面向5G通信基站管理系统性能评估的测试用例分布规律研究

李怿臻[1]，孙一丁[1]，邹金孜[1]，刘宇航[1]，秦涛[1]，李欢[2]，杨小燕[2]，梁涛涛[2]
（1. 西安交通大学智能网络与网络安全教育部重点实验室，710049，西安；
2. 华为技术有限公司西安研究所，710075，西安）

摘要：针对通信基站管理系统目前多采用融合人工经验的测试方法，无法模拟长周期且发生时间合理的测试用例问题，提出了一种基于基站告警日志的测试用例建模方法。首先，以基站日志产生时间间隔为研究对象，从数据整体、设备类型、日志类型、设备和日志类型等维度分析发现日志产生时间间隔服从幂律分布的规律；然后，对比分析最小二乘法、极大自然估计与最大后验估计3种不同的参数估计算法，以估算误差为依据，发现最小二乘法的拟合度更高且残差最小，拟合优度与平均绝对百分误差的平均值分别为0.96和3.5%；最后，采用最小二乘法对日志时间间隔进行幂指数估算，保留局点拟合优度大于0.7的告警组成测试用例。该模型应用于全国多个不同城市通信基站的告警日志间隔分布估算，实验结果表明：所提模型可以实现85%以上日志分布规律计算，为编排贴近实际运行环境的测试用例奠定基础，进一步提升测试结果的可靠度。

关键词：通信基站管理系统；测试用例；幂律分布；参数估计；最小二乘法
中图分类号：TP391.9 **文献标志码**：A
DOI：10.7652/xjtuxb202303018 **文章编号**：0253-987X(2023)03-0193-09

Research on Test Cases Distribution for Performance Evaluation of 5G Communication Base Station Management System

LI Yizhen[1], SUN Yiding[1], ZOU Jinzi[1], LIU Yuhang[1], QIN Tao[1],
LI Huan[2], YANG Xiaoyan[2], LIANG Taotao[2]
(1. MOE Key Lab for Intelligent and Network Security, Xi'an Jiaotong University, Xi'an 710049, China;
2. Xi'an Research Institute of Huawei Technology Co., Ltd., Xi'an 710075, China)

Abstract: The communication base station management system mostly adopts an empirical test method at present, which cannot simulate the test cases with a long period and reasonable occurrence time. To solve this problem, a test case modeling method based on the base station alarm log is proposed. Firstly, taking the log generation time interval of the base station as the research object, it is found that the log generation time interval obeys power-law distribution by analyzing data entirety, device type, log type, and device combination log type. Then, the least square method, maximum likelihood estimation, and maximum posterior estimation are compared and analyzed. Based on the estimation error, it is found that the least square method has the fitting of a higher degree and the smallest residual error, the value of goodness of fit is 0.96%, and mean

收稿日期：2022-09-01。 作者简介：李怿臻(1994—)，男，博士生；秦涛(通信作者)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(62172324，62102310)；陕西省重点研发计划资助项目(2023-YBGY-269)。
网络出版时间：2022-11-17 网络出版地址：https://kns.cnki.net/kcms/detail/61.1069.T.20221116.0920.002.html

absolute percentage error is 3.5%. Finally, the least square method is used to estimate the log time interval by power exponent, and the alarms with the value of goodness of fit greater than 0.7 are reserved to form test cases within a location. The model is applied to estimate the interval distribution of alarm logs of communication base stations in different cities throughout the country. The experimental results show that the proposed model can calculate more than 85% of log distribution law, which lays a foundation for arranging test cases close to the actual operating environment and further improves the reliability of test results.

Keywords: communication base station management system; test case; power-law distribution; parameter estimation; least square method

随着 5G 通信技术的成熟,各类终端规模增长迅速,给网络的高效管理和安全运行带来了挑战[1]。基站的高效管理是确保网络正常运转的基础,基站管理信息系统是进行基站维护、运营和信息管理的一个综合平台,进而实现网络负荷调优、设备故障监控等功能。管理系统上线运行前,需要进行系统化测试,确保投入运营后在系统功能、资源占用等方面的正确性,降低产生故障的风险[2]。通过部署前的测试,可以预先知道系统的承受力,为网络发展规划和管理信息系统的改进提供有力依据[3]。

目前的测试流程多采用融合专家经验的测试方法,其中最常见的测试方法是使用模拟器发送定时测试用例,或将历史数据中固定时间窗口的日志记录重复回放的方式用于模拟测试。文献[4-5]在网管系统稳定性测试中,预先设置发送间隔,采用 ROBOT VU 模拟 255 个用户以 100 条/s 的频率同时发送操作命令。同时,在测试网管系统处理告警极限时采用 NEF 模拟器以每秒不同数量的频率连续发送告警,直到系统不能正确处理或存在丢包现象。模拟器虽然能模拟长时间的测试环境,但生成测试用例的类型及各自数量并非来自于真实环境,与现网中的实际情况不符,仅仅只是人为并发的多个随机变量的无序列拼接。文献[6-7]在对 WEB 软件测试中,选择某一购物软件作为测试对象,通过回放历史数据模拟用户发送浏览、搜索、购买 3 种操作的发送请求,测试系统在短时间内对请求事务的响应时间。这种方法由于采用历史回放方式,生成的测试用例直接来自真实网络,因此在较短时间内能够完整地重现真实运营中的用户行为,与模拟器测试相比在测试用例发生时间编排方面与实际运行环境的相似性有所提高,但当历史数据发生时间较短或采集较为困难时,该测试方式不能做到长时间的模拟和不同测试用例的组合。同时,测试用例的发生是一个随机过程,在历史数据中连续发生的事件不一定在下一周期也按相同的时间序列产生,无法复现长周期的行为特征[8-9]。文献[10]为了测试通信系统的带宽,建立了服从泊松分布的数据模型,生成包长服从泊松分布的数据包注入处理系统,测试带宽是否满足要求。文献[11]发现数据产生间隔和数据大小服从泊松分布与指数分布,并依据分布模型产生测试用的流量。此类基于数据驱动的测试方法虽然出现较早,但尚未运用于通信基站管理系统的测试,并且分布规律可能并不适用于通信管理系统日志的建模。然而,测试用例的编排是确保测试质量的关键,与实际运行环境相似的测试用例才能提升测试结果可靠度,避免性能冗余与不足。

