

重点业务感知监测分析机器人 在网优智能化中的应用研究

Research on Application of Key Business Perception Monitoring and Analysis
Robot in Intelligent Network Optimization

程 乔,王映华,李 冉,刘文雯(中国联通广西分公司,广西 南宁 530028)

Cheng Qiao,Wang Yinghua,Li Ran,Liu Wenwen(China Unicom Guangxi Branch,Nanning 530028,China)

摘 要:

广西联通基于前期积累的网络优化经验,积极探索机器学习、软件编排等技术,开发了“重点业务感知监测分析机器人”,机器人能通过数据自动获取与数据处理、感知问题预测、智能定界分析、报告智能撰写输出、邮件推送这五大过程分析并优化重点业务感知,其应用效果显著,实现了网络优化从传统的“人工分析”到“智能分析”模式的变革,为未来的网优智能化进行积极探索。

关键词:

智能网优;机器人;指标预测;告警邮件推送
doi:10.12045/j.issn.1007-3043.2019.11.008
中图分类号:TN929.5
文献标识码:A
文章编号:1007-3043(2019)11-0033-06

Abstract:

Based on the network optimization experience accumulated in the early stage, Guangxi Unicom actively explores machine learning, software layout and other technologies, and develops a "key business perception monitoring and analysis robot", which can analyze and optimize the key business perception through the five processes of automatic data acquisition and data processing, perception problem prediction, intelligent boundary analysis, intelligent report writing and output, and email push. Its application effect is remarkable. It realizes the transformation of network optimization from the traditional "artificial analysis" to the "intelligent analysis" mode, and actively explores the intelligent network optimization in the future.

Keywords:

Intelligent network optimization; Robot; Index prediction; Alarm email push

引用格式:程乔,王映华,李冉,等.重点业务感知监测分析机器人在网优智能化中的应用研究[J].邮电设计技术,2019(11):33-38.

1 概述

移动互联网数据业务流量呈爆炸式增长态势,各类业务层出不穷,业务感知性能已成为运营商网络优化的重中之重,当前网络优化的工作模式在实时性、准确性、效率、成本等方面的弊端进一步凸显,即传统的网络优化以人工为主,无法适应日益复杂的网优工作需求,因此需要加快优化工作转型。本次广西联通结合现有的重点业务感知分析优化的工作场景,基于机器学习、软件编排、互联网等技术,充分挖掘丰富的

网络数据资源,并设计开发出“重点业务感知监测分析机器人”来优化重点业务感知,为未来的网优智能化转型积累优化经验。移动网络运维网优发展趋势如图1所示。

2 重点业务感知监测分析机器人实现流程

重点业务感知监测分析机器人的设计与实现,主要基于机器学习从历史数据中学习和建模,分析得出影响业务感知的规律特征,并通过软件设计编排等实现自动化、智能化的“机器工作”来代替“人工工作”,以提升工作效率和准确率。

