

# 智能可视与开放的无线网络大数据 Intelligent Visual and Open Large Data Assessment System for Wireless Network 评估体系

于洋(中国联通江苏分公司,江苏南京210019)  
Yu Yang(China Unicom Jiangsu Branch, Nanjing 210019, China)

## 摘要:

当前互联网时代信息爆发,旧有的评估体系亟需迭代,中国联通江苏分公司以智能化、可视化为主线,应用多种开源技术,通过建立自主、开放、合作的互联网生态获取多维数据,依托云虚拟化平台进行架构重耕,开发出了智能、可视、开放的新型大数据无线网络健康评估体系。

## 关键词:

智能;可视;大数据;自主;开放;合作  
doi:10.12045/j.issn.1007-3043.2018.12.004  
中图分类号:TP181  
文献标识码:A  
文章编号:1007-3043(2018)12-0017-06

## Abstract:

With the information outbreak in the current Internet era, the old evaluation system needs to be iterated urgently. China Unicom jiangsu branch takes the intelligent and visualization as the main line and uses a variety of open source technologies to obtain multi-dimensional data through the establishment of independent, open and cooperative Internet Ecology. Relying on cloud virtualization platform, an intelligent, visual and open health assessment system for new big data wireless network is developed.

## Keywords:

Intelligence; Visualization; Big data; Independence; Openness; Cooperation

**引用格式:**于洋. 智能可视与开放的无线网络大数据评估体系[J]. 邮电设计技术, 2018(12): 17-22.

## 1 概述

当下移动通信网络正由移动互联网向“互联网+”升级迈进,万物互联已经不再是理论概念,而是正在每一个人身边发生的事实。随着网络结构复杂度的增加,数据的数量和复杂度也随之爆发性地提高,给数据分析、理解和呈现带来了巨大挑战。各种来源的结构化、半结构化、非结构化数据需要经过人工收集整理,利用不同手段清洗降噪,统一标准化以后,才能进行分析处理。在建模、检验、部署的过程中,也需要人工的干预。在此之后,最终信息必须流经一个最紧

的瓶颈——人脑吸收和处理新信息的能力所能达到的速度。人类视觉系统不足以满足人类以数据本身的形式来工作的要求,迫切需要可视化工具。

江苏联通前期自主应用大数据处理技术,建立了对网络全方位评价的体检式健康评估体系。在实际应用过程中,也充分认识到了数据的收集、处理、分析和呈现各个环节存在的瓶颈,通过应用智能化与可视化技术,应用互联网数据,对现有体系进行迭代,建立了智能、可视、开放的新型无线网络健康指数评估体系。

### 1.1 互联网转型需要开放架构

互联网时代是一个数据大爆发的时代,数据已经成为最宝贵的潜在资源和财富,如何挖取互联网上公

收稿日期:2018-11-02

开数据的有用信息、获取互联网公司数据,与运营商大数据关联,形成合作新生态,是传统电信运营商进行互联网转型的一个关键。

### 1.2 人工效能瓶颈需要智能化

移动互联网的业务日趋丰富,不同的用户、业务类型、场景对于网络的需求不尽相同,这就需要运营商从用户体验的角度出发,采用迭代、循序渐进的方法建立快速响应的机制,对用户感知的波动能够敏感地做出响应。

智能化通过将日常大量的重复性工作进行自动化处理,利用聚类算法、机器学习等智能化技术,将数据处理结果进行关联分析。从而将运营商的优化和维护人员从简单、低层次的工作中解放出来,专注于深层次的优化方法研究。

### 1.3 信息处理瓶颈需要可视化

一幅图胜过千言万语,人类从外界获得的信息约有80%以上来自于视觉系统,当大数据以直观的可视化的图形展现时,分析者往往能够一眼洞悉数据背后隐藏的信息,并转化为知识以及智慧。因此,可视化技术作为服务于计算机与用户之间的沟通纽带,为用户提供关于数据和知识的直观信息。

