

基于 MySQL 的投诉稽核辅助分析工具研究与应用

张海波,姜莉莉,乔文通,段卫国,张 建

(中国移动通信集团设计院有限公司内蒙古分公司,内蒙古 呼和浩特 010010)

摘要: 文章分析了目前网优投诉稽核工作中的难点与痛点。为了解决这些问题,针对投诉系统页面中的不同数据特点,提出了投诉稽核辅助分析方案,使用辅助分析工具自动处理,既保证了数据的精确性,又极大拓展了稽核工作的深度与广度,将人员从繁重、重复性高的工作中解放出来,实现了高效率和高质量的双赢。

关键词: 投诉稽核辅助分析工具;MySQL 数据库;自动化智能化分析

中图分类号: TP309

文献标识码: B

文章编号: 2096-9759(2023)06-0103-05

0 引言

数据分析是网络优化管理及分析中最为频繁、重要的工作。近年来,为打造以客户满意度为驱动的投诉稽核工作模式,实现投诉稽核质量精细化分析,发现投诉稽核短板并不断跟踪,各省公司都十分重视投诉稽核分析工作。目前中国移动各省公司网络优化项目大致分为核心和非核心两个部分,核心部分又分为分析稽核、投诉稽核、费用稽核三大板块,以及诸如配合组巡、审计检查,专题问题分析等其他工作内容。投诉稽核工作月度工单量在 2020 年一度达到 21000 以上,2021 年压降工单量后,月度工单量仍能达到 5000 张以上,以集团投诉白皮书人均每月处理 500 张工单标准为参考,在投诉稽核上应配置最低 10 人以上的人力,但投诉稽核人员数量严重不足,工作负荷过高,促使我们致力于推动工作模式从依靠纯人工、过渡到半人工半自动、最终实现全部自动化智能化稽核。

1 方案概述

早期开展投诉稽核工作,需要人工在投诉系统页面中核查看户信息、投诉地点、涉及网元、性能指标等一系列数据,并根据核查结果,在 Excel 中依次记录,数据量庞大、统计维度广,既耗时又费力,再比如投诉位置与问题网元距离、关联工单情况等数据,完全依赖人工手动运算,效率低下而且极易出错,现有工作模式已经严重影响稽核进度和质量,为了提升工作效率,释放人力资源,借助 SQL 数据库脚本来替代人工操作。

以数据量为例,稽核工作的数据库内容涵盖 BMOP、性能指标、黑点信息、工单流转信息、逻辑关系信息、关联性能工单等,体量大小已经超过 100GB,通过个人 PC 端的 excel 软件处理完全不现实。我们研究的投诉稽核辅助分析工具的主要目标有三个,第一,解决人员严重不足、无法保障及时完成工作的进度问题;第二、抹平优化人员个体差异、能力高低、人为失误导致的工作质量问题;第三、拓展分析的广度与深度,不断提高工作质量,为后续拓宽工作范围奠定基础。

2 设计思路

投诉稽核辅助分析工具基于 MySQL 数据库技术进行开发,通过细化内蒙古移动网络集中投诉管理系统工单数据,根据稽核规范、现有工作模式以及前期工作经验,梳理数据逻辑关系、制定核查规则,大幅提升了投诉稽核工作的效率及质量。

投诉稽核核查范围包括工单逻辑性、合理性、合规性、关联指标改善度 4 个大类内容,134 个子项内容,使用辅助分析工具能够快速生成客观数据分析结果,有效支撑投诉稽核工作,总体设计思路如下:

数据处理基于开源数据库 MySQL,具有良好的扩展性;灵活、方便的数据分析方式,对于复杂查询,可以通过 SQL 语句实现,有效提高查询速度。各类分析均经过分类整理后固化在平台,以便后续调用。

基础数据源自内蒙古移动网络集中投诉系统,提取相关数据后,根据稽核要素、逻辑关系、各类指标的特征,定义查询条件和结果输出。逻辑关系可以通过编写 SQL 查询脚本实现,分析结果部分主要以表格方式呈现。

基于 WEB 的可视化报表呈现。针对所需要分析的数据特征,在 WEB 界面上实现数值、柱状图、饼图等多种类型的可视化报表呈现,并能够灵活的自定义各呈现结果的大小、位置。生产和管理人员能够直接生产报表,而不必等待 IT 开发。

