专题：通信与 AI 融合

智能运维中 KPI 异常检测的研究进展

王速[1]，卢华[2]，汪硕[1,3]，蔡磊[2]，黄韬[1,3]
（1. 北京邮电大学网络与交换国家重点实验室，北京 100876;
2. 广东省新一代通信与网络创新研究院，广东 广州 510663;
3. 网络通信与安全紫金山实验室，江苏 南京 211111）

摘　要：现有的网络监控和故障修复大多依赖规则系统或者人工处理，然而随着网络规模的不断增大和业务的多样化，这种方式难以满足要求。随着机器学习和深度学习等技术的快速发展，智能运维理论也取得了长足进步，利用人工智能技术提升网络运维智能化能力。KPI（key performance indicator）异常检测是智能运维的一项底层核心技术。针对 KPI 异常检测技术研究展开综述，对 KPI 数据和 KPI 异常进行了描述，并从单指标和多指标两个方面详细介绍了 KPI 异常检测技术的研究现状；分析了 KPI 检测的部署应用问题，讨论了未来的研究方向。

关键词：智能运维；KPI；异常检测
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2021105

Research progress of KPI anomaly detection in intelligent operation and maintenance

WANG Su[1], LU Hua[2], WANG Shuo[1,3], CAI Lei[2], HUANG Tao[1,3]
1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. Guangdong Communications & Networks Institute, Guangzhou 510663, China
3. Purple Mountain Laboratories, Nanjing 211111, China

Abstract: Existing network monitoring and fault repair mostly rely on rule systems or manual processing. However, the increase in network scale and the diversification of services make this approach difficult to deal with. With the rapid development of technology such as machine learning and deep learning, intelligent operation and maintenance theory has also made great progress, using artificial intelligence technology to enhance the intelligent ability of net-

收稿日期：2021−04−05；**修回日期：**2021−05−10
通信作者：卢华，luhua@gdcni.cn
基金项目：广东省重点领域研发计划基金资助项目（No.2018B010113001）；国家重点研发计划基金资助项目（No.2018YFB1800500）；国家自然科学基金资助项目（No.61902033）；北京市自然科学基金资助项目（No.4204105）
Foundation Items: The Research and Development Program in Key Areas of Guangdong Province (No.2018B010113001), The National Key Research and Development Program of China (No.2018YFB1800500), The National Natural Science Foundation of China (No.61902033), Beijing Natural Science Foundation (No.4204105)

work operation and maintenance. KPI (key performance indicator) anomaly detection is an underlying core technology of intelligent operation and maintenance. A survey on the KPI anomaly detection technology was given. Firstly, the KPI data and KPI anomalies were described. Then the research progress of single-dimensional KPI and multi-dimensional KPI anomaly detection were introduced. Then, the deployment and application problems of KPI anomaly detection were analyzed. Finally, future research directions were discussed.

Key words: intelligent operation and maintenance, KPI, anomaly detection

1　引言

网络信息化带动了整个世界经济的发展，促进了人类社会的进步。然而伴随着网络的大规模使用，企业的 IT 系统变得越来越复杂，传统的人工运维方式已经难以满足企业的智能化管理要求[1]。近年来，随着机器学习、深度学习等技术的发展，智能运维（artificial intelligence for IT operations，AIOps）的概念被提出，将 AI 与运维结合，利用人工智能技术自动监控和管理 IT 业务，提升运维效率。

智能运维包含很多关键场景和技术，涉及大型分布式系统的监控、分析、决策等。参考文献[2]将智能运维划分为 3 个阶段：针对历史事件，包含瓶颈分析、热点分析、故障传播图构建等技术；针对当前事件，包含异常检测、异常定位、故障根因分析等技术；针对未来事件，包含故障预测、趋势预测等技术。其中，KPI 异常检测是互联网服务智能运维的一个底层核心技术，上述大多数运维关键技术都依赖于 KPI 异常检测的结果[2]。本文研究智能运维中的 KPI 异常检测技术，从单维 KPI 异常检测和多维 KPI 异常检测两方面进行分析整理，总结和分析目前 KPI 异常检测领域的研究现状和研究趋势。

2　KPI 概述

为了保障基于 Internet 的应用服务的稳定性，运维人员通常会密切监控各种各样的 KPI（key performance indicator）。KPI 的异常往往反映了相关应用的故障，如外部攻击、服务器故障等。因此，高效、可靠的服务依赖于有效的异常检测手段。