基于以上问题,本文提出了一种面向通信管理系统性能评估的测试用例分布规律建模方法,在满足测试用例数量要求的同时兼顾发生时间的先后关系。首先,以日志产生时间间隔为研究对象,计算相邻告警产生的时间间隔,分析了累计 5 000 万条电信网管系统的告警日志,并统计其分布情况。分别从数据整体、设备类型、日志类型、设备和日志类型等 4 个维度分析发现日志产生间隔服从幂律分布。随后,采用最小二乘法、极大似然估计与最大后验估计 3 种算法估计分布参数,分析发现最小二乘法能够获得更小的残差,评价指标拟合优度与平均绝对百分误差的平均值分别是 0.96 和 3.5%。最后,使用该方法估算全国多个不同城市电信基站的告警日志间隔分布规律,实验结果表明,所提出的方法能够对数据集中 85%以上的告警日志建模。与传统方法相比,本文的方法在后期仿真测试阶段只需根据分布规律按测试需求的时间和规模产生测试用例,既可以模拟实际部署后的运行环境,又容易产生测

试用例。

1　告警日志规律挖掘

1.1　层级结构定义

通信基站网管系统的层级结构如图 1 所示，按照层级规模，将管理系统的对象分为局点、网元、告警 ID 这 3 个层面，具体定义如下。

(1)告警 ID：不同类型故障统一用不同告警类型表示，可以表示为 ID_1，ID_2，…，ID_n。

(2)网元类型：包括 2G、3G、4G、5G 网络的不同基站类型和无线站点，特定网元设备在运行中又会产生多种类型的告警，可表示为 $D_i=\{ID_1, ID_2, \cdots, ID_n\}$，其中 D_i 表示设备的网元类型。

(3)局点：按照物理位置划分，将某一个区域划分为一个局点，可用 $L=\{D_1, D_2, \cdots, D_n\}$。其中用 L 表示局点，局点内包括多种不同网元类型的设备。

