(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118643455 A
(43)申请公布日 2024.09.13

(21)申请号 202411124322.7

(22)申请日 2024.08.16

(71)申请人 安徽思高智能科技有限公司
地址 230088 安徽省合肥市高新区望江西路900号中安创谷科技园A1栋408

(72)发明人 张竞超 李扬 张泽锟 王健

(74)专利代理机构 武汉知产时代知识产权代理有限公司 42238
专利代理师 徐欢

(51)Int.Cl.
G06F 18/2433(2023.01)
G06F 18/25(2023.01)
G06F 18/213(2023.01)
G06F 11/30(2006.01)
G06F 18/241(2023.01)
G06N 3/049(2023.01)
G06N 3/0464(2023.01)
G06N 3/0895(2023.01)

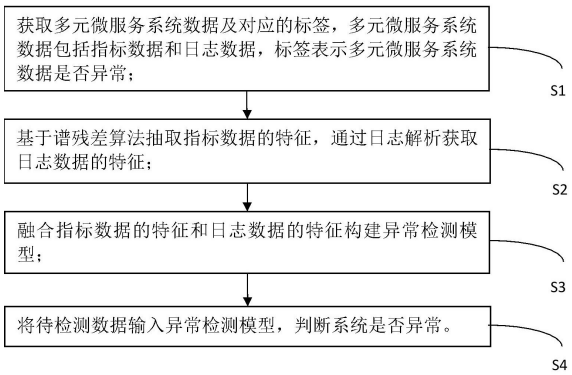
权利要求书3页 说明书7页 附图2页

(54)发明名称

一种微服务系统异常检测方法、存储介质、电子设备

(57)摘要

本发明公开了一种微服务系统异常检测方法、存储介质、电子设备，属于微服务系统检测领域，包括获取指标数据和日志数据的多元微服务系统数据及对应的标签；构建异常检测模型，使用指标数据的特征和日志数据的特征和相应的标签训练异常检测模型，异常检测模型包括：输入层、TCN模块、输出层，输入层将指标数据的特征和日志数据的特征拼接组成新特征，TCN模块捕获新特征的时间序列中的依赖关系后，输入输出层生成最终的异常分数；将待检测数据输入训练后的异常检测模型，得到系统异常情况。本发明将微服务系统指标数据和日数据融合，丰富了信息视角，为复杂微服务环境下的全面异常监测提供了有力解决方案。



CN 118643455 A

1.一种微服务系统异常检测方法,其特征在于,包括以下步骤:

S1、获取多元微服务系统数据及对应的标签,多元微服务系统数据包括指标数据和日志数据,标签表示多元微服务系统数据是否异常;

S2、通过谱残差算法抽取指标数据的特征,通过日志解析获取日志数据的特征;

S3、融合指标数据的特征和日志数据的特征构建异常检测模型;

具体为:

通过多层感知机将指标数据的特征和日志数据的特征映射到同一维度,得到指标数据隐层特征和日志数据隐层特征;

将指标数据隐层特征和日志数据隐层特征输入以时间卷积网络TCN为主干的网络中,得到指标数据高维隐层特征和日志数据高维隐层特征;

结合自动编码器将指标数据高维隐层特征和日志数据高维隐层特征压缩到低维嵌入空间,得到低维的指标数据的特征和低维的日志数据的特征;

根据低维的指标数据的特征和低维的日志数据的特征计算系统异常得分,基于系统异常得分判断系统是否异常;

S4、将待检测数据输入异常检测模型,判断系统是否异常。

2.根据权利要求1所述的一种微服务系统异常检测方法,其特征在于,步骤S1具体为:

S11、在系统正常运行阶段,收集指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$,$\mathbf{M}_{\mathrm{normal}}$和$\mathbf{L}_{\mathrm{normal}}$数据标签设为0,将指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$、指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$对应的标签设置为训练集$\mathbf{D}_{\mathrm{train}}$;

S12、设置故障注入参数,根据故障注入参数向系统随机注入故障,收集指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$,将故障注入时间段收集到的指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$的标签设为1,其他时间段收集到的指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$的标签设为0,将指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$、指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$对应的标签设置为测试集$\mathbf{D}_{\mathrm{test}}$;