## 2 系统实现

传统网优对网络健康进行评估的时候,多是从KPI类指标、MR类指标、路测类指标等单类指标或某几项指标进行分析,无法全面地呈现网络的健康情况,工作量较大,重点不突出。通过大数据处理,搭建网络健康度评估体系,将各项指标进行归类,以指标的重要程度、小区场景的归属等分别设定系数,汇总至平台,最终得到各地(市)网络健康度的量化评分,通过得分可以直接观察各地(市)网络健康以及其变化情况。

通过引入互联网数据、可视化与智能化技术,对系统做了以下迭代。

### 2.1 互联网化场景数据评估

在初期版本的网络健康评估中,现有场景由人工维护,工作量巨大且准确性不够,缺乏更细致的用户感知特性。

江苏联通在体系迭代过程中,采取两条腿走路的办法,一方面利用开源软件自主爬取互联网场景数据,与KPI+、MR等数据相结合进行场景分析;另一方面联合腾讯、阿里等互联网企业,获取特定场景的用

户感知数据,进行分析。

互联网场景数据的提取分为2类:一类是以互联网地图为主的地理信息,有助于提升全省场景数据维护的效率;另一类是用户活动的特性数据,如热门店铺、人口热力图等。将此类数据与MR分布进行关联,可以判别场景,分析网络覆盖与用户分布之间的关系。目前已提取超过50万个场景和电子围栏信息。

此外,随着BATJ入股中国联通,借助混改契机,江苏联通积极与腾讯、阿里等互联网企业开展深度合作。通过数据共享,联合开展用户高感知区域地理化分析评估。现已将腾讯王者荣耀卡顿小区比例由3.62%降至0.17%;阿里弱网格由4.58%降至1.32%,用户感知大为提高。

### 2.2 全流程智能化挖掘

#### 2.2.1 智能数据准备:爬虫机器人

江苏联通选择应用基于Webdrive的网页爬虫机器人技术,通过获取厂家网页客户端HTML元素,设置动作模式,基于事件/时间触发自动获取指定数据源数据,归档处理后置入指定目录。

同时由于原始数据通常会有大量的空值、无效值、偏离值等种种异常值,必须进行降噪清洗才能使用。通过将文本处理程序语句打包嵌套入爬虫机器人,自动剔除异常值并统一成标准格式,就形成了完整的数据准备模块,大大压缩了数据收集整理时间,由此前人工处理的15天缩短为30 min,实现了机器人数据自动采集清洗(见图1)。

#### 2.2.2 智能数据挖掘:封装 workflow

数据挖掘需要人工进行数据准备,在建模、检验、部署的过程中,也需要人工的干预。一旦需求有所变化,又要重新进行数据挖掘。定制性差,校验周期长,存在人工效能瓶颈。

将数据挖掘的数据准备、建模、检验、部署等工作流全部利用javascript、Python等语言封装,设置条件,定时自动执行,以达到机器人智能挖掘的目的。同时,形成了一个数据准备、数据挖掘、网络优化、二次评估检验数据挖掘成效、再调整数据挖掘建模的闭环流程。整个流程以终为始,不断改进提高,真正有效地将数据挖掘应用到网络评估优化中(见图2)。

