电信运营商的数据资产综述：数据、内联及外延

Overview of Telecom Operator's Data Assets: Data, Internality and Application

宋春涛，张　帆，王　勇，叶海纳（中国联通网络技术研究院，北京 100048）
Song Chuntao, Zhang Fan, Wang Yong, Ye Haina (China Unicom Network Technology Research Institute, Beijing 100048, China)

摘　要：
围绕电信运营商移动网络的数据资产进行梳理，同时讨论典型数据之间的内在联系，并制定出一种数据融合方案，进而研究数据应用的场景化驱动，给出了一种场景化应用的数据视图构建方法，并提出了基于场景化驱动的知识图谱构建方法，最后探讨了几种典型的数据外延性应用模式，为5G智能化时代的大数据应用提供一些思考。

关键词：
运营商；数据资产；内联性；数据视图；外延性
doi：10.12045/j.issn.1007-3043.2019.09.005
中图分类号：TP274
文献标识码：A
文章编号：1007-3043(2019)09-0020-05

Abstract:
It combs the data assets of mobile network of telecom operators, discusses the relationship among typical datas, and formulates a data fusion scheme. Then it studies the scene-driven data application, gives a method of constructing data view of scene-based application, and puts forward a scene-driven data view. Finally, several typical data extension application modes are discussed to provide some thoughts on the application of big data in the era of 5G intelligence.

Keywords:
Telecom operator; Data assets; Internal relationship; Data view; Extensibility

引用格式：宋春涛，张帆，王勇，等. 电信运营商的数据资产综述：数据、内联及外延[J]. 邮电设计技术，2019(9)：20-24.

0 引言

数据是电信运营商的金矿。电信运营商积累了丰富海量的数据资源，大数据金矿价值不断凸显，运营商正在逐步践行由通信服务运营向数据运营的必然转变。随着新兴业务及模式的不断涌现，运营商的“管道化”趋势日益显现，为了应对挑战，运营商应充分挖掘自己独特的全类型数据的价值。

全面理解电信运营商的各类网络运行及业务运营数据，梳理其内在的血缘及关联，形成具备运营能力的基础层，向构建支持面向场景的外延性应用数据视图，实现数据资产的价值映射，是运营商在数据技术方面的主要工作。

收稿日期：2019-06-06

本文主要围绕运营商移动网络的O域数据资产进行梳理，同时讨论典型数据之间的内在血缘及联系，进而研究数据应用的场景化驱动。

1 运营商的典型数据概览

电信运营商所掌握的数据具备基础性、多样性、全面性等特征，不仅包含着网络运行数据，同时包含着海量的运营数据，这些数据涵盖了移动网络从终端到接入网、从传输网到核心网、从业务平台到移动互联网各个网络及环节，基于多样化的采集方式，能获得海量的运行数据及业务数据，基于大数据分析方法和手段，结合移动互联、物联网等发展方向及应用走向，能够针对网络、用户、业务、终端进行深层次的数据挖掘和应用，本文重点关注运营商的移动网络数据。

总体上说，电信运营商的数据资源主要包括网络运行数据和业务运营数据两大类，即OSS域数据及BSS域数据，同时还包括M域、D域及E域等，这些数据均源于移动网络运行和业务运营，在两类数据之下则是纷繁复杂、形态多样的海量数据集合，具体数据内容存在巨大差异，但却有着内在联系，形成内联外延、纵合横通的多维数据魔方态，如图1所示。本文重点关注电信运营商OSS域数据。

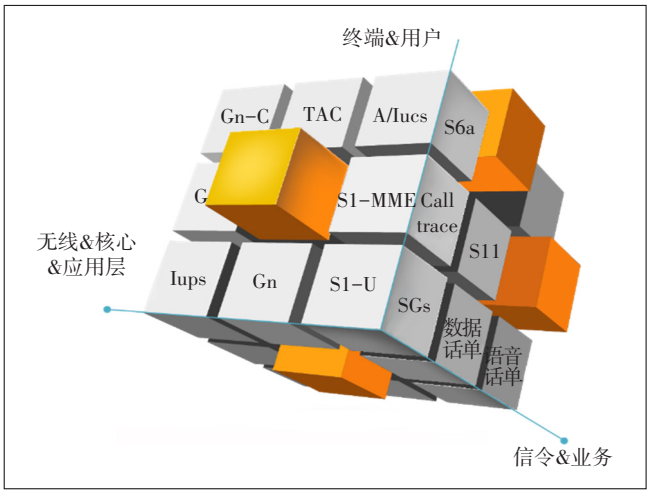

图1　移动运营商的多维数据魔方示意图

笔者总结典型的运营商移动网络数据洞察维度有以下几类：按网络制式划分的2G/3G/4G/5G网络及物联网等维度，按数据所属范畴划分的网络侧、用户侧及业务侧等维度，按网络分层划分的接入层、网络层、应用层等维度，按业务运行不同功能划分的控制面及用户面等维度。