2.1　KPI 数据

KPI 数据是一种通过定时采样获取的、具有特定意义的时间序列数据，格式为（时间戳，值）。KPI 可以被粗略地划分为两种类型[3]：服务 KPI 和机器 KPI。服务 KPI 通常是在互联网业务层面的分钟级的统计指标（例如每分钟的网页访问量、服务的在线用户数、服务响应用户的平均时间等），大多数是季节性的平滑 KPI，用来反映服务的质量和规模；机器 KPI 通常反映机器健康状况的秒级的统计指标（例如一台服务器的网卡吞吐率、内存利用率等），大多数是更复杂的细粒度 KPI。除物理意义外，KPI 在形状特征上也呈现多样性，大致可以分为周期性 KPI、稳定型 KPI 和持续波动型 KPI。

2.2　KPI 异常

KPI 序列往往可以呈现不同程度的异常模式，如图 1 所示，大致可以分为以下 3 类[4]。

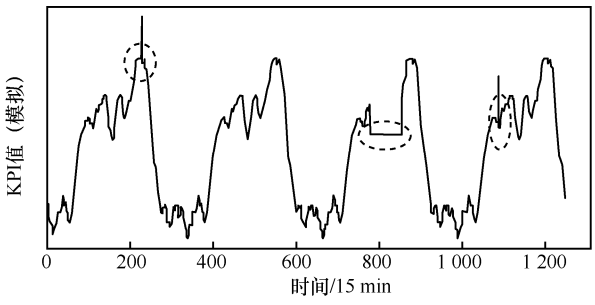


图 1　异常模式示意图
（从虚线圈左至右分别表示点异常、集合异常和峰值异常）

（1）点异常：点异常指在不考虑数据点的时间关系的情况下，某些超出了 KPI 序列正常范围的离群点。

（2）集合异常：集合异常指某些数据点的集合相对于整个 KPI 序列是异常的。集合异常的各个数据点本身可能不是异常的，但作为集合一起出现会被视为异常。

（3）峰值异常：峰值异常也可称为上下文异常，区别于点异常，在某些情况下，虽然数据点仍在 KPI 序列正常范围内，但与邻居点存在很大差异。

3 KPI 异常检测研究现状

异常检测指对不符合预期模式的数据的识别。异常检测作为数据挖掘领域的一项重要内容，在很多方面都有广泛的应用，例如图像异常检测、时间序列异常检测、日志消息异常检测等。在运维领域，KPI 异常检测是一项核心技术，运维团队需要实时监测分析网络设备以及服务的状况，及时发现风险，并尽快采取措施保障应用服务质量。根据场景和应用需求的不同，可以分为单维 KPI 异常检测和多维 KPI 异常检测两个类别。单维 KPI 异常检测关注指标级别的异常状况（KPI 的突增、抖动等），是目前主要的研究方向。区别于单维异常，多维 KPI 异常检测关注实体（例如服务器、航天器等工业设备）级别的异常。多维 KPI 异常示意图如图 2 所示，实体的异常事件通常会引起多个指标同时异常，且指标之间存在一定关联关系，因此多维 KPI 异常检测更具挑战性。

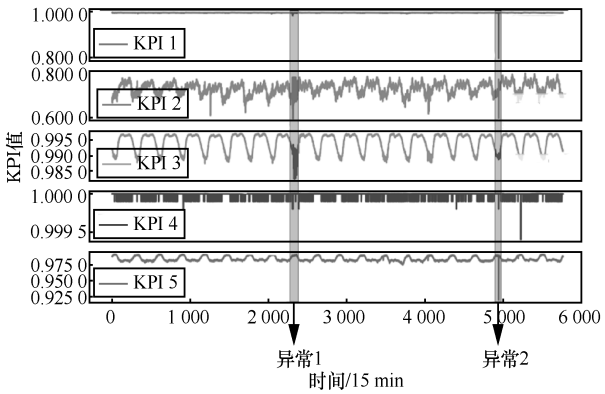


图 2　多维 KPI 异常示意图

3.1 单维 KPI 异常检测

目前，在单维 KPI 异常检测领域，学术界和工业界已存在大量的研究工作。通过对已有研究进行分类，大致可以分为两大类：传统方法和基于机器学习的方法。下面分别对这两类方法的研究现状进行分析和介绍。

3.1.1 *传统方法*

针对单维 KPI 数据的异常检测，主流方法包含基于规则的方法、基于统计学的方法、基于预测的方法、基于邻近度的方法等。

基于规则的方法具有目标明确、结果直观等特点，目前在工业界应用最为广泛。这种方法通常需要专家制定规则，尽管异常检测准确率较好，但对应用场景要求严苛。目前有很多基于规则的异常系统，例如针对亚马逊云服务（Amazon Web Services，AWS）资源和在 AWS 上运行的应用程序进行监控的 Amazon CloudWatch[5]系统。

