

面向数据缺失场景的网络性能评估方法

韩悦, 臧晓尧, 李海, 王杰

(国防科技大学信息通信学院, 湖北 武汉 430000)

摘要: 实时精准的网络性能评估对保障网络系统的稳定、高效运行具有重要意义。在现有的基于证据推理 (ER) 规则的网络性能评估研究中, 通常默认指标数据完整。然而, 在实际网络运行环境中, 受监控资源竞争、异常过滤等因素影响, 关键性能指标常呈现非随机缺失 (MNAR) 的特征, 导致评估结果出现偏差。针对这一挑战, 提出一种面向数据缺失场景的网络性能评估方法。首先, 通过构建一个同时包含数据分布与缺失机制的估计模型, 对传统期望最大化 (EM) 算法进行改进, 实现对网络中 MNAR 数据的精准估计; 其次, 将缺失数据的估计结果引入 ER 规则框架, 提出数据完整度的概念以修正指标权重, 以增强评估方法在数据缺失场景下的适应性; 再次, 通过建立以评估误差最小化为目标的参数优化模型, 进一步提高了评估精度; 最后, 结合真实网络运行场景开展对比实验, 以验证方法的有效性。

关键词: 证据推理规则; 数据缺失; 网络性能评估; 期望最大化算法

中图分类号: TP393.06

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025202

Network performance evaluation method for scenarios with missing data

HAN Yue, ZANG Xiaoyao, LI Hai, WANG Jie

Institute of Information and Communication, National University of Defense Technology, Wuhan 430000, China

Abstract: Real-time and accurate network performance evaluation is of great significance for ensuring the stable and efficient operation of network systems. Existing research on network performance evaluation based on the evidential reasoning (ER) rule typically assumes the completeness of indicator data. However, in practical network operating environments, key performance indicators often exhibit the behavior of missing not at random (MNAR) due to factors such as monitoring resource contention and anomaly filtering, leading to biased evaluation results. To address this challenge, a network performance evaluation method for scenarios with missing data was proposed. Firstly, by constructing an estimation model that simultaneously incorporated data distribution and missingness mechanisms, the traditional expectation maximization (EM) algorithm was improved to achieve accurate estimation of MNAR data in the network. Secondly, the estimated missing data was integrated into the framework of ER rule, where a data completeness metric was proposed and indicator weights were adjusted, thereby enhancing the adaptability of the evaluation method in scenarios with missing data. Then, a parameter optimization model aimed at minimizing evaluation error was established to further improve evaluation accuracy. Finally, the effectiveness of the method was validated through comparative experiments based on real-world network operation scenarios.

Keywords: evidential reasoning rule, missing data, network performance evaluation, expectation maximization algorithm

收稿日期: 2025-09-16; 修回日期: 2025-12-10

通信作者: 王杰, wangjie.23@nudt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62403474, No.U22B2005, No.42501502); 国防科技大学青年自主创新科学基金资助项目

Foundation Items: The National Natural Science Foundation of China (No.62403474, No.U22B2005, No.42501502), The Youth Independent Innovation Science Fund Project of National University of Defense Technology

0 引言

随着5G、物联网等新一代网络技术的迅猛发展,网络基础设施已成为支撑数字化社会运转的核心中枢^[1]。然而,在远程医疗诊断、智能交通调度、工业物联网等关键领域,网络服务质量的动态保障面临严峻挑战^[2]:一方面,网络运行状态受节点负载波动、链路时延变化等多维因素的耦合影响;另一方面,网络运行数据呈现出多源性、海量性等特征^[3],使传统评估方法难以实现精准的性能感知。网络性能评估的不足容易导致网络状态的误判,这不仅严重威胁到网络服务质量,还会极大影响用户体验^[4]。因此,实时、精准的网络性能评估可以为路由优化、负载均衡和资源调配提供科学决策支持,对维护网络系统稳定与安全具有重要意义。

在现有研究中,网络的性能评估方法可以分为基于数据、基于定性知识和基于半定量信息的方法^[5]。基于数据的方法主要依赖大量监测数据来实现性能评估,如神经网络和深度学习等。然而,该类方法存在不可解释、容易过拟合等问题^[6]。基于定性知识的方法依靠专家知识对网络系统进行建模,如专家系统和模糊推理等。该类方法在一定程度上弥补了定量数据的不足,但专家知识的模糊性和不完备性也会导致评估精度不理想^[7]。基于半定量信息的方法则是将定量数据和定性知识有机结合,通过融合多源信息开展网络性能评估^[8]。相比之下,该类方法继承了上述评估方法的优势,因而在网络性能评估领域潜力巨大。

作为半定量信息方法的典型代表,证据推理(ER, evidential reasoning)规则通过构建统一的分布式置信结构,并基于正交合成规则实现多源证据融合,形成结构化推理框架^[9-10]。由于其对不确定信息具有良好的处理能力,ER规则已被广泛应用于故障诊断、性能评估、决策优化等多个领域。文献^[11]提出了一种基于ER规则的时空域级联融合模型,有效降低了船舵故障诊断中的不确定性。文献^[12]提出了动态ER规则,合理揭示了无线传感器网络中噪声对性能的动态影响。文献^[13]提出了基于ER规则的维护时机决策模型,实现了工业设备全生命周期策略的优化。

然而,现有基于ER规则的研究多数假设数据集完整,即要求所有观测指标均无缺失数据。在实

际网络场景中,受监控资源竞争、异常过滤等因素影响,观测数据往往存在缺失^[14]。例如,在高负载状态下,节点因CPU资源耗尽而丢弃部分监控报文,导致吞吐量数据容易发生缺失。此类数据缺失问题若采用传统的直接删除法,将产生负面影响:一方面,在小样本评估场景下,数据直接删除会加剧评估的不确定性,导致模型陷入过拟合或高偏差状态;另一方面,当缺失数据涉及网络关联性指标或时序依赖性特征时,简单删除操作会破坏数据内在的时序关联结构,影响评估结果的准确性和维护决策的可靠性。因此,亟须发展适于复杂网络环境的缺失数据处理方法。

事实上,缺失数据的研究历经数十年发展已形成系统化框架,其大致分为3类,即完全随机缺失(MCAR, missing completely at random)、随机缺失(MAR, missing at random)与非随机缺失(MNAR, missing not at random)^[15]。其中,MCAR是指数据缺失是随机发生的,与自身属性以及其他属性取值无关;MAR是指缺失概率仅依赖于已观测变量,与缺失数据本身无关,只与完整属性取值有关;MNAR不仅与自身取值有关,也与缺失数据本身、完整属性取值有关,此类缺失是不可忽略的缺失^[16]。上述分类框架为理解缺失数据本质提供了理论基础。在实际网络的工作过程中,数据缺失往往服从MNAR机制^[17]。例如,在高负载状态下,传感器因资源过载更易触发数据丢失,此时缺失概率与负载状态(已观测变量)及数据本身幅值均显著相关。传统的删除法(如列删除法、行删除法等)不仅会加剧小样本评估的不确定性,更会人为弱化关键特征的关联性^[18]。相较于删除法,插补技术(如均值插补、贝叶斯多重插补等)通过建立缺失数据估计模型来保持数据完整性,但其插补过程依赖先验分布假设,难以适应复杂网络系统的动态演化特性^[19]。

针对上述缺失数据处理方法的局限性,期望最大化(EM, expectation maximization)算法提供了一种不需要先验分布假设的解决方案^[20]。然而,传统EM算法通常建立在MAR假设之上,难以处理实际网络中普遍存在的MNAR问题。为此,本文通过构建融合数据分布与缺失机制的估计模型,实现对传统EM算法的改进,使其能够适应指标缺失概率与自身取值相关的复杂场景。改进后的算法

依然遵循传统EM算法中期望步(E步)与极大化步(M步)交替迭代的过程。具体而言:在E步中,基于当前数据的分布参数和缺失机制计算缺失数据的条件期望;在M步中,更新包括缺失机制参数在内的全部模型参数,从而实现对MNAR数据的无偏估计^[21-22]。该方法不是依赖主观的先验假设,而是依据数据自身特征动态调整估计过程,显著提升了模型在复杂网络环境中的适应能力^[23]。

根据以上分析,针对网络性能评估中普遍存在的数据缺失问题,本文基于ER规则提出一种考虑缺失数据的网络性能评估方法,具体研究工作包括:针对网络中的MNAR机制,通过同时建模数据分布与缺失机制来改进传统EM算法,以提升缺失数据的估计精度;将估计结果与ER规则相结合,引入数据完整度以调整指标权重,以增强模型对不完整数据的适应能力;以最小化评估误差为核心目标,建立参数优化模型,实现参数的自适应调整;通过在某工业网络开展实验,验证了所提方法的有效性。

