(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117950830 A
(43)申请公布日 2024.04.30

(21)申请号 202311824636.3

(22)申请日 2023.12.27

(71)申请人 北京航空航天大学
地址 100191 北京市海淀区学院路37号

(72)发明人 吴文峻 陈睿博

(74)专利代理机构 北京永创新实专利事务所 11121
专利代理师 易卜

(51)Int.Cl.
G06F 9/48(2006.01)
G06F 9/50(2006.01)
G06F 18/2131(2023.01)
G06F 8/20(2018.01)
G06F 8/70(2018.01)
G06F 123/02(2023.01)

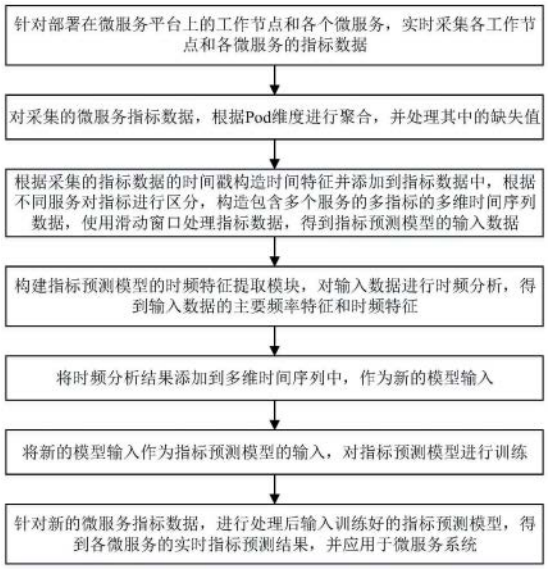
权利要求书3页 说明书6页 附图2页

(54)发明名称

一种基于时间序列时频分析的微服务指标预测方法

(57)摘要

本发明公开了一种基于时间序列时频分析的微服务指标预测方法,属于人工智能和微服务领域。首先采集微服务平台上各工作节点和各微服务的指标数据,并对指标数据进行处理;然后对指标数据构造包含多个服务的多指标的多维时间序列数据,使用滑动窗口处理指标数据,得到指标预测模型的输入数据。构建指标预测模型的时频特征提取模块,对输入数据进行时频分析,得到输入数据的主要频率特征和时频特征。最后将时频分析结果添加到多维时间序列中,作为新的模型输入,对指标预测模型进行训练,得到训练好的指标预测模型。本发明实现了对于非平稳序列的分析和特征提取,提高了指标预测模型的预测效率。



CN 117950830 A

1.一种基于时间序列时频分析的微服务指标预测方法,其特征在于,具体步骤如下:

步骤一、针对部署在微服务平台上的工作节点和各个微服务,实时采集各工作节点和各微服务的指标数据;

步骤二、对采集的微服务指标数据,根据Pod维度进行聚合,并处理其中的缺失值;

性能指标对应的m个时间序列数据在Pod维度聚合后由输入向量S表示:

$S=\{S_1,S_2,\ldots,S_i,\cdots,S_m\}$

其中第i个性能指标数据所对应的长度为n的特征向量S_i(1≤i≤m)表示为:

$S_i=\{S_{i1},S_{i2},\ldots,S_{in}\}$

步骤三,根据采集的指标数据的时间戳构造时间特征并添加到指标数据中,根据不同服务对指标进行区分,构造包含多个服务的多指标的多维时间序列数据,使用滑动窗口处理指标数据,得到指标预测模型的输入数据;

步骤四,构建指标预测模型的时频特征提取模块,对输入数据进行时频分析,得到输入数据的主要频率特征和时频特征;

首先,对各数据序列S_i进行快速傅里叶变换,得到其频率信息$Freqs_i=FFT(S_i)$;

随后,从各频率信息中选取幅值最高的M个频率分量作为对应数据的主要频率,表示为$Freqs_{main}=[fsi_1,fsi_2,\cdots,fsi_D]$;

最后,根据各个指标数据的主要频率设置对应的小波变换参数,并进行连续小波变换,从小波变换的结果中选取幅值最高的前N个频率,作为该序列的时频特征;

步骤五,将时频分析结果添加到多维时间序列中,作为新的模型输入;

将指标数据、时间特征和时频特征通过线性变换转化为相同的维度,基于数据本身特征构造Q,K,V矩阵;随后将时间特征与指标数据分别和时频特征相加,得到模型输入的时域数据和频域数据,分别表示为Q_tf,K_tf矩阵;