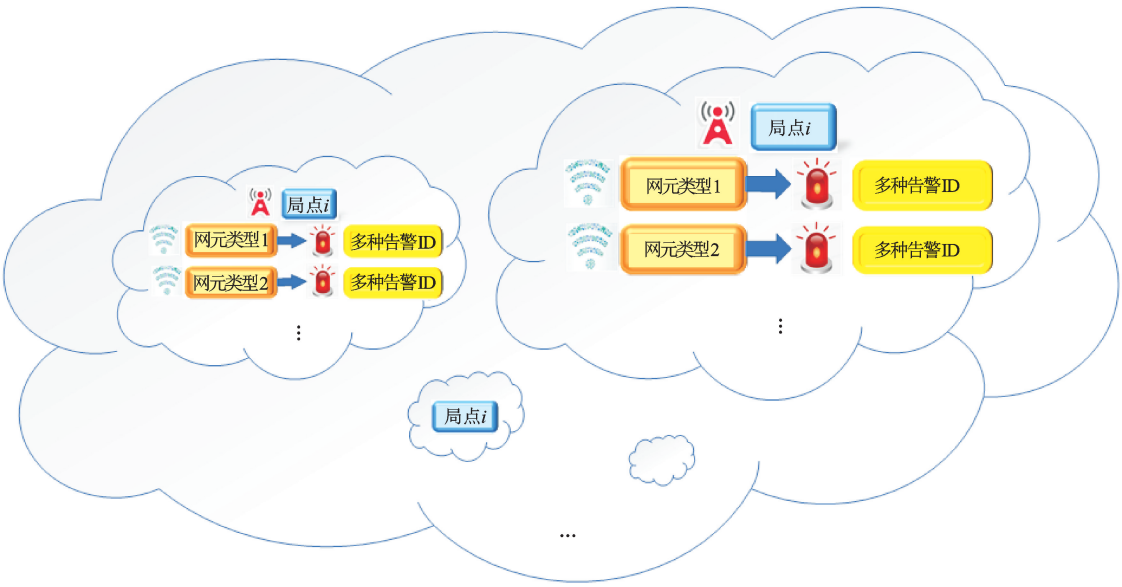


图 1　通信网管系统记录的层级结构

Fig. 1　Hierarchical structure of communication network management system records

1.2　告警发生间隔分析

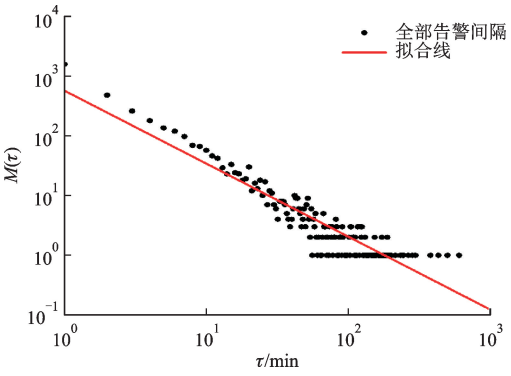
取某局点的数据进行分析，该局点包含“BTS3900”“BTS3900 LET”“BTS5900”“OSS”4 种网元类型，共计 30 种告警 ID。本文分别从数据整体、告警 ID、网元类型、网元类型组合告警 ID 4 个维度选取最主要的 ID 类型发生间隔。在真实的运行环境中，一个特定的网元所产生的告警类型对于管理系统是明确的，由于管理系统处理对象是告警 ID，网元类型是告警类型的上层结构，本文先选择告警 ID 再按多个维度分析。同时，为了保证数据规律的普适性，要求所选取的告警 ID 数据量在该局点所占比重足够大，能够代表大部分数据的趋势，因此在选择告警 ID 时需满足以下条件：

(1)所选取告警 ID 总量超过所有 ID 总量 80%，由数量最多 ID 依次减小选取，直至满足标准；

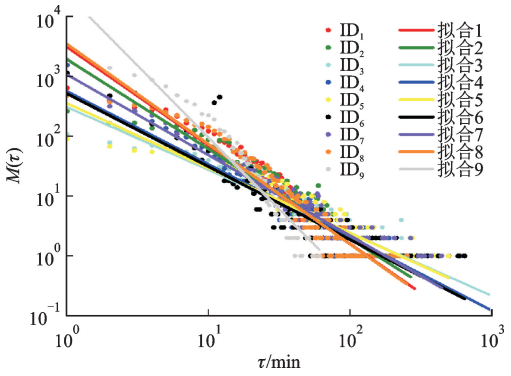
(2)以 1 小时间隔为时间粒度划定时间窗口，统计每个时间窗口内所有告警 ID 的发生数量，所选取的告警 ID 需满足在每个时间窗口内数量占该时间戳总数量的 80%以上；

(3)单独告警 ID 或某网元类型下全部类型告警 ID 数量少于设定阈值不做分析，如果样本数量过少，可能不存在特定的规律，则剔除这部分数据。

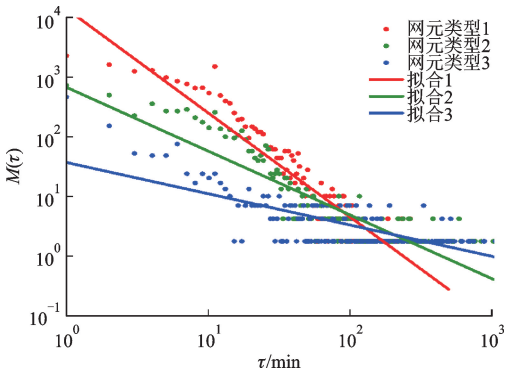
由此，从该局点筛选部分 ID 与网元类型作为主要研究对象。分析告警事件在不同层级结构的分布拟合情况，结果图 2 所示，图中 τ 表示相邻告警发生的时间间隔，$M(\tau)$ 表示间隔 τ 分钟发生的告警数量。其中图 2(a)表示不区分任何告警 ID 类型，根据发生时间统计所有告警发生间隔，其累积分布函数在双对数坐标系可以用一条直线拟合。不同网元类型可能存在相同的告警 ID，由图 2(b)可以看出，所选取的 9 种告警 ID 在不区分网元类型属性前提下的对数坐标系图，对不同 ID 可以用不同直线很好地拟合。图 2(c)、(d)中 3 种网元类型与 9 种网元类型组合 ID 编号的间隔分布同样可以被对数坐标系中多条直线拟合。



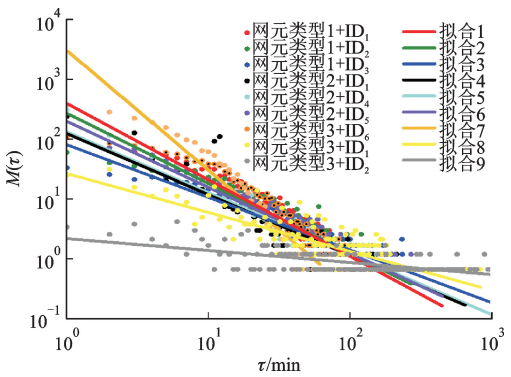
(a)不区分网元类型和告警 ID



(b)不区分网元类型



(c)不区分告警 ID



(d)区分网元类型和告警 ID

图 2 某局点累积分布函数对数坐标图

Fig. 2 Logarithmic graph of cumulative distribution function of a local point

采用拟合优度 R^2 来量化拟合效果的优劣，式(1)给出了 R^2 的计算方法

$$R^2=\frac{\sum_{i=1}^{n}(\hat{h}_i-\bar{h})^2}{\sum_{i=1}^{n}(\hat{h}_i-h_i)^2} \tag{1}$$

拟合优度的取值范围为[0,1]，一般其值越接近 1 说明拟合效果越好[12]。拟合优度的量化值如表 1 所示，由表 1 结果可以看出，不同维度的拟合优度均大于 0.7，部分拟合优度在 0.9 以上，在验证幂律分布拟合效果的同时也验证了其存在普遍性。