1 基础知识与问题描述

本节首先阐述ER规则的核心理论,然后明确本文需要解决的关键问题。

1.1 ER规则描述

在ER规则中,辨识框架由一系列满足相互排斥和穷尽的命题构成,可表示为 $\Omega = \{\theta_n | n = 1, 2, \dots, N\}$,其中元素 θ_n 表示第 n 个独立命题。为描述多个命题的不确定性,将辨识框架扩展至幂集空间 $P(\Omega)$,形式为

$$P(\Omega) = \{\emptyset, \theta_1, \theta_2, \dots, \theta_N, \{\theta_1, \theta_2\}, \dots, \{\theta_1, \theta_N\}, \{\theta_1, \dots, \theta_{N-1}\}, \Omega\} \quad (1)$$

其中, \emptyset 表示空集,不参与置信分配。在幂集框架下,给定证据 $e_l, l = 1, 2, \dots, L$ 的置信分布形式为

$$e_l = \{(\theta, \beta_{\theta,l}), \forall \theta \subseteq \Omega, \sum \beta_{\theta,l} = 1\} \quad (2)$$

其中, θ 表示幂集中的任意子集, $\beta_{\theta,l}$ 表示 e_l 对 θ 的置信度。对于 L 条证据而言,最终的融合结果可迭代计算为

$$p_{\theta,e(L)} = \begin{cases} 0, \theta = \emptyset \\ \frac{\widehat{m}_{\theta,e(L)}}{\sum_{D \subseteq \theta} \widehat{m}_{D,e(L)}}, \theta \neq \emptyset \end{cases} \quad (3)$$

其中, $p_{\theta,e(l)}$ 表示融合结果分配给 θ 的置信度, $\widehat{m}_{\theta,e(l)}$ 表示 l 条证据组合后未标准化的分配给 θ 的基本概率质量(BPM, basic probability mass),由式(4)决定。

$$\begin{cases} \widehat{m}_{\theta,e(l)} = (1 - r_l)m_{\theta,e(l-1)} + m_{P(\Omega),e(l-1)}m_{\theta,l} + \\ \sum_{B \cap C = \theta} m_{B,e(l-1)}m_{C,l} \\ \widehat{m}_{P(\Omega),e(l)} = (1 - r_l)m_{P(\Omega),e(l-1)} \end{cases} \quad (4)$$

其中, $m_{\theta,e(l-1)}$ 和 $m_{P(\Omega),e(l-1)}$ 分别表示组合 $(l-1)$ 条证据后分配给 θ 和 $P(\Omega)$ 的BPM,可以迭代计算为

$$m_{\theta,e(l-1)} = \begin{cases} 0, \theta = \emptyset \\ \frac{\widehat{m}_{\theta,e(l-1)}}{\sum_{D \subseteq \theta} \widehat{m}_{D,e(l-1)} + \widehat{m}_{P(\Omega),e(l-1)}}, \theta \neq \emptyset \end{cases} \quad (5)$$

$$m_{P(\Omega),e(l-1)} = \frac{\widehat{m}_{P(\Omega),e(l-1)}}{\sum_{D \subseteq \Omega} \widehat{m}_{D,e(l-1)} + \widehat{m}_{P(\Omega),e(l-1)}} \quad (6)$$

$$m_{\theta,l} = w_l \beta_{\theta,l} \quad (7)$$

其中, $m_{\theta,l}$ 表示第 l 条证据分配给 θ 的BPM。不断地运用式(2)~式(7),最终可以得到 L 条证据的融合结果。

1.2 问题描述

为准确评估缺失数据场景下的网络性能,构造本文方法的基本结构,如图1所示。所提方法包含三部分内容:缺失数据估计、网络性能评估和参数优化。根据上述分析,本文重点解决以下3个问题。

问题1:如何估计缺失数据。网络监测数据常服从MNAR机制,在此机制下,数据缺失的概率不仅依赖于已观测的数据,更与缺失数据本身有关。传统的EM算法因建立在MAR假设之上而不再适用。因此问题1的核心是建立一个同时包含数据分布与缺失机制的估计模型,具体可表示为

$$\hat{Z} \sim f(Z|X, \Theta, \Phi) \quad (8)$$

其中, Θ 和 Φ 分别为数据分布和缺失机制的参数集, X 为包含缺失数据的观测数据集, Z 为待估计的完整数据集, \hat{Z} 为模型输出的估计结果, f 表示本文所构建的参数估计模型。

问题2:如何融合多源数据。为获得网络性能

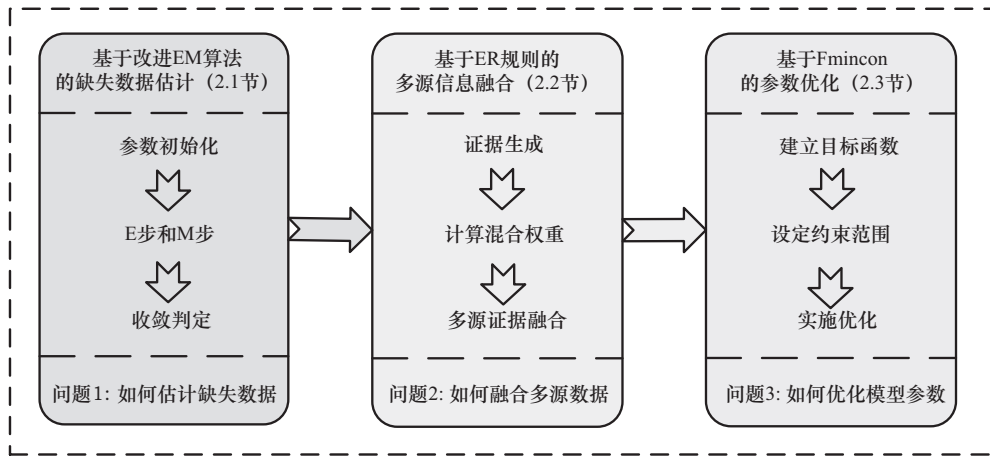


图1 网络性能评估方法的基本框架

的评估结果，问题2主要是根据问题1的缺失数据估计结果，建立一个基于ER规则的网络性能评估模型，实现多源信息的有效融合，如式(9)所示。

$$M = \psi(\hat{Z}, V', R) \quad (9)$$

其中， M 为网络性能的评估结果， V' 为所有的模型参数， R 为专家知识， ψ 为基于ER规则的非线性函数。

问题3：如何优化模型参数。为提升评估精度，需对模型参数进行优化。因此问题3主要是构建基于Fmincon的参数优化模型，如式(10)所示。

$$\begin{cases} \min f'(V) \\ \text{s.t. } V_{\text{lower}} \leq V \leq V_{\text{upper}} \end{cases} \quad (10)$$

其中， $f'(V)$ 为目标函数， V 为待优化参数， V_{lower} 和 V_{upper} 分别为优化参数需要满足的上下界约束。

2 考虑缺失数据的网络性能评估方法

为解决上述问题，本节提出了一种面向数据缺失场景的网络性能评估方法。首先，通过改进的EM算法估计观测数据中的缺失值；然后，基于ER规则，实现多源监测信息的有效融合；最后，构建参数优化模型，以提升评估精度。

2.1 基于改进EM算法的缺失数据估计

为处理MNAR机制，本节在现有EM算法的基础上，通过建立一个同时包含数据分布与缺失机制的估计模型，实现对缺失数据的无偏估计。假设网络系统包含 L 个关键性能指标 $\{e_1, e_2, \dots, e_L\}$ ，其监测数据在离散时间点 $t = 1, 2, \dots, T$ 上存在缺失，其中， T 为总观测时刻数。对于第 l 个指标 e_l ，其完整数据集为 $Z_l = \{z_l(t), t = 1, 2, \dots, T\}$ ，实际观测数据集为 $X_l = \{x_l(t), t = 1, 2, \dots, T\}$ ， $z_l(t)$ 和 $x_l(t)$ 表示相应的

观测数据。首先，定义状态变量 $\delta_l(t)$ 以表示 $x_l(t)$ 在 t 时刻是否存在缺失，即

$$\delta_l(t) = \begin{cases} 1, & x_l(t) \neq \text{NA} \\ 0, & x_l(t) = \text{NA} \end{cases} \quad (11)$$

假设指标 e_l 的完整数据服从正态分布，即满足 $z_l(t) \sim N(\mu_l, \sigma_l^2)$ ，其中， μ_l 和 σ_l 分别是指标 e_l 的均值和标准差。假设数据的缺失概率与数据值的本身相关，可以采用如下回归模型

$$\text{logit}(P(\delta_l(t) = 0)) = \phi_{0,l} + \phi_{1,l}z_l(t) \quad (12)$$

其中， $\text{logit}(\cdot)$ 表示回归函数， $\phi_{0,l}$ 和 $\phi_{1,l}$ 是指标 e_l 的缺失机制参数，参数 $\phi_{1,l} \neq 0$ 是MNAR机制存在的关键。在上述定义的基础上，基于改进EM算法的缺失数据估计包含如下步骤。

