

# 基于随机森林算法的地铁通信系统健康度动态评估方法

李三江<sup>1</sup> 江智麟<sup>1</sup> 林意为<sup>1</sup> 鲁怀科<sup>2</sup> 龚永军<sup>2</sup>

(1. 成都地铁运营有限公司, 610066, 成都; 2. 广西交控智维科技发展有限公司, 530001, 南宁)

**摘要** [目的] 地铁通信系统的健康度直接关系到地铁运营的效率和安全。针对现阶段地铁通信系统健康度评估方式存在评估结果片面性及评分标准主观性等问题, 需构建能够科学、准确地评估地铁通信系统运行状态和健康度的方法。[方法] 提出了一种基于随机森林算法的健康度动态评估方法, 结合深度学习进行通信健康度检测。设计评价指标体系, 收集相应数据, 在标签(真实健康度)不明确的情况下, 采用混合标签法确定标签并计算各个指标重要度; 适当地加入专家知识的影响, 根据指标权重和专家制定的指标自身评分标准, 计算通信系统各子系统健康度; 基于成都地铁维保数据进行了该方法试验验证。[结果及结论] 试验验证结果表明, 通过该方法可以得到有效的指标重要度, 可合理评估地铁通信系统健康度。

**关键词** 地铁; 通信系统; 健康度评估; 随机森林算法

**中图分类号** U283.2

**DOI:** 10.16037/j.1007-869x.2024.06.048

## Dynamic Health Evaluation Method for Metro Communication System Based on Random Forests Algorithm

LI Sanjiang<sup>1</sup>, JIANG Zhilin<sup>1</sup>, LIN Yiwei<sup>1</sup>, LU Huaike<sup>2</sup>, GONG Yongjun<sup>2</sup>

(1. Chengdu Metro Operation Co., Ltd., 610066, Chengdu, China; 2. Guangxi Jiaokong Zhiwei Technology Development Co., Ltd., 530001, Nanning, China)

**Abstract** [Objective] The health of metro communication system has a direct bearing on the efficiency and safety of metro operation. In view of the problems such as one-sided evaluation results and subjective evaluation criteria in current health evaluation method for metro communication systems, it is necessary to construct a method that can scientifically and accurately evaluate the operation status and health of the metro communication system. [Method] A dynamic health evaluation method based on random forests algorithm is proposed, and a communication health inspection is conducted in combination with deep learning. The evaluation index system is designed to collect the corresponding data. The mixed label method

is used to determine the label and calculate the importance of each index when the label (true health degree) is not clear. The health degree of each subsystem of communication system is calculated according to the index weight and the index self scoring standards established by experts, which properly incorporates the influence of experts' knowledge. Based on the maintenance data of Chengdu Metro, the method is tested and verified. [Result & Conclusion] The test verification results show that with the above method, effective index importance can be obtained and the health of metro communication system can be reasonably evaluated.

**Key words** metro; communication system; health evaluation; random forests algorithm

近年来,随着城市规模的不断扩张与人口流入量的增多,我国城市地下轨道交通得到高速发展<sup>[1]</sup>。地铁通信系统的评价也面临着诸多难题<sup>[2]</sup>。以成都地铁为例,现阶段各线路通信系统的健康度评估方式存在以下两个主要的问题:一是评估结果存在片面性,只能反映当时运行状态的好坏;二是评分标准存在主观性,无法完全客观判断评价指标对通信系统健康度的影响程度<sup>[3]</sup>。因此,构建一种能够科学、准确地反映地铁通信系统运行状态和健康度的模型迫在眉睫<sup>[4]</sup>。

由于历史运行数据保存不完善,地铁通信系统评价数据存在数量少以及正常与异常数据分布失衡等问题<sup>[5]</sup>。本文提出了一种基于随机森林的地铁通信系统健康度动态评估方法,评估过程分为4个阶段:第一阶段,设计评价指标体系,收集相应数据;第二阶段,在标签(真实健康度)不明确的情况下,采用混合标签法确定标签;第三阶段,基于随机森林算法计算各个指标重要度;第四阶段,适当加入专家知识的影响,根据指标权重和专家制定的指标自身评分标准,计算地铁通信系统健康度。

## 1 通信系统健康度评价指标体系

### 1.1 问题描述

地铁通信子系统的健康度,通常可以使用故障率来评估<sup>[6]</sup>。采用专家制定的指标自身评价标准对当时的数据进行评估,及时、准确地反映相应系统的健康度<sup>[7]</sup>。

