



(12) 发明专利

(10) 授权公告号 CN 114647465 B

(45) 授权公告日 2022.08.16

(21) 申请号 202210559572.8

李明 尹晓宇 宫帅 程航

(22) 申请日 2022.05.23

(74) 专利代理机构 南京钟山专利代理有限公司

(65) 同一申请的已公布的文献号

32252

申请公布号 CN 114647465 A

专利代理师 上官凤栖

(43) 申请公布日 2022.06.21

(51) Int.Cl.

(73) 专利权人 南京航空航天大学

G06F 9/448 (2018.01)

地址 211106 江苏省南京市江宁区将军大

G06K 9/62 (2022.01)

道29号

G06N 3/04 (2006.01)

专利权人 国家电网有限公司信息通信分公  
司

G06N 3/08 (2006.01)

国网安徽省电力有限公司信息通  
信分公司

(56) 对比文件

CN 114296975 A, 2022.04.08

国家电网有限公司

CN 113780470 A, 2021.12.10

(72) 发明人 李静 羊麟威 钱李烽 刘昕

审查员 郭俊

张攀 来风刚 都繁杰 周逸

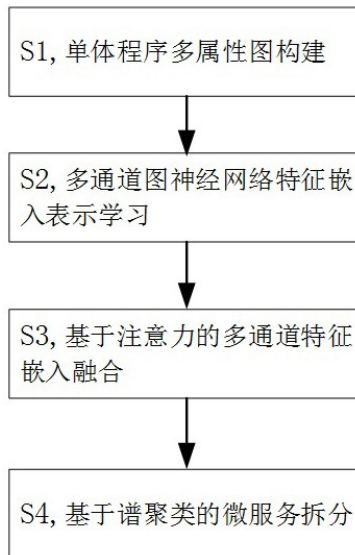
权利要求书4页 说明书14页 附图6页

(54) 发明名称

多通道注意力图神经网络聚类的单体程序  
拆分方法及系统

(57) 摘要

本发明公开了一种多通道注意力图神经网络聚类的单体程序拆分方法及系统,包括单体程序多属性图构建、多通道图神经网络特征嵌入表示学习、基于注意力的多通道特征嵌入融合、基于谱聚类的微服务拆分。通过图神经网络重构信息和聚类信息构建新的损失函数,实现了图注意力神经网络与聚类的联合学习框架,实现了提取微服务在功能性和模块性方面性能的提升。本发明结合单体程序多种属性信息,构建多通道图注意力网络,实现了更为高质量的特征嵌入表示,同时也提升了微服务提取方法的可扩展性,避免了微服务提取方法在应用中使用受限等问题。



1. 多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于,包括:

步骤S1:对JAVA单体程序中获取的类和类之间的多源关系建模并组合形成多通道信息数据对,基于多通道信息数据对和邻接矩阵构建单体程序多属性图;

步骤S2:将单体程序多属性图输入到多通道图卷积编码器获得对应的特征嵌入表示,利用多通道图卷积解码器重构多通道信息数据对并基于重构误差设计重构损失函数;

步骤S3:通过注意力机制计算特征嵌入表示中各结点的重要性,并自适应融合为新的特征嵌入表示;

步骤S4:通过图神经网络和聚类的联合学习框架对新的特征嵌入表示进行聚类,实现微服务的划分,并基于重构损失函数和聚类过程中产生的聚类损失函数,对图神经网络的参数进行优化。

2. 如权利要求1所述的多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于:所述步骤S1具体包括如下子步骤:

步骤S11:通过对JAVA应用程序的动态分析,获得运行时调用链追踪数据信息;

步骤S12:根据调用链追踪数据信息,得到类与调用链之间的关系子图 $X_{class-trace}$ ;

步骤S13:根据调用链追踪数据信息,得到类之间的矩阵属性子图 $X_{class-occurrence}$ ;

步骤S14:根据调用链追踪数据信息,得到类之间的继承关系子图 $X_{class-inheri}$ ;

步骤S15:根据调用链追踪数据信息,得到类之间的邻接矩阵 $A$ ;邻接矩阵定义如下:

$$A_{uv} = \begin{cases} 1, & \text{if } (v_a, v_b) \in E \\ 0 & \text{otherwise} \end{cases}$$