步骤六,将新的模型输入作为指标预测模型的输入,对指标预测模型进行训练,得到训练好的指标预测模型;

步骤七、针对新的微服务指标数据,进行处理后输入训练好的指标预测模型,得到各微服务的实时指标预测结果并应用于微服务系统,服务于系统指标的异常检测及服务扩缩容和服务调度场景。

2.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,所述工作节点的指标数据包括CPU利用率和内存利用率,所述微服务的指标数据包括各个微服务之间的请求调用响应时间以及请求次数。

3.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,所述缺失值处理的具体操作是:使用一阶移动平均法对缺失值进行填充,即取缺失值前后的两个正常数据的平均值代替缺失值。

4.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,所述时间特征的构造过程为:

首先,将指标数据所有的时间戳转化为标准格式的时间字符串“yyyy-MM-dd HH:mm:ss”,即年、月、日、时、分、秒;

然后,对具体的日期进行计算,得到如下七个特征,分别是:“年份year”、“月份month”、“当前时间是所在年份的第几天Day_Of_Year”、“当前时间是所在月份的第几天Day_Of_

Month”、“当前时间是所在星期的第几天Day_Of_Week”、“当前时间是所在天内的第几个小时Hour_Of_Day”、“当前时间是否是休息日Is_Holiday”。

5.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,使用滑动窗口处理指标数据为:针对包含m个时间序列的性能指标数据$S=\{S_1,S_2,...,S_m\}$,每个时间序列S_i的长度为n,使用长度为k的滑动窗口在这m个时间序列上得到n-k+1组数据。

6.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,步骤四中通过小波变换得到时频特征的过程具体为:

输入序列的每一维数据S_i的小波变换结果表示为$CWT(S_i)=[wi_1,wi_2,\cdots,wi_S]^T$,对于每个数据序列的小波变换结果按顺序进行拼接,得到输入序列的时频特征,表示为:$TimeFreq(S)=(tf_1,tf_2,\cdots,tf_S)^T$。

7.根据权利要求1所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,指标预测模型的训练过程具体为:

步骤601、将数据输入到指标预测模型中,在位置编码层中为每个时间序列$S_i(1\leqslant i\leqslant m)$中的每一个时间戳对应的指标数据值$S_{ij}(1\leqslant i\leqslant m,1\leqslant j\leqslant k)$计算相应的位置向量$P_{ij}$,然后将位置向量$P_{ij}$加到$S_{ij}$上得到$X_{ij}$;

位置向量P_{ij}的计算公式如下:

$$PE(pos,2i)=\sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

$$PE(pos,2i+1)=\cos\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

其中PE代表位置嵌入操作,pos表示当前计算的位置编码数据在输入序列中沿时间维度上的位置;i对应列索引,d_{model}为输出嵌入空间的维度;

步骤602、将添加了位置信息的输入数据X_{ij}输入到指标预测模型的编码器中的Attention层中,在考虑整个时间序列$S_i(1\leqslant i\leqslant m)$的环境下计算$X_{ij}$的Attention系数,然后将输出向量标准化后前向传播到解码器中;

步骤603、解码器接收到编码器的输出结果后,计算解码器输入与编码器输出的交叉注意力系数,并基于该交叉注意力系数和解码层输入计算得到解码层输出,即为预测结果$Y_{ij}(1\leqslant i\leqslant m,2\leqslant j\leqslant k+1)$;

步骤604、将解码器的输出结果$Y_{ij}(j=k+1)$与实际结果$S_{ij}(j=k+1)$对比,计算得到损失函数值,将该值从输出层向隐藏层反向传播到输入层,根据随机梯度下降函数按照梯度的负方向调整指标预测模型中的各个网络参数的值,重复上述训练过程,直至损失函数值最小,即得到训练好的指标预测模型;

损失函数值的计算公式为:

$$loss=\frac{1}{m}\sum_{i=1}^{m}\left(Y_{ij}-S_{ij}\right)^2。$$

8.根据权利要求7所述的一种基于时间序列时频分析的微服务指标预测方法,其特征在于,在计算Attention系数时,引入了注意力融合机制,通过融合时域和频域下的注意力系数,来刻画时间序列的内在关联关系;

时域和频域下的注意力系数分别为:

$$A_t = \sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

$$A_f = \sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

融合注意力系数为：

$$\text{Mix Attn}\left(Q_{tf}, K_{tf}, Q', K', V\right) = \text{softmax}\left(\sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right) + \sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)\right)V$$

