合能够打破数据壁垒，实现更多的数据碰撞，衍生百花齐放的应用。面对未来的 BI，能够依托现有大数据平台+AI，将网络中海量的 XDR 和 MR 数据关联融合并实现共享，基于指纹定位算法将从小区级的运营运维放大到用户级，从用户感知出发，完成一整套大数据处理建模、开发，从常规业务、用户、覆盖、容量、质量及 3G/4G 等维度进行数据深度关联处理，支撑面向规划、面向网络、面向客户、面向市场的相关工作中，是增效降本的有效方法，同



图 12 营业厅选址输出

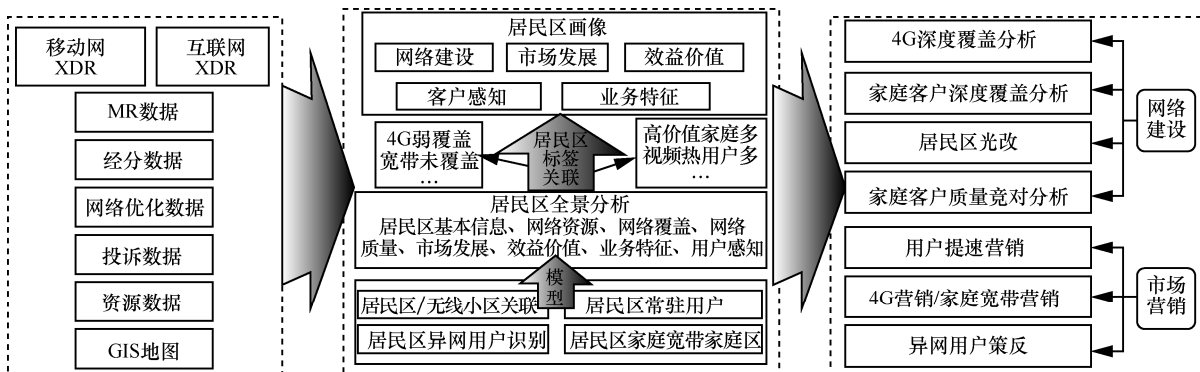


图 13 基于居民区及用户维度的应用



时将精准的用户信息应用到交通、公安等领域，也是未来运营商 ICT 发展的方向。

参考文献:

- [1] 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.
WEI J W, ZHANG Y Y. On big data based signaling monitoring platform[J]. Designing Techniques of Posts and Telecommunications, 2014(7): 47-52.
- [2] 王宜科, 梁松柏, 魏宁, 等. 基于MR的移动无线网络规划的可靠性分析[J]. 邮电设计技术, 2014(7):53-57.
WANG Y K, LIANG S B, WEI N, et al. Reliability analysis of mobile wireless network planning based on MR[J]. Designing Techniques of Posts and Telecommunications, 2014(7): 53-57.
- [3] 杨谦, 施文政, 林雨浓, 等. 基于MR定位技术的客服支撑综合应用系统的开发及应用[J]. 云南大学学报(自然科学版), 2011, 33(S2):264-268.
YANG Q, SHI W Z, LIN Y N, et al. Customer support integrated application system development and applications based on MR positioning technology[J]. Journal of Yunnan University(Natural Sciences), 2011, 33(S2): 264-268.
- [4] 周波. 基于Apache Flume的MR数据采集实现方案[J]. 电信科学, 2018, 34(Z1): 216-221.
ZHOU B. An implementation scheme for MR acquisition based on Apache Flume[J]. Telecommunications Science, 2018, 34(Z1): 216-221.
- [5] 叶树祥. 基于 XDR 大数据的高铁用户精准识别算法[J]. 信息通信, 2018, 186(6): 231-232.
YE S X. An algorithms of high-speed rail customer accurate recognition based on big data[J]. Information & Communications, 2018, 186(6): 231-232.
- [6] 谭晖, 王浩君, 席晨. 一种基于 DPI 数据的家宽投诉精细分析系统[J]. 信息通信, 2018(1): 254-255.
TAN H, WANG H J, XI C. An home broadband complaint analysis system based on DPI data[J]. Information & Communications, 2018(1): 254-255.
- [7] 杨岚兰, 石磊. 大数据技术及其在通信领域的应用[J]. 信息与电脑 (理论版), 2015(15): 68~69.
YANG L L, SHI L. Big data technology and application in telecom industry[J]. China Computer & Communication, 2015(15): 68-69.
- [8] 赵云山, 刘焕焕. 大数据技术在电力行业的应用研究[J]. 电信科学, 2014, 30(1): 57-62.
ZHAO Y S, LIU H H. Research on application of big data technique in electricity power industry[J]. Telecommunications Science, 2014, 30(1): 57-62.
- [9] 堵雯曦. 大数据技术在通信行业的应用探讨[J]. 江苏通信, 2015(3): 52-55, 59.
DU W X. Discussion on the application of big data in telecom industry[J]. Jiangsu Communication, 2015(3): 52-55, 59.
- [10] 张延彬. 基于移动通信行业的大数据服务研究[J]. 电信工程技术与标准化, 2016, 29(2): 44-47.
ZHANG Y B. Study on the service of mobile communication

industry based on big data[J]. Telecom Engineering Technics and Standardization, 2016, 29(2): 44-47.

- [11] 潘洋. 通信行业基于大数据的精确化服务应用研究[J]. 中国新通信, 2017, 19(6): 123.
PAN Y. Research on precision service application based on big data in telecom industry[J]. China New Telecommunications, 2017, 19(6): 123.
- [12] 王娟. 基于大数据的通信运营商商业化模型及应用研究[D]. 兰州: 兰州交通大学, 2016.
WANG J. Research on business model and application of telecommunications operators based on big data[D]. Lanzhou: Lanzhou Jiaotong University, 2016.
- [13] KIM N, JUNG K H, YONG S K, et al. Uniformly subsampled ensemble (USE) for churn management: theory and implementation[J]. Expert Systems with Applications, 2012, 39(15): 1183911845.
- [14] COUSSEMENT K, POEL D. Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques[J]. Expert Systems with Applications, 2008, 34(1): 313-327.
- [15] ALMANA A M, AKSOY M S, ALZHRANI R. A survey on data mining techniques in customer churn analysis for telecom industry[J]. Journal of Engineering Research and Applications, 2014, 4(5): 165-171.
- [16] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[作者简介]



刘韬 (1976—), 男, 中国联合网络通信有限公司安徽分公司数据分析师, 主要负责安徽联通大数据平台架构及搭建、B/O 域基础数据融合、数据分析、行业大数据应用研究等工作。



吴涛 (1973—), 男, 中国联合网络通信有限公司安徽分公司副总经理、大数据项目总负责人、数据分析顾问, 主要研究方向为数据分析与用户感知提升、精准化营销分析与数据挖掘、用户消费行为。



王斌 (1975—), 男, 中国联合网络通信有限公司安徽分公司网络发展部副总经理, 主要研究方向为安徽联通大数据平台架构、数据分析、用户消费行为数据分析、行业大数据应用等。