其中, $E$ 表示继承关系子图的边集合, $v_a$ 、 $v_b$ 分别表示结点 $a$ 和结点 $b$ , $A_{uv}$ 表示邻接矩阵 $A$ 的第 $u$ 行、第 $v$ 列的值;

步骤S16:将关系子图 $X_{class-trace}$ 、矩阵属性子图 $X_{class-occurrence}$ 和继承关系子图 $X_{class-inheri}$ 进行组合形成多通道信息数据对 $X_1$ 、 $X_2$ 和 $X_3$ ,其中, $X_1 = (X_{class-trace}, X_{class-occurrence})$ , $X_2 = (X_{class-trace}, X_{class-inheri})$ , $X_3 = (X_{class-occurrence}, X_{class-inheri})$ ,结合邻接矩阵 $A$ 构建单体程序多属性图。

3. 如权利要求2所述的多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于:所述步骤S2具体包括如下子步骤:

步骤S21:将步骤S1中得到的单体程序多属性图输入多通道图卷积编码器,学习得到三个通道的特征嵌入表示,模型描述如下:

$$f_m(G, X_m; \theta) \rightarrow Z_m$$

其中, $f_m$ 为第 $m$ 个通道信息数据与特征嵌入表示的映射关系, $G$ 为单体程序多属性图, $X_m$ 为第 $m$ 个通道信息数据对, $\theta$ 为图神经网络的参数, $Z_m$ 表示学习得到的特征嵌入表示;

第 $m$ 个图卷积编码器的第 $l$ 层的输出 $Z_m^{(l)}$ 为:

$$Z_m^{(l)} = \sigma(\tilde{D}_m^{-\frac{1}{2}} \tilde{A} \tilde{D}_m^{-\frac{1}{2}} Z_m^{(l-1)} W_m^l)$$

其中, $\tilde{A} = A + I_N$ 是具有自连接的邻接矩阵, $I_N$ 是单位矩阵; $\tilde{D}_m$ 是 $\tilde{A}$ 对应图的度矩阵,

$W_m^l$ 是第 $m$ 个图卷积编码器的权值参数,  $Z_m^{(l-1)}$ 表示第 $m$ 个图卷积编码器第 $l-1$ 层的输出,  $\sigma$ 表示非线性激活函数;

步骤S22:将步骤S21中得到的特征嵌入表示输入多通道图卷积解码器,重构多通道信息数据对;第 $m$ 个图卷积解码器的第 $l$ 层的输出  $\hat{Z}_m^l$ 为:

$$\hat{Z}_m^l = \sigma(\hat{D}^{-\frac{1}{2}} \bar{A} \hat{D}^{-\frac{1}{2}} \hat{Z}_m^{l-1} \hat{W}^{l-1})$$

其中,  $\bar{A} = 2I_n - A$ ,  $\hat{D} = 2I_n + D$ ,  $I_n$ 是单位矩阵,  $D$ 是 $A$ 对应图的度矩阵,  $\hat{W}^{l-1}$ 是第 $m$ 个图卷积解码器的权值参数;  $\hat{Z}_m^{l-1}$ 表示第 $m$ 个图卷积解码器第 $l-1$ 层的输出;第 $m$ 个图卷积解码器最后一层的输出就是重构后的第 $m$ 个通道信息数据对  $\hat{X}_m$ ;

重构 $m$ 通道视图邻接矩阵  $\hat{A}_m$ :

$$\hat{A}_m = \sigma(Z_m Z_m^T)$$

基于重构误差构建每个通道图神经网络的重构损失,重构损失函数 $L_{sa}$ 定义为:

$$L_{sa} = \min_{\theta_i} \sum_{m=1}^M \|X_m - \hat{X}_m\|_F^2 + \lambda \sum_{m=1}^M \|A - \hat{A}_m\|_F^2$$

其中,  $\theta_i$ 是图神经网络的参数,  $\lambda$ 是超参数,  $M$ 是通道数目。

4.如权利要求3所述的多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于:所述步骤S3具体如下:

通过注意力机制先对每个通道的特征嵌入表示进行非线性变换,再用一个共享注意力向量  $p_m \in R^{h \times 1}$  得到 $m$ 通道结点 $e$ 的注意力值  $w_m^e$ :

$$w_m^e = p_m^T \cdot \tanh(W_m \cdot (z_m^e)^T + b_m)$$

其中,  $W_m \in R^{h' \times h}$ 是权值矩阵,  $h'$ 表示共享注意力向量的维度,  $h$ 表示注意力机制的通道数;  $z_m^e \in R^{1 \times h}$ 表示 $m$ 通道 $e$ 结点的特征嵌入表示,  $b_m \in R^{h \times 1}$ 是偏置向量;对  $w_m^e$ 进行归一化:

$$\gamma_m^e = \text{soft max}(w_m^e) = \frac{\exp(w_m^e)}{\sum_{m=1}^M \exp(w_m^e)}$$

其中,  $\gamma_m^e$ 值表示 $m$ 通道 $e$ 结点的特征嵌入表示的重要性,获得学习后的权值  $\gamma_m = [\gamma_m^e] \in R^{n \times 1}$ ,  $n$ 表示嵌入特征维度;将三个通道的特征嵌入表示 $Z_m$ 进行融合获得新的特征嵌入表示 $Z$ :

$$Z = \sum_{m=1}^M \beta_m \cdot Z_m$$

其中,  $\beta_m = \text{diag}[\gamma_m]$ 。

5. 如权利要求4所述的多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于:所述步骤S4具体如下:

采用谱聚类算法对特征嵌入表示 $Z$ 进行聚类,使用高斯核定义仿射矩阵 $S$ ,如果 $e \neq o$ ,则

$$S_{eo} = \exp\left(\frac{\|z_e - z_o\|_2^2}{\Omega}\right), \quad e, o = 1, 2, \dots, N$$

否则 $S_{eo} = 0$ ;  $S_{eo}$ 代表结点 $e$ 和结点 $o$ 通过高斯核函数计算得到的相似度,  $N$ 表示结点数,  $z_e$ 表示结点 $e$ 的特征嵌入表示,  $\|\cdot\|$ 是向量 $z_e$ 和 $z_o$ 之间的欧几里得距离,  $\Omega$ 是尺度参数,  $S_{eo}$ 对称且非负;然后,运行谱聚类算法将 $N$ 个结点聚类到 $K$ 个不同划分中;

谱聚类后,构建簇的 $k$ - $d$ 树,找到与结点 $z_e$ 同一个簇中曼哈顿距离最近的邻居结点 $z_o$ ,定义聚类损失函数 $L_{clus}$ :

$$L_{clus} = \sum_e \min_o d(z_e, z_o), \quad e, o = 1, 2, \dots, u$$

其中,  $d(z_e, z_o)$ 表示结点 $e$ 和结点 $o$ 之间的曼哈顿距离 $|z_e - z_o|$ ;

综合 $L_{sa}$ 和 $L_{clus}$ 形成新的损失函数 $L$ :

$$L = \sum_{m=1}^M L_{sa}^m + \gamma L_{clus}$$

其中,  $\gamma > 0$ 是超参数;通过最小化损失函数 $L$ 训练并优化图神经网络的参数。

6. 多通道注意力图神经网络聚类的单体程序拆分系统,其特征在于,包括:多通道注意力网络模块、多通道图卷积神经网络模块、注意力机制模块和联合学习模块;

所述多通道注意力网络模块对JAVA单体程序中获取的类和类之间的多源关系建模并组合形成多通道信息数据对,基于多通道信息数据对和邻接矩阵构建单体程序多属性图;

所述多通道图卷积神经网络模块包括多通道图卷积编码器和多通道图卷积解码器,所述单体程序多属性图输入到多通道图卷积编码器获得对应的特征嵌入表示;所述多通道图卷积解码器重构多通道信息数据对并基于重构误差设计重构损失函数;

所述注意力机制模块通过注意力机制计算特征嵌入表示中各结点的重要性,并自适应融合为新的特征嵌入表示;

所述联合学习模块通过图神经网络和聚类的联合学习框架对新的特征嵌入表示进行聚类,实现微服务的划分,并基于重构损失函数和聚类过程中产生的聚类损失函数,对图神经网络的参数进行优化。

7. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序使计算机执行如权利要求1-5任一项所述的多通道注意力图神经网络聚类的单体程序拆分方法。

8. 一种电子设备,其特征在于,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行计算机程序时,实现如权利要求1-5任一项所述的多通道注意力图神经网络聚类的单体程序拆分方法。

## 多通道注意力图神经网络聚类的单体程序拆分方法及系统

### 技术领域

[0001] 本发明属于软件工程和人工智能领域,具体涉及一种多通道注意力图神经网络聚类的单体程序拆分方法及系统。

### 背景技术

[0002] 随着云计算技术的快速发展,为了充分利用云基础设施,灵活进行业务扩张和性能伸缩,降低维护成本,Amazon、IBM、Google等大型公司研究将单体应用软件系统迁移为基于微服务的架构,与单体架构这种将系统各个模块统一管理打包成单个应用程序的范式不同,微服务架构由许多独立的服务组成,服务之间通过轻量级通信协议进行动态交互,每个服务可独立修改、开发、部署和维护。因此,在增量迁移上云过程中,需要将单体应用替换到新的体系结构中。微服务提取或拆分是将现有代码重构为一组较小的独立代码组的过程,理想情况下不改变外部行为和语义。一旦移植到微服务,业务服务就可以独立地增强和扩展,提供敏捷性和改进的交付速度。目前大多数微服务拆分和重构过程是手工完成的,代价昂贵、耗时且易于出错,其质量通常与专家经验和知识密切相关,因此,迫切需要自动化的过程来将单体应用转换成微服务系统,微服务提取已成为软件工程和云计算领域研究的重要任务之一。

[0003] 近些年,已有许多工作研究单体应用系统的微服务提取方法,微服务提取的主要标准是每个微服务尽最大可能满足低耦合和高内聚标准。Mitchell等人于2006年提出一种基于源代码分析的软件聚类方法,通过对源代码中实体和关系图产生子系统分解,最后通过爬山优化搜索算法产生软件聚类。Levcovitz等人于2016年提出一种基于人工经验功能分解的微服务提取方法,通过评估代码、数据库表和SQL查询确定业务对象及关系,结合调用图推荐可能微服务。Gysel等人于2016年引入了一个基于16个耦合标准的服务分解工具,这些标准来源于行业和文献,通过创建加权图,利用聚类算法提取潜在的微服务。Mazlami等人于2017年提出一种通过抽取代码中的依赖结构信息,将软件系统建模表示成图结构,然后设计图切割算法切分图,每个子图则对应一个可选的服务提取方法。Chen等人于2017年使用自顶向下的方法,基于数据流进行候选服务提取,该方法需要根据用户对系统的自然语言描述构造一个数据流图来说明详细的数据。Baresi等人于2017年提出了一种基于OpenApi规范规定的功能语义相似性的方法,这种相似性取决于引用词汇表。Tyszberowicz等人于2018年提出了一种基于软件需求的功能分解的方法提取微服务的方法。然后使用图形分析工具确定微服务,Amiri等人于2018年提出一种使用业务流程模型和符号(BPMN)从业务流程中表示微服务的方法,实现过程中对结构依赖和数据对象依赖信息使用简单的矩阵加法进行了聚合。Munezero等人于2018年提出一种使用领域驱动设计(DDD)模式提取微服务的方法,需领域专家确定系统每个功能的责任边界。Jin等人于2019年以单体应用系统的运行轨迹作为输入,提出了一种基于多目标优化技术的服务拆分方法。Ding等人于2020年提出了一种场景驱动、自底向上的融合用户反馈的半自动化拆分方法,该方法从数据库拆分出发,向上映射到系统代码的拆分,考虑了数据关联度、数据共享度、拆分开销等不同

拆分因素。Zhang等人于2020年提出一种围绕负载均衡作为优化目标的基于运行轨迹的多目标优化的微服务提取方法。Brito等人于2021年提出一种基于主题建模的微服务提取方法,利用领域术语与聚类技术相结合生成一组基于原始单体应用软件的服务。Desai等人于2021年提出了基于深度学习图神经网络方法的微服务提取方法,是一种基于端到端的深度神经网络聚类联合框架,取得了较好的实验结果,并拓展了图神经网络的应用。Kalia于2021年提出了一种基于单体系统的运行轨迹中抽取出系统业务逻辑,使用分层聚类方法将单体应用划分为不同微服务的方法。