基于时域和频域融合后的注意力系数,结合输入的values计算得到注意力模块的最终输出。

一种基于时间序列时频分析的微服务指标预测方法

技术领域

[0001] 本发明属于人工智能和微服务领域,涉及一种微服务系统中服务指标预测的算法,具体是一种基于时间序列时频分析的微服务指标预测方法。

背景技术

[0002] 随着互联网的飞速发展,业务的复杂性和软件系统架构的复杂性也在逐渐变大,微服务架构是一种新的软件系统架构,旨在解决传统软件系统架构的复杂性问题,主要致力于设计和开发高可维护性和高可扩展性的软件系统架构。

[0003] 微服务架构为分布式应用系统的设计和实现提供了指导,遵循微服务架构的设计原则,业务中拆分的每一个微服务的功能都是具体的,且单个微服务的复杂度远小于整体的系统复杂度,所以开发人员可以快速开发或者更新某一个微服务。

[0004] 相比较传统的单体服务,各个微服务之间是相互独立的。所以,各个微服务可以单独开发和交付。在云原生时代,微服务以容器的形式部署在机器上,所以独立的微服务也可以被独立的调度部署和扩缩容。

[0005] 微服务架构有很多优秀的设计,但同时它也存在着一些问题。随着业务的发展和变化,微服务的数量也变得越来越多,为了能够及时对服务进行调度和调整,保证服务良好运行的同时确保服务质量,在服务调度过程中往往需要基于服务的指标情况确定调度策略,传统的调度方法一般基于历史数据或当前的服务指标进行调整,这种策略往往存在明显的滞后性。

发明内容

[0006] 针对以上问题,本发明提出一种基于时间序列时频分析的微服务指标预测方法,实现了对微服务指标的快速准确预测,为服务调度提供了重要依据,也帮助运维人员及时了解服务的状态变化,对实现服务的快速调度和及时调整具有重要意义。

[0007] 所述基于时间序列时频分析的微服务指标预测方法,具体步骤如下:

[0008] 步骤一、针对部署在微服务平台上的工作节点和各个微服务,实时采集各工作节点和各微服务的指标数据;

[0009] 工作节点的指标数据包括:CPU利用率、内存利用率等;

[0010] 微服务的指标数据包括:各个微服务之间的请求调用响应时间以及请求次数等;

[0011] 步骤二、对采集的微服务指标数据,根据Pod维度进行聚合,并处理其中的缺失值;

[0012] 缺失值处理的具体操作是:使用一阶移动平均法对缺失值进行填充,即取缺失值前后的两个正常数据的平均值代替缺失值。

[0013] 性能指标对应的m个时间序列数据在Pod维度聚合后由输入向量S表示:

[0014] $S=\{S_1,S_2,\ldots,S_i,\cdots,S_m\}$

[0015] 其中第i个性能指标数据所对应的长度为n的特征向量S_i(1≤i≤m)表示为:

[0016] $S_i=\{S_{i1},S_{i2},\ldots,S_{in}\}$

[0017] 步骤三,根据采集的指标数据的时间戳构造时间特征并添加到指标数据中,根据不同服务对指标进行区分,构造包含多个服务的多指标的多维时间序列数据,使用滑动窗口处理指标数据,得到指标预测模型的输入数据;

[0018] 时间特征为指标数据的时间戳提取后的结果,具体操作是:

[0019] 首先,将指标数据所有的时间戳转化为标准格式的时间字符串“yyyy-MM-dd HH:mm:ss”,即年、月、日、时、分、秒。

[0020] 然后,对具体的日期进行计算,得到如下七个特征,分别是:“年份year”、“月份month”、“当前时间是所在年份的第几天Day_Of_Year”、“当前时间是所在月份的第几天Day_Of_Month”、“当前时间是所在星期的第几天Day_Of_Week”、“当前时间是所在天内的第几个小时Hour_Of_Day”、“当前时间是否是休息日Is_Holiday”。

[0021] 使用滑动窗口处理指标数据为:针对包含m个时间序列的性能指标数据$S=\{S_1,S_2,...,S_m\}$,每个时间序列S_i的长度为n,使用长度为k的滑动窗口在这m个时间序列上得到n-k+1组数据;

[0022] 步骤四,构建指标预测模型的时频特征提取模块,对输入数据进行时频分析,得到输入数据的主要频率特征和时频特征;