S13、设置采样时间间隔,从训练集$\mathbf{D}_{\mathrm{train}}$和测试集$\mathbf{D}_{\mathrm{test}}$中采样训练集$\mathbf{D}_{\mathrm{train}}$的训练样本为$\mathbf{x}_{\mathrm{in}}$和测试集$\mathbf{D}_{\mathrm{test}}$的测试样本为$\mathbf{x}_{\mathrm{test}}$。

3.根据权利要求2所述的一种微服务系统异常检测方法,其特征在于,步骤S2中,通过谱残差算法抽取指标数据的特征,具体为:

基于所述指标数据,使用傅里叶变换将所述指标数据从空间域转换到频率域,得到指标数据的频谱信息;

计算指标数据的频谱的幅度谱;

对所述幅度谱进行谱残差变换,得到谱残差变换后的幅度谱;

对谱残差变换后的幅度谱进行傅里叶逆变换,将复数形式的频域结果转换回空间域,得到指标数据的显著图;

将指标数据的显著图归一化后纵向拼接,得到指标数据的特征。

4.根据权利要求3所述的一种微服务系统异常检测方法,其特征在于,

步骤S2中,通过日志解析抽取日志数据的特征,具体为:

使用Drain日志解析工具对所述日志数据进行聚类,每个聚类表示一个日志模板,得到日志模板集合$LTS=\{LT_1,LT_2,...,LT_N\}$,其中,N表示聚类簇集数;

预设一个时间间隔,统计预设时间间隔的不同时间段各个日志模板出现的频率,并转换为日志模板频次矩阵,将日志模板频次矩阵归一化,得到日志数据的特征X^L。

5.根据权利要求4所述的一种微服务系统异常检测方法,其特征在于,根据低维的指标数据的特征和低维的日志数据的特征计算系统异常得分,基于系统异常得分判断系统是否异常具体为:

通过以下方式计算系统异常得分:

$$s=\left(x'_{test}-\mu_{in}\right)^T\Sigma_{in}^{-1}\left(x'_{test}-\mu_{in}\right)$$

其中,s表示系统异常得分,μ_{in}和Σ_{in}^{-1}表示x'_{in}的均值和协方差矩阵,x'_{in}表示训练数据集D_{train}中的训练样本经过异常检测模型得到的低维特征,x'_{test}表示测试数据集D_{test}中的测试样本经过异常检测模型得到的低维特征,系统异常得分表示x'_{test}和x'_{in}之间的马哈拉比诺斯距离;

如果s大于预设的阈值Threshold,则测试样本x_{test}是异常值;相反,测试样本x_{test}是正常值。

6.根据权利要求5所述的一种微服务系统异常检测方法,其特征在于,以时间卷积网络TCN为主干的网络的损失函数定义为:

$$\mathcal{L}_{contrastive}=-\frac{1}{n}\sum_{i=1}^{n}\frac{sim\left(H_i^M,H_i^L\right)}{\sum_k sim\left(H_i^M,H_k^L\right)}$$

其中,$\mathcal{L}_{contrastive}$表示对比学习损失,$sim\left(H_i^M,H_i^L\right)$表示$H_i^M$和$H_i^L$之间的余弦相似度,$H_i^M$表示第i个指标数据隐层特征,$H_i^L$表示第i个日志数据隐层特征,$H_k^L$表示第k个日志数据隐层特征,n表示指标数据隐层特征个数。

7.根据权利要求6所述的一种微服务系统异常检测方法,其特征在于,自动编码器的损失函数表示为:

$$\mathcal{L}_{ae}=\|h-g(f(h))\|_2^2$$

其中,$\mathcal{L}_{ae}$表示自动编码器的损失函数,$\|\cdot\|_2^2$表示欧几里得范数的平方,$f(h)$表示对h编码的结果,$g(f(h))$表示对$f(h)$解码的结果,h表示高维隐层特征。

8.根据权利要求7所述的一种微服务系统异常检测方法,其特征在于,

通过最小化异常检测模型的整体损失函数优化整个异常检测模型,异常检测模型的整体损失函数表示为:

$$\mathcal{L}=\lambda\times\mathcal{L}_{contrastive}+(1-\lambda)\times\mathcal{L}_{ae}$$

其中,$\mathcal{L}$表示异常检测模型的整体损失函数,λ表示权重系数。

9.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有数据处

理程序,所述数据处理程序被处理器执行时实现如权利要求1至8中任一项所述的方法。

10.一种电子设备,其特征在于,包括处理器和存储器,所述处理器与所述存储器相互连接,其中,所述存储器用于存储计算机程序,所述计算机程序包括计算机可读指令,所述处理器被配置用于调用所述计算机可读指令,执行如权利要求1至8中任一项所述的方法。