上述不同维度的数据洞察，将移动网络数据划分成多维组合的数据魔方。为了便于理解，下面针对移动网络，从网络/用户/业务维度，结合数据采集方式、内容、所属域进行简单归类，移动网络数据主要分为网络域数据、用户域数据及业务域数据3部分。

a）网络域数据主要是与网络运行及维护相关联的数据，既包括与网络基础配置、运行监测、网络质量评估相关的数据类型，如基础资源及配置数据、性能统计数据、DT/CQT等测试数据、监控预警等，也包括体现用户与网络之间信令交互部分的数据，如移动性管理、移动业务管理、业务交互及保障等方面数据。

b）用户域数据则主要是用户与网络交互以及进行业务产生的各类记录数据，如用户在网运行的信令交互数据以及用户进行业务过程中产生的业务记录。

c）业务域数据主要是指用户在网享受服务时产生的业务记录数据，这也是在数据业务之外产生的价值数据。基于特定的DPI技术可以针对这些业务数据进行更加丰富的信息和特征挖掘。业务域数据是移动互联时代的特有数据。

2 移动网络数据的内联性

移动网络运行过程获取的各类数据之间存在必然的内联性。虽然这些数据来自于不同的网元、接口及系统，在网络运行中发挥着不同的作用，但是各类数据并不是孤立的，而是相互关联、相互解释，甚至相互影响的。

2.1 移动网络数据的内联性

要了解移动网络数据的内联性，首先必须梳理典型的可获取数据以及这些数据的关键特征和聚合维度。笔者将移动网络运行及大数据应用研究中常用的典型数据及特征做了整理，具体如表1所示。

表1　移动网络数据的典型数据类型及特征标签

特征标签	典型数据类型							
	工参	CM等参数类	话统	Counter	MR	DT/CQT	话单类	Call-Trace类
时间类	√	√	√	√	√	√	√	√
号码类					√	√	√	√
终端标识类					√	√	√	√
身份标识类					√	√	√	√
网元标识类	√	√	√	√	√	√	√	√
业务类型类				√	√	√	√	√
IP地址类		√				√	√	
URI类						√	√	
地理类	√	√	√	√	√	√	√	√
原因类			√	√	√	√	√	√
业务量类			√	√		√	√	√

移动网络运行数据的基本特征主要包括时间、地点、网元及身份标识、记录类型、记录业务量，这些特征组合也成为各维度数据相互关联的基础信息，可以作为数据融合的关键特征。

其中话单类数据种类繁多，包括语音、数据业务类、控制信令类等，有着巨大的挖掘价值。典型话单列表如表2所示。

表2是不同业务、不同信令节点、不同接口及不同专项的典型话单，这些话单中的海量信息产生并服务于网络运行及业务运营，同时也作为衍生物为大数据分析及挖掘提供海量价值资源。

表2　移动网络数据的典型话单列表

序号	话单名称	接口及话单类型
1	语音呼叫/SMS/LAU/切换/Paging	A/Iu-CS
2	附着&去附着/LAU/重定位/寻呼及业务请求/PDP激活&去激活&修改/Iu释放	Gb/Iu-PS
3	通用业务话单	Gn/S1-U用户面话单
4	特定协议话单（HTTP/HTTPS/DNS/RTSP/RTP/RTCP/VoIP/P2P/MMS/EMAIL/FTP/IM/STREAMING等）	
5	附着&去附着/TAU/业务请求/PDN连接&去连接	S1-MME信令话单
6	UE请求承载资源分配&修改/网络发起的EPS承载上下文激活&去激活&修改	
7	切换/上下文释放/S1关联/UE能力	
8	CSFB话单	专项话单
9	4G短信话单	
10	用户小区轨迹话单	

目前运营商基本完成了上述类型数据的采集，并将其应用于自身运营分析来支撑市场前端、网络运行及对外价值应用。相比传统的网络KPI分析，目前的应用已经开始关注更多其他的可能应用，如DPI深度解析、LBS计算、APP QoS保障、用户/终端识别等。但是，随着多样化场景及业务需求的爆发，针对用户、业务及终端的挖掘层次逐步加深，单一维度及表单的挖掘已经难以满足需求，数据融合成为运营商大数据发展的必由之路。