2.1 重点业务感知监测分析机器人的流程框架设计

收稿日期:2019-09-03

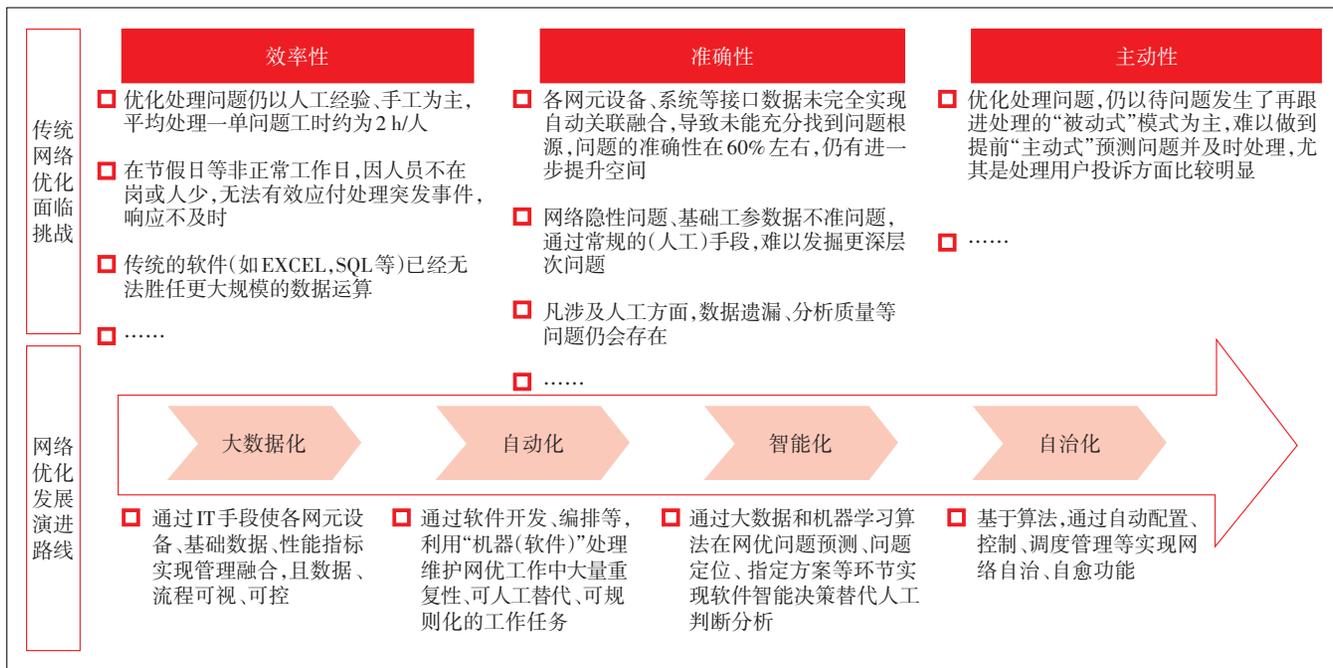


图1 AI+移动网络运维网优发展趋势

结合广西联通上网业务人数、流量、业务请求次数等情况,最终选择以微信、腾讯视频、淘宝、王者荣耀、高德地图等6类15种重点业务进行监测分析,如表1所示。

表1 重点业务监测指标

类别	重点业务标的	主要参考指标
即时通信	微信(含朋友圈、微信小视频)、QQ	TCP成功率、上/下行RTT时延
视频	腾讯视频、爱奇艺	TCP成功率、上/下行RTT时延、速率
游戏	王者荣耀	TCP成功率、上/下行RTT时延、速率
浏览下载	网易新闻、今日头条、腾讯新闻、百度、淘宝、京东	TCP成功率、HTTP成功率、上/下行RTT时延、速率
地图导航	高德地图、百度地图	TCP成功率、上/下行RTT时延
支付	微信支付、支付宝	上/下行RTT时延

重点业务感知监测分析机器人主要通过R语言与数据库进行数据交互和逻辑运算,实现对业务的KQI实时监控和提前预测,并快速对监控到的告警问题进行智能分析定界,然后自动输出报告,并下发告警邮件推送通知。本次以微信业务为例进行说明。其框架如图2所示。