一种微服务系统异常检测方法、存储介质、电子设备

技术领域

[0001] 本发明微服务系统检测领域,具体涉及一种微服务系统异常检测方法、存储介质、电子设备。

背景技术

[0002] 在当今数字化时代,微服务架构作为一种高度灵活、可扩展的软件设计模式,已经成为构建大规模应用和服务的主流选择。然而,随着微服务系统的规模和复杂性的增加,监测这些服务的运行状况变得愈发关键。微服务系统的复杂性主要表现在多样性的服务组合、分布式环境下的通信和协作,以及异构数据的处理。这些特征使得异常监测变得更为困难,传统的监测方法难以胜任。微服务系统的指标数据通常涉及到多个维度和模态,而融合这些异构的多模态数据具有重要意义。如何整合这些数据,建模时序关系,以及有效处理高维度和噪声是一项重要的挑战。

发明内容

[0003] 本发明的目的在于:为了解决微服务系统异构的多模态数据难以融合导致微服务系统监测困难的问题,本发明提出一种微服务系统异常检测方法,为微服务系统提供一种全面而高效的监测和异常检测解决方案,以促进系统的稳定性和可靠性,包括以下步骤:

S1、获取多元微服务系统数据及对应的标签,多元微服务系统数据包括指标数据和日志数据,标签表示多元微服务系统数据是否异常;

S2、通过谱残差算法抽取指标数据的特征,通过日志解析获取日志数据的特征;

S3、融合指标数据的特征和日志数据的特征构建异常检测模型;

具体为:

通过多层感知机将指标数据的特征和日志数据的特征映射到同一维度,得到指标数据隐层特征和日志数据隐层特征;

将指标数据隐层特征和日志数据隐层特征输入以时间卷积网络TCN为主干的网络中,得到指标数据高维隐层特征和日志数据高维隐层特征;

结合自动编码器将指标数据高维隐层特征和日志数据高维隐层特征压缩到低维嵌入空间,得到低维的指标数据的特征和低维的日志数据的特征;

根据低维的指标数据的特征和低维的日志数据的特征计算系统异常得分,基于系统异常得分判断系统是否异常;

S4、将待检测数据输入异常检测模型,判断系统是否异常。

[0004] 进一步地,步骤S1具体为:

S11、在系统正常运行阶段,收集指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$,$\mathbf{M}_{\mathrm{normal}}$和$\mathbf{L}_{\mathrm{normal}}$数据标签设为0,将指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$、指标数据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$对应的标签设置为训练集$\mathbf{D}_{\mathrm{train}}$;

S12、设置故障注入参数，根据故障注入参数向系统随机注入故障，收集指标数据$\mathbf{M}_{\text{testl}}$和日志数据$\mathbf{L}_{\text{test}}$，将故障注入时间段收集到的指标数据$\mathbf{M}_{\text{testl}}$和日志数据$\mathbf{L}_{\text{test}}$的标签设为1，其他时间段收集到的指标数据$\mathbf{M}_{\text{testl}}$和日志数据$\mathbf{L}_{\text{test}}$的标签设为0，将指标数据$\mathbf{M}_{\text{testl}}$和日志数据$\mathbf{L}_{\text{test}}$、指标数据$\mathbf{M}_{\text{testl}}$和日志数据$\mathbf{L}_{\text{test}}$对应的标签设置为测试集$\mathbf{D}_{\text{test}}$；

S13、设置采样时间间隔，从训练集$\mathbf{D}_{\text{train}}$和测试集$\mathbf{D}_{\text{test}}$中采样训练集$\mathbf{D}_{\text{train}}$的训练样本为$\mathbf{x}_{\text{in}}$和测试集$\mathbf{D}_{\text{test}}$的测试样本为$\mathbf{x}_{\text{test}}$。

[0005] 进一步地，步骤S2中，通过谱残差算法抽取指标数据的特征，具体为：

基于所述指标数据，使用傅里叶变换将所述指标数据从空间域转换到频率域，得到指标数据的频谱信息；

计算指标数据的频谱的幅度谱；

对所述幅度谱进行谱残差变换，得到谱残差变换后的幅度谱；

对谱残差变换后的幅度谱进行傅里叶逆变换，将复数形式的频域结果转换回空间域，得到指标数据的显著图；

将指标数据的显著图归一化后纵向拼接，得到指标数据的特征。

[0006] 进一步地，步骤S2中，通过日志解析抽取日志数据的特征，具体为：

使用Drain日志解析工具对所述日志数据进行聚类，每个聚类表示一个日志模板，得到日志模板集合$\mathbf{LTS}=\{\mathbf{LT}_1,\mathbf{LT}_2,...,\mathbf{LT}_N\}$，其中，N表示聚类簇集数；