#### 2.2.3 智能数据输出:微信机器人

数据分析完成后,需要将评估数据和分析结果通报至省市优化人员和分管领导,原先通过报表和文档邮件报送的方式较为原始,通过应用selenium、itchat

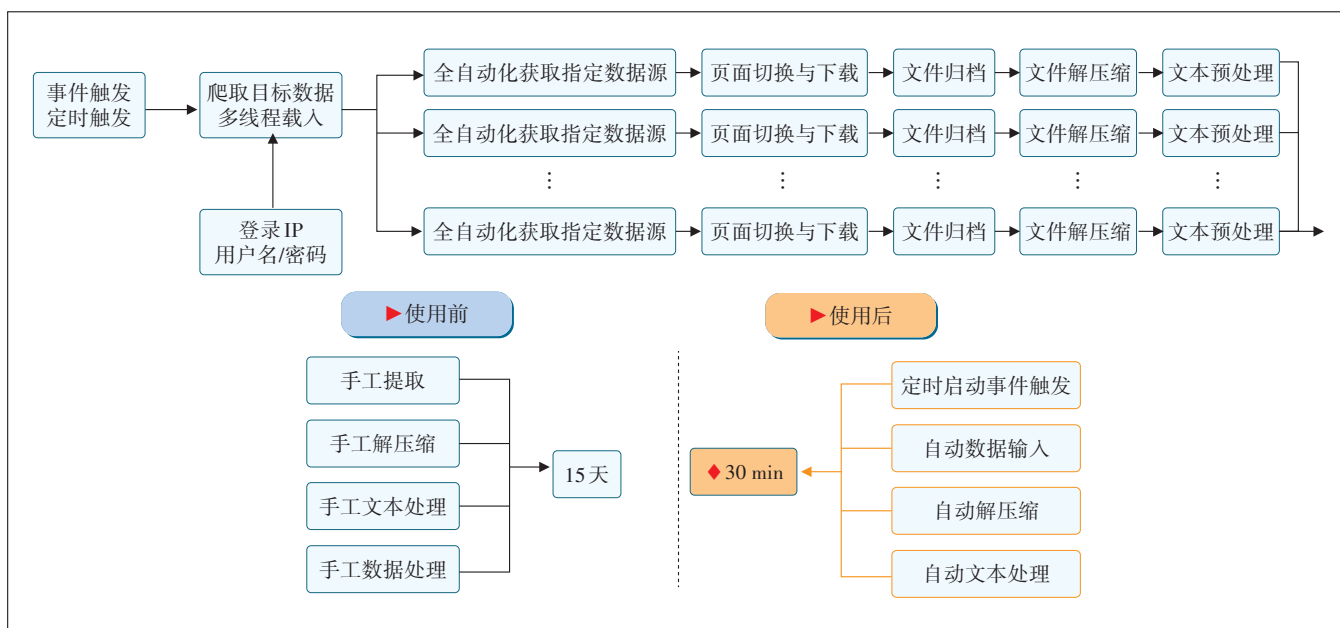


图1 基于Webdrive的网页爬虫机器人工具实现完整架构图

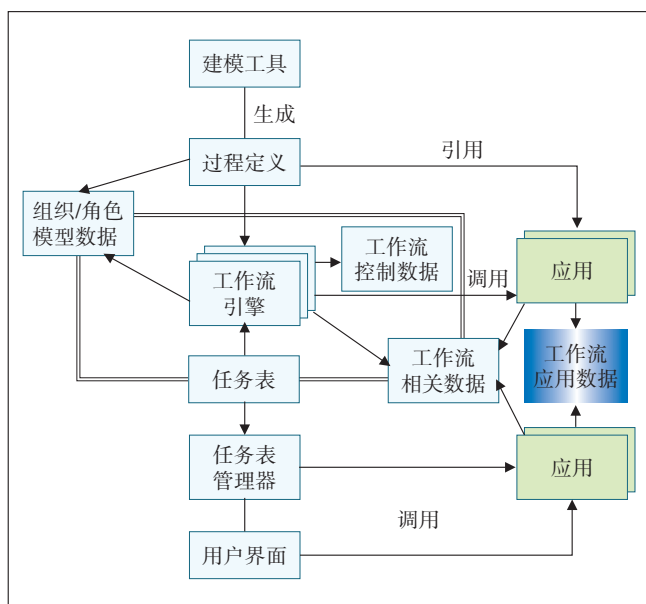


图2 智能数据挖掘

等当前较为流行的开源库,可以利用网页版微信 API, 编制微信机器人实现评估结果随时随地的查询。

微信机器人初始登录步骤如下(见图3)。

- 打开首页,分配一个随机 UUID。
- 根据该 UUID 获取二维码图片。
- 微信客户端扫描该图片,在客户端确认登录。
- 浏览器不停的调用一个接口,如果返回登录成功,则调用登录接口。
- 此时可以获取联系人列表,可以发送消息。然