2.1.1 数据库连接及数据源验证

数据源主要来自SI-U信令数据,包含了URI、RTT、IP_LEN等业务特征、业务感知有关的字段。通过搭建R语言与数据库的数据交互环境,并结合历史

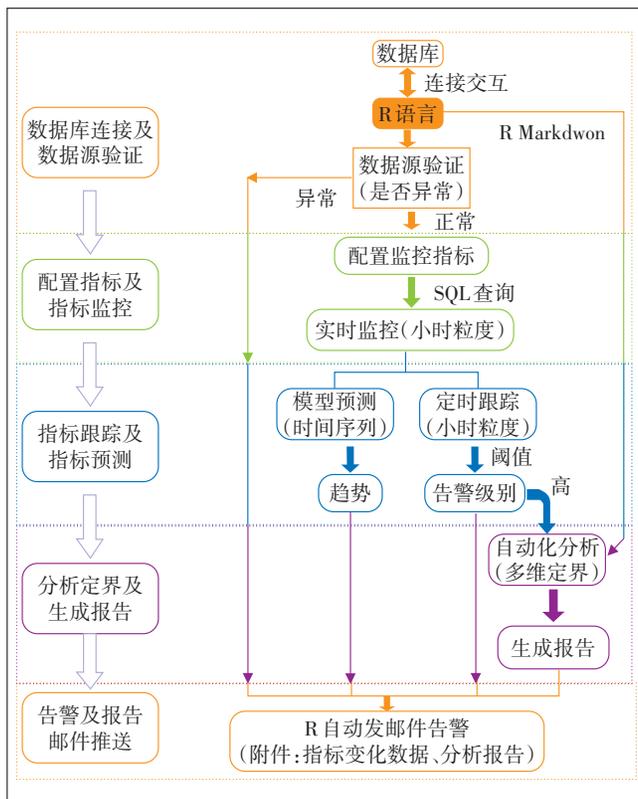


图2 重点业务感知监测分析机器人的流程框架设计图

数据源对当前入库的数据源的质量进行监控验证,避免因数据源异常影响系统整体的准确性。基于模型算法得出历史同时段的数据量波动规律,对比当天各

接口数据量差异,若超过设定的阈值则判断该时段数据源出现异常,系统会自动触发“数据源异常”的告警邮件推送。

2.1.2 指标配置及实时监控

在数据源无异常的前提下,通过R语言的逻辑算法对关键指标进行配置和实时监控。指标的配置即把关键指标统计逻辑转换成R语言能识别并统计的机器语言。实时监控即利用R语言设置的逻辑算法对每个关键指标持续统计和监控,及时发现指标告警信息,并把该时段的指标写入本地历史指标变化表中。

2.1.3 指标跟踪及指标预测

指标跟踪及指标预测模块主要针对本时段的指标设置告警规则,基于时间序列模型对未来一个时段的指标变化趋势进行预测。告警规则的设置主要是针对本时段统计到的指标设置不同的阈值以及对应的不同告警级别,达到告警条件则利用R语言发送内容为“本时段告警类别+级别+未来时段变化趋势”的邮件,其中告警类别指业务类别,如视频、游戏;级别如1级、2级,1级告警级别最高。而当告警级别高时则自动跳转到智能分析定界和自动生成报告模块,并发送“本时段告警类别+级别+未来时段变化趋势+自动定界报告”的告警邮件。

2.1.4 智能分析定界及生成报告

若指标的告警级别高,则进入智能多维分析定界并生成报告的模块。本模块主要是基于R语言中能自动化生成报告的Rmarkdown脚本实现。实现原理是利用R语言与数据库交互的数据,针对告警指标在多维度建立各自的逻辑算法聚类分析,最终定界出指标告警的原因并自动生成Rmarkdown报告。

2.1.5 告警及报告邮件推送

借助R语言的mailR包相关函数,把数据源是否异常、指标预测趋势、指标告警及级别、自动化定界报告等相关信息以邮件的形式自动发送给相关人员。

2.2 基于时间序列的业务性能预测模型选择

时间序列预测是一种基于历史数据的延伸预测,也称历史引伸预测法,通过编制连续的时间序列并分析其反映出来的发展过程、周期和趋势,进行类推或延伸,借以预测下一段时间可能达到的水平。结合历史数据与专家经验得知业务流量、性能指标、时间呈现规律性的周期变化,符合时间序列算法的应用场景。因此本次主要采用“时间序列预测的指数平滑法”进行建模。

2.2.1 指数平滑法原理与应用

指数平滑法即根据历史实际数据和预测值,用指数加权的办法进行预测。其中三次指数平滑法适用于有趋势和季节性的序列,故本次使用此算法,以微信业务上行RRT时延指标(TCP三步握手的第一步上行连接服务器时延)为例进行建模。