基于统计学的方法是一类经典方法，总体思想是对数据的分布做出某些假设，然后利用统计推断方法找出该假设条件下的异常。例如熟知的 3σ 准则，假定数据服从正态分布，如果某些值超过 3 倍标准差，那么可以将其视为异常点。3σ 准则具有方法简洁、运算速度快等优势，但由于其存在较强的假设条件，在很多数据中表现不佳。为此，Siffer 等[6]基于极端值理论（extreme value theory，EVT）提出了 SPOT 算法，该算法不对数据分布进行假设，对单峰时间序列和高吞吐数据流序列具有较好的检测效果。

基于预测的方法通过建模并预测 KPI 曲线，根据预测值和实际值之间的误差，采用 $k\sigma$、分位数等方法或基于预设定的阈值等进行异常检测。代表性的时序预测方法有 ARIMA[7]算法和 Holt-Winters[8]算法，主要适用于具有线性趋势和周期波动的 KPI 序列。此外，Facebook 开源了单变量时间序列预测框架 Prophet[9]，具体包含序列

建模、模型评估、问题呈现和可视化反馈 4 个模块。相比于 ARIMA 和 Holt-Winters 模型，Prophet 对于数据缺失和趋势变化具有较强的鲁棒性，并且通常能够很好地处理序列中的异常值。

基于邻近度的方法是指利用异常点和正常点之间相似度低的特点来识别异常，例如最简单的 K 近邻算法基于样本和其第 k 个邻居的距离判断样本是否异常。类似地，Breunig 等[10]提出了局部离群因子（local outlier factor，LOF）检测方法，通过分析样本点的邻域密度检测异常，Kriegel 等[11]提出了基于角度的离群点（angle-based outlier detection，ABOD）检测方法，通过计算每个样本与所有其他样本对所形成的夹角的方差检测异常。

3.1.2　基于机器学习的方法

在实际运维环境中，传统异常检测算法部署到特定服务仍然存在很大的挑战。由于 KPI 种类多、变化快、规律复杂，为保证服务质量，这些方案需要运维人员定期手动调整检测器的内部参数或者监测阈值，而且具体的参数也大多凭借经验而定。在此背景下，基于机器学习技术的智能化运维成为必然的方向。

根据 KPI 曲线是否具有异常标注，可以将异常检测算法大致分为无监督异常检测技术和监督异常检测技术。代表性的监督学习方法有雅虎开发的 EGADS 框架[12]和清华 Netman 实验室开发的 Opprentice 框架[13]。EGADS 和 Opprentice 均为有监督的集成学习方法，利用多种传统异常检测算法输出的异常分数作为特征并利用用户反馈作为标签来训练异常分类器，均在 KPI 异常检测上取得了很好的效果。然而，这两种方法严重依赖良好的人工标注，在实际应用中通常是不可行的。此外，基于多个异常检测器的集成学习分类器也面临运算开销大、正负样本不均衡等问题。因此，无监督学习方法成为 KPI 异常检测的主要研究方向。

基于无监督机器学习的异常检测方法的主要思想是利用单类标签（即正常 KPI 样本）确定“正常区域”，然后通过测量 KPI 观测值和“正常区域”的距离来推断异常，其理念是由于异常检测中正常样本远远多于异常样本，因此即使缺乏标签仍然可以对模型进行训练。Amer 等[14]验证了无监督条件下 SVM 算法的有效性，并对 one-class SVM 算法进行改进，使学习到的决策边界尽可能不受异常值的干扰。Yang 等[15]采用了一种全局最优的期望最大化（expectation maximization，EM）算法，用高斯混合模型（Gaussian mixture model，GMM）拟合数据并基于最大似然方法估计参数。Munz 等[16]提出了基于 K-means 聚类算法的异常网络流量检测方法，该方法也可视为一种样本点间的相似度度量方法。

此外，随着深度学习算法的不断突破，基于无监督深度学习的异常检测算法逐渐成为研究的热点，大多数方法基于生成模型。生成模型的目标是给定训练样本，希望能获得与训练数据具有相同分布的新数据样本，而异常检测需要学习正常模式，因此非常适合生成模型。其中，代表性的算法有变分自编码器（variational auto-encoder，VAE）[17]、生成式对抗网络（generative adversarial network，GAN）[18]、像素循环神经网络（pixel recurrent neural network，PixelRNN）[19]等。以应用较为广泛的 VAE 为例，图 3 表示 VAE 模型的架构，观测数据 x 由隐变量 z 产生，$z \to x$ 是生成模型 $p_\theta(x|z)$，$x \to z$ 是识别模型 $q_\varphi(z|x)$，从自编码器（auto-encoder，AE）的角度来看，$p_\theta(x|z)$ 类似于解码器，$q_\varphi(z|x)$ 类似于编码器。