步骤1 参数初始化。根据初始观测数据实现参数的初始化，如式(13)所示。

$$\begin{cases} \mu_l^{(0)} = \frac{\sum_{t=1}^T \delta_l(t) x_l(t)}{\sum_{t=1}^T \delta_l(t)} \\ (\sigma_l^2)^{(0)} = \frac{\sum_{t=1}^T \delta_l(t) (x_l(t) - \mu_l^{(0)})^2}{\sum_{t=1}^T \delta_l(t) - 1} \\ \phi_{0,l}^{(0)} = 0 \\ \phi_{1,l}^{(0)} = 0 \end{cases} \quad (13)$$

其中， $\mu_l^{(0)}$ 表示初始均值，其余参数含义与之类似。

步骤2 E步和M步。基于步骤1的计算结果，计算缺失数据的条件期望，并用这一期望值代替缺失数据，如式(14)所示。

$$\hat{z}_l^{(s)}(t) = \begin{cases} x_l(t), \delta_l(t) = 1 \\ E[z_l(t)|X_l, \delta_l(t), \Phi_l^{(s)}, \Phi_l^{(s)}], \delta_l(t) = 0 \end{cases} \quad (14)$$

其中, $m_{P(\Omega), e(l-1)} = \frac{\widehat{m}_{P(\Omega), e(l-1)}}{\sum_{D \subseteq \Omega} \widehat{m}_{D, e(l-1)} + \widehat{m}_{P(\Omega), e(l-1)}}$ 和

$\Phi_l^{(s)}$ 为指标 e_l 在当前状态下的分布参数和缺失机制参数, $\hat{z}_l^{(s)}(t)$ 为当前缺失数据的估计值, s 为迭代次数。然后, 使用当前完整数据集 $\hat{Z}_l^{(s)}$ 更新参数

$$\begin{cases} \mu_l^{(s+1)} = \frac{\sum_{t=1}^T \hat{z}_l^{(s)}(t)}{T} \\ (\sigma_l^2)^{(s+1)} = \frac{\sum_{t=1}^T (\hat{z}_l^{(s)}(t) - \mu_l^{(s+1)})^2}{T} \end{cases} \quad (15)$$

最后, 通过最大化以下对数似然函数以更新参数 Φ_l

$$\Phi_l^{(s+1)} = \operatorname{argmax}_{\Phi_l} \sum_{t=1}^T [\delta_l(t) \log(1 - p_l(t)) + (1 - \delta_l(t)) \log p_l(t)] \quad (16)$$

其中, $p_l(t)$ 为指标 e_l 在时刻 t 的缺失概率。

步骤 3 收敛判定。不断重复步骤 2, 直到参数变化量满足式(17)时, 终止迭代。

$$\|\Phi_l^{(s+1)} - \Phi_l^{(s)}\| + \|\Phi_l^{(s+1)} - \Phi_l^{(s)}\| < \varepsilon \quad (17)$$

其中, ε 为收敛阈值。收敛后的参数集为 $\hat{\Theta}_l$ 和 $\hat{\Phi}_l$, 最终得到完整数据集的估计值 \hat{Z}_l 。在上述推理过程中, 式(12)通过构建逻辑回归函数将缺失概率与数据本身相关联, 刻画了典型的 MNAR 机制。然而, 该模型在复杂缺失场景下的适应性仍需进一步探讨。一方面, 所提模型默认缺失机制具有无记忆性, 这可能导致对连续缺失事件的刻画能力不足。另一方面, 所提模型缺乏对不同缺失机制的区分能力, 这可能导致部分指标的缺失机制参数被错误估计, 进而影响整体数据补全的精度。为提升模型在上述场景中的适应性, 可以对式(12)的回归函数加以扩展。例如, 引入时间滞后项以捕捉历史缺失模式对当前缺失概率的影响, 或嵌入缺失机制识别逻辑以实现 MCAR 与 MNAR 类型的自适应区分, 从而增强模型对复杂缺失模式的刻画能力与适应性。

2.2 基于 ER 规则的多源信息融合

针对上述缺失数据估计结果, 本节开展基于

ER 规则的网络性能评估, 具体步骤如下。

步骤 1 证据生成。对于 L 个测试指标 $\{e_1, e_2, \dots, e_L\}$, 经过改进的 EM 算法填补后的完整数据集为 $\hat{Z} = \{\hat{z}_l(t), t = 1, 2, \dots, T, l = 1, 2, \dots, L\}$, 其中, $\hat{z}_l(t)$ 为指标 e_l 在时刻 t 的估计值, 即

$$\hat{z}_l(t) = \begin{cases} \text{估计值}, \delta_l(t) = 0 \\ z_l(t), \delta_l(t) = 1 \end{cases} \quad (18)$$

其中, $\delta_l(t)$ 由式(11)定义, $z_l(t)$ 为原始观测值。然后, 定义完备的评估等级空间为 $G = \{G_1 < G_2 < \dots < G_N\}$, 其中 $<$ 表示评估等级的严格递增关系, 且每个等级 G_n 对应参考值为 $F_l(G_n)$ 。

最后, 在不考虑局部无知和全局无知的前提下, 指标 e_l 在时刻 t 的置信分布表示为

$$\begin{aligned} B(e_l) &= \langle G_n, \hat{\beta}_{n,l}(t) \rangle_{n=1}^N \\ \text{s.t. } \sum_{n=1}^N \hat{\beta}_{n,l}(t) &\equiv 1 \end{aligned} \quad (19)$$

其中, $\hat{\beta}_{n,l}(t)$ 表示指标 e_l 在 t 时刻相对于等级 G_n 的置信度, 计算式为

$$\hat{\beta}_{j,l}(t) = \begin{cases} \frac{\hat{z}_l(t) - F_l(G_n)}{F_l(G_{n+1}) - F_l(G_n)}, j = n + 1, \\ F_l(G_n) \leq \hat{z}_l(t) \leq F_l(G_{n+1}) \\ F_l(G_{n+1}) - \hat{z}_l(t) \\ F_l(G_{n+1}) - F_l(G_n), j = n, \\ F_l(G_n) \leq \hat{z}_l(t) \leq F_l(G_{n+1}) \\ 0, \text{其他} \end{cases} \quad (20)$$

式(20)通过线性转化逻辑实现指标数据到等级置信度的映射, 具体而言: 当 $\hat{z}_l(t)$ 位于等级 G_{n+1} 和 G_n 对应的参考区间时, 采用线性转化方法计算置信度。此时, 相对于等级 G_n 的置信度 $\hat{\beta}_{n,l}(t)$ 随 $\hat{z}_l(t)$ 的增大而线性下降, 相对于等级 G_{n+1} 的置信度 $\hat{\beta}_{n+1,l}(t)$ 随 $\hat{z}_l(t)$ 的增大而线性上升。若 $\hat{z}_l(t)$ 不在相应区间内, 则其与 G_n 、 G_{n+1} 不存在关联, 故置信度为 0。

步骤 2 计算混合权重。为综合考虑数据质量对评估结果的影响, 本节引入混合权重的概念, 目的是平衡指标的重要性、数据完整性和数据可靠性。首先, 定义 d_l 为数据完整度以量化指标缺失的严重程度, 计算式为

$$d_i = \frac{\sum_{t=1}^T \delta_i(t)}{T} \quad (21)$$

其中, $d_i = 1$ 表示指标 e_i 在全部观测时刻均无数据缺失, 数据完整性最高; d_i 越小表示数据缺失越严重。其次, 定义数据缺失条件下指标 e_i 的可信程度为可靠度 r_i , 计算式为

$$r_i = \frac{T - y_i}{T} = \frac{T - \sum_{t=1}^T y_i(t)}{T} \quad (22)$$

其中, y_i 表示指标 e_i 不可靠数据点的数量; $y_i(t)$ 为二元标志, 计算式为

$$y_i(t) = \begin{cases} 0, \hat{\mu}_i - \varphi_i \hat{\sigma}_i \leq \hat{z}_i(t) \leq \hat{\mu}_i + \varphi_i \hat{\sigma}_i \\ 1, \text{其他} \end{cases} \quad (23)$$

其中, φ_i 是专家设定的可信区间参数, 用于识别异常值; $\hat{\mu}_i$ 与 $\hat{\sigma}_i$ 为 2.1 节改进的 EM 算法估计后的数据均值和标准差。最后, 假设 e_i 的初始权重为 w_i , 其值往往由专家根据证据的重要程度设定。结合以上分析, 混合权重 \tilde{w}_i 可定义为

$$\tilde{w}_i = \frac{d_i w_i}{1 + w_i - r_i} \quad (24)$$

显然, 当不考虑数据缺失时, 式(24)退化为 $\tilde{w}_i = \frac{w_i}{1 + w_i - r_i}$ 。该式与文献[10]中 ER 规则的权重归一化形式一致, 其中参数 $\frac{1}{1 + w_i - r_i}$ 起到归一化作用, 确保权重合理分配。进一步地, 当专家设定的初始权重与可靠度相等, 即 $w_i = r_i$ 时, 式(24)可再简化为 $\tilde{w}_i = w_i$, 此时 ER 规则退化为经典的 Dempster 组合规则。这表明在数据完全可靠且完整时, 专家权重可直接使用。然而, 在实际网络环境中, 数据缺失是不可避免的。通过引入 d_i , 式(24)能够动态调整混合权重。 d_i 越高, 数据越完整, 赋予的权重值就相应提升; 反之, 融合过程中数据缺失严重的指标的影响会被降低, 这有助于增强评估模型的适应性。

