

基于大数据和AI的网络效能

Network Effectiveness Evaluation System
Based on Big Data and AI

评价体系


王 健,杜福之(中国联通山东分公司,山东 济南 250001)

Wang Jian, Du Fuzhi (China Unicom Shandong Branch, Ji'nan 250001, China)

摘 要:

为了解决现有维护人员效率低、配置和结构不合理的突出问题,中国联通山东分公司采用大数据分析和AI智能算法等创新手段,构建基于网络资源、业务收入、用户规模、地理分布等多维度立体网络维护效能标杆模型,建立科学的网络维护效能量化评价体系,重构和增强网络线队能力,引领整体维护效率的提升,实现降本增效。

关键词:

效能评价;量化模型;大数据;AI
doi:10.12045/j.issn.1007-3043.2021.01.002
文章编号:1007-3043(2021)01-0008-05
中图分类号:TN929.5
文献标识码:A
开放科学(资源服务)标识码(OSID): 

Abstract:

In order to solve the outstanding problems of low efficiency, unreasonable configuration and structure of existing maintenance personnel, China Unicom Shandong Branch adopts innovative means such as big data analysis and AI intelligent algorithm to build a multi-dimensional network maintenance efficiency benchmarking model based on network resources, service income, user scale and geographical distribution, and it establishes a scientific quantitative evaluation system of network maintenance efficiency, reconstructs and enhances the ability of network line team, which leads the improvement of overall maintenance efficiency, and achieves cost reduction and efficiency.

Keywords:

Effectiveness evaluation; Quantitative model; Big data; AI

引用格式:王健,杜福之.基于大数据和AI的网络效能评价体系[J].邮电设计技术,2021(1):8-12.

1 概述

随着网络技术的不断演进和通信行业竞争日益激烈,网络运营管理面临由“维护网络、关注质量”向“经营网络、关注效能”的运维新时代转型问题,如何提高劳动效能、激发人员潜能,助力公司走出高质量、高效能发展之路成为网络运营管理的一项重要课题。中国联通山东分公司积极探索,以效能改善为目标,以业务为维护效能评价的核心驱动要素,面向战略和转型需要,基于现有人员情况,采用大数据分析和AI

智能算法等创新手段,构建基于网络资源、业务收入、用户规模、地理分布等多维度基础数据,建立网络维护运营量化配置模型,引导全省网络维护线优化人员配置、调整人员结构、推进复用高效、激发队伍活力。

2 网络运营维护效能体系存在的问题

2.1 维护力量配备与网络演进不匹配

随着网络智能化演进,网络自动化维护作业能力大幅提升,现有人员仍按照传统的维护作业计划配置,人员配置过剩、维护效率低的情况日益突出,急需一套适应网络技术和企业发展需要的效能配置模型。

2.2 人员缺乏统一的配置标准

收稿日期:2020-11-02

中国联通山东分公司现有维护人员7000余人,各市公司在实施网络维护划小改革后,仍存在人员冗余、配置不均的问题,现有人员配置多继承历史因素,各地(市)、县区之间维护效率差距较大,缺乏统一的效能评价和配置标准,效率提升优化没有明确的目标。

3 建立网络维护效能评价体系

3.1 统一评价原则

a) 统一标准原则。全省按照统一的职责界面、支撑流程、业务承载进行人员效能分析建模,便于纵向管理穿透。

b) 复合建模原则。从网络资源、用户规模、收入规模等多维度建立复合配置模型,利于综合量化分析。

c) 标杆引领原则。尊重人员现状,通过对标分析,开展全量评价,在全省平均生产效能基础上,运用效能取优的算法,激发员工活力。

d) 数据强支撑原则。充分依托互联网化的管理手段,测算关键数据有系统支撑的,均从各类维护管理支撑系统提取,实现信息化支撑使能,控制数据测算风险,避免人为因素导致的测算偏差。