3 功能设计

投诉稽核辅助分析工具能够实现基础信息核查,合理合规核查,逻辑关系核查,关联问题核查,处理质量核查等分析工作,通过 SQL 语句编辑核查规则实现复杂查询与判断,根据实际需求生成有效数据分析结果。

3.1 基础信息核查

投诉工单内用户号码、基础工参、网元信息的有效性核查:

(1) 投诉号码格式判断

主工单派单环节——“故障号码”,“客户投诉号码”格式有误,则输出结果:“投诉号码格式有误”。

(2) 分公司提供经纬度格式判断

主工单分公司环节——经纬度信息提供有误,则输出结果:“主工单分公司环节录入经纬度错误,无法核实投诉问题具体位置,计费工单派发无依据”。

3.2 合理合规核查

投诉工单内投诉处理时长、审批耗时合规核查。

(1) 投诉号码重复派单判定核查

以主工单派单环节——故障号码/客户投诉号码为基准,计算重复出现次数,目前暂时统计拥有中端优化子流程的工单统计重复次数,分别针对主工单和中端优化子流程的重复

收稿日期:2023-03-13

基金项目:2020 年江西省大学生创新创业训练计划重点项目(202111504002)。

作者简介: 张海波(1982-),男,山东文登人,咨询设计师,中级工程师,大学本科,主要研究无线网络优化;姜莉莉(1980-),女,辽宁盘锦人,咨询设计师,中级工程师,工程硕士,主要研究无线网络设计、规划、项目管理等工作;乔文通(1990-),男,辽宁人,高级咨询设计师,中级工程师,大学本科,主要研究无线网络设计、规划、项目管理等工作;段卫国(1982-),男,陕西府谷人,高级工程师,职业资格:注册咨询(投资)工程师,大学本科(学士),研究方向无线网规划咨询设计及网络优化;张建(1987-),男,内蒙古人,咨询设计师,中级工程师,大学本科,研究方向无线网络规划、设计及优化。

工单根据各自派单时间计算时间差小于 7 天的判定为重复派单, 输出结果: 主工单号码: 重复派单, 中端号码: 重复派单

(2) 投诉地点重复派单核查

以主工单分公司环节——地点关键词为基准, 计算重复出现次数, 目前暂时统计拥有中端优化子流程的工单统计重复次数, 分别针对主工单和中端优化子流程的重复工单根据各自派单时间计算时间差小于 7 天的判定为重复派单, 输出结果: 主工单投诉地点重复派单, 中端投诉地点重复派单。

3.3 逻辑关系核查

投诉问题定位与后续流转、投诉位置与网元距离、规划建设位置与投诉位置距离核查。

(1) 主工单_原因分类落点核查

以主工单分公司环节——原因分类落点结果进行判断, 如果原因分类落点为“非无线网络问题”且已派发中端优化子流程, 则输出结果: “非无线网络问题派发计费工单不合理”。

(2) 关键信息 ECI 合理性判断

如果“关键信息_ECI 与 GIS 距离”大于“中端_ECI 与 GIS 距离”, 且距离差大于 30 米, 则输出结果: “关键信息_ECI 非最近小区, 干扰智能分析结果”, 如果“中端_ECI 与 GIS 距离”大于“关键信息_ECI 与 GIS 距离”2KM 以上, 则输出结果: “中端优化确认服务小区与投诉点距离 (XXX 米), 超过投诉点与最近小区距离 (XXX 米) 2KM, 合理性存疑”。