步骤 3 多源证据融合。基于 ER 规则的解析形式, 网络性能的评估结果可以表示为

$$\hat{P}_{n,e(L)}(t) = \frac{\zeta \prod_{i=1}^L (\tilde{w}_i \hat{\beta}_{n,i}(t) + 1 - \tilde{w}_i) - \zeta \prod_{i=1}^L (1 - \tilde{w}_i)}{1 - \zeta \left[\prod_{i=1}^L (1 - \tilde{w}_i) \right]} \quad (25)$$

$$\zeta = \left[\sum_{n=1}^N \prod_{l=1}^L (\tilde{w}_l \hat{\beta}_{n,l}(t) + 1 - \tilde{w}_l) - (N-1) \prod_{l=1}^L (1 - \tilde{w}_l) \right]^{-1} \quad (26)$$

其中, $\hat{p}_{n,e(L)}(t)$ 表示评估结果相对于等级 G_n 的置信度。为获取定量的评估结果, 假设 $\hat{p}_{n,e(L)}(t)$ 的效用参考值为 $u(G_n)$, 则输出效用 $\hat{U}(t)$ 可表示为

$$\hat{U}(t) = \sum_{n=1}^N \hat{p}_{n,e(L)}(t) u(G_n) \quad (27)$$

2.3 基于 Fmincon 的参数优化

由于部分参数 (如参考值 $F_l(G_n)$ 、初始权重 w_l 、可信区间参数 φ_l 等) 通常依赖专家经验设定, 难以保证参数的合理性。因此, 需要建立参数优化模型以提升评估精度。首先, 定义待优化参数集为 $V = \{F_l(G_n), w_l, u(G_n), \varphi_l\}$, 为全面评估模型性能, 本文引入多个评价指标。

1) 均方误差 (MSE, mean square error): 衡量评估结果与真实值之间的平均平方偏差, 作为主要优化目标。

2) 均方对数误差 (MSLE, mean squared logarithmic error): 基于对数尺度衡量评估结果与真实值之间的差异。

3) 中位数绝对误差 (MedAE, median absolute error): 衡量评估结果与真实值之间偏差的稳健中心趋势。

4) 交叉熵损失 (CEL, cross entropy loss): 衡量评估结果的概率分布与真实分布之间的差异。

上述指标的具体计算式分别为

$$\left\{ \begin{array}{l} \text{MSE} = \frac{\sum_{t=1}^T (\hat{U}(t) - U(t))^2}{T} \\ \text{MSLE} = \frac{\sum_{t=1}^T (\log(\hat{U}(t) + 1) - \log(U(t) + 1))^2}{T} \\ \text{MedAE} = \text{median}(|\hat{U}(t) - U(t)|) \\ \text{CEL} = -\frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \sum_{l=1}^L \beta_{n,l}(t) \log(\hat{\beta}_{n,l}(t)) \end{array} \right. \quad (28)$$

其中, $\beta_{n,l}(t)$ 为网络的真实置信分布, $U(t)$ 为网络的真实性能状态。一般而言, 优化目标是最小化 MSE, 具体为

$$\begin{aligned}
 & \min \text{MSE}(V) \\
 & \text{s.t. } F_l^{\min}(G_n) \leq F_l(G_n) \leq F_l^{\max}(G_n) \\
 & \quad w_l^{\min} \leq w_l \leq w_l^{\max} \\
 & \quad u^{\min}(G_n) \leq u(G_n) \leq u^{\max}(G_n) \\
 & \quad \varphi_l^{\min} \leq \varphi_l \leq \varphi_l^{\max}
 \end{aligned} \tag{29}$$

其中, $F_l^{\max}(G_n)$ 和 $F_l^{\min}(G_n)$ 分别表示 $F_l(G_n)$ 的上下界约束, 其余参数含义与之类似。针对上述优化目标, 本节采用 MATLAB 的 Fmincon 函数进行参数优化。该方法通过内点算法高效处理非线性约束条件, 特别对等式约束提供精确优化。同时利用解析梯度加速收敛过程, 避免数值微分导致的累积误

差。通过参数优化过程, 可以有效降低模型对专家知识的依赖, 提高评估结果的准确性。

2.4 整体实施流程

结合前述内容, 本文提出一种缺失数据场景下的网络性能评估方法, 其实施流程如图 2 所示, 主要步骤阐述如下。

步骤 1 首先, 基于包含缺失数据的观测数据集 X 初始化参数 $\theta^{(0)}$; 然后, 根据式(13)~式(17), 经 E 步与 M 步迭代优化, 对参数进行更新; 最后, 当截止条件满足时, 输出估计后的完整数据集 \hat{Z} 。

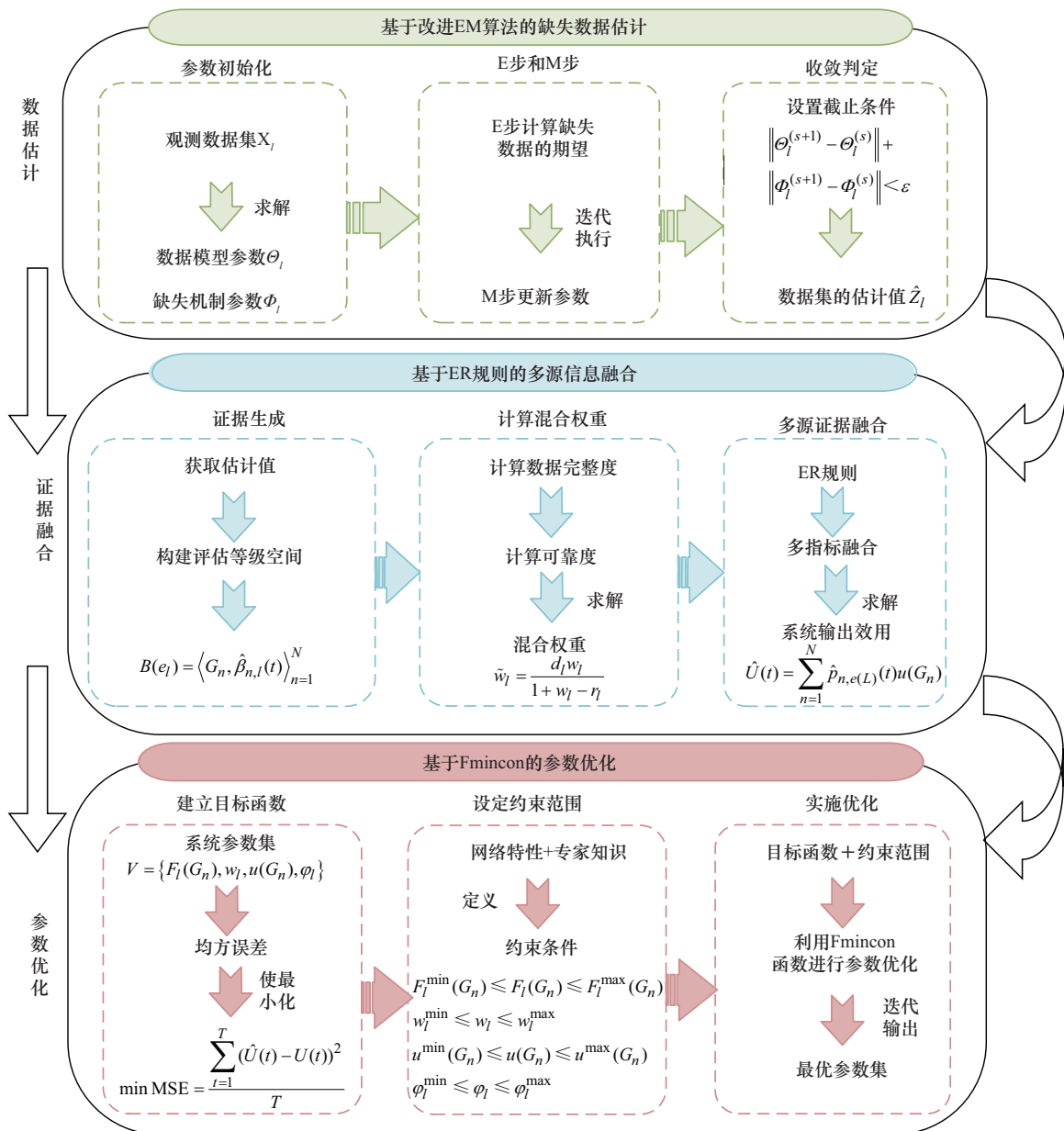


图2 网络性能评估方法的实施步骤

步骤 2 首先, 结合网络工作机理和实际工作环境, 设定指标参考等级, 并基于式(20)计算完整数据集 \hat{Z} 中各指标的置信分布 $\hat{\beta}_{n,l}(t)$; 然后, 依据数据完整度 d_l 与可靠度 r_l , 计算混合权重 \hat{w}_l ; 最后, 基于式(25)~式(27), 融合多个证据, 得到性能评估结果 $\hat{U}(t)$ 。