[0023] 首先,对各数据序列S_i进行快速傅里叶变换,得到其频率信息$Freqs_i=FFT(S_i)$;

[0024] 随后,从各频率信息中选取幅值最高的M个频率分量作为对应数据的主要频率,表示为$Freqs_{main}=[fsi_1,fsi_2,\cdots,fsi_D]$;

[0025] 最后,根据各个指标数据的主要频率设置对应的小波变换参数,并进行连续小波变化,从小波变换的结果中选取幅值最高的前N个频率,作为该序列的时频特征。

[0026] 输入序列的每一维数据S_i的小波变换结果表示为$CWT(S_i)=[wi_1,wi_2,\cdots,wi_S]^T$,对于每个数据序列的小波变换结果按顺序进行拼接,得到输入序列的时频特征,表示为:$TimeFreq(S)=(tf_1,tf_2,\cdots,tf_S)^T$。

[0027] 步骤五,将时频分析结果添加到多维时间序列中,作为新的模型输入;

[0028] 将指标数据、时间特征和时频特征通过线性变换转化为相同的维度,基于数据本身特征构造Q,K,V矩阵;随后将时间特征与指标数据分别和时频特征相加,得到模型输入的时域数据和频域数据,分别表示为Q_tf,K_tf矩阵。

[0029] 步骤六,将新的模型输入作为指标预测模型的输入,对指标预测模型进行训练;

[0030] 指标预测模型是基于Transformer模型构造的,包括编码层和解码层;

[0031] 指标预测模型的训练过程为:

[0032] 步骤601、将数据输入到指标预测模型中,在位置编码层中为每个时间序列S_i($1\leqslant i\leqslant m$)中的每一个时间戳对应的指标数据值S_{ij}($1\leqslant i\leqslant m,1\leqslant j\leqslant k$)计算相应的位置向量$P_{ij}$,然后将位置向量$P_{ij}$加到$S_{ij}$上得到$X_{ij}$;

[0033] 位置向量P_{ij}的计算公式如下:

[0034] $$PE(pos,2i)=\sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

[0035] $$PE(pos,2i+1)=\cos\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

[0036] 其中PE代表位置嵌入操作,pos表示当前计算的位置编码数据在输入序列中沿时间维度上的位置。i对应列索引,d_{model}为输出嵌入空间的维度。

[0037] 步骤602、将添加了位置信息的输入数据X_{ij}输入到指标预测模型的编码器中的Attention层中,在考虑整个时间序列S_i(1≤i≤m)的环境下计算X_{ij}的Attention系数,然后将输出向量标准化后前向传播到解码器中。

[0038] 在计算Attention系数时,引入了注意力融合机制,通过融合时域和频域下的注意力系数,来刻画时间序列的内在关联关系。

[0039] 时域和频域下的注意力系数分别为:

[0040] $$A_t = \sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

[0041] $$A_f = \sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

[0042] 融合注意力系数为:

[0043] $$\text{Mix Attn}\left(Q_{tf},K_{tf},Q',K',V\right)=\text{softmax}\left(\sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right)+\sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)\right)V$$

[0044] 基于时域和频域融合后的注意力系数,结合输入的values计算得到注意力模块的最终输出。

[0045] 步骤603、解码器接收到编码器的输出结果后,计算解码器输入与编码器输出的交叉注意力系数,并基于该交叉注意力系数和解码层输入计算得到解码层输出,即为预测结果Y_{ij}(1≤i≤m,2≤j≤k+1);

[0046] 步骤604、将解码器的输出结果Y_{ij}(j=k+1)与实际结果S_{ij}(j=k+1)对比,计算得到损失函数值,将该值从输出层向隐藏层反向传播到输入层,根据随机梯度下降函数按照梯度的负方向调整指标预测模型中的各个网络参数的值,重复上述训练过程,直至损失函数值最小,即得到训练好的指标预测模型。

[0047] 损失函数值的计算公式为:

[0048] $$loss=\frac{1}{m}\sum_{i=1}^{m}\left(Y_{ij}-S_{ij}\right)^2$$

[0049] 步骤七、针对新的微服务指标数据,进行处理后输入训练好的指标预测模型,得到各微服务的实时指标预测结果并应用于微服务系统,服务于系统指标的异常检测及服务扩缩容和服务调度场景。

[0050] 本发明的优点在于:

[0051] (1)本发明提出一种基于时间序列时频分析的微服务指标预测方法,通过信号处理中适用于非平稳序列分析的连续小波变换,实现对于非平稳序列的分析和特征提取,实现了对于非平稳序列的良好表达。