指数平滑法分为1个预测方程(式(1))和3个平滑方程(水平方程 l_t 、趋势方程 b_t 、季节性成分方程 s_t),其中 α 、 β^* 和 γ 为平滑参数, m 代表季节性周期。

$$\hat{y}_{t+h} = l_t + hb_t + s_{t-m+h_m^*} \quad (1)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (2)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \quad (3)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \quad (4)$$

其中式(1)中的 $h_m^* = [(h-1) \bmod m] + 1$,其确定了季节性指标的最后一个季节的样本($\lfloor u \rfloor$ 代表不超过 u 的最大整数)。式(2)是季节调整量($y_t - s_{t-m}$)和非季节性预测($l_{t-1} + b_{t-1}$)之间的加权平均。式(4)计算出了当前的季节性指标($y_t - l_{t-1} - b_{t-1}$)和上一周期的季节性指标的加权平均。

2.2.2 建模与效果检验过程

第一步对原始历史数据建立时间序列并对数化处理,其目的是为了消除数据源的异方差,减少数据的波动对预测结果的影响。

第二步采用HoltWinters算法(R语言包)对对数化处理后的时间序列进行预测建模。本次建模预测的 α 为0.60,即表明近期的观测值的影响权重比较大,大多数预测值与真实值相差不大,具有较高的可信度。

第三步对未来值进行预测及效果检验,首先预测未来 N 小时的上行RTT时延,其次对预测效果进行检验,检验延迟1~20阶中的预测误差是否非零自相关。图3呈现样本内预测误差在滞后9阶时超过置信边界略明显,其他基本在置信空间以内。接着采用Ljung-Box检验预测误差的随机性,经检验得出 $p=0.004919$,意味着置信度有99.51%,这样的值是可接受预测误差在1~20阶是非零自相关的,如图3所示(蓝色为预测部分)。因此按照以上基于HoltWinters算法的时间序列预测模型,可以有效预测微信业务性能变化趋势。