3.4 关联问题核查

投诉问题通过问题网元 ECI 关联性能工单处理情况, 更加全面的呈现投诉问题与性能的相关性, 确保问题处理质量。

(1) 对于“覆盖”类问题, 优化后 MR 覆盖率出现劣化, 判定不合格, 输出“覆盖类问题优化后 MR 覆盖率存在劣化”。

(2) 对于“低利用率”问题的判定, 现在只针对“高负荷”问题生效, 其余问题不再核查低利用率。

3.5 处理质量核查

依据质量稽核规范, 结合基础信息、合规性、逻辑性、关联性核查结果, 核查全流程质量, 确保投诉问题处理满足网络问题解决率评估要求。

4 功能实现

投诉稽核辅助分析工具实现过程如图 1 所示: 共涉及 4 个步骤: 原始数据采集、数据整理、数据导入编辑脚本、核查结果呈现。



图 1 投诉稽核辅助分析工具实现过程

4.1 数据库设计

投诉工单数据基于数据库基础数据关联, 根据投诉稽核核查规则编辑 SQL 语句脚本, 对投诉工单稽核要点做出核查, 生成所需信息并给出判断结果, 实现复杂的查询及集中投诉质量稽核自动化处理。创建表“投诉工单_中端优化处理时长统计”SQL 语句核查脚本如下:

```

DROP TABLE IF EXISTS `投诉工单_中端优化处理时长统计`;
CREATE TABLE `投诉工单_中端优化处理时长统计` (
  ID INT AUTO_INCREMENT PRIMARY KEY,
  `主工单编号` CHAR(55) DEFAULT NULL,
  `中端优化子流程_工单号` CHAR(55) DEFAULT NULL,
  `派单环节结束时间` datetime DEFAULT NULL,
  `审核环节结束时间` datetime DEFAULT NULL,

```

```

  `工单审核时长(小时)` DOUBLE DEFAULT NULL,
  `工单审核时长判断` CHAR(50) DEFAULT NULL,
  `第一次审核环节结束时间` datetime DEFAULT NULL,
  `中端优化环节结束时间` datetime DEFAULT NULL,
  `中端优化环节处理时长(小时)` DOUBLE DEFAULT NULL,
  `中端优化环节处理时长判断` CHAR(50) DEFAULT NULL,
  KEY `zhugongdan` (`主工单编号`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COL-

```

LATE=utf8mb4_0900_ai_ci;

```

INSERT INTO `投诉工单_中端优化处理时长统计` (`主工单编号`, `中端优化子流程_工单号`, `派单环节结束时间`, `审核环节结束时间`, `工单审核时长(小时)`, `工单审核时长判断`, `第一次审核环节结束时间`, `中端优化环节结束时间`, `中端优化环节处理时长(小时)`, `中端优化环节处理时长判断`)

```

```

SELECT aa.`主工单编号`, aa.`中端优化子流程_工单号`, `派单环节结束时间`, `审核环节结束时间`,

```

```

(ROUND ((UNIX_TIMESTAMP(`审核环节结束时间`)-UNIX_TIMESTAMP(`派单环节结束时间`))/(60*60),2)) "工单审核时长(小时)",

```

```

`工单审核时长判断`, `第一次审核环节结束时间`, `中端优化环节结束时间`,

```

```

(ROUND ((UNIX_TIMESTAMP(`中端优化环节结束时间`)-UNIX_TIMESTAMP(`第一次审核环节结束时间`))/(60*60),2)) "中端优化环节处理时长(小时)",

```

```

`中端优化环节处理时长判断`

```

```

FROM (

```

```

SELECT DISTINCT SUBSTRING_INDEX(`工单编号`,`-ZA`,1) "主工单编号", `工单编号` "中端优化子流程_工单号"

```

```

FROM `投诉工单详情-原始数据`

```

```

WHERE `工单名称` = "中端优化" ) aa

```

```

LEFT JOIN (

```

```

SELECT `主工单编号`, `中端优化子流程_工单号`, `第一次审核环节结束时间`, `中端优化环节结束时间`,

```

```

CASE

```

```

WHEN `中端优化子流程_工单号` IN (

```

```

SELECT DISTINCT `工单编号` "中端优化子流程_工单号"
FROM `投诉工单详情-原始数据`

```

```

WHERE `环节名称` IN ("搁置审批","规划审核(区)","白名单审核(分)","申诉审核","挂起审核(区)") )

```

```

THEN "工单存在审批环节, 需要手动确认时长"

```

```

WHEN (ROUND((UNIX_TIMESTAMP(`中端优化环节结束时间`)-UNIX_TIMESTAMP(`第一次审核环节结束时间`))/(60*60*24),2)) > 5 THEN

```

```

CONCAT ("中端优化环节处理周期超时 (", (ROUND((UNIX_TIMESTAMP(`中端优化环节结束时间`)-UNIX_TIMESTAMP(`第一次审核环节结束时间`))/(60*60),2)),"小时")

```

```

WHEN `第一次审核环节结束时间` IS NULL OR `中端优化环节结束时间` IS NULL THEN "时间统计异常, 需要手动确认"
ELSE NULL

```

```

END "中端优化环节处理时长判断"