[0052] (2)本发明通过在模型层面的结构调整,包括对于模型输入和注意力计算方法的调整,提高了模型的预测效率。

附图说明

[0053] 图1为本发明基于时间序列时频分析的微服务指标预测方法的流程图;

[0054] 图2为本发明的微服务指标预测模型的示意图。

具体实施方式

[0055] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整的描述,显然,所描述的实施例仅为本发明的一部分实施例,而不是全部的实施例,基于本发明中的实施例,本领域的普通技术人员在不付出创造性劳动的前提下所获得的所有其他实施例,都属于本发明的保护范围。

[0056] 本发明提出一种基于时间序列时频分析的微服务指标预测方法,通过准确预测服务指标的变化趋势,从而为服务调度提供参考依据,实现服务的提前调度,在保障服务质量的情况下,实现对服务资源的有效利用。

[0057] 一种基于时间序列时频分析的微服务指标预测方法,如图1所示,具体步骤如下:

[0058] 步骤一、针对微服务平台上的工作节点和各个微服务,通过开源监控工具Prometheus,实时采集工作节点的CPU利用率、内存利用率等指标数据和各个微服务之间的请求调用响应时间以及请求次数等指标数据;

[0059] 步骤二、对采集的微服务指标数据,根据Pod维度进行聚合,并处理其中的缺失值;

[0060] 缺失值处理的具体操作是:使用一阶移动平均法对缺失值进行填充,即取缺失值前后的两个正常数据的平均值代替缺失值。

[0061] 性能指标对应的m个时间序列数据在Pod维度聚合后由输入向量S表示:

[0062] $S=\{S_1,S_2,\ldots,S_i,\cdots,S_m\}$

[0063] 其中第i个性能指标数据所对应的长度为n的特征向量S_i(1≤i≤m)表示为:

[0064] $S_i=\{S_{i1},S_{i2},\ldots,S_{in}\}$

[0065] 步骤三、根据采集的指标数据的时间戳构造时间特征并添加到指标数据中,根据不同服务对指标进行区分,构造包含多个服务的多指标的多维时间序列数据,使用滑动窗口处理指标数据,得到指标预测模型的输入数据;

[0066] 时间特征为指标数据的时间戳提取后的结果,具体操作是:

[0067] 首先,将指标数据所有的时间戳转化为标准格式的时间字符串“yyyy-MM-dd HH:mm:ss”,即年、月、日、时、分、秒。

[0068] 然后,对具体的日期进行计算,得到如下七个特征,分别是:“年份year”、“月份month”、“当前时间是所在年份的第几天Day_Of_Year”、“当前时间是所在月份的第几天Day_Of_Month”、“当前时间是所在星期的第几天Day_Of_Week”、“当前时间是所在天内的第几个小时Hour_Of_Day”、“当前时间是否是休息日Is_Holiday”。

[0069] 使用滑动窗口处理指标数据为:针对包含m个时间序列的性能指标数据$S=\{S_1,S_2,\ldots,S_m\}$,每个时间序列S_i的长度为n,使用长度为k的滑动窗口在这m个时间序列上得到n-k+1组数据;

[0070] 步骤四,构建指标预测模型的时频特征提取模块,对输入数据进行时频分析,得到输入数据的主要频率特征和时频特征;

[0071] 首先,对各数据序列S_i进行快速傅里叶变换,得到其频率信息$Freqs_i=FFT(S_i)$;

[0072] 随后,从各频率信息中选取幅值最高的M个频率分量作为对应数据的主要频率,表示为$Freqs_ma_in=[fsi_1,fsi_2,\cdots,fsi_D]$;

[0073] 最后,根据各个指标数据的主要频率设置对应的小波变换参数,并进行连续小波变化,从小波变换的结果中选取幅值最高的前N个频率,作为该序列的时频特征。

[0074] 根据各个数据的主要频率范围计算得到进行小波变换时对应的尺度信息,通过这一步骤可以确保变换后每个数据的频率分量分布在其主要频率内,同样的,从小波变换的结果中选取幅值最高的前N个频率,作为该序列的时频特征。输入序列的每一维数据S_i的小波变换结果表示为$CWT(S_i)=[wi_1,wi_2,\cdots,wi_S]^T$,最后对于每个数据序列的变换结果按顺序进行拼接,得到输入序列的时频特征,表示为:$TimeFreq(S)=(tf_1,tf_2,\cdots,tf_S)^T$。