预设一个时间间隔，统计预设时间间隔的不同时间段各个日志模板出现的频率，并转换为日志模板频次矩阵，将日志模板频次矩阵归一化，得到日志数据的特征$\mathbf{X}^{\text{L}}$。

[0007] 进一步地，根据低维的指标数据的特征和低维的日志数据的特征计算系统异常得分，基于系统异常得分判断系统是否异常具体为：

通过以下方式计算系统异常得分：

$$\mathbf{s}=\left(\mathbf{x}'_{\text{test}}-\mu_{\text{in}}\right)^{\text{T}}\Sigma_{\text{in}}^{-1}\left(\mathbf{x}'_{\text{test}}-\mu_{\text{in}}\right)$$

[0008] 其中，s表示系统异常得分，μ_{in}和Σ_{in}^{-1}表示$\mathbf{x}'_{\text{in}}$的均值和协方差矩阵，$\mathbf{x}'_{\text{in}}$表示训练数据集$\mathbf{D}_{\text{train}}$中的训练样本经过异常检测模型得到的低维特征，$\mathbf{x}'_{\text{test}}$表示测试数据集$\mathbf{D}_{\text{test}}$中的测试样本经过异常检测模型得到的低维特征，系统异常得分表示$\mathbf{x}'_{\text{test}}$和$\mathbf{x}'_{\text{in}}$之间的马哈拉比诺斯距离；

如果s大于预设的阈值Threshold，则测试样本$\mathbf{x}_{\text{test}}$是异常值；相反，测试样本$\mathbf{x}_{\text{test}}$是正常值。

[0009] 进一步地，以时间卷积网络TCN为主干的网络的损失函数定义为：

$$\mathcal{L}_{\text{contrastive}}=-\frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}\frac{\text{sim}\left(\mathbf{H}_{\text{i}}^{\text{M}},\mathbf{H}_{\text{i}}^{\text{L}}\right)}{\sum_{\text{k}}\text{sim}\left(\mathbf{H}_{\text{i}}^{\text{M}},\mathbf{H}_{\text{k}}^{\text{L}}\right)}$$

[0010] 其中，$\mathcal{L}_{\text{contrastive}}$表示对比学习损失，$\text{sim}\left(\mathbf{H}_{\text{i}}^{\text{M}},\mathbf{H}_{\text{i}}^{\text{L}}\right)$表示$\mathbf{H}_{\text{i}}^{\text{M}}$和$\mathbf{H}_{\text{i}}^{\text{L}}$之间的余弦相

似度，H_i^M表示第i个指标数据隐层特征，H_i^L表示第i个日志数据隐层特征，H_k^L表示第k个日志数据隐层特征，n表示指标数据隐层特征个数。

[0011] 进一步地，自动编码器的损失函数表示为：

$$\mathcal{L}_{ae} = \|h - g(f(h))\|_2^2$$

[0012] 其中，$\mathcal{L}_{ae}$表示自动编码器的损失函数，$\|\cdot\|_2^2$表示欧几里得范数的平方，$f(h)$表示对h编码的结果，$g(f(h))$表示对$f(h)$解码的结果，h表示高维隐层特征。

[0013] 进一步地，通过最小化异常检测模型的整体损失函数优化整个异常检测模型，异常检测模型的整体损失函数表示为：

$$\mathcal{L} = \lambda \times \mathcal{L}_{contrastive} + (1-\lambda) \times \mathcal{L}_{ae}$$

[0014] 其中，$\mathcal{L}$表示异常检测模型的整体损失函数，λ表示权重系数。

[0015] 本发明还提出一种计算机可读存储介质，计算机可读存储介质上存储有数据处理程序，数据处理程序被处理器执行时实现上述的微服务系统异常检测方法。

[0016] 本发明还提出一种电子设备，包括处理器和存储器，所述处理器与所述存储器相互连接，其中，所述存储器用于存储计算机程序，所述计算机程序包括计算机可读指令，所述处理器被配置用于调用所述计算机可读指令，执行上述的微服务系统异常检测方法。