2.3 业务感知问题智能分析与自动化输出报告流程

广西联通基于前期积累的网络优化经验,设计编排了业务感知智能定界及自动化输出报告流程。

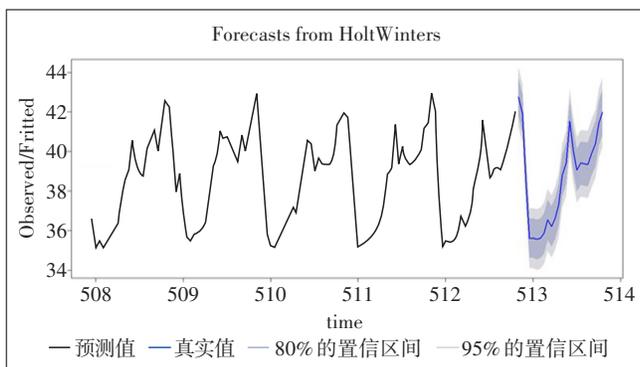


图3 未来值的预测及效果检验

2.3.1 业务感知五维分析定界法

在实时监控重点业务性能时,若某时段出现高告警级别(1级),系统将跳转至“五维分析定界”模块。主要思路如下。

a) 对业务多天的性能从时间维度进行分析,得出恶化时间和恶化程度。

b) 在业务维度排查该时段是否出现多个重点业务恶化,若仅出现单业务恶化告警,初步归类为“云(SP)”原因,并筛选该业务恶化所对应的IP。

c) 若多个重点业务同时恶化,则从设备维度排查SGW/PGW核心网问题,若问题集中则初步归类为“管原因”。

d) 排除非核心网因素后,则以地(市)区域维度排查,若问题集中在某地(市),则初步归类为“区域原因”,输出该地(市)恶化eNodeB列表。

e) 最终基于用户维度统计该时段受影响用户的占比,若受影响用户占比小于5%,则最终均定界为“端原因”。

基于业务感知五维分析定界法,通过端、管、云智能分析出业务感知恶化的原因。业务感知五维分析定界法主要流程如图4所示。

2.3.2 基于Rmarkdown的自动化报告输出

自动化报告输出主要是利用R语言的RMarkdown包对数据库的数据进行查询,针对告警指标在多维度建立各自的逻辑算法聚类分析,最终定界出指标告警的原因并自动生成报告。其中RMarkdown包可将R计算(代码块)与普通文本相结合,用简单灵活的方式来完成文档的标记并转换成PDF、word、html等格式文档输出。

2.4 实时告警邮件推送

告警邮件推送主要是通过R语言的mailR包设置邮件标题、正文以及附件等内容,并发送给指定的收

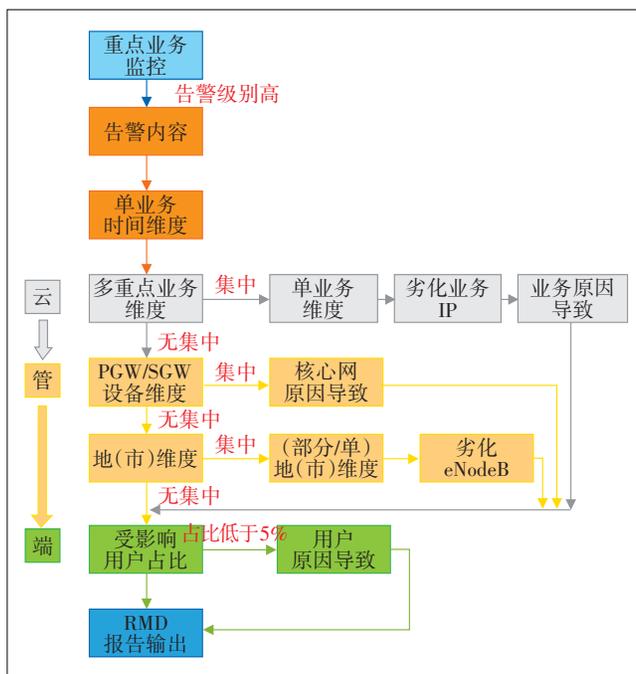


图4 业务感知五维分析定界流程图

件人。将该模块代码嵌入机器监控系统即可实时把告警分析内容通过邮件推送出去。

3 应用效果案例

基于机器学习的重点业务监控及预测分析系统,广西联通在2018年5月28日的实时监控业务感知中发现,微信业务当天12时上行RTT平均时延457.59 ms,且基于时间序列模型预测此项业务于当天13时将会继续恶化,预测值为483.23 ms,如图5所示。

基于上述结果,系统自动触发“重点业务(微信)出现上行RTT时延一级告警”,然后进入到智能化分析定界模块并自动生成报告,并于当天13:07:37将邮件推送给相关人员。其中报告涉及内容主要包括索引、告警内容、劣化业务时间维度分析、重点业务维度分析、核心网设备维度分析、地(市)维度分析、受影响用户分析、定界结论。

结合业务感知五维分析定界法得知:近3天的波动系数较小,未出现明显波动,该业务在5月28日12时开始恶化;未发现同时段其他重点业务出现明显时延高情况;微信业务请求次数TOP50涉及的IP中,IP为58.144.137.11至58.144.137.20地址段的时延均大于600 ms;该时段重点业务时延大于安全阈值(80 ms)的受影响用户数占比为64.56%;综上所述分析定界为云问题。