[0075] 步骤五,将时频分析结果添加到多维时间序列中,作为新的模型输入;

[0076] 将指标数据、时间特征和时频特征通过线性变换转化为相同的维度,基于数据本身特征构造Q,K,V矩阵;随后将时间特征与指标数据分别和时频特征相加,得到模型输入的时域数据和频域数据,表示为Q_tf,K_tf矩阵。

[0077] 步骤六、基于Transformer模型的主要思想构建并调整预测模型,将新的模型输入数据输入到指标预测模型中进行训练。

[0078] 指标预测模型结构如图2所示,模型由编码层和解码层构成,编码层输入为步骤三中处理后的数据,解码层输入为带预测数据之前长度为L的数据。

[0079] 训练的具体过程如下:

[0080] 步骤601、将数据输入到指标预测模型中,在位置编码层中为每个时间序列S_i(1≤i≤m)中的每一个时间戳对应的指标数据值S_{ij}(1≤i≤m,1≤j≤k)计算相应的位置向量P_{ij},然后将位置向量P_{ij}加到S_{ij}上得到X_{ij};

[0081] 位置向量P_{ij}的计算公式如下:

[0082] $$PE(pos,2i)=\sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

[0083] $$PE(pos,2i+1)=\cos\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

[0084] 其中PE代表位置嵌入操作,pos表示当前计算的位置编码数据在输入序列中沿时间维度上的位置。i对应列索引,d_{model}为输出嵌入空间的维度。

[0085] 步骤602、将添加了位置信息的输入数据X_{ij}输入到异常检测模型的编码器中的Attention层中,在考虑整个时间序列X_i(1≤i≤m)的环境下计算X_{ij}的Attention分数,然后将输出向量标准化后前向传播到解码器中;

[0086] 在计算Attention分数时,引入了注意力融合机制,模型中的注意力分为自注意力和交叉注意力,在本发明中,融合自注意力是基于输入时间序列的自相关系数和序列时频分解结果通过混合得到,融合交叉注意力的计算方式与自注意力类似,区别在于交叉注意力是计算输入序列与编码层输出序列的交叉注意力。

[0087] 基于时频分析结果计算频域下的时间序列自注意力和交叉注意力系数,基于原始时间序列的自相关系数计算时域下的序列自注意力和交叉注意力系数。

[0088] 通过SoftMax函数对时域和频域下的序列的自注意力系数和交叉注意力系数进行混合,得到融合注意力系数。

[0089] 时域和频域下的注意力系数分别为:

[0090] $$A_t=\sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

[0091] $$A_f=\sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)$$

[0092] 融合注意力系数为：

[0093] $$\text{Mix Attn}\left(Q_{tf},K_{tf},Q',K',V\right)=\text{softmax}\left(\sigma\left(\frac{Q'K'^{\mathrm{T}}}{\sqrt{d_k}}\right)+\sigma\left(\frac{Q_{tf}K_{tf}^{\mathrm{T}}}{\sqrt{d_k}}\right)\right)V$$

[0094] 基于时域和频域融合后的注意力系数，结合输入的values计算得到注意力模块的最终输出。

[0095] 步骤603、解码器接收到编码器的输出结果后，逐步生成预测结果Y_{ij}(1≤i≤m,2≤j≤k+1)；

[0096] 步骤604、将解码器的输出结果Y_{ij}(j=k+1)与实际结果S_{ij}(j=k+1)对比计算得到损失函数值，将该值从输出层向隐藏层反向传播到输入层，根据随机梯度下降函数调整指标预测模型中的各个网络参数的值，重复上述训练过程，直至损失函数值最小，即得到训练好的指标预测模型。

[0097] 损失函数值的计算公式为：

[0098] $$loss=\frac{1}{m}\sum_{i=1}^{m}\left(Y_{ij}-S_{ij}\right)^2$$

[0099] 步骤七、将训练好的指标预测模型用于微服务平台中的服务的实时的指标预测。

[0100] 具体过程如下：

[0101] 步骤701、实时采集微服务平台中微服务所在工作节点的指标数据和微服务的指标数据，包括CPU利用率、内存利用率和各个微服务之间的请求调用响应时间以及请求次数等；

[0102] 步骤702、将采集的指标数据信息S_{ij}(1≤i≤m,1≤j≤n)根据滑动窗口划分，然后输入到指标预测算法中，得出输出结果Y_{ij}(1≤i≤m,2≤j≤n+1)；