[0017] 本发明提供的技术方案带来的有益效果是：

本发明通过融合业务日志和指标，丰富了异常检测的信息视角，为异常的解释和定位提供了更多线索，提高了系统监测的全面性；通过傅里叶变换和谱残差变换对微服务指标数据进行特征抽取，不仅在时域上捕捉了复杂模式，还在频域上进行全面分析，本发明可以融合异构的多模态数据，提高了异常检测的准确性和全面性。

附图说明

[0018] 图1是本发明实施例一种微服务系统异常检测方法的流程图；

图2是本发明实施例微服务系统CPU利用率原始指标数据和微服务指标数据的特征；

图3是本发明实施例一示例性实施例中的一种电子设备的框图。

具体实施方式

[0019] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地描述。

[0020] 本发明实施例一种微服务系统异常检测方法的流程图如图1所示，包括以下步骤：

S1、获取多元微服务系统数据及对应的标签，多元微服务系统数据包括微服务系统指标数据和日志数据，标签表示多元微服务系统数据是否异常。

[0021] 进一步的实施例中，通过故障注入手段获取具有异常标签的微服务系统指标数据和日志数据，具体为：

S11、在系统正常运行阶段，收集指标数据M_{normal}和日志数据L_{normal}，M_{normal}和L_{normal}数据标签设为0，0表示正常，将指标数据M_{normal}和日志数据L_{normal}及指标数

据$\mathbf{M}_{\mathrm{normal}}$和日志数据$\mathbf{L}_{\mathrm{normal}}$对应的标签设置为训练集$\mathbf{D}_{\mathrm{train}}$。具体地,通过Prometheus和Node_exporter工具收集指标数据$\mathbf{M}_{\mathrm{normal}}$,通过FileBeat收集日志数据$\mathbf{L}_{\mathrm{normal}}$。

[0022] S12、设置故障注入参数,故障注入参数具体包括:主机CPU利用率、主机内存利用率、主机网络丢包率、服务响应时间延迟、容器CPU利用率、容器内存利用率、容器网络丢包率等。

[0023] 根据故障注入参数向系统随机注入故障,收集指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$,具体地,通过Prometheus、Node_exporter和FileBeat工具收集指标和日志数据;将故障注入时间段收集到的指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$的标签设为1,1表示异常,其他时间段收集到的指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$的标签设为0,0表示正常,将指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$及指标数据$\mathbf{M}_{\mathrm{testl}}$和日志数据$\mathbf{L}_{\mathrm{test}}$对应的标签设置为测试集$\mathbf{D}_{\mathrm{test}}$。

[0024] S13、设置采样时间间隔,具体地,设置时间间隔为5s,从训练集$\mathbf{D}_{\mathrm{train}}$和测试集$\mathbf{D}_{\mathrm{test}}$中采样训练集$\mathbf{D}_{\mathrm{train}}$的训练样本为$\mathbf{x}_{\mathrm{in}}$和测试集$\mathbf{D}_{\mathrm{test}}$的测试样本为$\mathbf{x}_{\mathrm{test}}$。

[0025] S2、基于谱残差算法抽取指标数据的特征,通过日志解析获取日志数据的特征。

[0026] 进一步的实施例中,通过谱残差算法抽取微服务指标数据的特征,具体为:

(1)基于上面获得的指标数据,使用傅里叶变换将指标数据从空间域转换到频率域,得到指标数据的频谱信息。

[0027] $$\mathrm{A}(\mathrm{f}) = \mathrm{Amplitude}(\mathfrak{F}(\mathrm{x}_{\mathrm{m}}))$$

[0028] $$\mathrm{P}(\mathrm{f}) = \mathrm{Phrase}\,(\mathfrak{F}(\mathrm{x}_{\mathrm{m}}))$$

[0029] $$\mathrm{L}(\mathrm{f}) = \log\,(\mathrm{A}(\mathrm{f}))$$

[0030] 其中,$\mathfrak{F}$表示傅立叶变换,x_{m}表示一维时序数据,即指标数据,$\mathrm{A}(\mathrm{f})$表示频域中的振幅,$\mathrm{P}(\mathrm{f})$表示频域中的相位,Amplitude(.)和Phrase(.)分别表示取频域中的振幅和相位,$\mathrm{L}(\mathrm{f})$表示对振幅频$\mathrm{A}(\mathrm{f})$做log计算的结果。