表 1 不同维度的分布拟合量化结果

Table 1 The distribution of different dimensions fits the quantitative results

图号	R^2		
	最大值	最小值	平均值
图 2(a)	0.94	0.94	0.94
图 2(b)	0.95	0.81	0.89
图 2(c)	0.90	0.75	0.84
图 2(d)	0.93	0.73	0.86

2 分布参数估算方法

2.1 幂律分布模型

幂律分布与现实生活息息相关，在很多学科中都能看到它的存在，例如金融股票、计算机科学、社会人群收入统计、生物医学和其他学科[13-14]。意大利经济学家帕累托研究人口收入分配后提出的帕累托定律，生物学家克莱伯研究动物体重与代谢率发现的克莱伯定律，都是常见的幂律分布理论[15]。

幂律分布定义如下，对于随机变量 x，如果 x 的概率密度函数是 $P(x)\sim x^{-\alpha}$，$\alpha>0$，则称 $P(x)$ 服从幂指数为 $-\alpha$ 的幂律分布，x 代表发生间隔，概率密度函数可用如下公式表示

$$P(x)=Cx^{-\alpha} \tag{2}$$

对式(2)左右同时取对数，可得到

$$\ln P(x)=\ln C-\alpha\ln x \tag{3}$$

式(3)表示在双对数坐标系中是一条直线。因此，幂律分布常用的判断方法是统计随机变量 x，用对数坐标系中的散点图代表它的概率分布，如果其图像能够被一条直线拟合，该随机变量就服从幂律分布[15]，这与 1.2 节中告警发生间隔的验证结果一致。

2.2 最小二乘法

最小二乘法(least square method，LSM)是一

种数学优化方法[16]。通过求原始数据与拟合数据的误差平方和的最小值，估算匹配函数的未知参数。简而言之，LSM 能够为一组数据找到最佳拟合直线或曲线的方法，因此常用于解决回归问题[17]。由于在对数坐标系中，概率分布图近似为一条斜率为负数的斜线，因此本文尝试采用最小二乘法找到最优拟合直线，解决幂律分布参数计算问题。

假设样本集 $D=\{(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)\}$，需要找到一个拟合函数 $\hat{y}=Cx^{-\alpha}$，使得函数 $\hat{y}$ 的输出尽可能贴近原始值 y_i，若 Q 为拟合值与原始值之间的残差平方和，则有

$$Q=\sum_{I=1}^{n}(\hat{y}-y_i)^2=\sum_{i=1}^{n}(Cx^{-\alpha}-y_i)^2 \quad (4)$$

根据最小二乘法的原理，当 Q 达到最小值时，函数参数 $\hat{\alpha}$、$\hat{C}$ 即为最优解，如以下算法 1 步骤所示。

步骤 1　定义函数拟合形式：$F(p,x)=Cx^{-\alpha}$。

步骤 2　定义拟合残差形式：$E(p,x,y)=F(p,x)-y$。

步骤 3　以数组的形式导入 $x\leftarrow M(\tau)$，$y\leftarrow\tau$。

步骤 4　初始化所求参数 α、C，确定待拟合形状。

步骤 5　计算函数值与真实值之间残差 E。

步骤 6　循环步骤 4、步骤 5，更新参数 α、C，直到残差 E 达到最小值，跳出循环，转到步骤 7。

步骤 7　输出告警 ID 对应的参数 α、C。

2.3　极大似然估计

极大似然估计（maximum likelihood estimation，MLE），一般又称为最大似然估计，是常见的概率模型未知参数求解方法[18]，概率模型是多个随机变量之间的概率描述，简而言之就是刻画随机变量关系的函数[18-19]。当概率模型函数的参数 θ 已知时可以用来描述样本集 D 每个样本 $(x_1,x_2,\cdots,x_n)$ 发生的概率。如果每个样本已知，概率模型参数 θ 根据样本出现的不同结果可以取不同值，刻画不同结果出现的概率分别是多少，那么该函数就称为概率模型的似然函数[20]。

构造极大似然函数前要选择适合样本集 D 的概率分布模型，若模型选择错误即使再多的样本也难以拟合 D 的分布，估计的结果也会不准确[21]。在上节规律挖掘中已经验证告警间隔服从幂律分布，因此由式(1)可知概率模型的未知参数就是幂指数 α。参数计算的具体步骤如下。

首先，构建似然函数 $L(\alpha)$，设概率模型的密度函数为 $f(x_i;\alpha)$，α 为未知参数

$$L(\alpha)=\prod_{i=1}^{n}f(x_i;\alpha) \quad (5)$$

其中 $x_1,x_2,\cdots,x_n$ 为观测值，等式左右两边同时取对数，求 $\ln L(\alpha)$ 的导数并令导数为 0，解得的 $\hat{\alpha}$ 即发生概率最大时 α 的估计值，也就是幂律分布的幂指数。详细步骤如以下算法 2 所示。