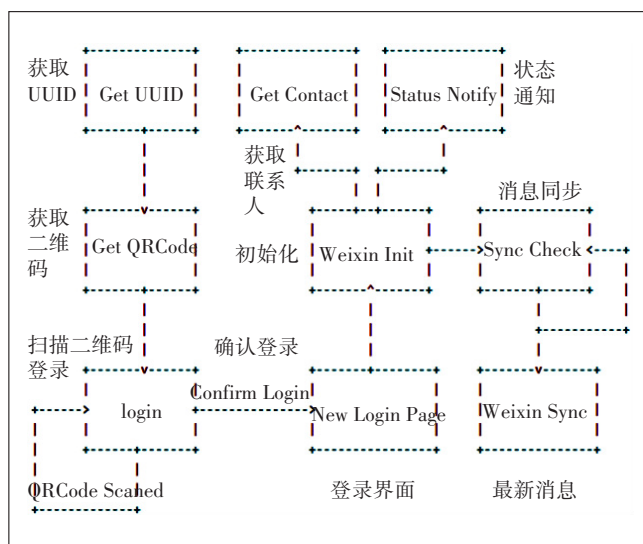


图3 微信机器人初始化流程图

后不断调用同步接口。

f) 如果同步接口有返回,则可以获取新消息,然后继续调用同步接口。

登录后,即可通过处理模块进行文本/语音自动识别,回复文本、图片等结果,还能够执行程序,建立通知群等功能,实现智能、自动的结果输出。主要实现框架如下(见图4)。

- 微信适配层:用于解析和封装微信消息。
- 主业务处理模块:处理解析后的消息。
- DAO 模块:数据访问对象(DataAccess Ob-

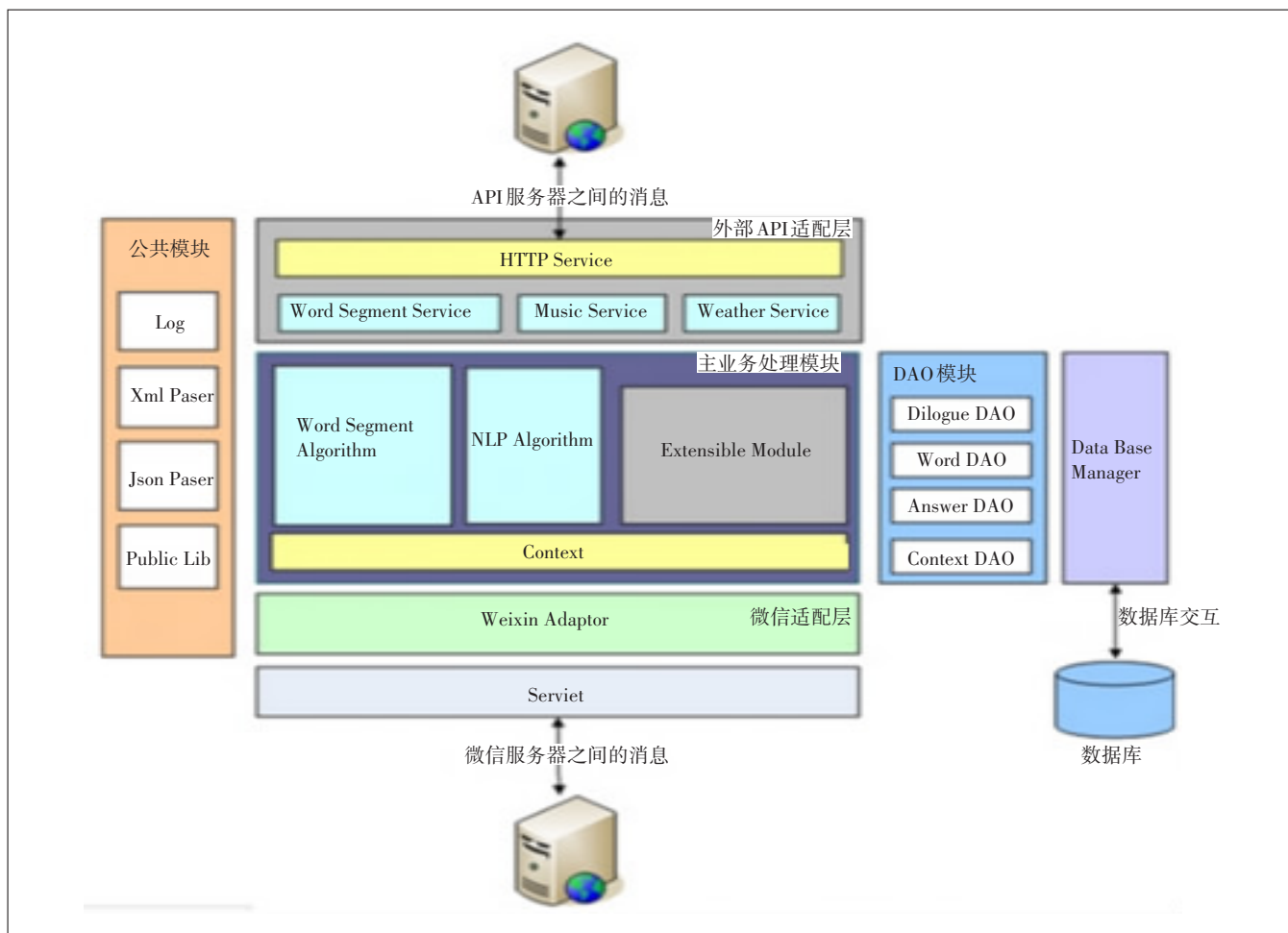


图4 微信机器人框架图

jects), 隔离业务逻辑和数据库操作。

d) 外部API模块: 通过url访问来获取使用第三方提供的API, 比如天气查询、搜歌、分词等。

e) 公共接口: 日志、公共库、Xml和Json解析。

目前, 江苏联通已将微信机器人用于日常指标发布和各种专项保障中, 结合可视化呈现技术, 获得了管理人员与一线优化人员的一致好评(见图5)。

### 2.3 可视化分析呈现

在生产生活中, 人的创造性不仅取决于人的逻辑思维, 而且取决于人的形象思维。重要的决定常常是来自决策者观察数据后产生的直觉。海量的数据只有变成可视化的形式, 才能激发人的形象思维。

江苏联通在应用大数据进行网络评估优化的过程中, 深感传统的分析输出形式不能直观地呈现网络问题, 非一线优化人员难以理解。常常需要借助Word、PPT等工具软件进行图表可视化再处理, 缺乏实时呈现能力, 耗时长、效果差, 在实现全流程智能化、