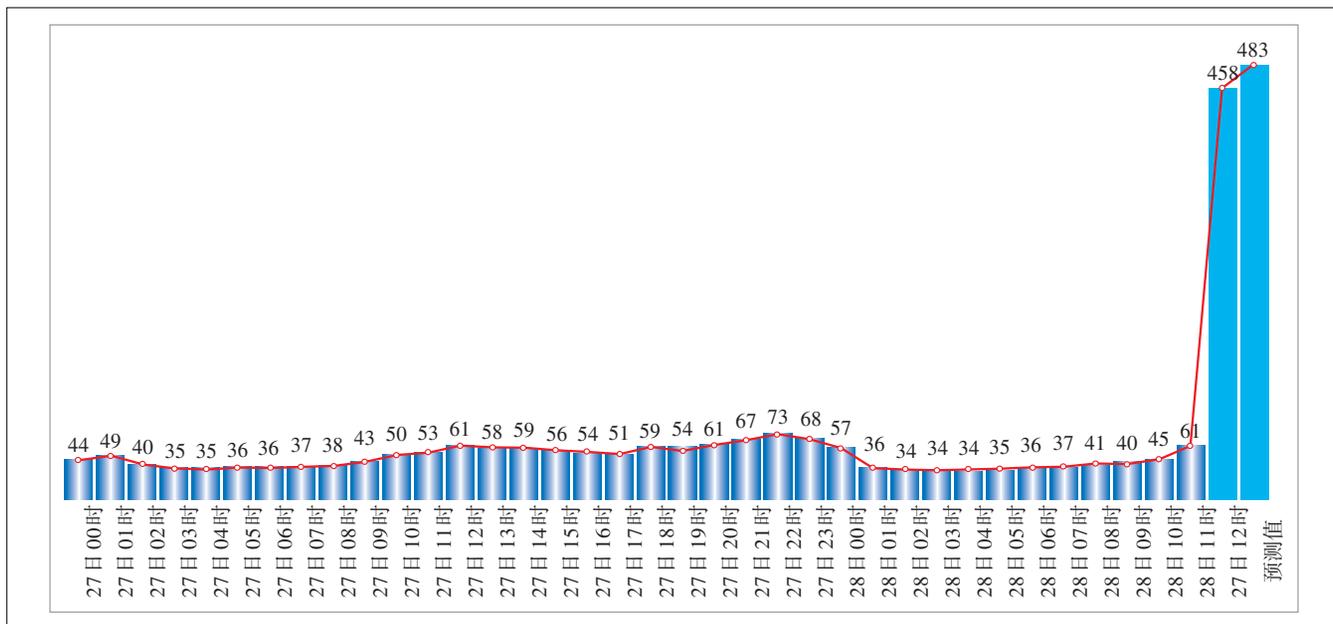


图5 重点业务上行RTT时延跟踪与预测

结合实际人工拨测和信令问题分析结果,发现主要感知问题为用户微信业务中的朋友圈图片打不开或出现白屏现象,通过信令分析得知上行RTT时延超过100 ms。

广西联通与腾讯团队迅速成立联合工作组,开展微信业务感知优化提升工作。在提供分析数据并联合微信团队优化服务器响应机制后,上行RTT时延从原来的457 ms下降至36 ms,微信朋友圈白屏现象恢复正常,有效提升了微信用户的上网感知。

广西联通目前基于重点业务感知监测分析机器人的部署,为微信、王者荣耀游戏、淘宝、腾讯视频、新浪微博等多个重点业务提供监控保障,累计处理了32起因业务性能突发恶化而影响用户感知的事例,有效提升了对突发性问题的应变能力和响应能力,其应用效果显著。

4 智能网优的探索与展望

未来AI+网优智能化的应用场景有着广大的开拓探索空间,结合不同的机器算法特点可用于智能场景的覆盖评估、智能动态阈值隐形故障挖掘、场景的参数智能优化、智能业务或负荷预测、基站画像管理等领域。广西联通以“聚类算法管理小区业务画像”和“基于MR-MDT数据发掘小区天馈问题”的应用场景作为试点,对网优智能化应用场景进行持续探索研究,其成果初显成效。