2.2 移动网络数据的融合

移动网络数据融合，不仅仅是按照时间、网元ID、用户ID等进行相互关联，还涉及到数据血缘梳理、数据清洗、数据转换、结构化、数据分级、数据校准、数据关联等。数据的关联和融合是一个系统性工程，虽然移动运营商数据大多为结构化数据，但仍需要诸多步骤来实现数据的真正融合。

笔者结合实际工作总结移动网络数据融合的步骤如下。

a）数据采集及汇聚：主要是将移动网络各系统、平台、接口及测试等数据进行采集并汇聚，比如目前某运营商的OSS系统将全国各本地网的各类数据进行采集并通过北向接口实现统一的ESB汇聚，供上层应用使用，采集、解析及汇聚非本文重点讨论内容，可参考相关技术文档。

b）数据清洗：目前移动网络数据受限于采集、解析及传输等能力，尚且存在一些质量问题，如采集范围及能力不足带来的数据缺失、疏密度不同、数据值异常、多源数据冲突等，需要针对这些问题，借助ETL、机器学习、异构模型等进行清洗与比对、滤重、映射与互补等，提升原始数据质量。

c）数据分级：因为移动网络原始数据涉及网元、实体、接口等较多，与之相对应维度的数据及表单也较多，同时在某些表单中实际包括多个层级和维度的数据，如A/Iu-CS口话单生成是按照一次语音呼叫的全流程生成的记录，这其中既包括该次呼叫各信令节点信息（时间、事件、时延、成败、掉话等）、又包含不同等级实体（如网络、小区及上联网元、用户等），这些数据实际上是以一次呼叫关联起来的多层数据，在实际应用中则需要结合场景化需求进行数据分级，如分解为用户级、小区级、事件级等。数据分级是为了满足面向应用的场景化需求，也是后面进行域变换的数据基础。

d）域变换：从网元、实体、接口等观察分析数据，主要是基于传统网络运行及维护工作需求，随着数据价值的不断发掘及市场前端部门需求的不断增加，当前维度已经远不能满足需求，必须针对新的前端及外部应用需求进行数据变换，因此有必要将网元、实体、接口域数据变换至用户、终端及业务域。变换方法主要是从既有分类话单中按照用户、终端、业务、网络等维度进行数据梳理及抽取，如将A/Iu-CS呼叫话单按用户进行数据抽取，形成用户域中的呼叫话单，将S1-MME信令话单按终端进行抽取获得终端域中的信令话单，将S1-HTTP按业务大小类进行抽取，获得业务域中的数据业务话单。域转换中的数据抽取不是简单的维度变换和数据拷贝，而是基于场景化需求和数据逻辑性构建，会涉及大量的抽取模型和数据处理工作，此处不再详述。

e）数据关联：完成域变换之后，在同一域内仍然存在大量并行的数据信息及表单，这些并行数据从不同角度对该域主体进行属性和行为的解释，但分散的形态无法满足上层应用和非专业人士的引用，必须进行数据关联，这也是数据融合的重要一步。数据融合的基本方法仍是以组合关键标识的方式进行，如时间点（段）、用户标识、用户网络标识（ID）、终端标识、小区标识、业务标识等，在不同数据维度和层级上进行数据关联操作。数据关联实际上是对数据之间的血缘、归属、解释、行为关系进行梳理，并形成相应的关系拓扑，而关联既可以是逻辑上的，也可以是数据库中的实体关联。

f）数据视图：海量及庞杂的运营商数据能够满足

众多专业的内外部应用需求，但是在实际使用时，则因为专业知识及需求差异，需要针对不同部门、行业、场景制定便于其理解和引用的数据视图集合，以灵活适配多样化的应用场景。

2.3 移动网络数据视图

数据视图作为面向不同专业及行业应用的可理解数据集，是由核心业务的主题驱动的，基于业务诉求来构建的面向需求，而非面向数据域的主题数据集合。

构建数据视图的关键在于主题，不同的主题应用需求决定视图的数据内容，主要基于用户、业务、网络、终端等大维度，如常见的视图有自然人视图、政企视图、渠道视图、小区视图、产品视图、终端视图等（见图2），这些视图向下来源于运营商的各类底层数据，向上满足不同部门和专业的业务需求。