图5 微信机器人应用展示

自动化的过程中, 是一个瓶颈。

在对呈现痛点进行思考研究后,通过体系迭代,应用开源可视化工具,建立了与数据库直接连接,自动实时滚动输出的大屏监控呈现。该方案可以实现对全省网络按指标、区域、时间、重要等级等多维度的可视化直观呈现。

## 2.4 智能决策

### 2.4.1 机器学习智能化场景动态门限

在前期评估中,通过横向对比,发现某些区域整体早上的话务模型与早晚的话务模型相差较大,而其他热点区域话务模型基本无变化。这是因为不同场景、不同时段的话务模型存在区别,网络健康评估门限“一刀切”造成的,迭代后的网络健康度评估平台通过历史大数据机器自动对比进行机器学习,根据不同时段不同场景动态地调整评估门限,使得网络健康度评估更为准确。

机器学习步骤如下。

Step1: 获取每场景/每小区  $n \times 24$  h 的健康指数分布。

Step2: 去除异常样本。

Step3: 使用聚类算法分组,将指标分布特征相似的场景/时刻划分为一组。

Step4: 为同一组内的场景/时刻设定同样的门限。

Step5: 按照统计学原理和可配置的算法准则确定门限。

### 2.4.2 智能决策支持网络问题定位

移动通信网络问题纷繁芜杂,同一问题可能导致多项 KPI 异常,同一 KPI 异常也可能由不同问题引起,兼之要考虑区域、场景、时段、用户等各个维度,人工判断问题耗时长,对人员技能和经验要求较高,一直制约着评估后的优化处理效率。