4.1 基于聚类算法的流媒体业务的小区画像管理与重点监控保障

结合中国联通总公司通报的2I流媒体业务播放小区列表,基于SPSS Modeler大数据分析建模软件对流媒体播放次数进行“K-Means”聚类分析,最终按播放次数来划分出“高”“中”“低”3个类别。如本次聚类有319个小区(播放次数≥185)的播放次数比较大,定义为“高”类别,流媒体业务的小区画像聚类算法建模如图6所示。对某地(市)为“高”类别的流媒体业务小区试点进行重点监控和保障,累计完成12个重点小区性能恶化问题(如故障告警,高负荷均衡等)的高优先响应处理。通过聚焦重点业务小区的感知保障处理,有效避免潜在用户投诉,维护网络口碑。

4.2 基于MDT_MR大数据发掘基站小区天馈问题

基于爬虫获取场景图层信息,结合地理、工参、MDT等数据,建模计算出用户位置与正北方向的夹角 $\angle\theta$,并与小区工参天线方向角 $\angle\alpha$ 对比,通过异常的偏差判断用户位置是否在天线合理范围内,其中MDT_MR发掘小区天馈问题算法原理如图7所示。在某地(市)试点累计发掘了17处疑似天馈异常问题并现场验证,其准确性为70%以上。

