



智能分析和智能响应技术在移动投诉处理中的研究与应用

吴强, 许樱, 徐小雅

(中国电信股份有限公司上海分公司, 上海 200120)

摘要: 移动互联网和5G新技术的快速普及, 对用户响应速度的要求不断提高, 促使投诉处理向自动化、智能化方向发展。针对用户对移动通信服务的主要需求, 融合移动投诉智能分析和智能响应、移动用户定位、海量数据特定记录搜索、中间件、Hadoop与传统关系型数据库互操作、TB级用户呼叫记录实时采集、GIS、网优数据翻译等关键技术, 以用户需求为基础, 以业务和技术驱动为主线, 对大数据存储和挖掘、移动投诉智能分析和响应等关键技术进行基础研究和集成应用创新, 构建一个多业务、快速智能、可靠易用的移动投诉智能分析和智能响应系统, 支撑电信在移动投诉处理中的创新与规模化应用。

关键词: 5G; 大数据; 智能分析; 智能响应; 移动; 投诉

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025089

0 引言

移动互联网和新一代移动通信技术的快速普及, 使投诉随着用户数和数据量飞速增长, 对数据分析效率和分析准确性的要求都在提高, 这促使投诉处理向自动化、智能化方向发展。其核心是各种类型数据资源能面向需求和应用的融合、分析、搜索和应用。建立支持多业务、多服务手段的多制式、多维度综合投诉(预)处理服务应用, 能够带动电信企业用户服务提升、网络质量发展, 形成从资源、核心技术、平台到应用产业化的创新生态链, 具有较大的经济和社会效益, 主要解决了以下问题。

(1) 前后端联动, 创建面向多维数据服务手段的投诉处理技术体系。建立以资源库(数据、算法、规范等)为核心, 以中间件、数据挖掘等

关键技术为支撑, 以应用为纽带的具有多维综合数据服务能力的基本技术体系。基于资源开发集成了面向电信客服热线、智能终端客户端、投诉中心工位等多场景应用, 为用户提供快捷的投诉处理服务。

(2) 基于Hadoop的PaaS云计算架构、支持多维度投诉智能分析和智能响应集成研究。对投诉成因智能分析和响应、海量数据中特定记录的快速搜索、移动用户定位、中间件、GIS、App等技术进行集成和应用创新, 提升系统的智能响应能力、人机交互能力、多业务服务能力, 形成从核心技术、平台到应用产业化的创新生态链。

(3) 移动投诉智能分析和智能响应关键技术研究。针对覆盖语音和数据业务、华为中兴和朗讯等多厂商数据的智能分析, 研究不同数据的模

块化分析、多数据关联分析等关键技术，融合应用多数据调用关键技术，研究实现自动化、智能化的投诉成因判断。针对投诉地址描述模糊、客服和用户缺少网优专业知识的情况，研究模糊地址识别、网优数据翻译等关键技术。

(4) 移动用户定位关键技术研究。融合基础三角定位算法、自优化学习算法、大数据训练等关键技术，研究实现移动用户的高精度位置定位服务。

(5) 海量数据中特定记录搜索关键技术研究。对在千亿行数据中搜索用户唯一投诉呼叫记录进行研究，实现通过用户呼叫记录搜索规则、用户行为模式分析和用户定位分析等关键技术的数据搜索模型。

1 技术方案

以电信复杂无线环境下对用户投诉快速分析和解决的需求为背景，围绕自动化、智能化解决

用户日常投诉和移动互联网 App 投诉的需求，通过建立“无线投诉（预）处理系统”，形成自动化、智能化（预）处理移动投诉的相关算法和服务，并建立多维度的综合网络优化技术支撑体系，提升企业客服、市场、网优等部门服务用户的能力。移动投诉智能分析和智能响应系统的网络架构如图1所示。