通过人工移植专家经验,建立知识库和方法库,结合评估体系数据,搭建针对网络问题智能决策支持系统,可以直接给出网络问题类型判断和优化建议,大大减轻了一线优化人员的工作负担,也使网络问题的定位更加准确。

决策支持系统仅强调其支持作用,而不企图代替或者独立于决策者进行自动决策,本身既不试图对一个具体决策问题提供最终的决策结果,也不必事先注入确定的分析、求解步骤。主要好处是在非程序化决策问题求解过程中,人机各尽自己的优势来处理问题中适合各自处理的部分,在需要人的见识和判断的地方,决策人员(用户)被高度卷入,系统通过对数据和

模型的存取与运行,实现人机一体的决策求解。

系统构成主要由知识库、方法库、数据库、模型库构成(见图6)。

a) 模型库。模型库是共享资源,模型库中的模型可以重复使用,通过模型库可以将多个模型组合起来构成更大的模型。模型库的模型除了数学模型之外,还含有数据处理模型、图形/图像模型、报表模型、智能模型等。

b) 知识库。知识库中以一定形式存放着网络问题分析的相关知识,分为如下几种类型。

(a) 问题识别规则。

(b) 模型选择、构造、解释等规则。

(c) 专家领域的知识。

知识获取的任务是从知识源抽取知识,并转化为相应的计算程序,开发支援知识获取的技术和系统的自学能力。知识获取方式有人工移植、机器感知、机器学习等。现阶段使用人工移植专家经验的方式,构建网络问题识别的产生式规则,进行分析判断。后期将通过逐步应用机器学习的决策树等算法,将事实规则化,知识构造化,以便组成便于推理使用的表示方式,自动获得知识,进行知识积累。

c) 方法库。具有特定功能的模块化程序单位称之为方法,方法的可扩充集合称之为方法库。方法库是一种软件工具库,储存着一整套数值方法和非数值方法,例如方差方法、微分方程、数学和概率统计函数等。设立方法库的目的有2个,一是对数据库中的数据提供分析手段,二是为模型提供方法上的支持。

d) 数据库。数据库用于存储系统数据,基于现有的虚拟化云平台,可直接利用评估体系数据库,读取原始指标和评估数据。

## 3 结束语

江苏联通应用多种开源工具,实现了自主可控的网络优化,建立了智能可视与开放的大数据无线网络健康评估体系,截至2018年4月底,梳理各类网络问题42 387个,闭环40 390个,闭环率95.28%,KPI+各地(市)得分均值由87.2提升至89.8,网络质量提升明显。