5 结束语

广西联通基于前期积累的网络优化经验,积极探索机器学习、软件编排等技术,通过开发的重点业务

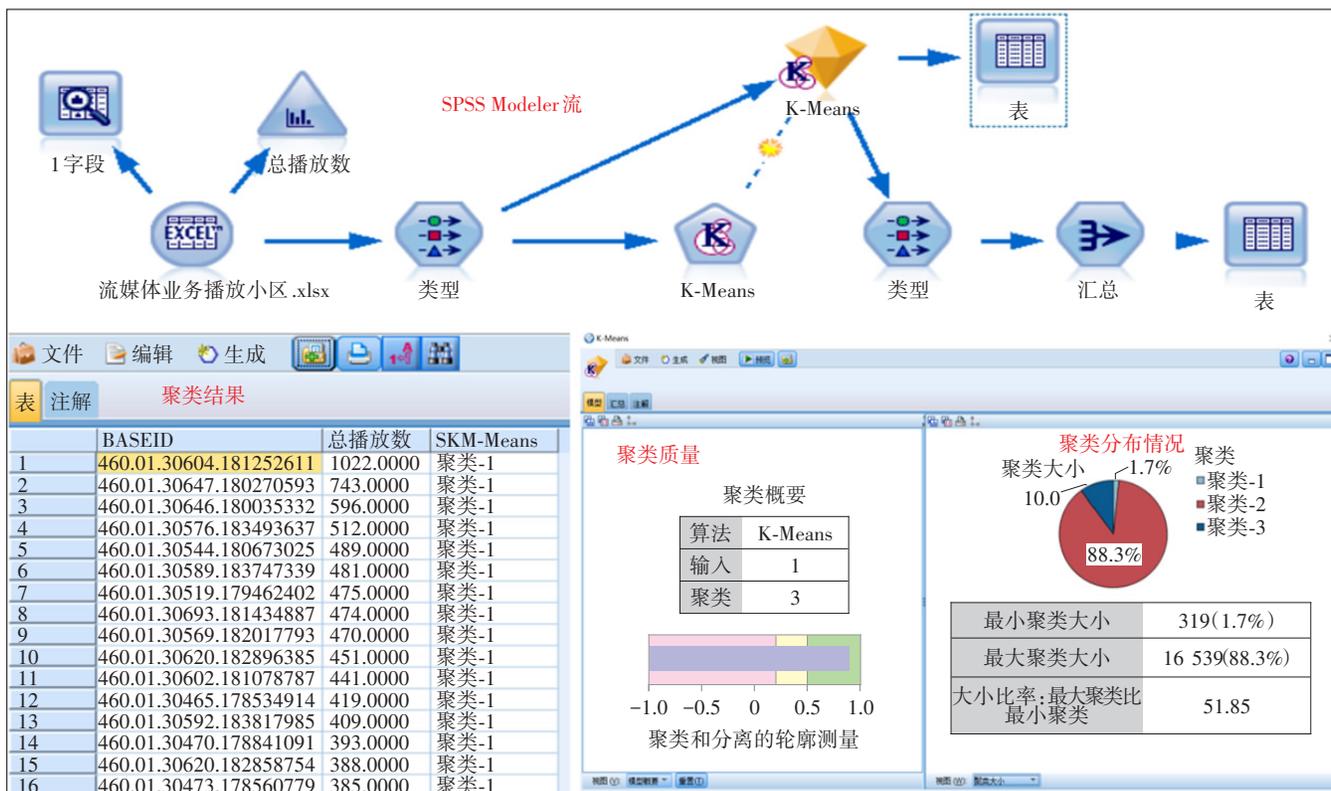


图6 基于聚类算法的流媒体业务的小区画像管理

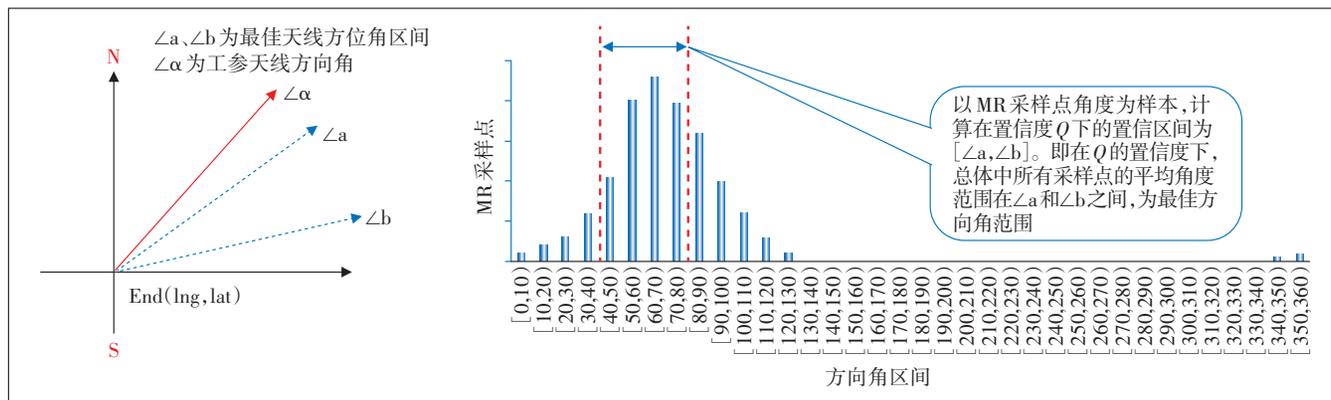


图7 基于MDT_MR大数据挖掘基站小区天馈问题原理

感知监测分析机器人实现了15类重点业务感知的智能洞察与优化分析,在未来的网优智能化转型上进行了积极探索,具有一定的推广价值。

参考文献:

[1] 薛毅,陈立萍. 统计建模与 R 软件[M]. 北京:清华大学出版社, 2007:32-40.
 [2] ZIEBARTH N L, ABBOTT K C, IVES A R. Weak population regulation in ecological time series[J]. Ecology Letters, 2010, 13(1): 21 - 31.
 [3] FANG D, ZHANG Y, SPICHER K. Forecasting accuracy analysis

based on two new heuristic methods and Holt-Winters-Method[C]// 2016 IEEE International Conference on Big Data Analysis (ICBDA). IEEE, 2016.

[4] 徐涛. 基于视觉注意机制的服务机器人深度感知方法研究[D]. 北京:北京工业大学, 2017.

作者简介:

程乔,高级工程师,硕士,主要从事数据业务端到端优化和大数据分析应用支撑工作;王映华,学士,主要从事数据业务优化及流量经营市场支撑工作;李冉,学士,主要从事移动通信网络的规划与设计工作;刘文雯,学士,主要从事大数据分析和市场服务支撑工作。