系统处理的移动互联网数据和语音数据每天的数据量达 10 TB、百亿条。基于 Hadoop 架构，通过系统建设，扩展了多业务、多系统的数据处理需求和与关系型数据库互操作需求，实现了底层海量数据采集、存储和计算能力。制定大数据能力调用的统一接口技术规范，确保相关关键技术顺利实现与系统的集成；实现上层应用系统与底层 Hadoop 系统之间的松耦合，使一套大数据支撑平台可以支持多个上层应用系统的访问。系统大数据支撑能力从逻辑架构上可分为4个层次，如图2所示。

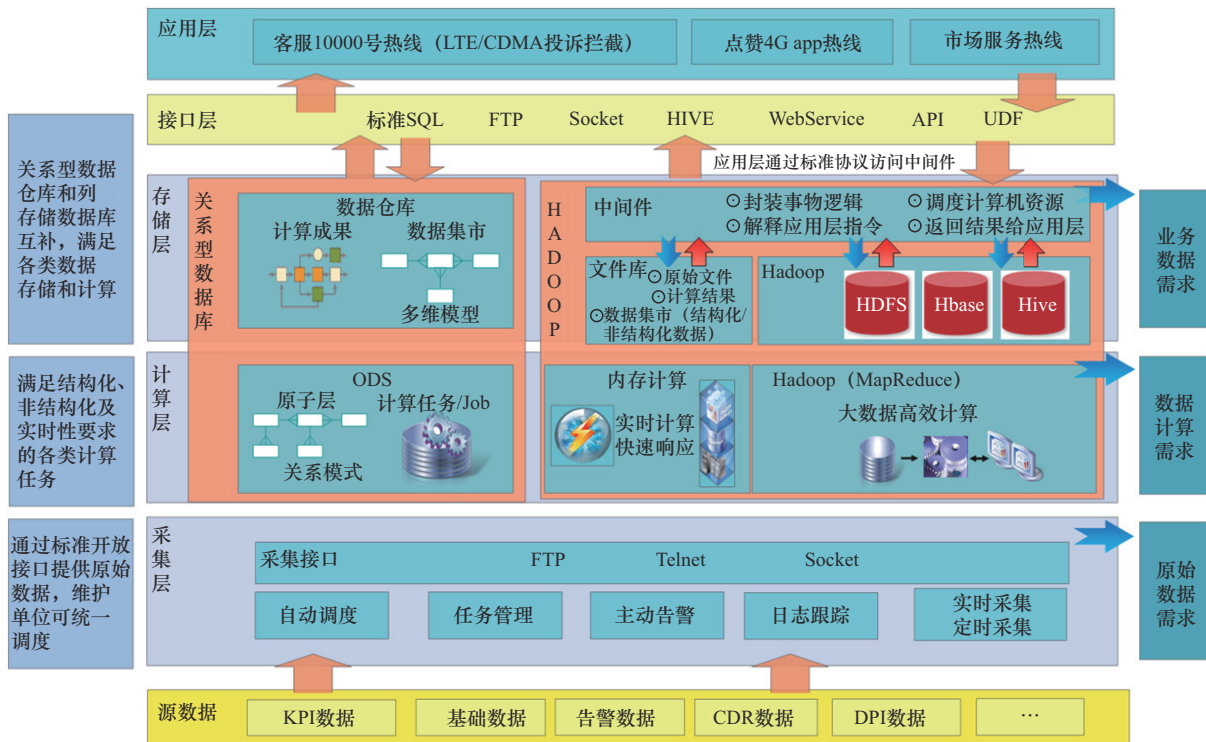


图1 移动投诉智能分析和智能响应系统的网络架构

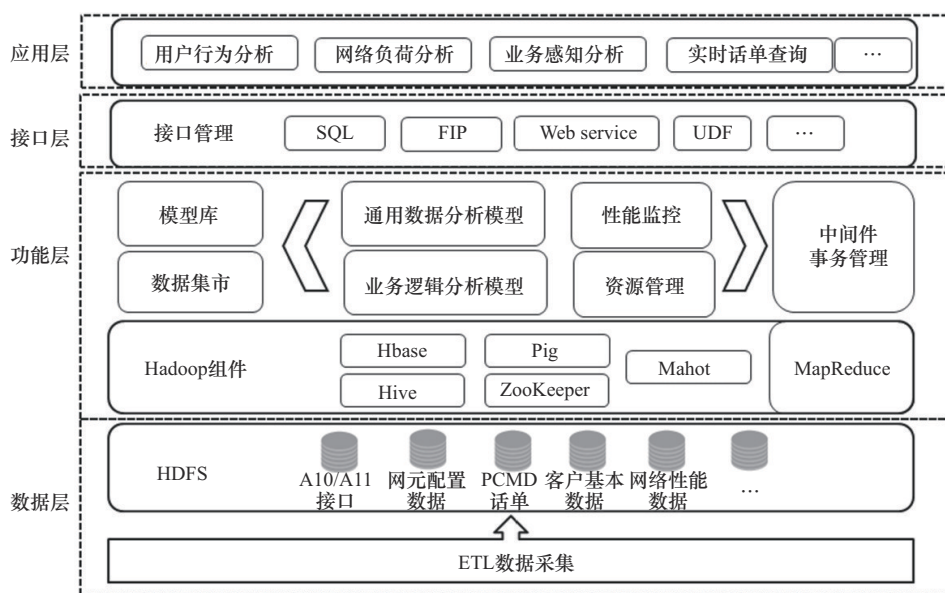


图2 大数据支撑系统架构

(1) 应用层

用户通过应用层使用系统的各个功能。10000号客服热线、电信点赞4G App 热线用户通过PC、智能手机终端等多种方式获得投诉处理信息服务,包括投诉原因查询、投诉解决方案、应答指导等信息服务。

(2) 接口层

平台提供主流的标准接口(消息接口、文件接口、HTTP接口等)支撑应用层调用底层大数据计算成果、支持Hadoop数据和传统关系型数据库数据的互调用实现智能分析等。

(3) 功能层

功能层主要接收接口层下发的大数据需求并完成数据的查询、计算和统计。功能层集成了系统最多的研究成果,包括移动投诉智能分析算法、移动用户通话记录定位算法、海量数据中特定记录搜索算法、中间件、可变字段日志型数据查询性能优化等关键技术,为系统的应用打下基础。