步骤 1　定义拟合函数及目标值

$$t,w,t=F(p,x)=Cx^{-\alpha}$$

步骤 2　定义似然函数形式为幂律分布并取对数。

步骤 3　以数组的形式导入 $x\leftarrow M(\tau)$，$y\leftarrow\tau$。

步骤 4　最大化对数似然函数，确定最大似然时 w 及与 w 相对应的 β。

步骤 5　循环步骤 3、步骤 4，更新参数 w、β，直到 w 达到最小值，跳出循环，转到步骤 6。

步骤 6　输出告警 ID 对应的参数 α、C。

2.4　极大后验估计

极大后验估计（maximum a posteriori probability estimate，MAP）与 MLE 类似，两种方法都认为参数是未知的固定值，从原理上讲都是已知样本发生的结果，估计该结果发生概率最大时的模型参数，不同之处在于 MAP 属于概率中的贝叶斯学派，而 MLE 属于频率学派，MAP 考虑了参数本身的概率分布，当参数的先验概率是均匀分布时，两种估计方法相同[22-23]。

$$\alpha_{\mathrm{MAP}}=\arg\max L(\alpha)p(\alpha) \quad (6)$$

式中 $L(\alpha)$ 就是 MLE 中的似然函数，与 MLE 不同之处在于使似然函数取最大值时考虑了参数的概率。因此，当数据量足够大时 $p(\alpha)$ 的影响会越来越小，MAP 与 MLE 参数估计结果也会更加接近。由于本文实验数据较为充足，因此两种方法的估计结果也会较为接近，实验结果对比部分恰好证明了这一事实。详细的过程如算法 3 所示。

步骤 1　定义拟合函数及目标值 $t,w,t=F(p,x)=Cx^{-\alpha}$。

步骤 2　定义先验函数形式与幂律分布函数，并对似然函数取对数。

步骤 3　定义最大后验概率函数 $h(w)$ 形式。

步骤 4　以数组的形式导入 $x\leftarrow M(\tau)$，$y\leftarrow\tau$。

步骤 5　最大化对数似然函数，求解 w。

步骤 6　循环步骤 4，步骤 5，更新参数 w，直到 w 达到最小值，跳出循环，转到步骤 7。

步骤 7　输出告警 ID 对应的参数 α、C。

3 实验结果分析

3.1 实验数据

本文实验数据来源于某公司电信网络基站管理系统的告警日志，涉及杭州、福州、内蒙古自治区赤峰、呼和浩特等地区，包括 2019 年和 2020 年，8 个局点不同月份累计三千多万条日志记录。日志主要内容包含网元类型、告警 ID、发生时间等十几种信息。数据集整体数据量统计情况如表 3 所示。

表 2 各局点数据量统计

Table 2 Statistics of data quantity of each local point

编号	网元类型数	告警类型数	告警总数
局点 1	4	13	604 907
局点 2	4	11	462 694
局点 3	2	13	394 688
局点 4	5	9	67 044
局点 5	4	13	76 063
局点 6	4	10	144 690
局点 7	3	8	89 568
局点 8	7	21	2 723 861

首先需要对数据进行清洗与处理，完成主要数据列缺失值的删除和提取网元类型、告警 ID 和发生时间属性。由于某些类型的告警 ID 数量较少无法拟合或不服从幂律分布，因此，在进行计算时设定日志数量阈值 $\mu=2\,000$，当告警数量小于该阈值时不进行参数计算。结合 1.2 节中验证幂律分布的原则，选取数量较多且符合幂律分布的告警 ID 进行后续工作。

3.2 拟合效果评价标准

本文从拟合情况与计算误差 2 个维度对 3 种参数计算方法结果展开分析。评估指标包括拟合优度 R^2 和平均绝对百分比误差(mean absolute percentage error，用符号 E_{MAPE} 表示)。E_{MAPE} 用来描述拟合值与真实值平均绝对误差的百分比，优点在于评估值的量纲和原值保持一致，数值的大小直接反映数量估算误差占原数据比重的百分比，范围$[0,+\infty)$，数值越小，说明计算方法拥有更好的精确度[24-25]，计算公式如下

$$E_{\mathrm{MAPE}}=\frac{1}{M}\sum_{i=1}^{n}\left|\frac{h_i-\hat{h}_i}{h_i}\right|\times 100\% \tag{7}$$

3.3 参数估计方法对比

不同参数估计方法均有各自优点，最小二乘法可以对多种分布拟合，通用性较强，极大似然估计法要先确定数据的分布函数形式，而最大后验估计优点是加入了先验概率，但人为经验选取的预设参数不一定准确。因此，选取局点 1 中的告警日志，对 3 种参数估计方法的结果进行对比，通过残差和拟合度 2 个维度指标选取最优参数计算方法，计算结果如下图。

图 3 是局点 1 中部分网元类型组合告警 ID 细粒度的计算结果，横坐标用编号 1～10 来表示不同网元类型的告警 ID，左侧纵坐标表示拟合优度。可以看出 LSM 获得的 R^2 最大且均大于 0.9，而 MLE 和 MAP 得到的结果比较相近，但它们与 LSM 的结果相比拟合准确度不稳定，在某些告警数据上差距不大，但大多数告警计算结果小于 LSM。E_{MAPE} 表示所有告警间隔数量拟合值与真实值误差占原数据比重平均值的百分比，根据图中右侧纵坐标能够看出 LSM 在不同告警 ID 分布拟合中获得的 MAPE 最小，说明与实际间隔的数量差距较小，其中 E_{MAPE} 最小值为 2.49%，最大值为 4.92%，而其他两种方法最小值与最大值分别为 3.21%和 7.43%，无论是误差上限还是下限，LSM 都拥有更小的估算误差。