某一个地铁通信子系统有多种设备,会产生不同类型的数据,这些数据从不同程度上可以反映出该子系统健康状况。设定指标数据集  $E = [e_1, e_2, \dots, e_u]$ , 其中包含  $u$  个指标,通过建立模型  $f$ , 确定每一个指标的重要度  $I$ :

$$I = f(E) \quad (1)$$

$$I = \{i_1, i_2, \dots, i_u\} \quad (2)$$

式中:

$i_u$ ——第  $u$  个指标的重要度。

### 1.2 评价指标体系

由于不同通信子系统之间及同一子系统之间均存在厂家设备差异<sup>[8]</sup>,需结合地铁通信系统运行设备的相关数据特征及专家知识经验,建立面向不同通信子系统的通用有效的健康度评价指标体系。

以成都地铁通信系统为例,该系统主要包括传输系统、无线通信系统、程控交换系统及电源系统 4 个子系统,该系统的健康度评价指标体系归纳为性能指标、状态指标与告警指标 3 种类型<sup>[9]</sup>。

### 1.3 混合标签处理

针对原始健康度评分置信度无法得到客观衡量与保证的问题,在确定各专用通信子系统的健康度评价指标体系后,采集相应的数据,通过利用数据自身信息作为特征实现科学的标签化处理,避免伪标签造成逻辑错误。主要步骤如下:

步骤 1 确定评价指标集  $F$ 。评价指标集由性能指标集  $P$ 、状态指标集  $S$ 、告警指标集  $A$  等 3 类子指标集组成,其中每一类子指标集由所属同一类型评价指标组成。

步骤 2 确定每一个指标阈值范围,基于历史数据,联合行业标准、设备厂家标准以及设定的异常判断标准,划分各个指标的多级阈值或状态范围  $T_{i,r}$ :

$$T_{i,r} = \{t_0, t_1, \dots, t_s\}, \quad i \in F \quad (3)$$

式中:

$i$ ——指标类型,  $F = (P, S, A)$  中的一种;

$T_{i,r}$ —— $i$  类指标的第  $r$  个指标的阈值或状态

集合;

$t_s$ ——该指标第  $s$  个等级阈值范围,  $s$  值越大越异常。

步骤 3 确定每一个指标自身标签  $l_{ij}$ 。根据多级阈值范围确定各个指标对应的数值标签。阈值级数决定指标标签数值范围,从 0 开始计数,标签数值越大,表明数据异常程度越高。

步骤 4 计算每一类指标综合阈值范围  $T_i$ 。确定某条评价数据中各个指标的自身标签后,对于同一类型指标,将其下所有指标的最大标签值进行相加,并通过向上取整的计算方式获取均值,作为此类型指标的综合标签阈值。

步骤 5 确定每一类指标综合标签。若所属同一类指标下的全部指标自身标签相加之和小于其综合标签阈值,设定对应类型指标的综合标签为 0,表示正常,否则为 1,表示异常。

步骤 6 构造每一条评价数据综合标签。每条评价数据的综合标签为各类指标综合标签之和。

## 2 通信系统健康度评估

### 2.1 随机森林算法

在本文的地铁通信系统数据集中,以各个指标作为特征,以健康等级作为分类类别构建随机森林。基于随机森林模型预测健康度等级,并且根据基尼系数计算出每个特征对模型预测健康度等级的贡献,也就是每个指标的重要度。基尼系数计算方法:

$$G(p) = \sum_{k=1}^K p_k(1 - p_k) = 1 - \sum_{k=1}^K p_k^2 \quad (4)$$

式中:

$G(p)$ ——基尼系数;

$K$ ——类别总数;

$p_k$ ——类别  $k$  的样本权重。

第  $b$  棵决策树在节点  $m$  上的重要性即节点  $m$  分枝前后的基尼系数变化量  $V_{bm}$  为:

$$V_{bm} = G_{bm} - G_{bl} - G_{br} \quad (5)$$

式中:

$G_{bm}, G_{bl}, G_{br}$ ——分枝后 3 个新节点的基尼系数。

如果特征  $V_{jbm}$  在决策树  $b$  中出现的节点在集合  $M$  中,那么第  $j$  个决策树单一重要度  $V_j$  为:

$$V_j = \sum_{b=1}^n \sum_{m \in M} V_{jbm} \quad (6)$$

式中:

$n$ —随机森林中树的总数;

$V_{jbm}$ ——集合  $M$  中的每一个值。

最后将所有求得的重要性进行量纲一化:

$$Y_j = \frac{V_j}{\sum_{b=1}^B V_i} \quad (7)$$

式中:

$B$ ——决策树总数;

$Y_j$ ——单一重要度  $V_j$  在总和重要度的占比。

## 2.2 权值动态更新算法

各个指标重要度的更新流程如图1所示。对于地铁通信子系统的健康度进行评估,需设置其中每个指标的重要度。若只考虑从数据出发计算指标重要度,由于地铁通信系统的特殊性,则存在严重的数据不平衡情况。因此,除了基于随机森林算法计算出指标重要度,还要结合专家知识进行更新。对于任意的一个指标  $X_j$  的重要度  $I_j$ ,由通过随机森林算法计算出的重要度值  $Y_j$  和专家确定的重要度  $E_j$  组成,取值范围为  $[0, 100]$ ,计算公式为:

$$I_j = \alpha Y_j + (1 - \alpha) E_j \quad (8)$$

式中:

$\alpha$ —— $(0, 1)$  区间的一个常数。

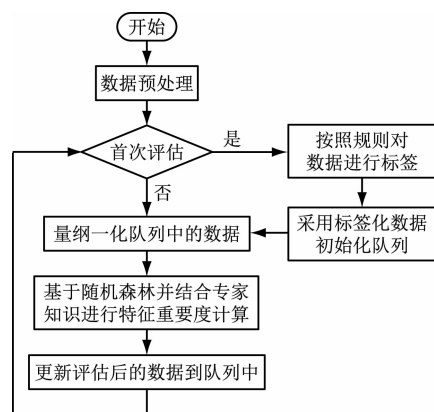


图1 权值动态更新算法流程图

Fig. 1 Flowchart of weight dynamic update algorithm

## 2.3 通信系统健康度评价

1) 单个网元的每个指标自身评分。基于历史数据,联合行业标准、设备厂家标准以及设定的异常判断标准,对各个指标划分多级阈值范围。根据指标状态(范围)均匀地评分,最差为0分,依次累进。例如无线系统中的天馈驻波比  $X$ ,可分为3个范围: $X \leq 1.5$  为正常,得分为满分; $1.5 < X \leq 2.0$  为亚正

常,可得一半分数; $X > 2.0$  则此项指标得分为0。

2) 单个网元的每个子系统评分。单个网元的每个子系统的分数由特征重要度以及每一个指标的自身评分确定。最终子系统分数  $H_D$  为:

$$H_D = \sum_{j=1}^n I_j p_j \quad (9)$$

式中:

$p_j$ ——指标自身评分比例。

根据子系统分数  $H_D$ ,确定单个网元子系统健康等级,如表1所示。标签等级越小,评价分数越高意味着健康等级越高,即系统不出故障或是系统的故障概率小。

表1 网元单个通信子系统健康度评价等级划分

Tab. 1 Classification of health evaluation levels for single communication subsystem of network element

标签等级	健康度等级	$H_D$ 取值范围
0	优	90 ~ < 100
1	良	75 ~ < 90
2	中	60 ~ < 75
3	差	< 60

3) 单条线路的子系统评分。地铁通信子系统的每一条线路的健康度评分是来自每一个网元的该子系统评分的平均值,计算公式为:

$$H = \frac{\sum_{w=1}^Q H_{Dw}}{Q} \quad (10)$$

式中:

$H$ ——子系统评分的平均值;

$H_{Dw}$ ——每一个子系统分数;

$Q$ ——某一线路中网元的个数。

## 3 试验验证

### 3.1 试验数据

试验数据来源于成都地铁维保公司的通信系统数据,包含传输系统、无线通信系统、程控交换系统及不间断电源系统4个子系统的数据。各个子系统数据指标以及数据量如表2所示。不间断电源系统暂无状态指标。本文以数据量最多的传输系统为试验示例。