(4) 数据层

数据层主要负责源数据的采集、ETL清洗分类,并根据数据特点和业务需求分类存储在不同

的区域中。采集的数据每天达到10 TB级、百亿条,数据类型包括文本、消息、数据库格式文件等多种类型,采集层通过TB级用户通话记录实时采集等关键技术实现采集数据的完整性和低时延。采集程序还集成移动用户通话记录定位算法、海量数据中特定记录搜索算法等关键技术完成用户位置定位,减轻功能层数据计算负荷。

2 关键技术研究

无线投诉(预)处理系统涉及的数据资源种类、类型、范围、结构差异性较大,处理的数据量海量、信息的表达方式有文字、GIS、图表等多种形式;服务的对象有客服热线(人工/自动)、App、投诉中心工位等多种形式,对面向语音业务、数据业务的智能分析技术、移动用户定位技术及海量数据特定记录搜索以及海量数据的采集、存储、高性能计算和任务管理等技术的研究十分关键。通过相关技术的研究,可以转换各种类型的数据资源为可服务的数据资源,并且可以有效识别用户需求,为用户提供最适合的数据服务。

2.1 独特的中间件技术

Hadoop 是开源系统，在发展中新的组件层出不穷。但是电信应用系统对稳定性要求较高，不能经常变换接口。因此，本文创新性地提出中间件封装复杂接口技术解决该问题：对上层应用系统封装统一接口，如 HTTP 或 Web service；对底层 Hadoop 不同组件封装不同的 ODBC、JDBC、HQL、FTP 等接口。这样既满足了 Hadoop 组件快速发展的需求，及时升级组件或添加新组件功能，又保证了上层应用开发和使用的稳定性。