自主优化能力的建立,既实现了人员由CT向ICT的转型,又节省了建设网络支撑平台的时间成本与金钱成本,累计节省人员与开发费用两千多万元。

此外,通过与互联网企业的深度合作,实现了互联网企业与江苏联通的共赢发展,打造出独有的、持

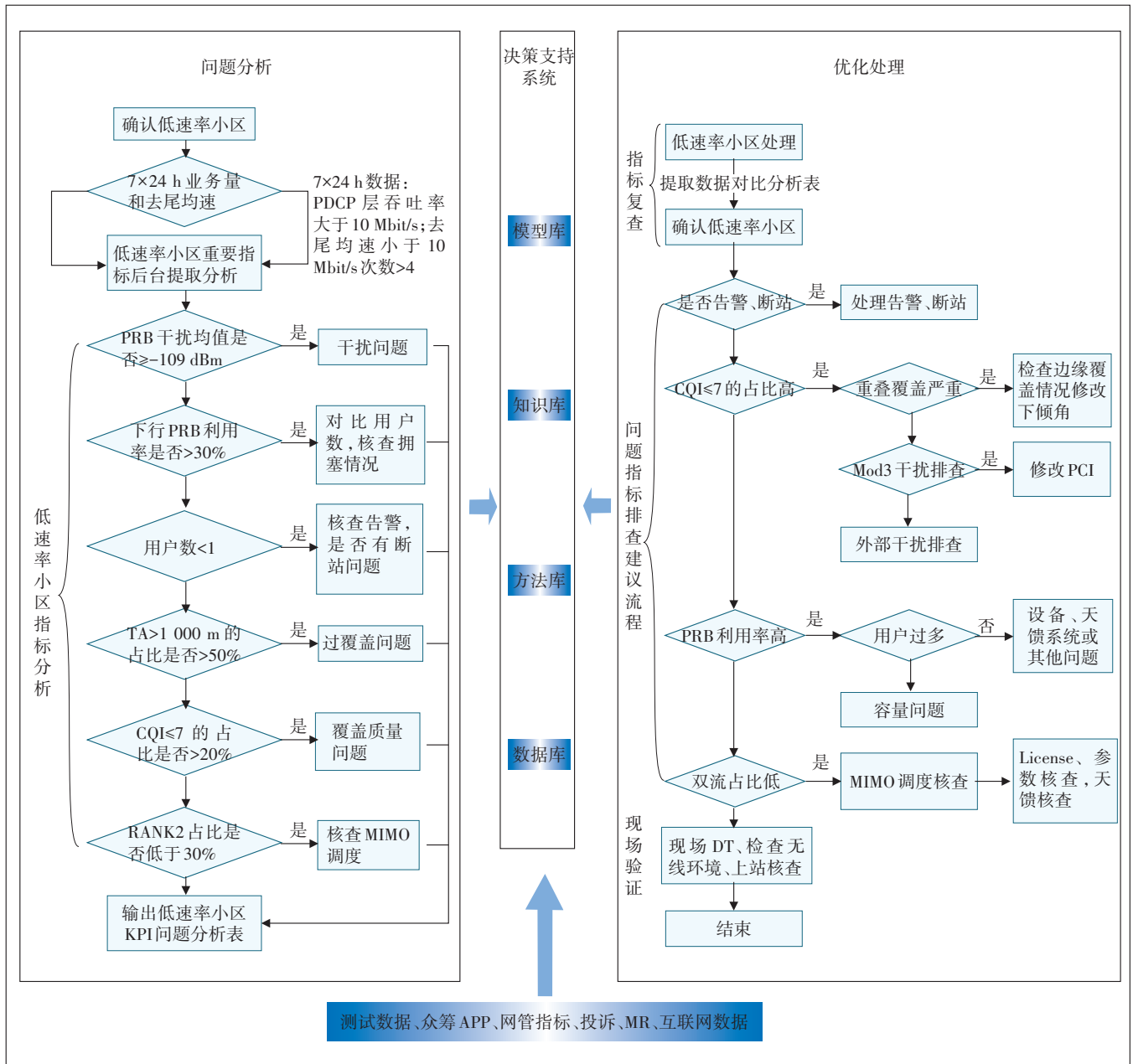


图6 决策支持系统构成示意图

续的新生态系统,网络评估优化更加贴近感知,实现了良好的社会效益。

参考文献:

[1] MEHMED K. 数据挖掘——概念、模型、方法和算法[M]. 陈茵,程雁,译. 北京:清华大学出版社,2003.  
[2] 陈丽萍. 基于决策树的面向对象分类方法研究[D]. 阜新:辽宁工程技术大学,2012.  
[3] 侯优优,隋严峰. 网络优化中的大数据应用[J]. 互联网天地,2014(1):34-37.  
[4] IKTOR M S, KENNETH C. 大数据时代[M]. 礁杨燕,周涛,译. 杭

州:浙江人民出版社,2013.

[5] 桑培罗布. 多元统计分析与数据可视化方法在调查分析中的应用[D]. 北京:中央民族大学,2011.

作者简介:

于洋,毕业于南京邮电大学,学士,主要从事移动通信网络优化工作。