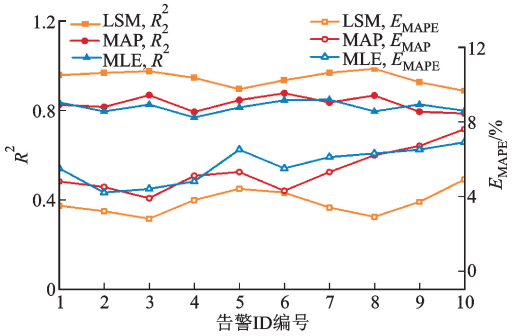


图 3 不同估计方法拟合误差对比

Fig. 3 Comparison of fitting errors of different estimation methods

图 4 是该局点全部告警 ID 使用 3 种估计方法的 R^2 与 E_{MAPE} 分布情况，从橘色箱体也能够看出 LSM 拥有更好的拟合度，其最小值均大于其他 2 种方法的最大值，且箱体较扁计算误差波动不大，这与图 3 中的结果一致。蓝色箱线表示 3 种方法的 E_{MAPE} 分布情况，其中 LSM 的 E_{MAPE} 上限和下限分别是 2.9%和 4.8%，所有告警的 E_{MAPE} 均小于 5%，而其他两种方法的 E_{MAPE} 中位数均超过 5%，且箱体较长，说明评价指标数值范围较宽，在不同告警中误差计算波动较大。为保证 LSM 的泛化性，又在其他局点的验证中发现 LSM 依旧获得较好的结果，分析原因由于数据自身的特点。当数据类型组成复杂、数量分布不平衡或难以达到海量的数据时，采用直接分布拟合的方式要优于概率估计，因为现实中

基站产生的网元类型与告警类型都是随机的，而单独局点的告警数据也只连续采集了几个月，难以满足概率估计的理想条件。因此，综合 E_{MAPE} 和 R^2 指标，本文中 LSM 是较为适合的幂指数估算方法。

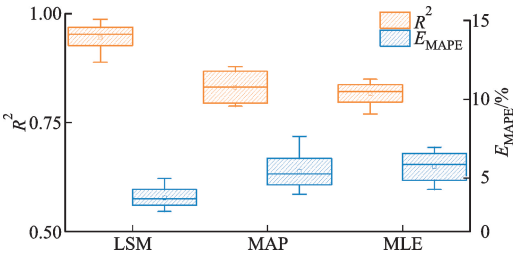


图 4　残差分布对比

Fig. 4　Residual distribution comparison

4　最小二乘法估计与结果分析

本文使用最小二乘法对 8 个局点的网元类型和告警 ID 分别计算其相邻告警产生时间间隔的幂指数，所有局点的指标计算如图 5，4-2 所示。在计算幂指数的过程中，因为某些告警 ID 或网元类型产生的告警数量过少不服从幂律分布，因此在计算中剔除该部分异常数据，多个局点的网元类型与告警 ID 合计 369 组数据，其中告警数量小于阈值的有 76 组，经过筛选后剩余 293 个分组。

拟合优度 R^2 表示最小二乘法计算的参数分布与原始数据的拟合程度，当 R^2 大于 0.7 时认为该拟合是有效的。图 5 为所有局点告警数据的 R^2 结果，R^2 在 0.7 以上的有 263 组。局点 1 和 8 的 R^2 整体上结果较好，大部分值均在 0.9 以上，由于这两个局点的数据量较多，呈现的分布规律更加明显，所以拟合效果也更好。局点 2 和局点 3 的 R^2 整体分布区间较小，主要集中在 0.88 附近，分析原因可能是由于这两个局点虽然数据量足够多，但是包含多种类型复杂的网元类型与告警 ID，因此虽然分布拟合区间分散不大但 R^2 数值区别较大。图 6 描述了多个局点平均绝对误差百分比的箱线图，所有局点的 E_{MAPE} 值均小于 5%，结果最好的局点 8 所有告警 E_{MAPE} 值均在 2%至 3%。从整体告警数据能够看出大部分数据拟合程度与残差均符合预期标准，能够获得较好的结果。

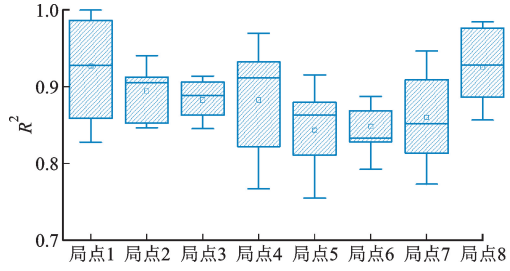


图 5　多局点 R^2

Fig. 5　The R^2 of multi-location

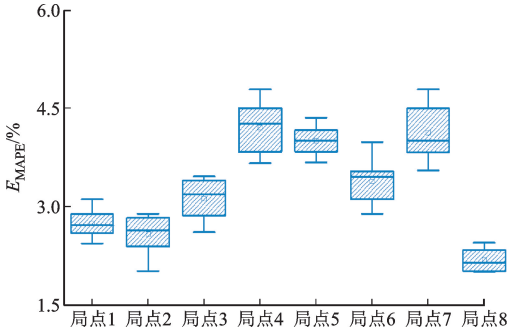


图 6　多局点 MAPE

Fig. 6　The MAPE of multi-location

表 3 为所有局点的幂律拟合情况，告警数表示该局点告警总数量，符合数表示符合幂律分布的告警数量，幂律占比表示符合幂律分布数据在总数量中的占比，E_{MAPE} 与 R^2 分别代表该局点评价指标的均值。从表中能够看出局点 1、局点 2、局点 3 与局点 8 的告警总数最多且幂律拟合占比均达到 90%以上，从侧面也说明数据量越多时分布规律也更明显，这也是本文在计算分布参数时数据量需满足一定阈值的原因。