```

4.2 数据采集整理

(1) 数据采集

操作人员需登录内蒙古移动网络集中投诉系统, 通过选择派单时间导出相应待处理工单。



图2 原始数据采集登录页面 1

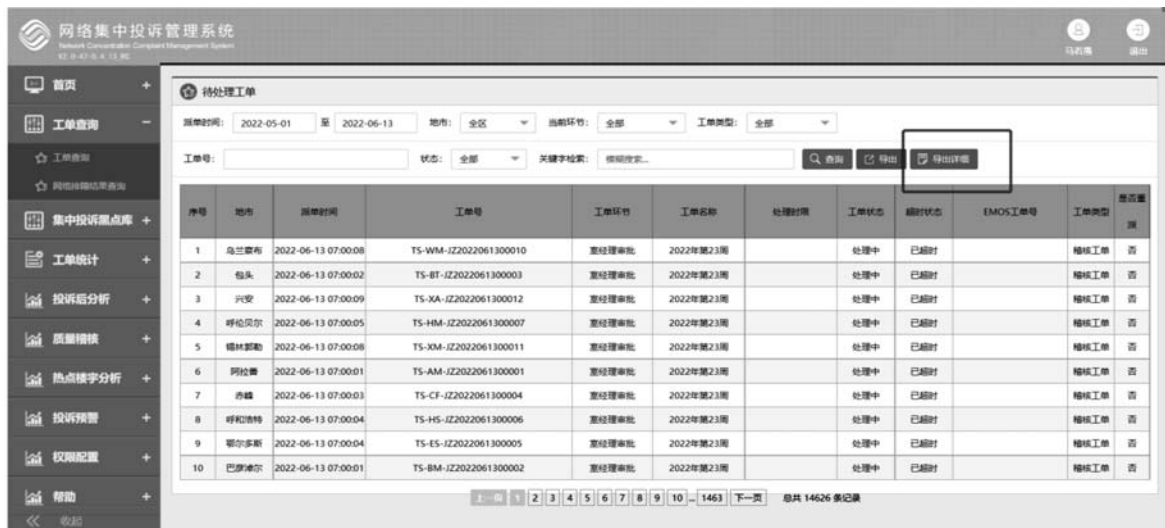


图3 原始数据采集登录页面 2

(2)数据整理

操作人员将得到的待处理工单数据按照需求定义数据类型、长度等,做好数据整理和准备。创建“投诉工单详情-原始数据”表 SQL 语句脚本:

```
DROP TABLE IF EXISTS `投诉工单详情-原始数据`;
CREATE TABLE `投诉工单详情-原始数据` (
  `工单编号` varchar(255) DEFAULT NULL,
  `工单名称` varchar(255) DEFAULT NULL,
  `地市` varchar(255) DEFAULT NULL,
  `派单时间` datetime DEFAULT NULL,
  `工单状态` varchar(255) DEFAULT NULL,
  `工单环节` varchar(255) DEFAULT NULL,
  `环节名称` varchar(255) DEFAULT NULL,
  `环节开始时间` datetime DEFAULT NULL,
  `环节结束时间` datetime DEFAULT NULL,
  `字段名称` varchar(255) DEFAULT NULL,
  `字段内容` varchar(5555) DEFAULT NULL,
  KEY `paidanshijian_index` (`派单时间`),
```

```
KEY `gongdanleixing_index` (`工单名称`),
KEY `huanjiemingcheng_index` (`环节名称`),
KEY `zidianming_index` (`字段名称`),
KEY `工单名称` (`工单名称`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

4.3 数据导入数据库设定脚本处理数据

操作人员只需要在电脑安装 MySQL 数据库,并在 workbench 内连接 MySQL 服务器创建需求表单导入基础数据,之后根据稽核要素、逻辑关系、各类指标的特征和工作需求,定义查询条件和结果输出,通过 MySQL 数据库,编辑 SQL 语句脚本,对投诉工单稽核要点做出核查。

4.4 导出最终核查结果

在 workbench 内连接 MySQL 服务器完成上述核查流程后,运行 SQL 语句脚本就可导出相应核查结果。

5 实施效果

成果应用前,稽核人员需在投诉系统页面中核查用户信息、投诉地点、涉及网元、性能指标等一系列基础数据,并根据



图4 SQL 语句脚本处理过程界面



图5 核查结果导出界面

DZ	EA	EB	EC	ED	EE	EF	EG	EH	FI	中 小 指 标
最近 规划 建站 编号	最近的 规划 建站 经纬度	最近的 规划 建站 经纬度	与最近 规划 建站 距离	经纬度 计数	投诉 半径 (米)	投诉 面积 (平方 米)	中端 ECl 与GS 距离	中端 ECl 与GS 距离	效果 评估 、指标 判定 汇总	
886 TS-YCL-X	116.0531	43.91359	220814.4	3	188798	111981.1	377558		负载不均(南区无线利用率差值 58.98%)	关
893 TS-YCL-T	120.6413	42.8761	171405.2	3	1078	3.65	2148		负载不均(南区无线利用率差值 34.96%)	关
894 TS-YCL-H	111.4768	39.94478	98324.24	3	2507	19.75	4456		负载不均(南区无线利用率差值 52.19%)	关
933 TS-YCL-B	110.6561	40.49677	93132.59	3	758	1.81	1514		弱覆盖(MR 64.54%)	关
963 TS-ZD-XA	121.986	46.07363	8862.75	4	337	0.36	412		弱覆盖(MR 76.33%)	关
974 TS-YCL-V	106.999	39.17446	301458	3	117	0.04	23		低切换(77.48%)、高负荷	关
979 TS-ZD-XA	121.7369	46.4898	334996.3	3	112	0.04			负载不均(南区无线利用率差值 49.99%)	关
980 TS-YCL-B	109.639	40.84023	54401.43	3	112	0.04	223		负载不均(南区无线利用率差值 39.27%)	关