步骤 3 首先, 以评估误差最小为目标建立优化函数, 如式(28)所示; 然后, 根据参数的物理意义, 设置相应的约束条件, 如式(29)所示; 最后, 采用 MATLAB 中的 Fmincon 函数对参数集进行优化求解, 实现参数的自适应调整。

3 案例分析

为了验证本文所提网络性能评估方法的有效性, 本节选取某工业网络测试平台进行实验验证。

3.1 实验背景描述

在本次实验中, 选取以下性能指标。

- 1) 传输时延 (TD, time delay): 端到端数据包传输时延。
- 2) 能量消耗 (EC, energy consumption): 单位吞吐量下的能量损耗。
- 3) 网络吞吐量 (NT, network throughput): 有效数据传输速率。

在上述指标中, TD 反映了网络的路径效率和

拥塞状况, 其测量方法为在端设备嵌入高精度时间戳探针, 记录数据包从源节点 MAC 层发出到目的节点 MAC 层接收的时间差; EC 指标表征了单位数据吞吐量的能耗强度, 是评估网络能效的关键参数, 通常采用高精度功率探针串接的方式进行直接测量; NT 用于衡量网络承载有效业务数据的能力, 对工业应用的容量需求至关重要, 其测量方法是通过网络测试仪注入全帧长流量, 并且在目的端口统计单位时间内成功转发的数据量。本次实验通过连续 400 组独立测试, 模拟传感器故障、通信丢包等实际场景, 采集 TD、EC、NT 的监测数据集, 如图 3 所示。实验结果表明, TD 和 EC 呈现出稳定上升趋势, 且波动范围分别控制在 $\pm 0.3 \text{ ms}$ 与 ± 0.12 内; NT 则表现出明显的下降趋势, 且波动范围达 $\pm 1.4 \text{ Gbit/s}$, 这种差异源于 NT 对网络负载变化的高度敏感性。

3.2 缺失数据估计

在数据处理阶段, 严格按照数据缺失率为 10% 的原则, 对采集到的原始数据集进行剔除, 具体剔除情况如图 4 所示。各指标最终的实际缺失数据量如表 1 所示。然后, 基于式(11)~式(17)对剔除后产生的缺失数据进行估计, 如图 5 所示。经计算, 此时的估计误差为 2.349×10^{-3} , 基本满足后续的性能评估需求。