表 3　各局点幂律分布统计

Table 3　Statistics of power law distribution at each point

编号	告警数	符合数	幂律比/%	E_{MAPE}/%	R^2
局点 1	604 907	603 567	98.78	2.82	0.91
局点 2	462 694	462 025	97.86	2.64	0.87
局点 3	394 688	386 794	96.00	3.01	0.85
局点 4	67 044	58 572	87.36	4.33	0.79
局点 5	76 063	67 622	88.90	3.98	0.81
局点 6	144 690	125 143	86.49	3.26	0.83
局点 7	89 568	76 616	85.54	4.01	0.78
局点 8	2 723 861	2 603 463	95.58	2.15	0.93

5　后续应用讨论

通过前文分析，可以发现告警事件的幂律分布具有一定的普遍性，本文进一步给出了局点 1 中部分网元类型的告警 ID 采用最小二乘法的幂指数计算结果，见表 2，从中可以看到，告警 ID 级别的拟合结果更优。

表 4 部分幂指数计算结果

Table 4 Partial power exponent calculation results

网元类型	告警 ID	数量	幂指数	R^2
网元类型 1	ID_1	49 335	2.566 49	0.93
网元类型 2	ID_2	55 810	3.020 70	0.88
网元类型 3	ID_3	78 312	2.261 20	0.91
网元类型 4	ID_4	99 777	2.279 07	0.91

待挖掘出告警日志的分布规律之后,在基站管理系统的测试中,就可以按照图 1 描述的层级结构对不同的测试场景、不同的测试环境、不同的测试时长进行组合,按照需求对测试用例进行编排,进一步提升测试的灵活性和发生时间合理性。对管理系统评估的辅助作用有以下几点。

(1)告警 ID 粒度的逼真度仿真。在运行环境收集数据的支持下,利用本文方法可以得到某一告警 ID 的幂律分布值,利用该分布值可以约束该告警 ID 测试用例在时间维度的编排。

(2)网元设备粒度的逼真度仿真。在计算得到不同告警 ID 的幂律分布值之后,可以通过组合不同的告警 ID,仿真不同类型的网元设备。

(3)局点粒度的逼真度仿真。通过组合不同网元设备,可以按需求配置网元类型和数量,仿真局点环境。

(4)时间粒度的逼真度仿真。通过幂律分布约束不同的告警类型,可以实现不同时间粒度的事件仿真,实现更长的高逼真度测试。

由此,对本文工作驱动的测试方式和传统的模拟器与回放驱动的方法的优劣进行了对比分析,结果如表 5 所示。表中"数量"表示测试用例的个数,"时长"代表模拟测试时间,场景表示根据不同运行场景对需要测试的告警 ID 进行组合,"环境"代表测试用例是否来源于真实运行环境,尽可能贴近实际运行环境的测试用例才能提升测试结果的可靠度。"逼真度"表示测试结果的可靠性。由表 5 结果可见,本文工作驱动的测试方法在灵活性、发生时间合理性等方面均有一定优势,提升了通信基站管理系统测试环境的逼真度。通过测试可以发现实际部署后可能出现的隐藏问题。

表 5 不同驱动测试方法对比

Table 5 Comparison of different driving test methods

驱动方法	测试是否符合要求				逼真度
	数量	时长	场景	环境	
模拟器	是	是	是	否	低
回放	是	否	部分是	是	高
本文方法	是	是	是	是	高

6 结论与展望

本文围绕改进通信管理系统中现有测试方法的不足进行了探索。从数据整体、设备类型、日志类型、设备和日志类型等多个维度对通信网络日志分布进行了分析,发现数据服从幂律分布;并使用不同方法计算分布参数,比较后发现最小二乘法能够获得较小的估算误差。最后,采用 8 个局点的数据对最小二乘法进行验证,发现不同局点 85%以上的数据服从幂律分布,由此推断幂律分布可以推广到其他局点告警日志数据建模中,具有较好的泛化性,并为后续测试用例的产生和编排奠定了基础。在下一步工作中,本文将利用幂律参数生成测试用例并挖掘不同场景下告警类型产生的关联关系,将分析结果用于 5G 基站管理系统的运维测试,进一步提升测试环境与实际运行环境的逼真度。

参考文献:

[1] MAJCHRZAK T A, KAINDL H, GRØNLI T M. Introduction to the minitrack on software development for mobile devices, the internet-of-things, and cyber-physical systems [C]//Proceedings of the 55th Hawaii International Conference on System Sciences. Piscataway, NJ, USA: IEEE, 2022: 7750-7751.

[2] SHI Gengyuan, WANG Chaokun, HUANG Bingyang, et al. UniTest: a universal testing framework for database management systems [C]//Database Systems for Advanced Applications. Cham, Germany: Springer International Publishing, 2021: 96-104.

[3] 朱向雷, 王海弛, 尤翰墨, 等. 自动驾驶智能系统测试研究综述 [J]. 软件学报, 2021, 32(7): 2056-2077. ZHU Xianglei, WANG Haichi, YOU Hanmo, et al. Survey on testing of intelligent systems in autonomous vehicles [J]. Journal of Software, 2021, 32(7): 2056-2077.

[4] 张朔. 软件测试方法及在综合网管系统测试中的应用与研究 [D]. 北京: 华北电力大学, 2007: 15-50.

[5] 邓庆. 软件测试在基站话务量上的应用 [D]. 西安: 西安电子科技大学, 2013: 21-65.