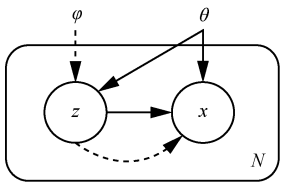


图 3　VAE 模型的架构

基于 VAE 算法，Xu 等[20]提出了无监督异常检测机制 Donut，Donut 针对季节性的 KPI 具有较

好的检测效果，并具有可靠的 KDE（kernel density estimation）理论解释。Li 等[21]在 Donut[20]机制的基础上进行了扩展，改进了 Donut 无法处理与时间相关的异常的缺陷，并更好地处理了 KPI 数据缺失的问题。此外，当面临更复杂的机器 KPI，这些 KPI 表现的非高斯的噪声和更复杂的数据分布使 Donut 等机制难以进行建模。针对此问题，Chen 等[22]提出了 Buzz 机制，通过 Wasserstein 距离[23]和分区分析的方法度量数据分布和生成分布的距离，并将模型转化为贝叶斯网络，另外将 VAE 作为生成模型来产生样本并采用另一个神经网络作为判别器来进行对抗训练，在机器 KPI 的异常检测上取得了较好的效果。

3.2 多维 KPI 异常检测

与单维 KPI 相比，多维 KPI 序列的维数多、数据量大、指标间相互关联等特征带来了更高的研究难度。目前国内外针对多维 KPI 异常检测的研究总体上有两种思路[24]：第一种思路是将多维 KPI 序列按维度切分成多个单维 KPI 序列，基于单维 KPI 检测方法寻找异常；第二种思路是直接分析多维 KPI 序列，例如将多维序列按时间切分成多个子序列，同时结合聚类等算法发现异常。第一种思路的研究相对成熟，但丢失了各维度 KPI 之间的关联性信息，而且为每个 KPI 维护单独的模型也带来更高的成本，因此第二种思路的研究更具有价值。从算法层面，类似于单维 KPI，多维 KPI 异常检测方法也可以分为传统方法和基于机器学习的方法两大类。

3.2.1 传统方法

与单维 KPI 序列类似，针对多维 KPI 数据异常检测的传统方法也可以大致分为基于规则的方法、基于统计学的方法、基于预测的方法和基于邻近度的方法。

基于规则的方法大多考虑多维序列的每条 KPI，并依据固定阈值、同比（与历史周期对比）、环比（与相邻时间对比）等方法判定异常。ADTK[25]是一个可用于单维和多维 KPI 序列异常检测的软件包，提供了一组具有统一 API 的通用检测器，包含基于规则的算法和一部分无监督算法，并支持可视化 KPI 序列和异常事件。

基于统计学的方法大多假设多维 KPI 的各个维度相互独立，利用统计推断方法计算各维度的异常度再进行相加。例如，Goldstein 等[26]提出了 HBOS（histogram-based outlier score）算法，基于各个特征之间的独立性假设，分别为每个维度做数据分布直方图并根据数据的频率计算异常程度。

基于预测的方法通常为各个维度分别构建预测模型，然后通过计算实际值与预测值之间的误差来识别异常，ARIMA[7]模型及其变体是常用的有效方法。这类方法可以很好地捕捉 KPI 序列的时间依赖性，但对噪声非常敏感，在噪声严重时可能会增加很多假阳性结果。

基于邻近度的方法通过定义距离度量并计算数据之间的距离来检测离群点，与单维 KPI 序列类似，主要包含 *K* 近邻、LOF[10]、ABOD[11]等算法。但对于多维 KPI 序列而言，此类方法计算复杂度过高，难以应用于大规模数据。

3.2.2 基于机器学习的方法

监督模型严重依赖准确的类别标签的特点使其难以应用在实际的运维环境中。无监督机器学习模型通过学习 KPI 数据的固有特征来识别异常，代表性的算法有 PCA、SVM 以及聚类算法（包括 DBSCAN、GMM、*K*-means 等）。陈兴蜀等[27]基于 ARIMA 预测算法得到多维特征偏移向量，然后基于 SVM 算法进行分类判别，实现了多维时间序列的在线检测。Hu 等[28]基于降维算法（PCA、SVD）将多维时间序列转化为一维序列，并利用滑动窗口将降维后的序列划分为多个子序列，然后对各个子序列进行特征提取，最后基于 one-class SVM 算法进行在线异常检测。