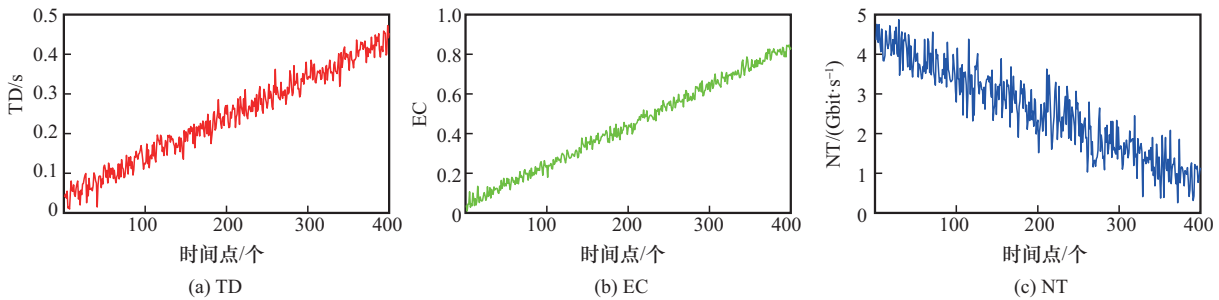


图 3 网络评估指标的原始数据

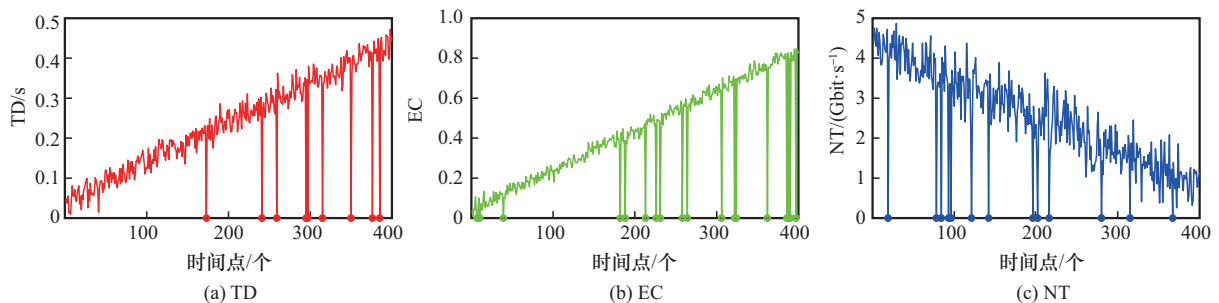


图 4 性能指标的缺失数据

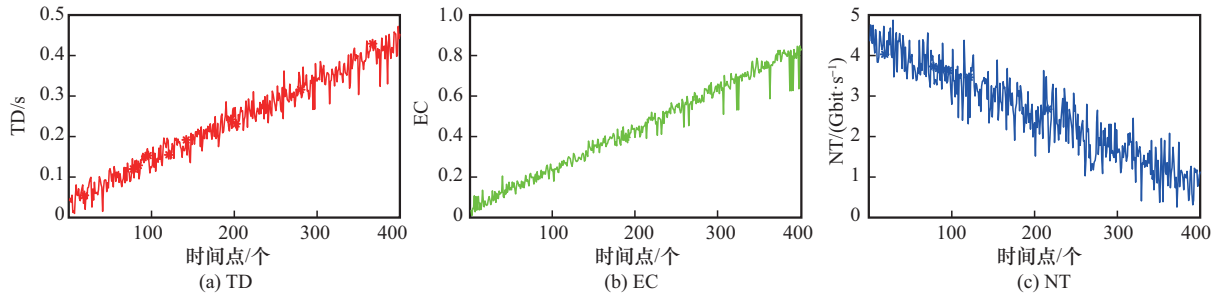


图5 估计后的完整数据

表1 各指标的缺失数据量

性能指标	缺失数据量/个
TD	9
EC	18
NT	13

3.3 评估模型的建立

针对上述缺失数据估计结果,建立基于ER规则的网络性能评估模型,具体步骤如下。

步骤1 构造置信分布。根据估计结果以及专家经验,设立各指标的参考等级及参考值,如表2所示。

表2 性能指标参考值

性能指标	θ_1 (高)	θ_2 (中)	θ_3 (低)
TD	0.01	0.25	0.55
EC	0.01	0.50	0.90
NT	5	2	0.25

基于式(20),将各指标的数据转化为置信分布形式,如图6~图8所示。

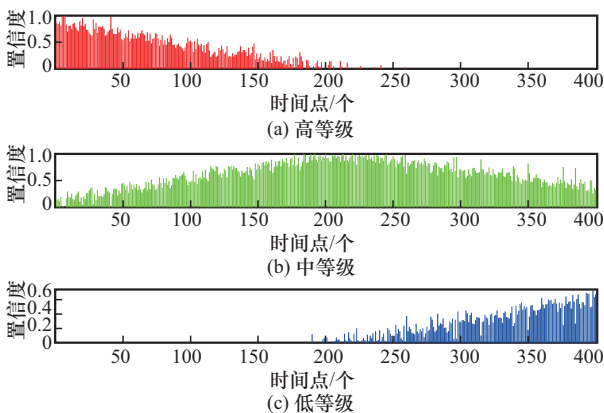


图6 TD的置信分布

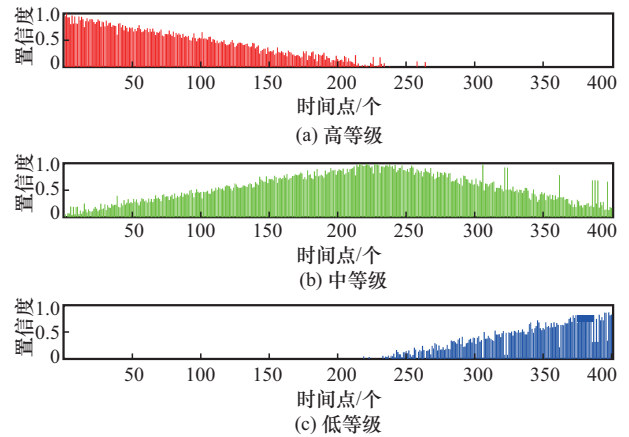


图7 EC的置信分布

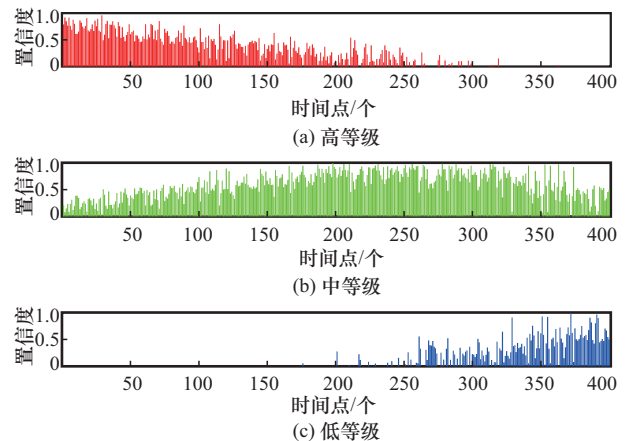


图8 NT的置信分布

步骤2 计算混合权重。首先,根据式(21)计算出3个指标的数据完整度 d_1 、 d_2 、 d_3 分别为0.977 5、0.955 0、0.967 5。其次,根据专家经验以及指标的重要性,将3个评估指标的初始权重值分别设定为 $w_1 = 0.8$ 、 $w_2 = 0.9$ 、 $w_3 = 0.6$ 。然后,基于式(22)和式(23)计算出数据缺失条件下3个指标的可靠度 r_1 、 r_2 、 r_3 分别为0.380 0、0.285 0、0.397 5。最后,根据式(24)得到混合权重值 \tilde{w}_1 、 \tilde{w}_2 、 \tilde{w}_3 分别为0.550 7、0.532 2、0.482 7。

步骤3 多源证据融合。根据式(25)和式(26)对指标进行融合, 得到评估结果的置信度分布结果, 如图9所示。根据图9可知, 网络性能整体稳定维持于中等级。对于高等级, 在前200个时间点内的置信度总体呈现下降趋势, 最大值为0.961 6, 之后基本稳定在0.090 0附近。对于低等级, 置信度随着时间点的增加总体呈现增大趋势, 并逐渐超过0.500 0, 最大值为0.850 9。上述结果说明, 网络性能随运行时间的增加逐渐从高退化至低。

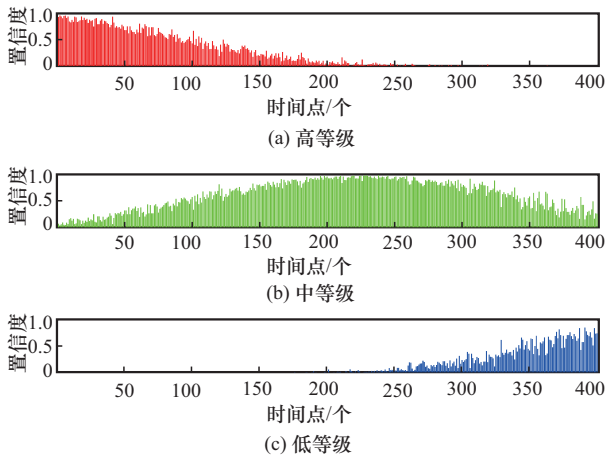


图9 评估结果的置信度分布

需要说明的是, 获取网络真实的性能状态在实际场景中存在难度: 一方面, 现实网络环境中存在复杂的随机干扰, 难以剥离无关因素对性能的影响; 另一方面, 性能状态的真值缺乏统一的量化标准, 不同业务场景下的性能判定维度也存在差异。为实现性能状态的量化, 本实验将网络真实性能状态设计为近似阶梯形的分级形式, 对应“好、中、差”3个等级, 如图10中的折线所示。这种划分的依据为观测指标的恶化程度: 在任一时刻, 若TD、EC、NT这3项指标均无明显劣化, 则判定性能状态为“好”; 若仅有部分指标出现小幅恶化、未突破关键阈值, 则判定为“中”; 若多数或全部指标发生显著恶化、已无法满足基础业务需求, 则判定为“差”。此外, 在ER规则框架下, 为实现性能状态的量化分析, 本实验将“好、中、差”这3个等级的效用值设定为1、0.6和0, 如表3所示。需要强调的是, 这一数值设定仅为验证方法的量化基准, 并非唯一标准。事实上, 也可采用2、1和0等其他数值序列, 只要保持等级间单调递减的关系即可, 并不影响实验结论。

性能等级	效用值
好	1
中	0.6
差	0

在上述假设下, 基于式(27), 得到评估结果的期望效用, 如图10中的曲线所示。在图10中, 网络性能的实际评估结果总体呈现下降趋势, 与图9的变化趋势基本相符。此时可计算出评估结果的各项误差指标, 如表4所示。

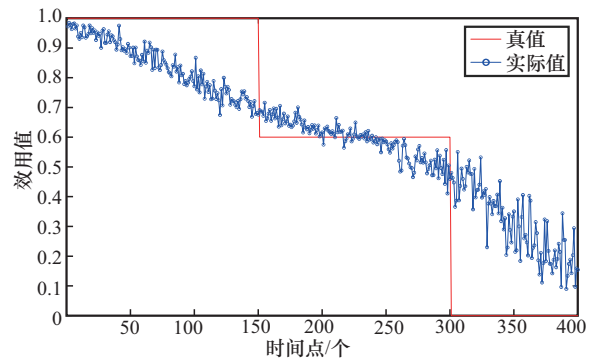


图10 网络性能评估的期望效用

误差指标	数值
MSE	0.041 4
MSLE	0.023 1
MedAE	0.114 9
CEL	0.578 8

3.4 参数优化

为提升评估精度, 需要进行参数优化。在实验中, 采用偶数组数据作为训练数据, 剩余数据为测试数据。各参数的约束条件及优化结果如表5所示, 优化前后模型的评估误差如表6所示, 优化后的评估结果如图11所示。

对比优化前后的参考值可以发现, 整体参考值的变化幅度较小, 说明专家给出的初始参考值较为准确。相比之下, w_3 从0.600 0降低至0.097 7, 优化效果最为突出。优化后各误差指标均得到改善, 表6中的“相对下降幅度”一列表明参数优化有效降低了专家知识的不确定性, 提升了模型的评估精度与参数校准能力。此外, 优化后模型

输出的置信分布与真实分布更为接近, 整体评估效能显著增强。

表 5 参数约束条件与优化结果

参数类别	参数名称	约束条件	优化结果
TD 参考值	θ_1 (高)	[0, 0.1]	0.099 5
	θ_2 (中)	[0.2, 0.3]	0.229 9
	θ_3 (低)	[0.5, 0.6]	0.500 0
EC 参考值	θ_1 (高)	[0, 0.1]	0.096 5
	θ_2 (中)	[0.4, 0.6]	0.460 2
	θ_3 (低)	[0.8, 0.95]	0.800 8
NT 参考值	θ_1 (高)	[4.8, 5.5]	4.973 2
	θ_2 (中)	[1.7, 3]	2.311 5
	θ_3 (低)	[0, 0.4]	0.278 9
等级效用值	好	[0.7, 1]	0.999 5
	中	[0.4, 0.7]	0.636 9
	差	[0, 0.4]	0.000 3
初始权重	w_1	[0, 1]	0.660 5
	w_2	[0, 1]	0.807 5
	w_3	[0, 1]	0.097 7
可信区间参数	φ_1	[0, 1]	0.533 6
	φ_2	[0, 1]	0.500 0
	φ_3	[0, 1]	0.562 6

表 6 优化前后模型的评估误差

评估指标	优化前	优化后	相对下降幅度
MSE	0.041 4	0.024 0	42.03%
MSLE	0.023 1	0.015 7	32.03%
MedAE	0.114 9	0.113 4	1.31%
CEL	0.578 8	0.380 5	34.26%

表 7 各模型参数设置

评估模型	融合方法	缺失处理方法	关键参数说明
模型 1	ER 规则	直接删除	缺失数据不进行填补, 直接参与后续计算; 数据完整度与可靠度基于含缺失数据的原始数据集计算
模型 2	ER 规则	均值插补	缺失值使用各指标历史均值填充, 填充后进行 ER 融合
模型 3	ER 规则	随机森林插补	插补参数: 树数量为 100, 最大迭代次数为 10, 收敛容差为 1×10^{-4} ; 森林结构: 特征子集数量为 2
模型 4	ER 规则	多重插补	插补参数: 迭代次数为 10, 生成插补集数量为 5; 初始填补采用列均值, 插补过程使用基于其他变量的简单线性回归预测, 最终结果取 5 组均值
模型 5	模糊推理	改进 EM 算法	模糊系统: 输入和输出变量各设 3 个模糊集, 采用三角形隶属函数; 规则库: 27 条规则, 基于“恶化指数”映射, 规则强度取输入隶属度最小值; 推理方法: 模糊推理, 最大值聚合
模型 6	三层前馈神经网络	改进 EM 算法	网络结构: 输入层有 3 个节点, 输出层有 1 个节点, 隐藏层分别有 3 个和 2 个节点; 训练设置: 学习率为 0.1, 最大训练轮次为 200, 训练目标误差为 5×10^{-2} , 采用 SCG 训练算法