该技术有以下主要独特功能。

- (1) 复杂接口封装特性。
- (2) 性能保证特性：中间件可监控系统软/硬件，协调各组件的资源占用并监控主要性能资源的消耗，保障 Hadoop 性能达到最优。
- (3) 开放平台特性：通过 API 提供服务，标准接口可以极大降低上层应用系统开发的难度和时间。
- (4) 支持业务逻辑封装特性：对经常使用的通用应用逻辑计算任务进行封装，第三方应用系统可以直接调用，实现了不同应用共享业务逻辑计算模型。

中间件访问技术如图3所示。

中间件技术搭建了一个上层应用与 Hadoop

的桥梁，提供给上层应用系统功能非常清晰的标准接口，应用功能不必理解 Hadoop 各个组件的特点和使用方法，而中间件根据配置好的应用功能特征自动调用不同的 Hadoop 资源来满足应用的需求。

2.2 移动投诉智能分析算法关键技术点

随着移动互联网和新一代移动通信技术的快速普及，相应的投诉伴随用户数、数据量的飞速增长也在不断增长，通过自动化、智能化的数据分析，异常通话成因判断，可大量避免人为分析失误、提高投诉的分析效率。

移动投诉智能分析实现了移动投诉分析的自动化和智能化，能及时、准确地响应客服热线的投诉信息解答，保证快速解决投诉，提升用户感知度。移动投诉智能分析结构如图4所示。

无线投诉智能分析的总体思路是“号码发起、逐层判断、关联汇总”。算法以投诉号码触发分析，以用户呼叫记录为主线数据，对用户行为、网元状态进行用户、告警、参数、覆盖、干扰和邻区等方面的全面分析、关联判断，得到用户投诉原因，并可给出相应的解决方案。