近几年，针对多维 KPI 序列的无监督深度模型

得到了广泛的研究。为了检测航天器异常，Hundman 等[29]将长短期记忆（long short-term memory，LSTM）网络应用于多元 KPI 序列预测，并使用平滑化后的预测误差确定异常。类似于自然语言处理中的 seq2seq 模型，Malhotra 等[30]提出了一种基于 LSTM 的编码器—解码器模型，目的是重构 KPI 序列的正常点的特征，并基于重构误差进行多维 KPI 异常检测。不同于以上两项基于误差的异常检测方法，Park 等[31]将 LSTM 和 VAE 模型组合起来，具体是将 VAE 中的前馈神经网络替换成 LSTM，并采用重建概率来判定异常。由于 VAE 学习的是隐变量 z 的分布，因此模型对噪声有较好的鲁棒性，且避免了过拟合。之后，Su 等[32]提出了 OmniAnomaly 机制，将 VAE 和 GRU（gate recurrent unit）结合，很好地捕获了多维 KPI 序列的时序性和指标依赖关系。具体地，OmniAnomaly 利用 GRU 捕获多维 KPI 序列复杂的时间依赖性，并基于 VAE 将观测值（即 x 空间）映射到随机变量（即 z 空间），另外提出了随机变量连接技术来建模随机变量之间的依赖关系。此外，受计算机视觉领域的启发，Zhang 等[33]提出了一种基于卷积 LSTM（conv-LSTM）的自编码器（auto-encoder）架构 MSCRED，与之前工作不同，MSCRED 没有直接将多元 KPI 序列输入模型，而是通过计算多维 KPI 序列指标间的协方差构造了分辨率特征矩阵，再将特征矩阵作为 AutoEncoder 结构的输入。

4　KPI 检测部署应用存在的问题和未来展望

针对 KPI 异常检测已经产生了很多的研究成果，但由于 KPI 的复杂性、多样性，在实际部署应用中仍存在很多难题，例如周期多样性、概念漂移、海量 KPI 序列的异常检测以及结果的可解释性等。

4.1　周期多样性

由于工作日、休息日、节假日或者商业活动对用户行为的影响，KPI 可能在不同的时期内差异很大。例如实际运维中很常见的“节假日效应”问题是指受节假日（例如国庆节、春节、中秋节）的影响，KPI 可能呈现与平时差异很大的模式，但这些模式却是正常的。因此，如何设计通用化的模型适应 KPI 的多种周期性是业务上面临的重要问题之一。传统方法（例如 Holt-Winters 和 ARIMA 异常检测器）需要将 KPI 的周期长度作为其输入参数，通常需要为模型手动配置为日、周、月等周期。这类检测器的性能受人为因素影响较大，且某些情况下，固定的周期长度难以表征业务 KPI 复杂的周期性。此外，基于变分自编码器（variational auto-encoder，VAE）的无监督异常检测算法 Donut 在平稳的季节性 KPI 上表现出良好的检测性能。然而，由于 Donut 将滑动窗口内的数据点作为模型输入，并没有时间或日期信息，因此也很难从训练数据中提取多种周期性的特征。

针对此问题，Zhao 等[34]提出了 Period 机制，主要分为离线周期性异常检测和在线适应性异常检测两个部分。在离线模块，将历史 KPI 按天切分成多个子序列，基于聚类的方法将这些子序列划分到不同的聚类簇并进行序列拼接，然后为各个聚类簇训练相应的模型。在在线模块，新的 KPI 序列根据具体日期划分到各个簇中，并基于相应的模型进行异常检测。另外在聚类部分，Period 基于 SBD（shape-based distance）算法进行距离度量，并考虑了周期漂移的影响。总体而言，Period 机制思路简单，能有效解决复杂周期性对异常检测效果的影响。

当然，Period 也存在一些问题，例如，Period 假设 KPI 的基本季节性是一天，因此将整个 KPI 直接按天切分成子序列，无法处理任意周期性（例如每年）的 KPI。此外，Period 也不适用于所有类型的 KPI，例如有趋势性和周期性的 KPI 需进行去趋势化预处理。还有针对具有多种周期性的多维 KPI 序列的异常检测还有待进一步研究。