### 3.2 试验结果

#### 3.2.1 传输系统指标重要度计算

传输系统原始数据共计2 572条,对原始数据

采用混合标签法进行处理,得到每条数据的健康度等级标签,其结果表明各种标签之间数据量差别比较大。

为了尽可能准确地计算各指标重要度,采用 SMOTE (合成少数类过采样技术) 采样方式对原始数据进行均衡化后,基于随机森林算法预测数据标签。在设置专家知识时,参考了原传输系统的评价模型,其中首次差和重要度设置为 20,前次差和重要度设置为 10,一、二、三级告警的重要度分别设置为 30、20、10,越限状态重要度设为 10。

首先按时间顺序将原始数据分为两部分,各包含 1 286 条数据,采用混合标签法对数据进行标签化。之后的每一次迭代添加 128 条数据并丢弃原队列中的前 128 条数据,采用最近一次的指标重要度划分健康度等级数据。

在迭代的过程当中,为减少主观因素对结果的影响,式(8)中  $\alpha$  首次取值为 0.5,之后每次取值为 0.4。由于无越限状态的异常数据,在首次评估时计算重要度为 0,最终的重要度仅仅由专家知识决定。其余各个指标的重要度计算结果如表 3 所示。相较于首次计算,终次计算的首次差和与前次差和的重要度均有提高,尤其是后者,提升了 10 dBm 左右。指标重要度与迭代次数关系如图 2 所示,当迭代至第 5 次时,结果趋于稳定,最终的首次差和、前次差和、一级告警次数、二级告警次数及三级告警次数为表 3 中的终次计算结果。

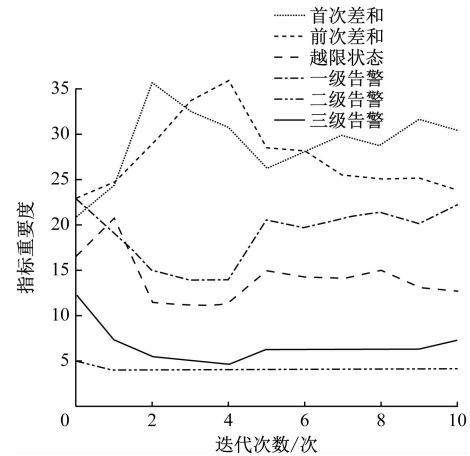


图 2 指标重要度与迭代次数关系

Fig. 2 Relationship between index importance and iteration times

相对于专家知识来说,最终的结果使得首次差和略微增加,前次差和的比重增大近两倍,各级告警次数均有不同程度的减少。从试验结果来看,最终的指标重要度曲线都将趋于平稳,可见模型具有较高的鲁棒性。

### 3.2.2 评分实践

根据不同轮次的指标重要度计算结果和评分标准进行评分实践,例如在第 2 次迭代中,钟楼站的 2021 年 1 月 28 日的的数据当中,仅有首次差和发生异常,根据  $3\sigma$  定律,此时为亚正常状态,扣除掉该指标重要度一半的分数,即扣除 12 分,其余指标均不扣分,最终健康度评价分数为 88 分。节选的部分评价数据如表 4 所示。

表 2 各通信子系统数据集

Tab. 2 Dataset of communication subsystems

通信子系统	性能指标	状态指标	告警	数据量/次
传输系统	光功率首次差和、光功率前次差和	越限状态	当日一级告警发生次数、当日二级告警发生次数、当日三级告警发生次数	2 572
无线通信系统	驻波比、基站温度、前向功率波动值	主用 TSC 模块故障状态、备用 TSC 模块故障状态、BR 发射状态、风扇工作状态	当日一级告警发生次数、当日二级告警发生次数、当日三级告警发生次数	1 950
程控交换系统	MPU 占用率、TPU 占用率、APU 占用率	关键状态检查项数据包括运行年限和系统承载量状态	当日一级告警发生次数、当日二级告警发生次数、当日三级告警发生次数	388
不间断电源系统	总电压、一般负载各组总电压、温度均值、温度最小值、温度最大值		当日一级告警发生次数、当日二级告警发生次数、当日三级告警发生次数	570

注:MPU—主处理单元;TPU—张量处理单元;APU—加速处理器;TSC—基站控制器;BR—基站发信机。

原有的评价方案着重关注光功率的变化大小,根据经验划分范围并设定指标重要度,本文模型基

于历史数据和专家知识,从数据角度分析所有指标,提升了评价指标的客观性和准确性。比如金科



表3 基于随机森林的指标重要度

Tab.3 Index Importance based on Random Forest

项目	首次差和/ dBm	前次差和/ dBm	越限状态/ dBm	一级告警/ 次	二级告警/ 次	三级告警/ 次
专家知识特征重要度	20	20	20	20	15	5
首次计算	21	27	10	18	14	10
终次计算	24	32	8	16	16	4