[0031] (2)计算指标数据的频谱的幅度谱,即频谱的模。

[0032] $$\mathrm{AL}(\mathrm{f}) = \mathrm{h}_{\mathrm{q}}(\mathrm{f}) \cdot \mathrm{L}(\mathrm{f})$$

[0033] 其中,$\mathrm{AL}(\mathrm{f})$表示对$\mathrm{L}(\mathrm{f})$进行均值滤波后的结果,$\mathrm{h}_{\mathrm{q}}(\mathrm{f})$是q×q的窗口函数,即滤波器,用于对幅度谱进行平均,具体表示为:

$$\mathrm{h}_{\mathrm{q}}(\mathrm{f}) = \frac{1}{\mathrm{q}^2}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

[0034] (3)对所述幅度谱进行谱残差变换,取每个频率分量的幅度谱的对数,减去平均滤波,以突出输入指标中的显著特征,得到谱残差变换后的幅度谱。

[0035] $$\mathrm{R}(\mathrm{f}) = \mathrm{L}(\mathrm{f}) - \mathrm{AL}(\mathrm{f})$$

[0036] 其中,$\mathrm{R}(\mathrm{f})$表示谱残差变换后的幅度谱。

[0037] (4)对谱残差变换后的幅度谱进行逆变换,通常使用傅里叶逆变换,将复数形式的频域结果转换回空间域,得到指标数据的显著图。

[0038] $$M(x_m) = \| \mathfrak{F}^{-1}(\exp(R(f) + iP(f))) \|$$

[0039] 其中,$\mathfrak{F}^{-1}$表示傅立叶逆变换,i为虚数单位,$\| \cdot \|$表示向量的模,$M(x_m)$表示指标数据的显著图,即通过谱残差算法得到的指标显著图向量表达。

[0040] (5)将指标数据的显著图进行归一化,并纵向拼接,得到指标数据的特征。

[0041] 参考图2,图2是本发明实施例微服务系统CPU利用率原始指标数据和微服务指标数据的特征。

[0042] 进一步的实施例中,通过日志解析抽取日志数据的特征,具体为:

(1)使用Drain日志解析工具对所述日志数据进行聚类,每个聚类表示一个日志模板,得到日志模板集合$LTS = \{LT_1, LT_2, ..., LT_N\}$,其中,N表示聚类簇集数。

[0043] (2)预设一个时间间隔,统计预设时间间隔的不同时间段各个日志模板出现的频率,并转换为日志模板频次矩阵,其中日志模板频次矩阵表示各个日志模板在不同时间段出现的频率,例如第a行的第b列表示在第a个时间段内日志模板b出现的次数。将日志模板频次矩阵归一化,得到日志数据向量X^L,即日志数据的特征。

[0044] S3、融合指标数据的特征和日志数据的特征构建异常检测模型。

[0045] 具体为:

通过多层感知机将指标数据的特征和日志数据的特征映射到同一维度,得到指标数据隐层特征和日志数据隐层特征表示为:

$$H^M = MLP(X^M)$$

[0046] $$H^L = MLP(X^L)$$

[0047] 其中,H^M为指标数据隐层特征,H^L为日志数据隐层特征,$MLP()$表示多层感知机,X^M表示指标数据的特征,X^L表示日志数据的特征。

[0048] 将指标数据隐层特征和日志数据隐层特征输入以时间卷积网络TCN为主干的网络中,得到高维隐层特征h。

[0049] 结合自动编码器将高维隐层特征h压缩到低维嵌入空间,得到低维的指标数据的特征和低维的日志数据的特征。

[0050] 通过以下方式计算系统异常得分:

$$s = (x'_{test} - \mu_{in})^T \Sigma_{in}^{-1} (x'_{test} - \mu_{in})$$

[0051] 其中,s表示系统异常得分,μ_{in}和Σ_{in}^{-1}表示x'_{in}的均值和协方差矩阵,x'_{in}表示训练数据集D_{train}中的训练样本经过异常检测模型得到的低维特征,x'_{test}表示测试数据集D_{test}中的测试样本经过异常检测模型得到的低维特征,系统异常得分表示x'_{test}和x'_{in}之间的马哈拉比诺斯距离。