4.2 概念漂移

运维领域的概念漂移（concept drift）是指KPI序列受活动发布、业务变更、资源调度等一些特殊事件的影响产生的水平漂移，例如资源占用率的突然上升或下降，如图 4 所示。概念漂移可以分为预期的概念漂移和意外的概念漂移。意外的概念漂移被视为一种异常，可能会带来额外的经济损失，此时服务提供商可能会采用切换流量（switch traffic）、回滚版本（rollback version）、重启实例（restart instance）等操作将服务恢复到发生漂移之前的状态来及时止损。预期的概念漂移不属于异常，但由于数据分布的改变，导致原有的异常检测器在一段时间内会产生大量的报警。因此，研究概念漂移旨在避免发生有预期的漂移后异常检测器长时间的精度损失。

针对 KPI 的概念漂移问题，Ma 等[35]提出了StepWise 机制，StepWise 分为 3 部分：首先基于奇异频谱变换（singular spectrum transform，SST）算法和极值理论（extreme value theory，EVT）检测是否发生概念偏移；然后区分此偏移是否符合预期，针对不符合预期的概念偏移，运维工程师要快速处理该异常；最后对于符合预期的漂移，设计了具有鲁棒性的线性模型（robust linear model，RLM）拟合新的数据分布，从而迅速适应原有异常检测算法和参数。总体而言，StepWise是解决 KPI 概念漂移的一个有效方法，但在实际运用过程中，对于概念漂移检测和新数据拟合方面，仍需针对特定业务精心设计，很难找到某个方法适应所有的场景。

4.3 海量 KPI 序列的异常检测

海量 KPI 序列的异常检测一直是运维中的一个重要难题。大多数的异常检测算法（例如有监督的 EDAGS[12]、Opprentice[13]、无监督的 Donut[20]等）都假定需要为每个 KPI 训练和维护单独的模型。面对海量的 KPI，由于模型选择、参数调整、

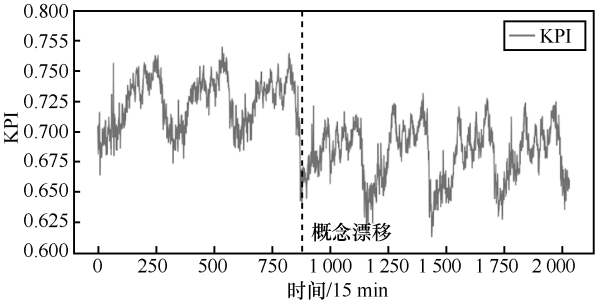


图 4 概念漂移示意图

模型训练等过程带来的巨大开销，为所有的 KPI 分别训练异常检测器是不可行的。

Li 等[36]提出了 ROCKA 机制，基于时间序列聚类的方法解决这一挑战。具体地，ROCKA 提取每个 KPI 的基础形状（称为基线），采用 SBD 算法作为距离度量，并基于 DBSCAN 算法根据基线形状的相似性将 KPI 聚类成几个簇，然后为每个簇训练异常检测模型。对于新的 KPI，根据与各个簇之间的相似度为其分配适当的模型。基于 ROCKA，Bu 等[37]提出了 ADS 机制，在聚类 KPI 后，利用聚类簇的中心和新的 KPI 序列进行半监督学习，从而快速准确地为新的 KPI 分配模型。

ROCKA 和 ADS 提出了解决海量 KPI 异常检测的一个新方向。但现实中很多 KPI 数据相似性不高，很难被聚类，而且聚类带来的异常检测精度下降的影响也是不容忽视的。另外，多维 KPI 的时间高维度、指标高维度、指标存在关联等特点为海量异常检测带来了更大的难度，这也是未来研究的一个重要方向。

4.4 异常结果的可解释性

在多维 KPI 异常检测领域，一个重要的问题是如何对实体级别的异常进行合理解释，即为什么观测值被认定为异常。有效的结果解释有助于对异常原因的分析并加速后续的故障排除。然而，目前的异常检测大多基于无监督深度算法，属于黑盒模型，可解释性较差。无监督模型大多数基于重构误差或重构概率来判断异常，一个直观的

想法是寻找误差最大或重建概率最低的几个 KPI 维度作为实体级别异常的原因[31-32]。此外，结合时序异常检测和图两个领域，利用图挖掘的方式找出时间序列的内在联系也是分析异常原因的一个方向[38]，有研究人员提出构建有效的故障传播图挖掘特征之间的关联[39-41]。然而，利用故障传播图挖掘准确的因果关系也存在着很大的挑战。由于故障发生往往遵循涟漪效应，会从一个维度故障传播到其他维度或者整个系统状态，在有标签或存在专家经验的多元时间序列中容易出现误标签，即一开始的故障仍认为是正常，直到系统发生异常才被检测到。另外，涟漪效应的传播和发生依赖于初始根因，当初始根因不同，则其传播关系和关联性会发生一定的变化，造成构建故障传播图得到的因果关系不稳定。针对此问题，未来可以尝试利用不同时间序列的趋势变化以及时间先后特征构造动态的依赖关系。