3.2 运用互联网化思维方式开展运维效能分析建模

充分考虑人员灵活复用、机构差异设置、能力锯齿分布的现状,消除单专业维护偏离度,发挥运维线整体协同效应,按照以下步骤对市公司本部(含城区)、县公司分别进行整建制建模分析。

a) 优选关键效能参数。按照维护动作聚焦、数据准确性验证、误差数据淘汰的步骤,甄选网络维护效能评价的关键效能参数。

b) 开展大数据效能分析。根据现有人员配置情况,通过大数据分析算法和AI智能建模,进行效能评价分析,制定效能标杆。

c) 聚类标杆配置模型。根据大数据分析效能评价结果,运用AI机器学习,结合多维度参数智能模拟测算人员配置。

3.3 县公司聚类分析建模

全省共123个县公司,由于样本量较大,采用聚类分析建模,创造性地建立W-IUT分析模型,确定标杆规则。

采用多维度星型建模方式,建立W-IUT效能分析模型,依托大数据分析,量化县公司维护产能。按照

紧密衔接业务支撑需求、综合考量维护管理难度的原则,选取收入规模、用户规模、地域分布3个元素作为聚类特征,将全省县区公司分为3类。按照工作量强相关、维护可量化、系统强支撑的原则,以设备量、光缆长度、机房数量等主要影响因素为效能评价模型,3类县(区)公司分别建模。W-IUT效能分析模型如图1所示。

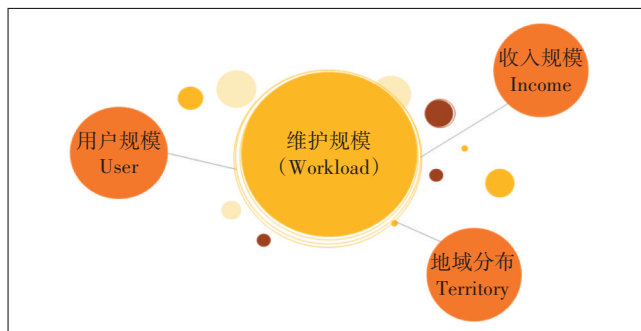


图1 W-IUT效能分析模型

3.3.1 对特征数据进行标准化,运用标准化算法将各维度特征数据统一到同一维度

收入规模、用户规模、地域分布3种聚类特征间的数据分布相差很大,如果直接进行分析会突出数值较高的指标在综合分析中的作用,从而使各个参数以不等权参与运算,因此,在聚类前,先对数据进行预处理,对各参数进行无量纲化处理,消除量纲影响和变量自身数值大小的影响,将不同维度的数据去量纲后归纳到同一维度上进行评价。综合考虑,采用数据处理中的标准化(Standard)算法,对原始数据进行标准差标准化,得到均值为0,标准差为1的数据序列,标准化后的数据符合标准正态分布,使分类结果更加合理。

$$X' = (X - X_{\text{mean}}) / \delta$$

式中:

X_{mean} —— 样本数据的均值

δ —— 样本数据的标准差

特征数据计算规则如表1所示。

表1 特征数据计算规则

调节系数	计算规则
用户规模	等效维护用户数=固网网上用户数(宽带)+固网网上用户数(固话) \times 0.25+固网网上用户数(IPTV) \times 0.5+移网网上用户数
收入规模	根据所支撑的市场营销中心归集收入规模
地域分布	覆盖面积/(核心机房数 \times 100+汇聚机房 \times 20+综合接入机房数+自有基站机房数)

使用 sklearn 进行标准化,代码如下:

```
from sklearn.preprocessing import StandardScaler
st = StandardScaler()
df2 = st.fit_transform(df1)
df2 = pd.DataFrame(df2)
df2.columns = ['user', 'area', 'income']
```

处理后得到标准化后的用户规模、地域分布、业务收入数据。

3.3.2 县(区)网格聚类,运用标准化后的特征数据对县(区)进行分类

在3个特征标准化基础上,采用K-means++算法进行聚类分析。

步骤1:随机选取一个样本作为第1个聚类中心c1。

步骤2:计算每个样本与当前已有聚类中心最短距离(即与最近一个聚类中心的距离),用 $D(x)$ 表示。这个值越大,表示被选取作为聚类中心的概率较大。最后,用转盘法选出下一个聚类中心。