项目研究解决了众多类型数据的自动化、关联化、智能化分析难题：用户 CFC/CFCQ/TCSI 判断、Release Count 判断、参数核查、邻区核

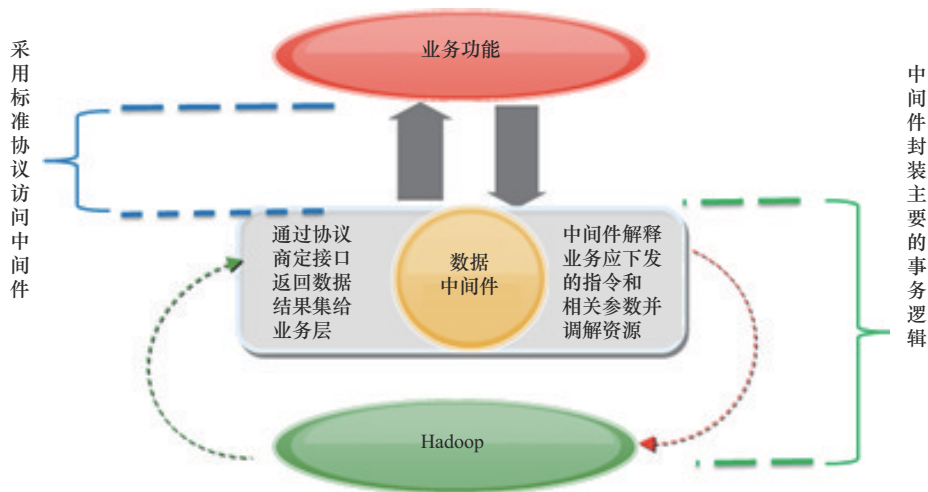


图3 中间件访问技术

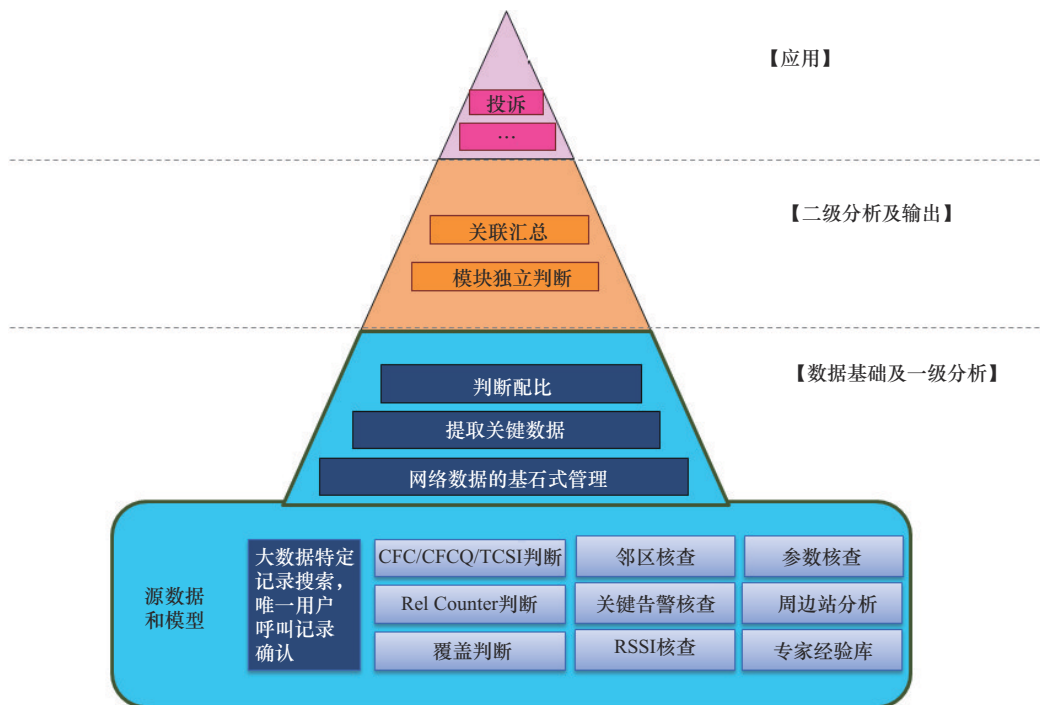


图4 移动投诉智能分析结构

查、关键告警核查、RSSI核查、覆盖判断、周边站分析和专家经验分析。

移动投诉智能分析运行逻辑如图5所示。

2.3 Hadoop 与传统关系型数据库互操作架构设计

在Hadoop大数据技术之前，电信已经建设

了很多基于关系型数据库的应用系统。为了保障原有的技术成果，需要大数据支撑系统能够集成这些系统。项目组提出针对不同应用场景集成Hadoop和关系型数据库。

根据CAP定理、数据结构和业务需求，项目组将大数据量的日志型数据存储于Hadoop系统，

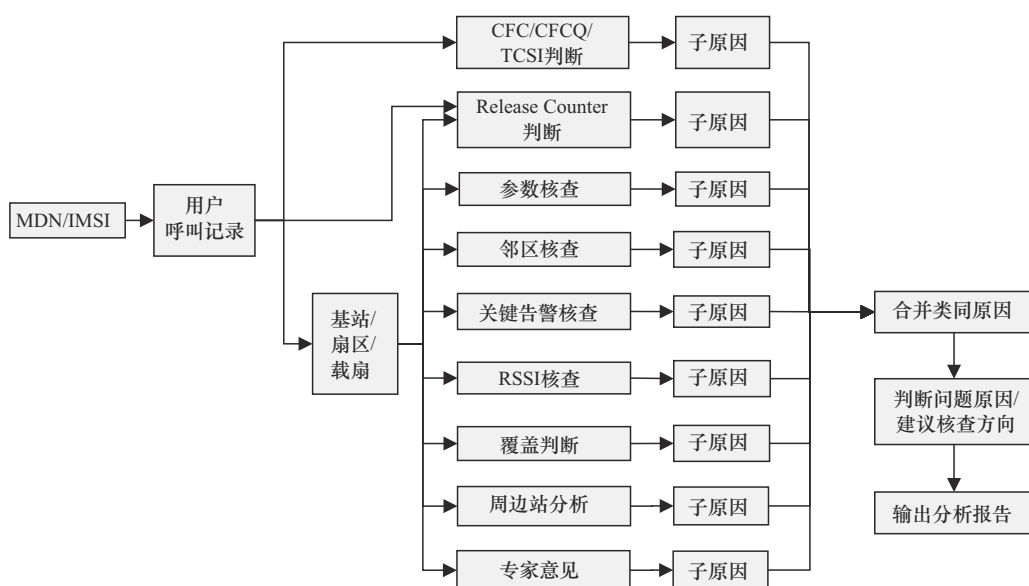


图5 移动投诉智能分析运行逻辑

数据量较小的结构化数据存储传统的关系型数据库，确保用户所得信息的正确性与实时性。

(1) 通过中间件实现 Hadoop 与传统关系型数据库互访

中间件实现了 SQL 适配能力，通过 API 方式识别基于 SQL 的编程命令，并转换成 Hadoop 可执行的 HQL 语句；中间件业务适配能力，识别应用后将业务分配给相应组件处理。

(2) 通过标准数据接口实现数据调用

系统制定了一系列个性化的接口规范，支持 SQL、FTP、WebService、socket、oracle、HIVE、UDF 等数据接口，并通过灵活的 Hadoop 数据与传统 oracle 数据库数据的互操作设计满足不同业务需要。

数据调用逻辑设计分为周期性统计需求数据调用逻辑和非周期性数据查询或统计需求数据调用逻辑，具体介绍如下。

(1) 周期性统计需求数据调用逻辑