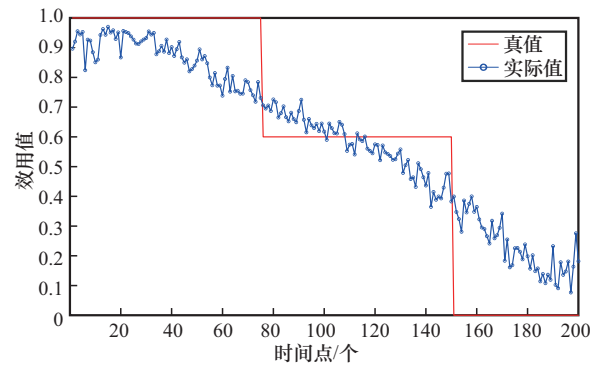


图 11 优化后的评估结果

3.5 对比实验

为了进一步说明所提模型的有效性, 本节分别与不考虑缺失数据的评估模型 (记为模型 1)、基于均值估计的评估模型 (记为模型 2)、基于随机森林插补的评估模型 (记为模型 3)、基于多重插补的评估模型 (记为模型 4)、基于模糊推理的评估模型 (记为模型 5) 和基于神经网络的评估模型 (记为模型 6) 进行对比。为确保对比的公平性, 所有模型均采用相同的训练集与测试集。同时, 模型 1~模型 4 均以相同的含缺失数据的原始数据集作为输入, 模型 5 和模型 6 则使用经过改进 EM 算法估计后的完整数据集作为输入。此外, ER 规则相关参数在各模型中保持一致, 上述所有模型的参数设置如表 7 所示。

各模型在测试集上的评估结果如图 12 和表 8 所示。

根据表 8, 从核心优化目标 MSE 来看, 本文所提模型在精度上显著优于其他对比模型。

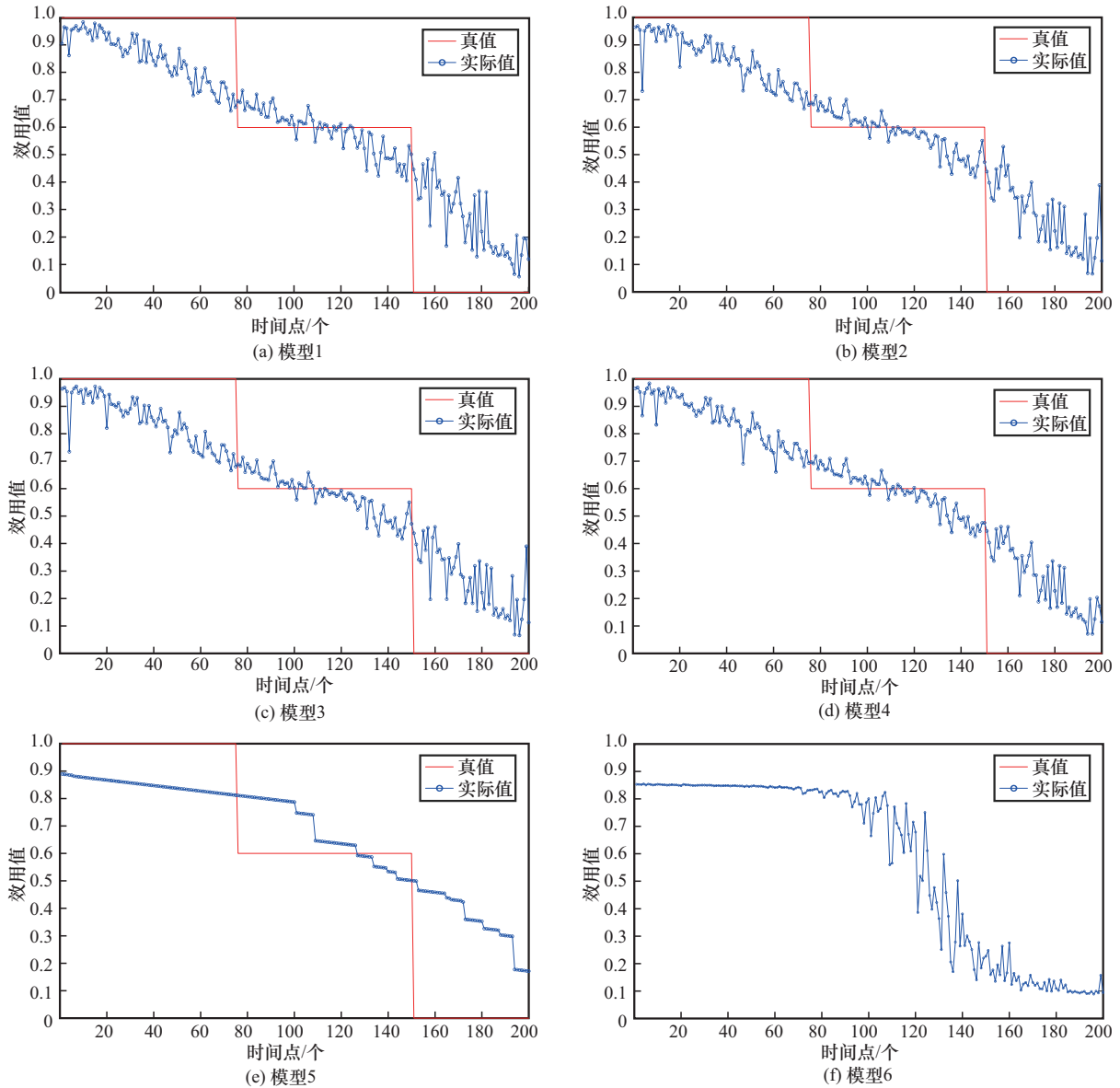


图 12 各模型的评估结果

表 8

各模型的评估结果对比

评估模型	所提模型	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6
MSE	0.024 0	0.034 7	0.035 7	0.034 5	0.034 3	0.051 2	0.033 1
MSLE	0.015 7	0.019 3	0.020 2	0.019 5	0.019 5	0.029 4	0.018 9
MedAE	0.113 4	0.115 1	0.118 7	0.118 8	0.120 5	0.163 2	0.128 4
CEL	0.380 5	0.483 6	0.503 8	0.495 5	0.479 8	—	0.713 6

1) 模型 1 和模型 2 的 MSE 分别为 0.034 7 和 0.035 7, 分别比本文模型高约 44.58% 和 48.75%。这说明在 MNAR 机制下, 简单删除法会导致信息损失, 而均值插补法则无法反映网络的动态变化, 两者均会引入显著偏差。

2) 模型 3 和模型 4 的 MSE 虽优于模型 1 和模型 2, 但仍高于本文模型。这说明随机森林法虽能处理非线性关系, 但其缺失数据填充策略相对简单, 且模型本身不具备处理信息不确定性的机制。多重插补虽考虑了填补不确定性, 但其插补过程通

常基于MAR假设,使评估精度不高。

3) 模型5的MSE最高,主要原因在于其预设的模糊规则库难以完备地覆盖复杂多变的网络场景,对缺失数据和不确定性的处理较为粗糙,导致评估能力不足。模型6的MSE比本文模型高出37.92%,其性能受限源于在有限样本与存在缺失的条件下易发生过拟合。在CEL指标上,本文模型也明显优于其他模型,体现了其概率校准的可靠性。

综合MSE、MSLE、MedAE与CEL这4项指标来看,本文基于改进EM算法的缺失数据估计、ER规则融合及参数优化等步骤,使所提出的评估模型在复杂缺失数据场景下具有更优的评估性能。

3.6 复杂度分析

除了评估精度外,算法的计算效率和适用范围也是实际应用中的关键考量。本节针对改进EM算法的计算复杂度与适用规模进行补充分析,以明确其在大规模网络性能评估中的应用边界。在网络性能指标数量为 L 、观测时间点数为 T 时,算法复杂度分析如下:在E步中,需要计算每个缺失数据的条件期望,其复杂度为 $O(L \times T)$;在M步中,更新分布参数和缺失机制参数需要对所有数据点进行遍历计算,复杂度同样为 $O(L \times T)$ 。设算法收敛所需迭代次数为 K ,则总复杂度为 $O(K \times L \times T)$ 。综上所述,算法的计算负担与指标数量 L 和观测时间点数 T 成线性关系,具备一定的扩展性。

在本次案例中,即 $L = 3$ 且 $T = 400$ 时,算法在普通工作站上的平均运行时间小于8s,基本能够满足评估的实时性需求。然而,对于需要处理成百上千维指标或超高频采样的极端场景,计算时间可能成为瓶颈。为此,本文方法适用于指标数量在2~50、样本量在100~10 000的网络性能评估场景。对于更高维或更大规模的数据,可以通过指标筛选或结合分布式计算框架来提升处理效率。