步骤3:重复步骤2,直到选出 k 个聚类中心。

使用 sklearn 进行聚类,使用 pyecharts 进行绘图,代码如下:

```
from sklearn.cluster import KMeans
from pyecharts import Scatter3D, Page
from IPython.display import IFrame
km = KMeans(n_clusters=3).fit(df2)
df['分类-归一化'] = km2.labels_
page = Page()
scatter3D = Scatter3D(title="全省县(区)聚类-标准化", width=1000, height=600, title_pos='center')
scatter3D.add("分类 1", np.array(df1[km1.labels_==0]))
scatter3D.add("分类 2", np.array(df1[km1.labels_==1]))
scatter3D.add("分类 3", np.array(df1[km1.labels_==2]))
scatter3D.set_xaxis3d_name('等效用户数')
scatter3D.set_yaxis3d_name('地域分布')
scatter3D.set_zaxis3d_name('收入')
scatter3D.set_legend_orient('vertical')
scatter3D.set_legend_pos('80%')
page.add(scatter3D)
```

```
page.render('./1.html')
```

```
IFrame(src='./1.html', width=1200, height=800)
```

生成效果如图2所示。

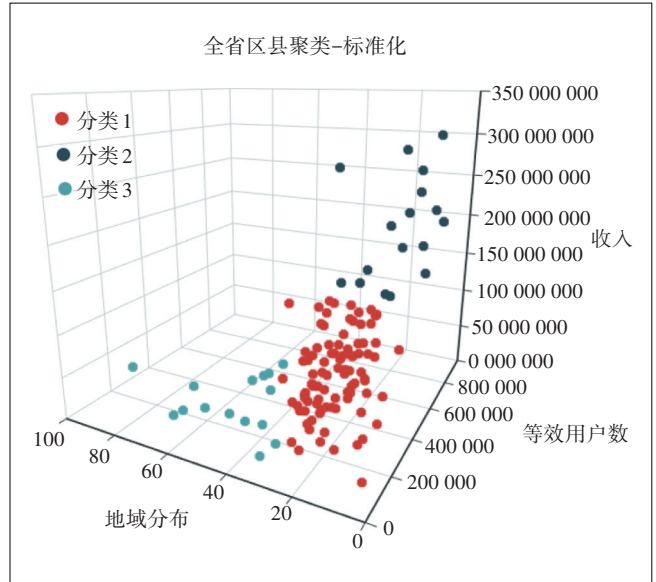


图2 区县聚类

通过以上算法,将全省123个县(区)网格按照类别特征相近的原则,划分为3类。第1类,收入和等效用户数较少,机房密度较大;第2类,收入和等效用户数较多,机房密度较大。第3类,收入和等效用户数较少,机房密度较小。

3.3.3 关键因子建模,在各分类中使用关键因子确定标杆

按照W-IUT效能分析模型,在对123个县(区)网格进行聚类分析后,根据与日常维护生产动作强相关、数据可量化的原则,选取归纳网格日常维护具体工作量化特征,优选5个关键建模因子,作为大数据分析建模的基础,分别在3个类别的县(区)中进行AI建模。县区关键因子如表2所示。