5　结束语

在当今社会不断智能化的进程中，智能运维势在必行。KPI 异常检测是智能运维中的一项基础且重要的任务，异常定位、根因分析等运维问题都依赖于异常检测的结果。本文从单维 KPI 和多维 KPI 两个层面，回顾了面向 KPI 异常检测的研究进展和成果，并将现有方法分为基于规则的方法、基于统计的方法、基于预测的方法、基于相似度的方法以及基于机器学习的方法。此外，本文总结分析了异常检测模型在部署应用中面临的几个问题并提出可能的优化方向。未来 KPI 异常检测应重点考虑实际应用中的难点问题，另外可以考虑其他领域的交叉和借鉴，例如可以尝试结合自然语言处理中的预训练和微调的方法应对海量 KPI 异常检测的时间和指标高维度问题，以及基于因果推断的方法挖掘多维 KPI 间的动态依赖关系等。

参考文献：

[1] 李青. 基于大数据分析的云资源池告警信息关联方案[J]. 电信科学, 2020, 36(10): 159-171.
LI Q. Alarm information association scheme of cloud resource pool based on big data analysis[J]. Telecommunications Science, 2020, 36(10): 159-171.

[2] 裴丹, 张圣林, 裴昶华. 基于机器学习的智能运维[J]. 中国计算机学会通讯, 2017, 13(12): 67-73.
PEI D, ZHANG S L, PEI C H. Intelligent operation and maintenance based on machine learning[J]. Communications of CCF, 2017, 13(12): 67-73.

[3] CHEN W X, XU H W, LI Z Y, et al. Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE[C]//Proceedings of IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 1891-1899.

[4] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection[J]. ACM Computing Surveys, 2009, 41(3): 1-58.

[5] Amazon Web Services. Amazon cloudwatch alarm[EB]. 2018.

[6] SIFFER A, FOUQUE P A, TERMIER A, et al. Anomaly detection in streams with extreme value theory[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2017: 1067-1075.

[7] ZHANG Y, GE Z H, GREENBERG A, et al. Network anomography[C]//Proceedings of the 5th ACM SIGCOMM conference on Internet measurement - IMC '05. New York: ACM Press, 2005: 30-30.

[8] YAN H, FLAVEL A, GE Z H, et al. Argus: End-to-end service anomaly detection and localization from an ISP's point of view[C]//Proceedings of 2012 IEEE INFOCOM. Piscataway: IEEE Press, 2012: 2756-2760.

[9] TAYLOR S J, LETHAM B. Forecasting at scale[J]. The American Statistician, 2018, 72(1): 37-45.

[10] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[C]//Proceedings of the 2000 ACM SIGMOD International Conference on Management of data - SIGMOD '00. New York: ACM Press, 2000: 93-104.

[11] KRIEGEL H P, S HUBERT M, ZIMEK A. Angle-based outlier detection in high-dimensional data[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08. New York: ACM Press, 2008: 444-452.

[12] LAPTEV N, AMIZADEH S, FLINT I. Generic and scalable framework for automated time-series anomaly detection[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2015: 1939-1947.