[0103] 尽管上面对本发明说明性的具体实施方式进行了描述，以便于本技术领域的技术人员理解本发明，且应该清楚，本发明不限于具体实施方式的范围，对本技术领域的普通技术人员来讲，只要各种变化在所附的权利要求限定和确定的本发明的精神和范围内，这些变化是显而易见的，一切利用本发明构思的发明创造均在保护之列。

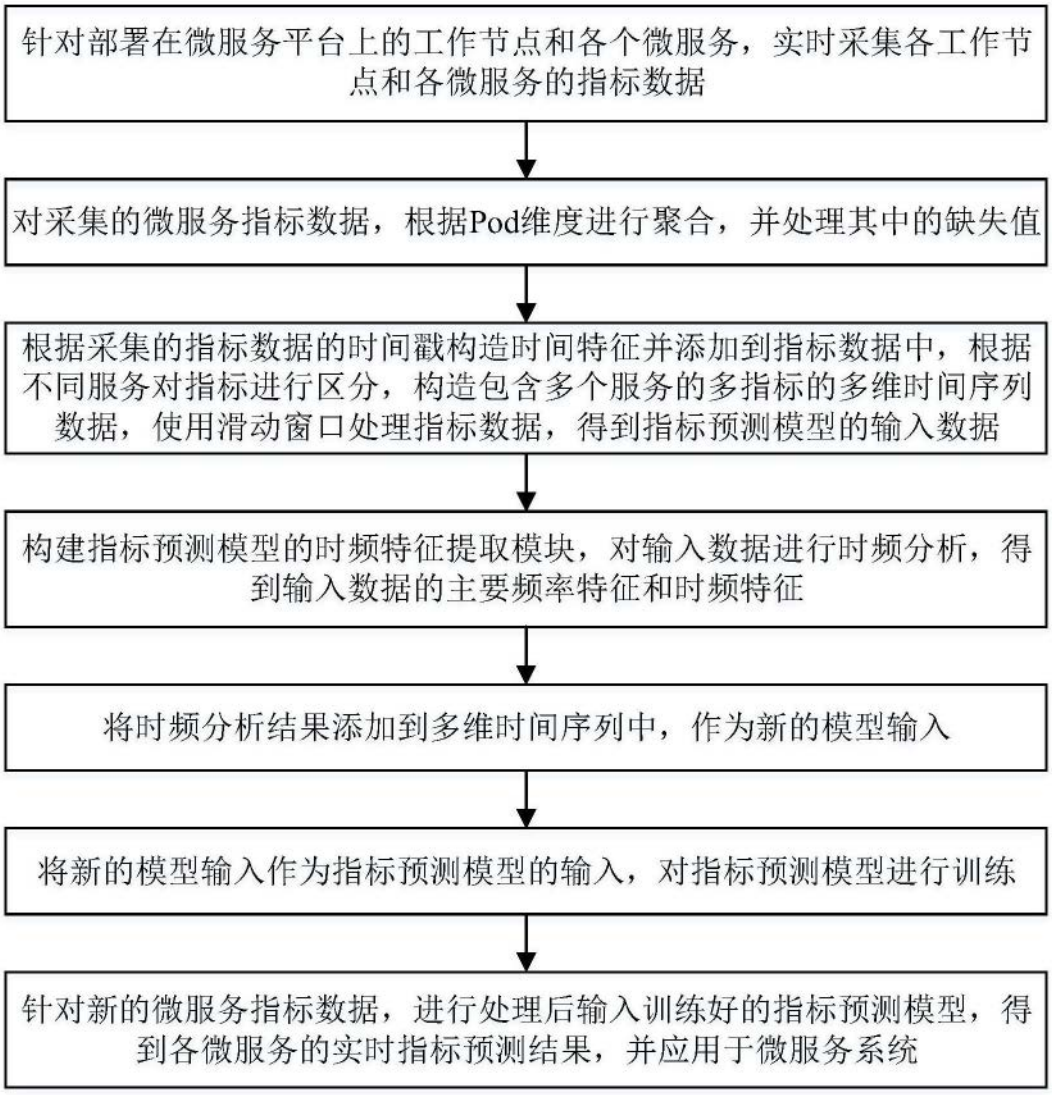

针对部署在微服务平台上的工作节点和各个微服务，实时采集各工作节点和各微服务的指标数据
对采集的微服务指标数据，根据Pod维度进行聚合，并处理其中的缺失值
根据采集的指标数据的时间戳构造时间特征并添加到指标数据中，根据不同服务对指标进行区分，构造包含多个服务的多指标的多维时间序列数据，使用滑动窗口处理指标数据，得到指标预测模型的输入数据
构建指标预测模型的时频特征提取模块，对输入数据进行时频分析，得到输入数据的主要频率特征和时频特征