以效能优化为引导,取效能较优的80%县(区)作为标杆,进行机器学习建模,将各县区网格的Worklord关键五因子输入后,直接输出每个县(区)的结果。

使用多元线性模型,分别对3类县区公司进行建模,每类县区公司选取效能较优的80%作为标杆。

```
import pandas as pd
```

表2 县区关键因子

关键建模因子	客户专线数量(网元出租、数据、光纤、独享互联网专线、裸光纤)	宽带端口数	移网小区数	本地网光缆皮长	等效机房数量
数据来源	资源管理系统	号线资源管理系统	移网配置系统	资源管理系统	资源管理系统

```
import numpy as np
from sklearn.linear_model import LinearRegression
df1 = df[df['分类-标准化'] == 0]
df2 = df[df['分类-标准化'] == 1]
df3 = df[df['分类-标准化'] == 2]
lin1 = LinearRegression().fit(x1,y1)
lin2 = LinearRegression().fit(x2,y2)
lin3 = LinearRegression().fit(x3,y3)
y_pre1 = lin1.predict(x1)
y_pre2 = lin2.predict(x2)
y_pre3 = lin3.predict(x3)
df1['模型结果'] = y_pre1
df2['模型结果'] = y_pre2
df3['模型结果'] = y_pre3
df1['人员变动比例'] = (df1['模型结果'] - df1['
总计'])/df1['总计']
df2['人员变动比例'] = (df2['模型结果'] - df2['
总计'])/df2['总计']
df3['人员变动比例'] = (df3['模型结果'] - df3['
总计'])/df3['总计']
x1t=df1.sort_values(by='人员变动比例',ascending
=False).iloc[:82,:].iloc[:,3:8]
x2t=df2.sort_values(by='人员变动比例',ascending
=False).iloc[:15,:].iloc[:,3:8]
x3t=df3.sort_values(by='人员变动比例',ascending
=False).iloc[:14,:].iloc[:,3:8]
y1t=df1.sort_values(by='人员变动比例',ascending
=False).iloc[:82,:]['总计']
y2t=df2.sort_values(by='人员变动比例',ascending
=False).iloc[:15,:]['总计']
y3t=df3.sort_values(by='人员变动比例',ascending
=False).iloc[:14,:]['总计']
lin1 = LinearRegression().fit(x1t,y1t)
lin2 = LinearRegression().fit(x2t,y2t)
lin3 = LinearRegression().fit(x3t,y3t)
y_pre1t = lin1.predict(x1)
y_pre2t = lin2.predict(x2)
y_pre3t = lin3.predict(x3)
df1['前80%模型结果'] = y_pre1t
df2['前80%模型结果'] = y_pre2t
df3['前80%模型结果'] = y_pre3t
全省县区效能分析如图3所示。
```

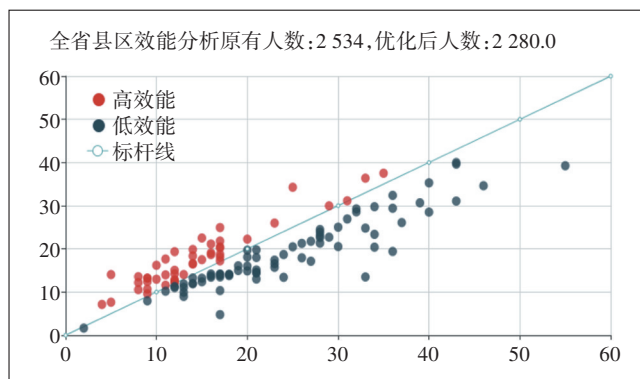


图3 全省县区效能分析

经过测算,需减配人员的县区网络共70个,不增不减的县区网络10个,需增配人员的县(区)网络43个。

3.4 市公司本部效能分析建模

市公司本部,包括本部核心侧维护人员和城区末梢维护人员,考虑到AI智能分析样本量较小,采用分类方式易产生较大偏差,直接采用多维度复合测算的方式。

3.4.1 采用多维度复合归一化算法

市区效能分析包括市本部核心侧维护人员和城区末梢维护人员,采用多维度复合测算的方式,梳理其维护职责、工作内容、量化特征,优选11个关键建模因子 $[x_1, x_2 \dots x_{11}]$,使用线性归一化算法,确定各因子的系数 $[a_1, a_2 \dots a_{11}]$,去除数据的单位限制,将其转化为无量纲的纯数值,便于不同单位和量级的11个因子进行比较和加权,根据效能指标计算人员配备比例。地(市)效能指标计算方式如下:

$$Y' = \sum_{n=1}^{11} a_n x_n$$

11个关键建模因子如表3所示。

根据不同专业维护特性,按照强相关的原则,针对不同因子钻取不同口径数据。宽带端口数、本地网光缆皮长、等效机房数量中的汇聚机房、综合接入机房和自有基站机房采用城区数据,其他因子采用全市数据。