4 结束语

本文针对现有ER规则未考虑数据缺失问题的不足,提出一种面向数据缺失场景的网络性能评估方法,核心贡献如下:针对网络场景中常见的MNAR机制,构建融合数据分布与缺失机制的联合估计模型,并通过改进传统EM算法,实现对缺

失数据的精准估计;在ER规则框架下,引入数据完整度作为指标权重的调节依据,增强了ER规则在数据缺失场景下的适用性;构建以评估误差最小化为目标的参数优化模型,实现参数的自适应调整,从而提升了评估结果的准确性。

然而,本文仍存在一定局限性:一是计算复杂度较高,改进后的EM算法需同步估计数据分布与缺失机制参数,当评估指标繁多或数据规模较大时,迭代优化过程中的计算量会显著增加,可能影响模型评估的实时性;二是对缺失机制的假设依赖性较强,尽管模型针对MNAR机制构建了估计框架,但实际网络环境中的缺失机制往往更为复杂,模型对于极端缺失或混合缺失场景的适应性仍有待进一步验证。未来研究将重点围绕提升算法效率、增强模型对多样化缺失机制的适应性展开。

参考文献:

- [1] 黄韬,霍如,刘江,等.未来网络发展趋势与展望[J].中国科学(信息科学),2019,49(8):941-948.
HUANG T, HUO R, LIU J, et al. Development trends and prospects of future networks[J]. Science in China (Information Sciences), 2019, 49(8): 941-948.
- [2] 南洋,陈琳.基于客观权重确定的数据中心网络性能评估方法[J].计算机应用,2015,35(11):3055-3058,3091.
NAN Y, CHEN L. Performance evaluation method based on objective weight determination for data center network[J]. Computer Applications, 2015, 35(11): 3055-3058, 3091.
- [3] 孙剑文,张斌,常禾雨,等.网络异常检测中的流量表示研究[J].通信学报,2025,46(1):192-209.
SUN J W, ZHANG B, CHANG H Y, et al. Research on traffic representation in network anomaly detection[J]. Journal on Communications, 2025, 46(1): 192-209.
- [4] 张冬艳,胡铭曾,张宏莉.基于测量的网络性能评价方法研究[J].通信学报,2006,27(10):74-79,85.
ZHANG D Y, HU M Z, ZHANG H L. Study on network performance evaluation method based on measurement[J]. Journal on Communications, 2006, 27(10): 74-79,85.
- [5] 韩晓霞,王杰,陈媛,等.基于证据推理规则的给排水管道安全性评估[J].科学技术与工程,2021,21(32):13758-13764.
HAN X X, WANG J, CHEN Y, et al. Safety assessment of water supply and drainage based on evidential reasoning rule[J]. Science Technology and Engineering, 2021, 21(32): 13758-13764.
- [6] 黄赞,张帆,郭威,等.一种基于数据标准差的卷积神经网络量化方法[J].电子学报,2023,51(3):639-647.
HUANG Y, ZHANG F, GUO W, et al. A quantification method of convolutional neural network based on data standard deviation[J]. Acta Electronica Sinica, 2023, 51(3): 639-647.
- [7] 史宏艳,罗敏霞.基于区间值相似度的模糊推理算法[J].电子学报,2022,50(11):2738-2745.
SHI H Y, LUO M X. Fuzzy inference method based on interval-valued similarity measure[J]. Acta Electronica Sinica, 2022, 50(11): 2738-2745.

- [8] CHENG C, QIAO X Y, LUO H, et al. A semi-quantitative information based fault diagnosis method for the running gears system of high-speed trains[J]. IEEE Access, 2019, 7: 38168-38178.
- [9] KURT Y B, UFLAZ E, AKYUZ E, et al. Analysis of human reliability in detecting GPS spoofing on ECDIS in congested waterways under evidential reasoning and HEART approach[J]. Computers & Security, 2025, 151: 104316.
- [10] 周志杰, 唐帅文, 胡昌华, 等. 证据推理理论及其应用[J]. 自动化学报, 2021, 47(5): 970-984.
ZHOU Z J, TANG S W, HU C H, et al. Evidential reasoning theory and its applications[J]. Acta Automatica Sinica, 2021, 47(5): 970-984.
- [11] XU X B, HUANG W D, ZHANG X L, et al. An evidential reasoning-based information fusion method for fault diagnosis of ship rudder[J]. Ocean Engineering, 2025, 318: 120082.
- [12] HAN Y, YU R F, WANG J, et al. Evidential reasoning rule with dynamic reliability for performance assessment of wireless sensor network[J]. ICT Express, 2025, 11(1): 19-25.
- [13] SUN Y J, FENG Z C, ZHOU Z J, et al. New evidential reasoning-rule based optimal maintenance time determining method considering dynamic parameter boundary[J]. IEEE Transactions on Industrial Informatics, 2025, 21(2): 1239-1248.
- [14] 詹兆康, 胡旭光, 赵浩然, 等. 基于多变量时空融合网络的风机数据缺失值插补研究[J]. 自动化学报, 2024, 50(6): 1171-1184.
ZHAN Z K, HU X G, ZHAO H R, et al. Study of missing value imputation in wind turbine data based on multivariate spatiotemporal integration network[J]. Acta Automatica Sinica, 2024, 50(6): 1171-1184.
- [15] 邓建新, 单路宝, 贺德强, 等. 缺失数据的处理方法及其发展趋势[J]. 统计与决策, 2019, 35(23): 28-34.
DENG J X, SHAN L B, HE D Q, et al. Processing method of missing data and its developing tendency[J]. Statistics and Decision, 2019, 35(23): 28-34.
- [16] 郑智泉, 陈妍, 王孟孟, 等. 不同缺失率下的数据填补算法稳定性研究[J]. 统计与决策, 2023, 39(8): 12-17.
ZHENG Z Q, CHEN Y, WANG M M, et al. Research on stability of data filling algorithms with different miss rates[J]. Statistics and Decision, 2023, 39(8): 12-17.
- [17] 李聪, 骆志刚. 基于数据非随机缺失机制的推荐系统托攻击探测[J]. 自动化学报, 2013, 39(10): 1681-1690.
LI C, LUO Z G. Detecting shilling attacks in recommender systems based on non-random-missing mechanism[J]. Acta Automatica Sinica, 2013, 39(10): 1681-1690.
- [18] 熊中敏, 郭怀宇, 吴月欣, 等. 缺失数据处理方法研究综述[J]. 计算机工程与应用, 2021, 57(14): 27-38.
XIONG Z M, GUO H Y, WU Y X, et al. Review of missing data processing methods[J]. Computer Engineering and Applications, 2021, 57(14): 27-38.
- [19] 宋枝璘, 郭磊, 郑天鹏. 认知诊断缺失数据处理方法的比较: 零替换、多重插补与极大似然估计法[J]. 心理学报, 2022, 54(4): 426-444.
SONG Z L, GUO L, ZHENG T P. Comparison of missing data handling methods in cognitive diagnosis: zero replacement, multiple imputation and maximum likelihood estimation[J]. Acta Psychologica Sinica, 2022, 54(4): 426-444.
- [20] 刘钰, 王方刚, 张静文, 等. 多径信道下基于 EM 算法的盲 LDPC 编码器识别研究[J]. 通信学报, 2018, 39(9): 43-48.
LIU Y, WANG F G, ZHANG J W, et al. EM-based blind LDPC identification in multipath channels[J]. Journal on Communications, 2018, 39(9): 43-48.
- [21] 陈佳磊. 用于非随机缺失数据的深度生成式插补模型的研究[D]. 长春: 吉林大学, 2024.
CHEN J L. Research on deep generative imputation models for non-random missing data[D]. Changchun: Jilin University, 2024.
- [22] 陈晨. 非随机缺失机制下参数分数插补法的应用研究[D]. 北京: 北京林业大学, 2022.
CHEN C. Applied research on parametric fractional imputation method under non-random missing mechanism[D]. Beijing: Beijing Forestry University, 2022.
- [23] 周鑫. 基于 EM 算法的 G0 分布参数最大似然估计[J]. 电子学报, 2013, 41(1): 178-184.
ZHOU X. An EM algorithm based maximum likelihood parameter estimation method for the G0 distribution[J]. Acta Electronica Sinica, 2013, 41(1): 178-184.

[作者简介]



韩悦 (1977-), 男, 陕西西安人, 博士, 国防科技大学副教授, 主要研究方向为未来通信网络、智算网络等。



臧晓尧 (1998-), 女, 满族, 吉林蛟河人, 国防科技大学硕士生, 主要研究方向为证据推理、网络态势感知等。



李海 (1997-), 男, 湖南湘乡人, 博士, 国防科技大学讲师, 主要研究方向为空间信息网络、资源调度等。



王杰 (1995-), 男, 安徽滁州人, 博士, 国防科技大学讲师, 主要研究方向为信息融合、网络态势感知等。