将时频分析结果添加到多维时间序列中，作为新的模型输入
将新的模型输入作为指标预测模型的输入，对指标预测模型进行训练
针对新的微服务指标数据，进行处理后输入训练好的指标预测模型，得到各微服务的实时指标预测结果，并应用于微服务系统

图1

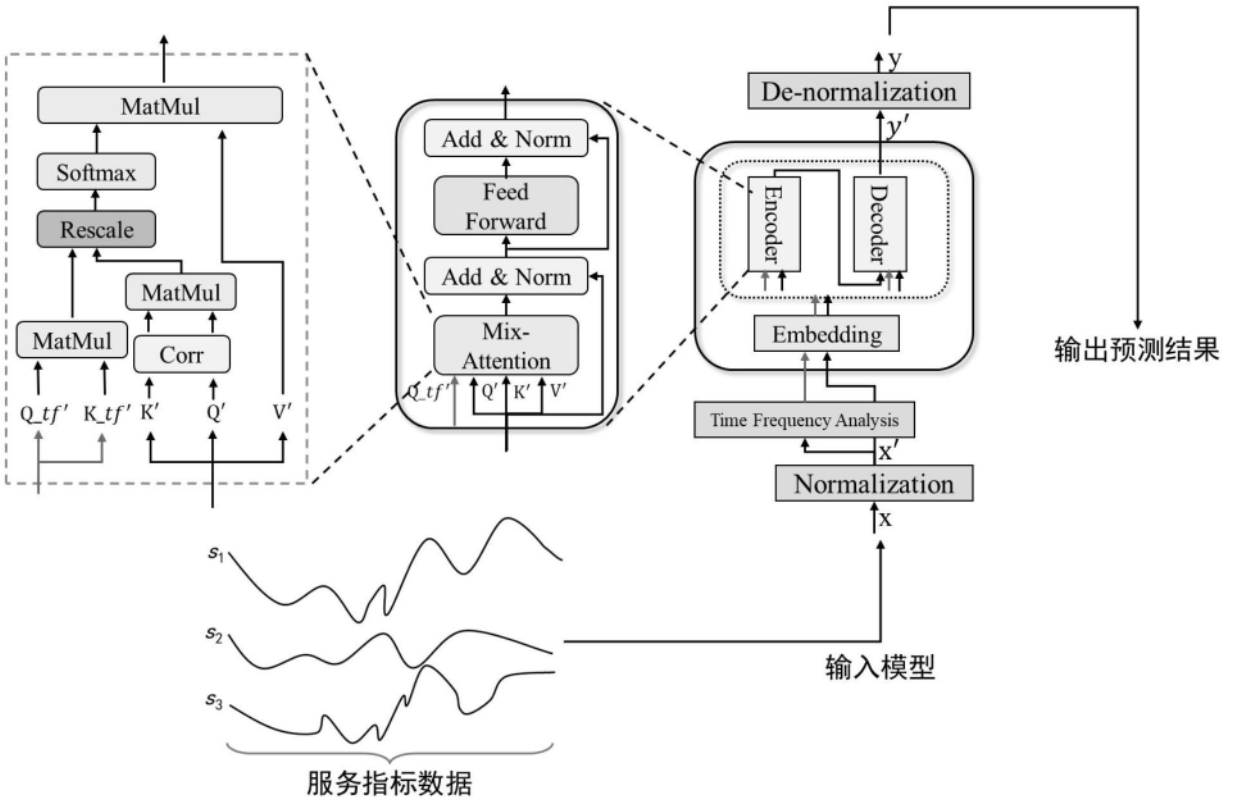

MatMul
Softmax
Rescale
MatMul
MatMul
Corr
Q_tf′ K_tf′ K′ Q′ V′
Add & Norm
Feed Forward
Add & Norm
Mix-Attention
Q_tf′ Q′ K′ V′
y
De-normalization
y′
Encoder
Decoder
Embedding
Time Frequency Analysis
x′
Normalization
x
输出预测结果
s1
s2
s3
服务指标数据
输入模型

图2