表4 节选部分数据的评分

Tab.4 Partial data scoring

迭代 轮次	时间	站点	首次差和/ dBm	前次差和/ dBm	越限状态	一级告警 次数/次	二级告警 次数/次	三级告警 次数/次	原有评分/ 分	本模型 评分/分
2	2021-01-28	钟楼站	1.4	0	0	0	0	0	80	88.0
4	2021-03-18	金科北路站	0.2	0.1	0	0	0	0	80	100.0
5	2021-04-15	惠王陵站	0.6	0	0	0	0	1	80	98.0
7	2021-06-17	团结新区站	1.5	0.4	0	0	0	0	80	70.5

北路站 2021 年 3 月 18 日的原有评分为 80 分,主要是由于光功率的首次差和大于了限定的阈值,本文模型通过统计分析历史数据,认为首次差和为 0.2 dBm 与前次差和为 0.1 dBm 都不足以对传输子系统的稳定性造成影响,故评分为 100 分。根据成都地铁维保公司的实际应用来看,本文模型得出的评分标准能够准确、快速反映系统的健康度,具有较高的实用价值。

#### 4 结语

本文提出了一种基于随机森林算法的地铁通信系统健康度评价方法,综合考虑影响设备健康状态的关键因素,通过构建科学合理的综合评估模型实现对设备健康度评估,避免评估的主观性。根据评估结果能够对设备的检修决策进行优化,提高地铁通信系统维护与管理工作的效率与质量。创造性地提出了动态更新方式,同时加入专家知识,可以在一定程度上减少数据不均衡带来的误差。试验结果表明,通过该方法可以得到有效的指标重要度,具有较高的准确性和鲁棒性。

#### 参考文献

- [1] 姚云峰,伍逸夫,冯玉光,等. 装备健康状态评估方法研究[J]. 现代防御技术, 2012, 40(5): 156.  
YAO Yunfeng, WU Yifu, FENG Yuguang, et al. Health condition assessment of equipment[J]. Modern Defence Technology, 2012, 40(5): 156.
- [2] 刘君,赵立进,黄良,等. 基于 TOPSIS 和灰色关联分析的变

压器状态评价方法[J]. 电力科学与技术学报, 2019, 34(4): 63.

LIU Jun, ZHAO Lijin, HUANG Liang, et al. State evaluation method for power transformer based on the topsis and grey relational analysis[J]. Journal of Electric Power Science and Technology, 2019, 34(4): 63.

- [3] 伏玉明,刘伯鸿,宋爽. 基于模糊综合评判的转辙机健康评估研究[J]. 铁道科学与工程学报, 2017, 14(5): 1070.

FU Yuming, LIU Bohong, SONG Shuang. Research on health assessment methods for switch machine based on fuzzy comprehensive evaluation[J]. Journal of Railway Science and Engineering, 2017, 14(5): 1070.

- [4] 杨华强,王立琼,安向昕,等. 基于主客观赋权的设备多属性健康度建模[J]. 机械设计与研究, 2021, 37(4): 154.

YANG Huaqiang, WANG Liqiong, AN Xiangxin, et al. Multi-attribute health model of equipment based on subjective and objective integrated weighting[J]. Machine Design & Research, 2021, 37(4): 154.

- [5] 雷云鹏,姜萍萍,范东明. 城轨计轴设备健康状态评估系统[J]. 铁道通信信号, 2019, 55(11): 90.

LEI Yunpeng, JIANG Pingping, FAN Dongming. Health assessment system of axle counter in urban rail transit[J]. Railway Signalling & Communication, 2019, 55(11): 90.

- [6] 马元,王鹏,孙瑞. 基于动态权值的转辙机健康度评估方法[J]. 铁道通信信号, 2021, 57(7): 25.

MA Yuan, WANG Peng, SUN Rui. Health evaluation method of switch machine based on dynamic weight[J]. Railway Signalling & Communication, 2021, 57(7): 25.

- [7] 杨利伟,邢雯雯,张莉平,等. 基于 GA 优化 BP 神经网络模型的污水管道系统健康状况评估[J]. 给水排水, 2021, 57(9): 123.