[0052] 如果s大于预设的阈值Threshold,则测试样本x_{test}是异常值;相反,测试样本x_{test}是正常值。

[0053] 进一步的实施例中,以时间卷积网络TCN为主干的网络的损失函数定义为:

$$\mathcal{L}_{contrastive} = -\frac{1}{n}\sum_{i=1}^{n}\frac{sim\left(H_i^M, H_i^L\right)}{\sum_k sim\left(H_i^M, H_k^L\right)}$$

[0054] 其中,$\mathcal{L}_{contrastive}$表示对比学习损失,$sim\left(H_i^M, H_i^L\right)$表示$H_i^M$和$H_i^L$之间的余弦相似度,$H_i^M$表示第i个指标数据隐层特征,$H_i^L$表示第i个日志数据隐层特征,$H_k^L$表示第k个日志数据隐层特征,n表示指标数据隐层特征个数。

[0055] 进一步的实施例中,自动编码器的损失函数表示为:

$$\mathcal{L}_{ae} = \|h - g(f(h))\|_2^2$$

[0056] 其中,$\mathcal{L}_{ae}$表示自动编码器的损失函数,$\|\cdot\|_2^2$表示欧几里得范数的平方,$f(h)$表示对h编码的结果,$g(f(h))$表示对$f(h)$解码的结果。

[0057] 进一步的实施例中,通过最小化异常检测模型的整体损失函数优化整个异常检测模型,异常检测模型的整体损失函数表示为:

$$\mathcal{L} = \lambda \times \mathcal{L}_{contrastive} + (1-\lambda) \times \mathcal{L}_{ae}$$

[0058] 其中,$\mathcal{L}$表示异常检测模型的整体损失函数,λ表示权重系数。

[0059] S4、将待检测数据输入异常检测模型,判断系统是否异常。

[0060] 在Sock-shop和Online-boutique两个数据集上验证本发明所提出的方法在故障注入场景下异常检测的效果,并与经典异常检测方法K最近邻算法(K-Nearest Neighbors, KNN)和支持向量机(support vector machines, SVM)的性能比较如表1所示,比较指标为:精确率(Precision)、召回率(Recall)F1分数(F1-Score)。

[0061] 表1

数据集/模型	Sock-shop			Online-boutique		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
本发明	82.75%	80.30%	81.48%	79.12%	78.24%	78.75%
KNN	78.92%	77.85%	78.38%	76.30%	75.18%	75.76%
SVM	73.20%	72.40%	72.80%	70.55%	68.93%	69.70%

[0062] 在一示例性实施例中,包括一种计算机可读存储介质,计算机可读存储介质存储有计算机程序,计算机程序被处理器执行时实现上述的微服务系统异常检测方法。

[0063] 请参阅图3,在一示例性实施例中,还包括一种电子设备,包括至少一处理器、至少一存储器、以及至少一通信总线。

[0064] 其中,存储器上存储有计算机程序,计算机程序包括计算机可读指令,处理器通过通信总线调用存储器中存储的计算机可读指令,执行上述的微服务系统异常检测方法。