981 TS-ZD-XA	121.7369	46.4898	221840.4	3	75	0.02	150		负载不均(南区无线利用率差值 43.72%)	关
995 TS-ZD-HS	111.7033	39.91557	443.11	3	96	0.03	14		弱覆盖(MR 78.64%)	关
1009 TS-YCL-B	110.6561	40.49677	72022.39	3	1740	9.51	2008		高负荷、负载不均(南区无线利用率差值 75.26%)	关
1017 TS-RG-XH	116.4226	45.65855	301974.3	3	2	0	4		低切换(75.54%)	关
1021 TS-ZD-XA	121.7369	46.4898	333396.4	4	273	0.23	541		负载不均(南区无线利用率差值 78.53%)	关
1025 TS-YCL-T	120.6413	42.8761	159073	3	159	0.08	316		负载不均(南区无线利用率差值 35.15%)	关
1037 TS-ZD-XA	121.7369	46.4898	230093.8	3	6449	130.66	12885		低切换(77.02%)	关
1042 TS-YCL-B	109.3445	40.58315	54684.25	4	21266	1420.76	41831		负载不均(南区无线利用率差值 36.21%)	关
1048 TS-RG-XH	116.4226	45.65855	443684.7	3	804	2.03	1585		负载不均(南区无线利用率差值 57.30%)	关
1051 TS-YCL-B	109.639	40.84023	21946.4	3	14	0	28		负载不均(南区无线利用率差值 64.80%)	关
1055 TS-YCL-B	110.6561	40.49677	53835.67	3	741	1.72	1374		负载不均(南区无线利用率差值 76.92%)	关
1056 TS-ZD-HS	111.4768	40.01219	83821.07	4	128814	52128.59	256481		负载不均(南区无线利用率差值 55.81%)	关
1097 TS-YCL-T	119.6655	45.55587	50941.98	3	11	0	16		弱覆盖(MR 77.42%)	关
1120 TS-YCL-B	109.3445	40.58315	154305.7	4	542	0.92	1079		高负荷、负载不均(南区无线利用率差值 56.09%)	关
1123 TS-YCL-R	110.6561	40.49677	123752	3	705	1.56	1130		负载不均(南区无线利用率差值 60.83%)	关

图6 最终核查结果呈现

核查结果,在 Excel 表格中依次记录相关数据,此项工作既耗时又费力,而且在人工手动操作过程中极易出现错误,另外基础数据中的部分数据,例如投诉位置与问题网元之间距离、相

关联工单情况,完全依靠人工手动运算,工作效率相当低。投诉稽核辅助分析工具使用后替代纯人工手动处理,实现集中投诉质量稽核工作智能化,达到自动化换人、降本增效、提高

基于知识图谱的医疗问答系统研究

李 伟,王竣生,秦 鹏

(六盘水师范学院 计算机科学学院,贵州 六盘水 553000)

摘要: 随着人工智能技术的不断发展,越来越多的自然语言处理技术应用到医疗行业。如何从海量医疗数据中提炼信息,并根据用户的问题给出针对性较强的回答,是推进医疗智能化的关键问题。文章研究利用 BiLstm+CRF 模型处理医疗领域问答相关数据,基于图数据库 Neo4j 构建一个医疗知识图谱,并在此基础上构建一个问答系统,实现医疗知识的自动问答服务。实验结果表明,该系统可以为用户提出的问题查找匹配正确答案并返回给用户。

关键词: 医疗知识;知识图谱;问答系统

中图分类号: TP311

文献标识码: A

文章编号: 2096-9759(2023)06-0107-03

Research on Medical Q&A System Based on Knowledge Graph

LI Wei, WANG Junsheng, Qin Peng

(School of Computer Science, Liupanshui Normal University, Liupanshui, Guizhou 553000, China)