3.4.2 以效能优化为引导,取效能较优的14个市公司作为标杆

通过机器学习建模,将各市区11个维度关键因子输入后,得到初步分析结果。

Python代码如下:

```
import pandas as pd
```


表3 市区关键因子

关键建模因子	数据来源
等效用户数	营服系统+综合网管系统
客户专线数量	资源管理系统
宽带端口数	号线资源管理系统
移网小区数	移网配置系统
本地网光缆皮长	资源管理系统
干线光缆皮长	通信能力报表
核心网用户数	综合网管系统
等效传输设备数	综合网管系统
等效数据设备数	综合网管系统
等效机房数量	资源管理系统
2019年完成收入	营服系统

```
import numpy as np
x = df.iloc[:, 2:]
y = df['总计']
a = x.sum()
b = x*a[0]/a
bb = b/np.array(y).reshape(17, 1)
sq = []
for bbcolumn in bb.columns:
    sq.append(y[bb.sort_values(by=bbcolumn, ascending=False).iloc[:, 14, :]].index].sum()/b.iloc[bb.sort_values(by=bbcolumn, ascending=False)].iloc[:, 14, :].index, :][bbcolumn].sum())
df['计算结果'] = (b*sq).sum(axis = 1)
17个市公司需减配人员的有10个市公司, 不增不减的有1个市公司, 需增配人员的有6个市公司。
```

4 运用效能评价体系引领整体效能提升

以“全面提升网络运营效能、提升人均维护效能、优化运维人员配置、激发基层员工活力”为目标, 结合网络维护划小改革, 制定效能优化的合理路径。

a) 现有人员数量>优化目标。

路径一: 精简现有人员至优化目标。结合维护认证上岗工作的开展, 淘汰认证不通过的低效能、低技能人员, 让不适应维护工作的人员退出维护专业线, 提升维护质量和维护效率。

路径二: 按照优化目标人数核减绩效包。继续实施量化薪酬抢包, 按照优化目标人数绩效薪酬的一定比例核减绩效包, 以工作量核定绩效, 实现多劳多得,

不劳不得, 解决“闲人”的问题。

b) 现有人员数量<优化目标。

路径一: 补充人员至优化目标。通过校园招聘、社招、其他人员转岗、外包等方式, 将通过维护认证的人员补充维护专业线, 满足工作需求。

路径二: 按照优化目标人数核增绩效包。按照需补充的人员对应的绩效薪酬, 取一定比例投入抢包薪酬池, 鼓励现有人员继续发挥“能人”效应, 高效运转, 实现企业员工双赢。

5 结束语

山东联通顺应互联网化转型的大趋势, 充分挖掘现有人员、资源、业务数据, 创新性地建立基于大数据分析和AI智能算法的运维效能评价模型, 通过多维度分析建模, 制定各市、县人员配置标杆, 引导全省维护人员优化维护效能, 预计全省运维人员劳动生产效率提升10%, 全面提升网络维护管理效益, 助力公司高效能发展。

参考文献:

- [1] PETER HARRINGTON. 机器学习实战[M]. 北京: 人民邮电出版社, 2013.
- [2] 加文·海克. scikit-learn 机器学习[M]. 2版. 北京: 人民邮电出版社, 2019.
- [3] 斋藤康毅. 深度学习入门 基于Python的理论与实现[M]. 北京: 人民邮电出版社, 2018.
- [4] 弗朗索瓦·肖莱. Python 深度学习[M]. 北京: 人民邮电出版社, 2018.
- [5] 阿曼多·凡丹戈. Python 数据分析[M]. 2版. 北京: 人民邮电出版社, 2018.
- [6] 韦斯·麦金尼. 利用Python进行数据分析[M]. 2版. 北京: 机械工业出版社, 2018.
- [7] WESLEY CHUN. Python 核心编程[M]. 3版. 北京: 人民邮电出版社, 2016.
- [8] 埃里克·马瑟斯. Python 编程 从入门到实践[M]. 3版. 北京: 人民邮电出版社, 2020.
- [9] 杰克·万托布拉斯. Python 数据科学手册[M]. 北京: 人民邮电出版社, 2018.

作者简介:

王健, 毕业于山东大学, 学士, 主要从事网络运营与优化工作; 杜福之, 毕业于南京邮电大学, 工程硕士, 主要从事网络运营与优化工作。