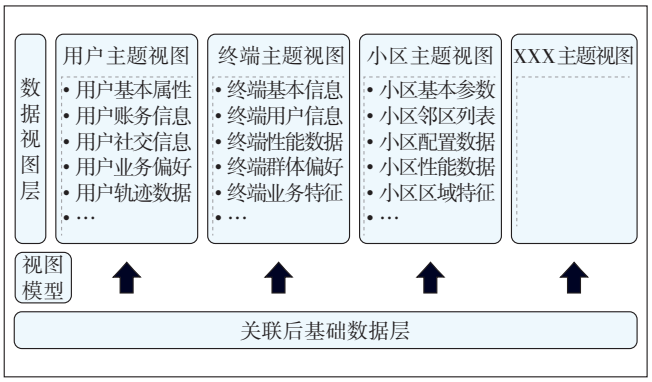

图2　移动网络数据的视图层构建

笔者认为移动网络数据视图层是运营商数据体系的重要中台层，这一层的型特征如下。

a）完成了底层数据的解析、域变换及数据聚合，即面向不同专业用户完成了数据解释。

b）针对某专业需求，无需掌握并遍历全集数据，仅需进入该专业相关视图即可。

c）数据视图作为数据资产流通及变现的基础，也为数据资产的生产和应用提供量化基础。

d）解决跨域数据无法打通、无法看全问题，有助于数据使用者、数据运营者及数据生产者的身份梳理和定位。

数据视图实际上是近年流行的数据中台的主要内容构成，中台则是视图的结构承载。数据视图的构建，在具体实现上可以是物理上的，也可以是逻辑上的，这需要根据实际的应用需求和资源条件来综合考量。

3 移动网络数据的外延性

数据资产的最终目的是应用，在完成中台级数据汇聚和视图构建后，数据尚未完成其价值的体现，数据价值的体现则主要是通过其数据的外延性应用来实现的。

3.1 移动网络数据的知识图谱

根据网元、网管、系统、接口等不同采集方式及传统应用方式的不同，数据被人为整理为小区、用户、终端、业务等维度，然而实际情况是，以业务为牵引的各单元通过有机的协作来实现业务运行及业务保障，在该过程中产生的数据是自然内联并难以分割的。典型的协作单元及其包含的内容如图3所示。

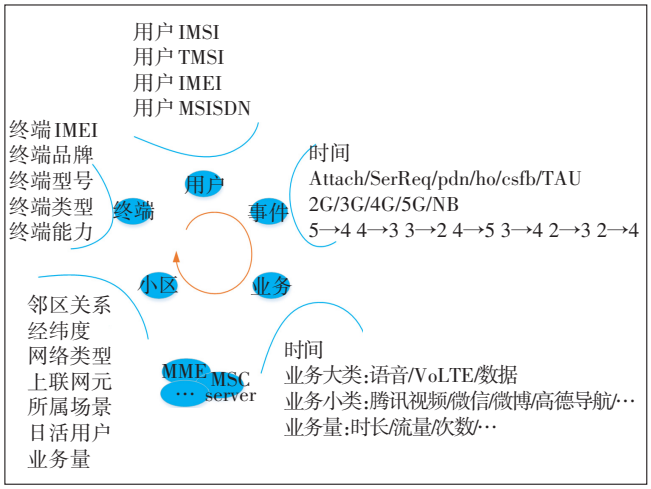

图3　移动网络数据的典型主体及内容示意图

在实际应用中，不同行业及产品对于运营商数据的需求是不同的，在保障数据内联完整性和灵活性的前提下，引入针对性的行业知识图谱是一个较好的解决方案。

从Google搜索开始，到现今的智能机器人、风险控制、证券投资、智能医疗、自适应教育、推荐系统等，知识图谱正在更广泛的方向上获得重视和应用，电信运营商在这方面也有相关的实践和应用，但目前主要是基于B域中用户的属性、资料、交往圈等抽取相应的实体和关系并构建知识图谱，典型应用方向为诈骗社区发现、风控、征信等领域。

不同于B域主要基于用户维度的视图和图谱构建，O域数据因为其网络侧特有的定位、业务识别、多类型终端、信令交互数据等，可以构建区别于B域的多类垂直行业及产业的知识图谱，以实现不同的场景及

产品需求，典型的行业应用有智慧城市规划、智慧交通、工业互联网、智慧医疗等。

移动网络数据的行业知识图谱的构建并不是必要的，同时也不是万能的，这需要依据实际产品和业务需求而构建，因此笔者将O域数据的行业知识图谱构建作为数据外延性的一种补充形式。