Abstract: With the continuous development of artificial intelligence technology, more and more natural language processing techniques are applied to the medical industry. How to extract information from the vast amount of medical data and give more targeted answers based on the user's questions is a key issue in advancing medical intelligence. In this paper, we use the BiLstm+CRF model to process question and answer related data in the medical field, construct a medical knowledge graph based on the graph database Neo4j, and build a question and answer system based on it to realise an automatic question and answer service for medical knowledge. The experimental results show that the system can find the exact answer to the question asked by the user and return it to the user.

Keywords: medical knowledge; knowledge graph; question and answer system

1 引言

随着信息技术的快速发展,网络已经成为人们日常生活中不可缺少的一部分,而网络中的数据也变得越来越丰富,如何从网上数据中获取有用信息,并将其应用于用户问答需求,已成为一个亟待解决的问题。知识图谱作为一种新型的数据库技术,在大数据分析、人工智能等领域有广阔的应用前景,而知识图谱与问答系统相结合,可为用户提供高质量、个性化和准确的问答服务。

知识图谱是由 Google 公司在 2012 年发布的^[1],主要应用在 Google 搜索引擎中,其最初的目的是借助知识图谱来优化 Google 的搜索结果,为用户提供含有完整的搜索结果,方便用户能够尽快获取到所需的信息,为用户提供更好的搜索体验。Siddhant Garg^[2]等人提出了一种 TANDA 的有效技术,主要用

于自然语言任务的预训练变换模型的精调。Shen^[3]等人提出每个医学答案对应一个有效问题的分布,着重研究了如何整合结构化知识和非结构化知识来生成上述医疗问答对。崔洁^[4]等人提出采用自顶向下的方式构建乳腺肿瘤知识图谱,为乳腺肿瘤疾病知识学习与推理奠定了数据基础。

本文研究基于知识图谱,设计一个医疗知识问答系统。首先在网上抓取医疗领域问答相关数据,然后基于图数据库 neo4j 构建一个医疗知识图谱,并在此基础上构建一个问答系统,来实现医疗知识的自动问答服务。

2 相关技术

2.1 医疗知识图谱构建

知识图谱的建设^[5],一般包括数据获取、数据管理和数据存储等过程。医疗知识图谱要先定义医疗领域相关的实体、

收稿日期:2023-03-10

基金项目:六盘水师范学院校级基金项目(LPSSYZDZK202204)。

作者简介:李伟(1985-),山东聊城人,硕士,副教授,研究方向为数据挖掘、大数据技术。

工作效率和质量,提升用户感知的目的。按稽核人员每天人工处理 230 单计算,使用工具前每单耗时 10 分钟,通过投诉稽核辅助分析工具处理,投诉稽核自动化处理工单总计耗时约 20 分钟,效率提升: $230 \times 10 / 20 = 115$ 倍以上,且通过对比分析,准确率达到了 100%。

6 结语

本工具基于 MySQL 数据库技术架构进行开发,操作人员只需要在电脑安装 MySQL 数据库,并在 workbench 内连接 MySQL 服务器创建需求表导入基础数据进行运算便可得到所需结果。本工具具有较强的可移植性,关联维度的数据可从各省的网络优化管理平台大数据获取,适用于其他省开展应

用部署,各省可在符合《中国移动网络数据安全管理办法》的基础上,利用集中投诉分析管理系统中的各种大数据信息,开展投诉工单分析和集中呈现关联等大数据分析应用,可参考本成果投诉稽核辅助分析工具,通过开发相关算法、功能、工具,实现快速识别、智能分析相关问题,提升问题定位准确性及分析效率。

参考文献:

- [1] 张伟丽,江春华,魏劲超. MySQL 复制技术的研究及应用[J]. 计算机科学. 2012, (z3). 168-170.
- [2] 王锐. 搭建 MySQL 数据库主从库平台实现数据备份[J]. 电脑编程技巧与维护. 2011, (19). 30-31.