(下转第 275 页)

位过程中误差包络的上界和下界。本研究在河北北京车智能制造基地试车线对该方法进行了工程验证。验证结果表明:在车速为 80 km/h 时,列车最短追车距离仅为 100 m,与采用原算法的追车距离相比,缩短量最高达 30 m。由此可见,该列车定位误差估算方法能够在保证 SIL4 安全等级的前提下有效缩短虚拟编组追车距离,有效降低传统误差包络上下界计算的保守性,为追车场景中领头车和跟随车的 MA 及 EBI 等参数计算提供依据。

## 参考文献

- [1] 唐涛, 罗啸林, 刘宏杰, 等. 城轨列车虚拟编组安全防护与运行控制技术研究进展[J]. 科技导报, 2023, 41(10): 31.  
TANG Tao, LUO Xiaolin, LIU Hongjie, et al. Research review of the protection and operation technology for virtually coupled train sets in metros [J]. Science & Technology Review, 2023, 41(10): 31.
- [2] NOLD M, CORMAN F. Dynamic train unit coupling and decoupling at cruising speed: systematic classification, operational potentials, and research agenda[J]. Journal of Rail Transport Planning & Management, 2021, 18: 100241.
- [3] 刘宏杰, 唐涛, 张艳兵, 等. 城轨虚拟编组关键性能指标及技术探讨[J]. 都市轨道交通, 2023, 36(1): 28.  
LIU Hongjie, TANG Tao, ZHANG Yanbing, et al. Discussion on the key performance indicators and technologies of virtual coupling in metros[J]. Urban Rapid Rail Transit, 2023, 36(1): 28.
- [4] GALLO F, DI FEBBRARO A, GIGLIO D, et al. A mathematical programming model for the management of carriages in virtually

coupled trains[C] //2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). Rhodes, Greece. New York: IEEE, 2020: 1.

- [5] PARK J, LEE B H, EUN Y. Virtual coupling of railway vehicles: gap reference for merge and separation, robust control, and position measurement[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(2): 1085.
- [6] DI MEO C, DI VAIO M, FLAMMINI F, et al. ERTMS/ETCS virtual coupling: proof of concept and numerical analysis[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(6): 2545.
- [7] ZHANG Z, SONG H, WANG H, et al. Cooperative multi-scenario departure control for virtual coupling trains: a fixed-time approach[J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 8545.
- [8] WANG Q, CHAI M, LIU H, et al. Optimized control of virtual coupling at junctions: a cooperative game-based approach[J]. Actuators, 2021, 10(9): 207.
- [9] QUAGLIETTA E, WANG M, GOVERDE R M P. A multi-state train-following model for the analysis of virtual coupling railway operations[J]. Journal of Rail Transport Planning & Management, 2020, 15: 100195.

· 收稿日期:2023-09-21 修回日期:2024-01-24 出版日期:2024-06-10  
Received:2023-09-21 Revised:2024-01-24 Published:2024-06-10  
· 第一作者:范莹,高级工程师,fanyingnew@163.com  
通信作者:王志卓,助理工程师,wangzhizhuo@vip.qq.com  
· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议  
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

## (上接第 269 页)

- YANG Liwei, XING Wenwen, ZHANG Liping, et al. Evaluation of the health status of sewage pipeline system based on GA optimized BP neural network model[J]. Water & Wastewater Engineering, 2021, 57(9): 123.
- [8] 邵威, 许厚棣, 金贤龙. 基于模糊准则的雷达健康评估方法研究[J]. 雷达科学与技术, 2021, 19(2): 152.  
SHAO Wei, XU Houdi, JIN Xianlong. Research on fuzzy rules based health assessment method of radar[J]. Radar Science and Technology, 2021, 19(2): 152.
- [9] 朱毅, 陈兴蜀, 陈敬涵, 等. 基于模糊综合评价模型的 DNS

健康度评估[J]. 信息安全, 2018(4): 65.

ZHU Yi, CHEN Xingshu, CHEN Jinghan, et al. DNS health assessment based on fuzzy comprehensive evaluation[J]. Netinfo Security, 2018(4): 65.

· 收稿日期:2023-10-30 修回日期:2024-01-25 出版日期:2024-06-10  
Received:2023-10-30 Revised:2024-01-25 Published:2024-06-10  
· 通信作者:李三江,高级工程师,86253160@qq.com  
· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议  
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license