3.2 移动网络数据的外延应用

移动网络数据的外延性，实际上是在数据视图基础上更加精细化应用的一步，不同专业及部门的需求一般都是某一视图的子集，或基于某一视图的定制化数据提取，如图4所示。

典型的外延应用方式主要有以下3种：

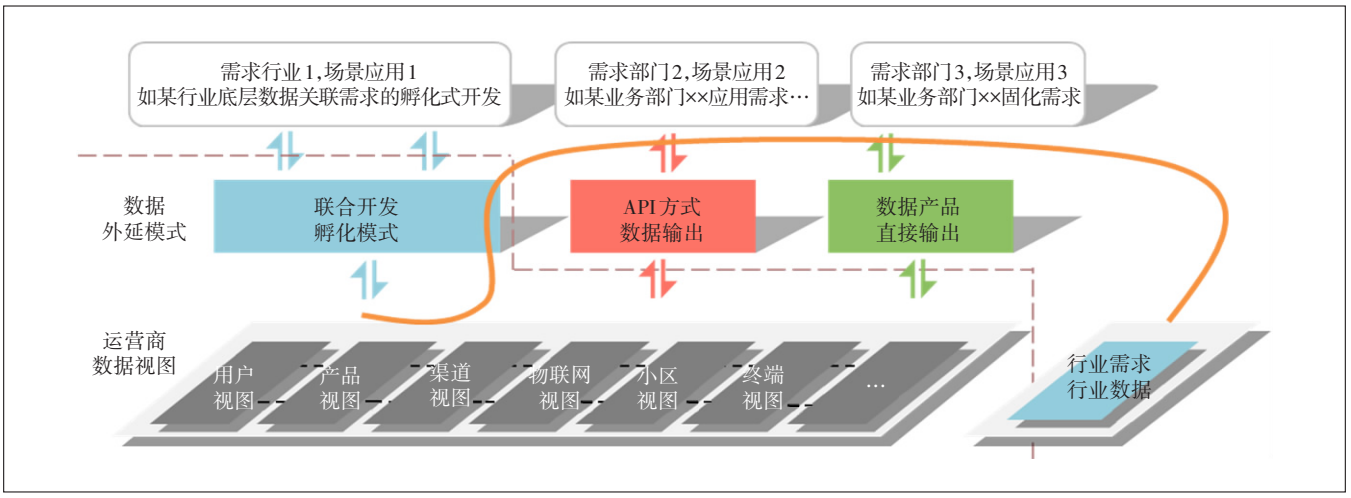

图4　移动网络数据外延性应用的典型模式

a）深层数据关联，运营商需导入行业数据，并与自身数据进行关联和整合，同时提供强大的孵化环境，以深度数据合作方式实现数据对外应用与价值变现。

b）以行业需求为驱动，运营商制造满足行业需求的中间层数据并以API等方式为行业应用赋能，实现数据的外延性应用。

c）以行业需求为驱动，运营商直接生产满足行业需求的结果性数据，以数据产品方式直接输出，实现数据的外延性应用。

以上3种典型的数据外延性应用适用于不同的应用场景及不同数据等级需求，能够在规避数据隐私的情况下，以灵活的方式满足多样化的行业需求，既能为中小型企业及前端部门提供数据能力，又能提供数据处理及应用的资源环境。

4 总结

电信运营商成为信息产业的基础设施提供者的大趋势逐渐显现，随着新兴业务及融合应用需求的不断涌现，不断挖掘自身具备的全类型数据的多样化价值，快速转变为数据运营商，并为信息产业的持续繁荣及智慧未来全面赋能，已经成为电信运营商责无旁贷的重要使命。因此本文围绕运营商移动网络的数据资产进行梳理，同时讨论典型数据之间的内在血缘及联系，进而研究数据应用的场景化驱动，并探讨了几种典型的数据外延性应用模式，为5G智能化时代的开启提供一些参考。

参考文献：

[1] 移动网络DPI信令采集设备技术规范：T22-306(2016)[S]. 北京：中国联通，2016.

[2] 塞西亚. LTE/LTE-Advanced-UMTS长期演进理论与实践[M]. 北京：人民邮电出版社，2012.

[3] 沈嘉. 3GPP长期演进(LTE)技术原理与系统设计[M]. 北京：人民邮电出版社，2008.

[4] E-UTRAN S1 Application Protocol (S1AP) (Release 13)：3GPP TS 36.413[S/OL]. [2019-05-06]. ftp://3gpp.org/Specs/.

[5] Radio Resource Control (RRC) Protocol specification (Release 11)：3GPP TS 25.331[S/OL]. [2019-05-06]. ftp://3gpp.org/Specs/.

作者简介：

宋春涛，高级工程师，硕士，主要从事移动网络大数据分析挖掘及应用研究等工作；张帆，工程师，硕士，主要从事移动网络大数据分析挖掘及应用、机器学习等工作；王勇，高级工程师，硕士，主要从事移动网络优化分析、端网协同研究及大数据分析挖掘等工作；叶海纳，高级工程师，博士，主要从事移动网络大数据分析挖掘及应用研究、数据产品业务需求分析及产品设计。