[0065] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

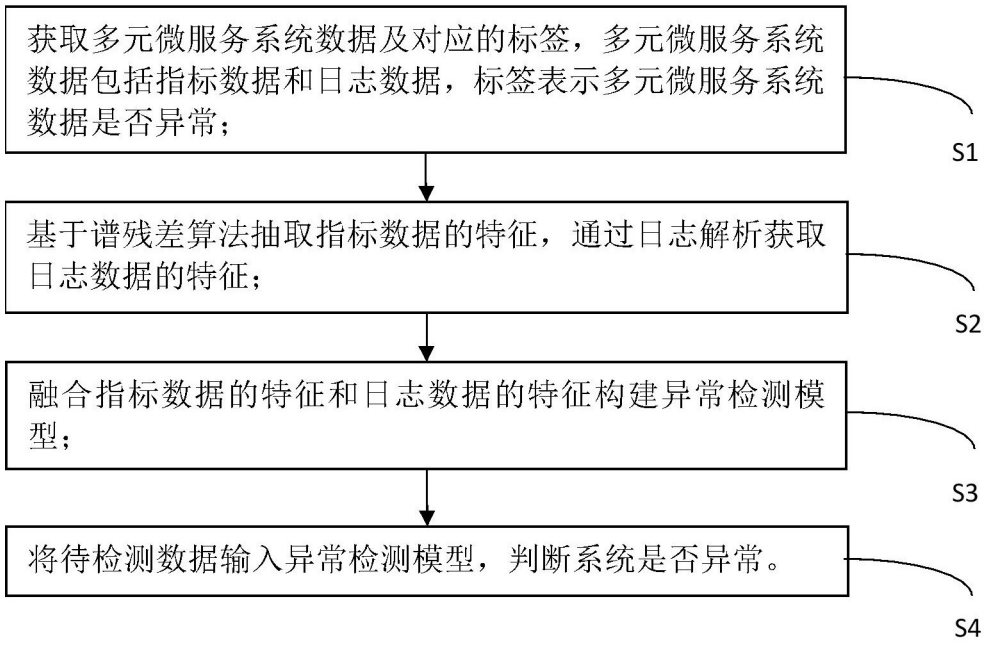

获取多元微服务系统数据及对应的标签，多元微服务系统数据包括指标数据和日志数据，标签表示多元微服务系统数据是否异常；
S1
基于谱残差算法抽取指标数据的特征，通过日志解析获取日志数据的特征；
S2
融合指标数据的特征和日志数据的特征构建异常检测模型；
S3
将待检测数据输入异常检测模型，判断系统是否异常。
S4

图1

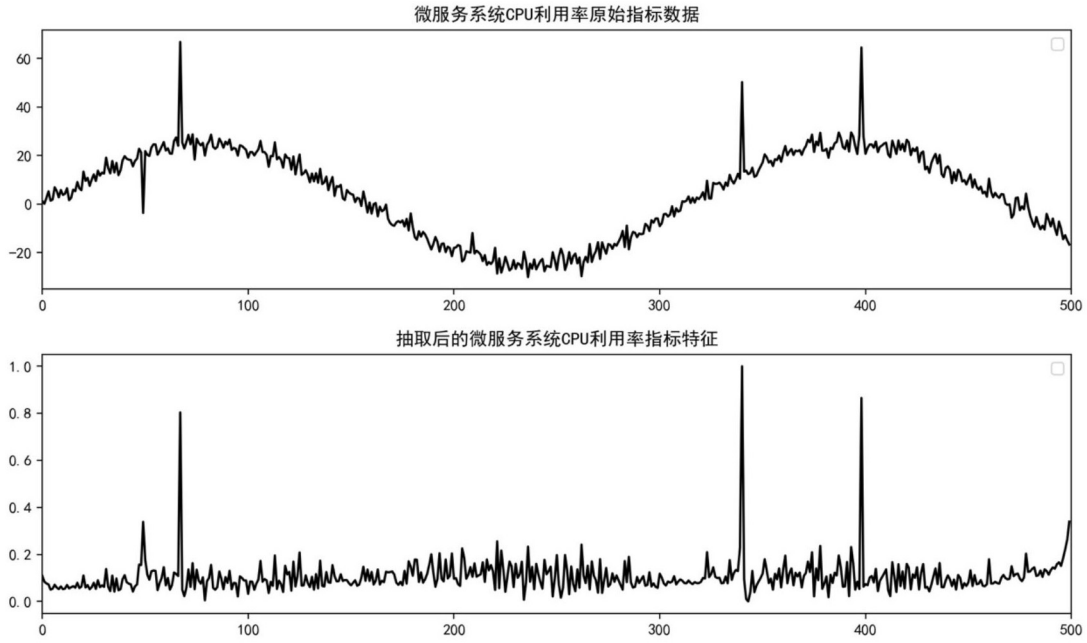

微服务系统CPU利用率原始指标数据
60
40
20
0
-20
0
100
200
300
400
500
抽取后的微服务系统CPU利用率指标特征
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500

图2

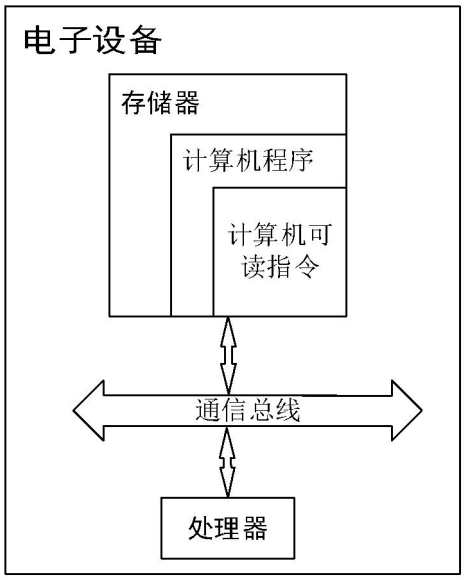


图3