[13] LIU D P, ZHAO Y J, XU H W, et al. Opprentice: towards practical and automatic anomaly detection through machine learning[C]//Proceedings of the 2015 Internet Measurement Conference. New York: ACM Press, 2015: 211-224.
[14] AMER M, GOLDSTEIN M, ABDENNADHER S. Enhancing one-class support vector machines for unsupervised anomaly detection[C]//Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description - ODD '13. New York: ACM Press, 2013: 8-15.
[15] YANG X W, LATECKI L J, POKRAJAC D. Outlier detection with globally optimal exemplar-based GMM[C]//Proceedings of the 2009 SIAM International Conference on Data Mining. Philadelphia: Society for Industrial and Applied Mathematics, 2009: 145-154.
[16] MUNZ G, LI S, CARLE G. Traffic anomaly detection using k-means clustering[C]//Proceedings of GI/ITG Workshop MMBnet. [S.l.: s.n.], 2007: 13-14.
[17] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special Lecture on IE, 2015, 2(1): 1-18.
[18] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. [S.l.: s.n.], 2014: 2672-2680.
[19] OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[J]. arXiv preprint arXiv:1601.06759, 2016.
[20] XU H W, FENG Y, CHEN J, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications[C]//Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18. New York: ACM Press, 2018: 187-196.
[21] LI Z Y, CHEN W X, PEI D. Robust and unsupervised KPI anomaly detection based on conditional variational autoencoder[C]//Proceedings of 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC). Piscataway: IEEE Press 2018: 1-9.
[22] CHEN W X, XU H W, LI Z Y, et al. Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE[C]//Proceedings of IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. [S.l.:s.n.] 2019: 1891-1899.
[23] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
[24] 胡珉, 白雪, 徐伟, 等. 多维时间序列异常检测算法综述[J]. 计算机应用, 2020, 40(6): 1553-1564.
HU M, BAI X, XU W, et al. Review of anomaly detection algorithms for multidimensional time series[J]. Journal of Computer Applications, 2020, 40(6): 1553-1564.
[25] Arundo Analytics. Anomaly detection toolkit [EB]. 2019.
[26] GOLDSTEIN M, DENGEL A. Histogram-based outlier score (hbos): a fast unsupervised anomaly detection algorithm[J]. KI-2012: Poster and Demo Track, 2012: 59-63.
[27] 陈兴蜀, 江天宇, 曾雪梅, 等. 基于多维时间序列分析的网络异常检测[J]. 工程科学与技术, 2017, 49(1): 144-150.
CHEN X S, JIANG T Y, ZENG X M, et al. Network anomaly detector based on multiple time series analysis[J]. Advanced Engineering Sciences, 2017, 49(1): 144-150.
[28] HU M, JI Z W, YAN K, et al. Detecting anomalies in time series data via a meta-feature based approach[J]. IEEE Access, 2018(6): 27760-27776.
[29] HUNDMAN K, CONSTANTINOU V, LAPORTE C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 387-395.
[30] MALHOTRA P, RAMAKRISHNAN A, ANAND G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[J]. arXiv preprint arXiv:1607.00148, 2016.
[31] PARK D, HOSHI Y, KEMP C C. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1544-1551.
[32] SU Y, ZHAO Y J, NIU C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 2828-2837.
[33] ZHANG C X, SONG D J, CHEN Y C, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019(33): 1409-1416.
[34] ZHAO N W, ZHU J, WANG Y, et al. Automatic and generic periodicity adaptation for KPI anomaly detection[J]. IEEE Transactions on Network and Service Management, 2019, 16(3): 1170-1183.
[35] MA M H, ZHANG S L, PEI D, et al. Robust and rapid adaption for concept drift in software system anomaly detection[C]//Proceedings of 2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE). Piscataway: IEEE Press, 2018: 13-24.
[36] LI Z H, ZHAO Y J, LIU R, et al. Robust and rapid clustering of KPIs for large-scale anomaly detection[C]//Proceedings of 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Piscataway: IEEE Press, 2018: 1-10.
[37] BU J H, LIU Y, ZHANG S L, et al. Rapid deployment of anomaly detection models for large number of emerging KPI streams[C]//Proceedings of 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC). Piscataway: IEEE Press, 2018: 1-8.
[38] CHENG Z Q, YANG Y, WANG W, et al. Time2Graph: revisiting time series modeling with dynamic shapelets[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 3617-3624.
[39] SHEN Z Y, CUI P, LIU J S, et al. Stable learning via differentiated variable decorrelation[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 2185-2193.
[40] HUANG Q B, WEI J L, CAI Y, et al. Aligned dual channel graph convolutional network for visual question answering[C]//Proceedings of the 58th Annual Meeting of the Associa-

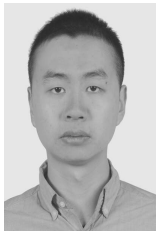
tion for Computational Linguistics. Online. Stroudsburg: Association for Computational Linguistics, 2020: 7166-7176.

[41] CHANG H, RONG Y, XU T Y, et al. A restricted black-box adversarial framework towards attacking graph embedding models[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 3389-3396.

[作者简介]

王速（1997− ），男，北京邮电大学硕士生，主要研究方向为软件定义网络、网络人工智能。

卢华（1976− ），男，广东省新一代通信与网络创新研究院网络技术创新中心主任，主要研究方向为核心网、软件定义网络、P4 可编程、虚拟化等。



汪硕（1991− ），男，博士，北京邮电大学讲师，主要研究方向为软件定义网络、数据中心网络、确定性网络、网络人工智能等。

蔡磊（1981− ），男，广东省新一代通信与网络创新研究院研究员，主要研究方向为可编程交换技术、网络流量控制技术等。

黄韬（1980− ），男，博士，北京邮电大学教授，主要研究方向为路由与交换、软件定义网络、网络试验设施等。