- 小数据周期性统计需求→SQL 接口→oracle+ SQL 接口→客户端
- 大数据周期性统计需求→SQL 接口→SQL 适配层→HADOOP→SQL 适配层→SQL 接口→oracle

(2) 非周期性数据查询或统计需求数据调用逻辑

- 小数据查询业务需求→SQL 接口→oracle > SQL 接口→客户端
- 大数据查询业务功能→SQL 接口→SQL 适配层→HBase→SQL 适配层→SQL 接口→客户端
- 小数据统计业务功能→SQL 接口→oracle→SQL 接口→客户端
- 大数据统计业务功能→SQL 接口→SQL 适配层→HIVE→SQL 适配层→SQL 接口→客户端

2.4 端到端数据拼接方法的关键技术点

电信端到端的话单数据主要包括 Um 接口的 CDR 数据、核心网 CHR 数据的高层的 DPI 数据。这三类数据是多对多的关系，很难一一对应。

基于各类话单数据的原始字段值通过一系列的关键字段对应、时间差范围和置信空间等技术，实现了单一用户和单一网元的端到端话单的拼接并得到系统私有的键值，极大地方便了各类算法的数据调用，为智能优化分析、关联分析打下了基础。

所使用的原始字段包括 Month and Day、Connection Duration、Primary Record Creation Time、start_time、Mobile ID Type、Mobile ID、imsi、imei、ESN、bsid 等。系统私有的键值包括 Um_key、CDR_key、DPI_key 等。

2.5 可变字段日志型数据的优化方法

电信许多系统输出日志的字段越来越多，这些日志文件往往是前一些字段相同，但后一些字段根据日志输出所有者不同输出的字段也不同。

根据可变字段日志型文件的结构特性以及应用上的特征，对数据中协议（A11、LATLNG、PPP、TCP/UDP 和 UNK）字段为键值进行第一层分区，以时间戳进行二层分区，缩小访问的数据量，从而提高整个可变字段日志型数据的查询效率。