[6] DURELLI V H S, DURELLI R S, BORGES S S, et al. Machine learning applied to software testing: a systematic mapping study [J]. IEEE Transactions on Reliability, 2019, 68(3): 1189-1212.

[7] QIU Kun, ZHENG Zheng, TRIVEDI K S, et al. Stress testing with influencing factors to accelerate data race software failures [J]. IEEE Transactions on Reliability, 2020, 69(1): 3-21.

[8] QU Liangqiong, BALACHANDAR N, ZHANG Miao,

et al. Handling data heterogeneity with generative replay in collaborative learning for medical imaging [J]. Medical Image Analysis, 2022, 78: 102424.

[9] 罗娜，李爱平，吴泉源，等. 基于概要数据结构可溯源的异常检测方法 [J]. 软件学报，2009，20(10)：2899-2906.
LUO Na, LI Aiping, WU Quanyuan, et al. Sketch-based anomalies detection with IP address traceability [J]. Journal of Software, 2009, 20(10): 2899-2906.

[10] 甘忠海. 测试用泊松分布数据流的设计及实现 [J]. 集成电路应用，2005(3)：34-35.
GAN Zhonghai. Design and implementation of Poisson distribution data flow for test [J]. Application of IC, 2005(3): 34-35.

[11] 岳丽霞. 基于业务的配电宽带无线通信网络性能测试系统研究与实现 [D]. 成都：西南交通大学，2014：10-62.

[12] 商立群，李洪波，侯亚东，等. 基于特征选择和优化极限学习机的短期电力负荷预测 [J]. 西安交通大学学报，2022，56(4)：165-175.
SHANG Liqun, LI Hongbo, HOU Yadong, et al. Short-Term power load forecasting based on feature selection and optimized extreme learning machine [J]. Journal of Xi'an Jiaotong University, 2022, 56(4): 165-175.

[13] 张重阳，张海. 百度百科词条编辑时间间隔的幂律分布研究 [J]. 计算机仿真，2020，37(2)：269-274.
ZHANG Chongyang, ZHANG Hai. Power law distribution of editing time interval of Baidu encyclopedia entries [J]. Computer Simulation, 2020, 37(2): 269-274.

[14] ANTIPOV D, BUZDALOV M, DOERR B. Lazy parameter tuning and control: choosing all parameters randomly from a power-law distribution [C]//Proceedings of the Genetic and Evolutionary Computation Conference. New York, NY, USA: Association for Computing Machinery, 2021: 1115-1123.

[15] SHARMA S, PENDHARKAR P C. On the analysis of power law distribution in software component sizes [J]. Journal of Software: Evolution and Process, 2022, 34(2): e2417.

[16] CASTELANI E V, LOPES R, SHIRABAYASHI W V I, et al. A robust method based on LOVO functions for solving least squares problems [J]. Journal of Global Optimization, 2021, 80(2): 387-414.

[17] 于振华，黄山阁，杨波，等. 新型冠状病毒肺炎传播动力学模型构建与分析 [J]. 西安交通大学学报，2022，56(5)：43-53.
YU Zhenhua, HUANG Shange, YANG Bo, et al. Dynamics modeling and analysis of COVID-19 [J]. Journal of Xi'an Jiaotong University, 2022, 56(5): 43-53.

[18] PENG Jiangtao, LI Luoqing, TANG Yuanyan. Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(6): 1790-1802.

[19] NETTASINGHE B, KRISHNAMURTHY V. Maximum likelihood estimation of power-law degree distributions via friendship paradox-based sampling [J]. ACM Transactions on Knowledge Discovery from Data, 2021, 15(6): 106.

[20] 李斐. 指数总体删失数据的极大似然估计 [J]. 高等数学研究，2021，24(1)：28-30.
LI Fei. Maximum likelihood estimation of missing data for exponential population [J]. Studies in College Mathematics, 2021, 24(1): 28-30.

[21] 刘鹏飞，王绍臣，周望. 总体协方差矩阵不存在情形下的经验似然 [J]. 中国科学(数学)，2022，52(9)：1089-1094.
LIU Pengfei, WANG Shaochen, ZHOU Wang. Empirical likelihood without population covariance matrix [J]. Scientia Sinica(Mathematica), 2022, 52(9): 1089-1094.

[22] KESHVARI S, ENSAN F, SADOGHI YAZDI H. ListMAP: listwise learning to rank as maximum a posteriori estimation [J]. Information Processing & Management, 2022, 59(4): 102962.

[23] XIE Qi, ZHAO Qian, XU Zongben, et al. Color and direction-invariant nonlocal self-similarity prior and its application to color image denoising [J]. Science China: Information Sciences, 2020, 63(12): 222101.

[24] 王知雨，王斌，王朝晖. 采用非线性最小二乘法的超级电容等效电路模型参数辨识 [J]. 西安交通大学学报，2020，54(4)：10-18.
WANG Zhiyu, WANG Bin, WANG Chaohui. A parameter identification method for an equivalent circuit model of supercapacitor using nonlinear least squares [J]. Journal of Xi'an Jiaotong University, 2020, 54(4): 10-18.

[25] 平昱恺，黄鸿云，江贺，等. 目标检测模型的决策依据与可信度分析 [J]. 软件学报，2022，33(9)：3391-3406.
PING Yukai, HUANG Hongyun, JIANG He, et al. Decision basis and reliability analysis of object detection model [J]. Journal of Software, 2022, 33(9): 3391-3406.

（编辑　刘杨）