2.6 TB 级用户通话记录实时采集关键技术点

项目采集的数据源包括两种移动通信制式（频分多址和时分多址）、两种数据结构（结构型和非结构型）、两种流类型（实时流和非实时流）、3 种设备厂商（朗讯、华为、中兴）等格式数据，每天采集的数据量达 10 TB、百亿条。采集数据复杂、庞大，遇到很多疑难问题。

项目组通过 TCP 存储传输机制优化、话单数据库结构优化、SQL Loader 数据库连接时长优化、长话单滑动窗入库优化等技术，解决 TB 级数据实时传输话单丢失严重的问题。



3 实施效果

项目主要完成以下成果。

(1) 前后端联动完成了创建面向多维数据服务手段的投诉处理技术体系。建立以资源库(数据、算法、规范等)为核心,以中间件、数据挖掘等关键技术为支撑,以应用为纽带的具有多维综合数据服务能力的基本技术体系。基于资源开发集成了面向电信客服热线、智能终端客户端、投诉中心工位等多场景应用,为用户提供快捷的投诉处理服务。

(2) 完成了基于Hadoop的PaaS云架构、支持多维度投诉智能分析和智能响应集成。基于Hadoop架构,以大数据挖掘为基础,实现支持面向用户的投诉智能分析和智能响应、人机交互和多媒体数据处理及共享的技术架构。

(3) 完成了移动投诉智能分析和智能响应关键算法和模型建立。构建基于模块化分析的多类型数据自主分析和二次关联方法的分析模型,实现投诉成因智能分析并覆盖语音/数据等业务、制式及华为、中兴、朗讯等厂商,实现模糊地址识别和高精度定位技术、网优数据翻译。

(4) 完成移动用户定位。提出融合基础三角定位算法、基于优化数据的自优化学习算法、大数据训练等关键技术的高精度定位算法,实现用户模糊地址准确定位。

(5) 完成海量数据中特定记录搜索关键技术研发。建立以用户呼叫记录搜索规则为主、结合用户行为模式分析和定位分析的数据搜索模型,实现在千亿行数据中快速搜索到用户唯一投诉呼叫记录。

4 结束语

本文主要研究针对用户对无线通信服务的主要需求,融合移动投诉智能分析和智能响应、移动用户定位、海量数据特定记录搜索、中间件、Hadoop与传统关系型数据库互操作、TB级用户呼叫记录实时采集、GIS、网优数据翻译等关键技术,构建一个多业务、快速智能、可靠易用的移动投诉智能分析和智能响应系统,拓展了信息服务手段。

本文的创新方法在某市电信客服热线、App热线、投诉中心工位、大型乐园保障等信息服务平台进行规模应用,为企业节约1 311万元。通过本文的创新方法,可以显著提升客户感知,取得较好的社会效益和经济效益。

参考文献:

- [1] ERICH G, RICHARD H, RALPH J, et al. 设计模式:可复用面向对象软件的基础[M]. 李英军, 马晓星, 蔡敏, 等译. 北京:机械工业出版社, 2019.
- [2] IAN G, YOSHUA B, AARON C. 深度学习[M]. 赵申剑, 黎晓君, 符天凡, 等译. 北京:人民邮电出版社, 2017.
- [3] THOMAS H.C, CHARLES E.L, RONALD L.R, et al. 算法导论(原书第3版)[M]. 殷建平, 徐云, 王刚, 等译. 北京:机械工业出版社, 2012.

[作者简介]

吴强(1975-),男,中国电信股份有限公司上海分公司高级工程师,主要研究方向为蜂窝移动通信网络的优化、维护、建设和规划、大数据技术和AI算法模型等。

许樱(1981-),女,中国电信股份有限公司上海分公司工程师,主要研究方向为IDC机房设计、数据治理、大数据技术等。

徐小雅(1999-),女,中国电信股份有限公司上海分公司助理工程师,主要研究方向为移动蜂窝网络建设项目工程监理、数据治理、大数据技术等。