[0004] 综上所述,近几年伴随人工智能技术的发展,微服务提取工作取得了很大的进展,尽管进行了大量的研究,但仍没有单一的提取方法可以推荐于普遍应用中。另外,大多数方法存在只适用于一个数据集的情况,在多个数据集上缺乏可扩展性,或者缺少自动化实现的方法。

## 发明内容

[0005] 本发明针对现有技术中的不足,提供一种多通道注意力图神经网络聚类的单体程序拆分方法及系统,通过注意力机制,实现多通道信息的有效融合,提升多通道特征的代表能力。

[0006] 为实现上述目的,本发明采用以下技术方案:

[0007] 多通道注意力图神经网络聚类的单体程序拆分方法,其特征在于,包括:

[0008] 步骤S1:对JAVA单体程序中获取的类和类之间的多源关系建模并组合形成多通道信息数据对,基于多通道信息数据对和邻接矩阵构建单体程序多属性图;

[0009] 步骤S2:将单体程序多属性图输入到多通道图卷积编码器获得对应的特征嵌入表示,利用多通道图卷积解码器重构多通道信息数据对并基于重构误差设计重构损失函数;

[0010] 步骤S3:通过注意力机制计算特征嵌入表示中各结点的重要性,并自适应融合为新的特征嵌入表示;

[0011] 步骤S4:通过图神经网络和聚类的联合学习框架对新的特征嵌入表示进行聚类,实现微服务的划分,并基于重构损失函数和聚类过程中产生的聚类损失函数,对神经网络的参数进行优化。

[0012] 为优化上述技术方案,采取的具体措施还包括:

[0013] 进一步地,所述步骤S1具体包括如下子步骤:

[0014] 步骤S11:通过对JAVA应用程序的动态分析,获得运行时调用链追踪数据信息;

[0015] 步骤S12:根据调用链追踪数据信息,得到类与调用链之间的关系子图 $X_{class-trace}$ ;如果类 $i$   $class_i$ 出现在入口点 $p$ 所在的执行追踪中,则 $X_{class-trace}(i, p) = 1$ ,否则 $X_{class-trace}(i, p) = 0$ ;

[0016] 步骤S13:根据调用链追踪数据信息,得到类之间的矩阵属性子图 $X_{class-occurrence}$ ;若有 $q$ 条调用链同时包含类 $i$ 与类 $j$ ,则 $X_{class-occurrence}(i, j) = q$ ;

[0017] 步骤S14:根据调用链追踪数据信息,得到类之间的继承关系子图 $X_{class-inheri}$ ;如果两个类通过继承关系相关,则 $X_{class-inheri}(i, j) = 1$ ,否则 $X_{class-inheri}(i, j) = 0$ ;

[0018] 步骤S15:根据调用链追踪数据信息,得到类之间的邻接矩阵 $A$ ;邻接矩阵定义如下:

$$[0019] \quad A_{uv} = \begin{cases} 1, & \text{if } (v_a, v_b) \in E \\ 0 & \text{otherwise} \end{cases}$$

[0020] 其中,  $E$ 表示继承关系子图的边集合,  $v_a$ 、 $v_b$ 分别表示结点 $a$ 和结点 $b$ ,  $A_{uv}$ 表示邻接矩阵 $A$ 的第 $u$ 行、第 $v$ 列的值;

[0021] 步骤S16:将关系子图 $X_{class-trace}$ 、矩阵属性子图 $X_{class-occurrence}$ 和继承关系子图 $X_{class-inheri}$ 进行组合形成多通道信息数据对 $X_1$ 、 $X_2$ 和 $X_3$ ,其中,  $X_1 = (X_{class-trace}, X_{class-occurrence})$ ,  $X_2 = (X_{class-trace}, X_{class-inheri})$ ,  $X_3 = (X_{class-occurrence}, X_{class-inheri})$ ,结合邻接矩阵 $A$ 构建单体程序多属性图。

[0022] 进一步地,所述步骤S2具体包括如下子步骤:

[0023] 步骤S21:将步骤S1中得到的单体程序多属性图输入多通道图卷积编码器,学习得到三个通道的特征嵌入表示,模型描述如下:

$$[0024] \quad f_m(G, X_m; \theta) \rightarrow Z_m$$

[0025] 其中,  $f_m$ 为第 $m$ 个通道信息数据与特征嵌入表示的映射关系,  $G$ 为单体程序多属性图,  $X_m$ 为第 $m$ 个通道信息数据对,  $\theta$ 为图神经网络的参数,  $Z_m$ 表示学习得到的特征嵌入表示;

[0026] 第 $m$ 个图卷积编码器的第 $l$ 层的输出  $Z_m^{(l)}$ 为:

$$[0027] \quad Z_m^{(l)} = \sigma(\tilde{D}_m^{-\frac{1}{2}} \tilde{A} \tilde{D}_m^{-\frac{1}{2}} Z_m^{(l-1)} W_m^l)$$

[0028] 其中,  $\tilde{A} = A + I_N$ 是具有自连接的邻接矩阵,  $I_N$ 是单位矩阵;  $\tilde{D}_m$ 是  $\tilde{A}$ 对应图的度矩阵,  $W_m^l$ 是第 $m$ 个图卷积编码器的权值参数,  $Z_m^{(l-1)}$ 表示第 $m$ 个图卷积编码器第 $l-1$ 层的输出,  $\sigma$ 表示非线性激活函数;

[0029] 步骤S22:将步骤S21中得到的特征嵌入表示输入多通道图卷积解码器,重构多通道信息数据对;第 $m$ 个图卷积解码器的第 $l$ 层的输出  $\hat{Z}_m^l$ 为:

$$[0030] \quad \hat{Z}_m^l = \sigma(\hat{D}_m^{-\frac{1}{2}} \bar{A} \hat{D}_m^{-\frac{1}{2}} \hat{Z}_m^{l-1} \hat{W}^{l-1})$$

[0031] 其中,  $\bar{A} = 2I_n - A$ ,  $\hat{D}_m = 2I_n + D$ ,  $I_n$ 是单位矩阵,  $D$ 是 $A$ 对应图的度矩阵,  $\hat{W}^{l-1}$ 是第 $m$ 个图卷积解码器的权值参数;  $\hat{Z}_m^{l-1}$ 表示第 $m$ 个图卷积解码器第 $l-1$ 层的输出;第 $m$ 个图卷积解码器最后一层的输出就是重构后的第 $m$ 个通道信息数据对  $\hat{X}_m$ ;

[0032] 重构 $m$ 通道视图邻接矩阵  $\hat{A}_m$ :

$$[0033] \quad \hat{A}_m = \sigma(Z_m Z_m^T)$$

[0034] 基于重构误差构建每个通道图神经网络的重构损失,重构损失函数  $L_{sa}$  定义为:



$$[0035] \quad L_{sa} = \min_{\theta_i} \sum_{m=1}^M \|X_m - \hat{X}_m\|_F^2 + \lambda \sum_{m=1}^M \|A - \hat{A}_m\|_F^2$$

[0036] 其中,  $\theta_i$  是图神经网络的参数,  $\lambda$  是超参数,  $M$  是通道数目。

[0037] 进一步地, 所述步骤S3具体如下:

[0038] 通过注意力机制先对每个通道的特征嵌入表示进行非线性变换, 再用一个共享注意力向量  $p_m \in R^{h \times 1}$  得到  $m$  通道结点  $e$  的注意力值  $w_m^e$ :

$$[0039] \quad w_m^e = p_m^T \cdot \tanh(W_m \cdot (z_m^e)^T + b_m)$$

[0040] 其中,  $W_m \in R^{h \times h}$  是权值矩阵,  $h'$  表示共享注意力向量的维度,  $h$  表示注意力机制的通道数;  $z_m^e \in R^{1 \times h}$  表示  $m$  通道  $e$  结点的特征嵌入表示,  $b_m \in R^{h \times 1}$  是偏置向量; 对  $w_m^e$  进行归一化:

$$[0041] \quad \gamma_m^e = \text{soft max}(w_m^e) = \frac{\exp(w_m^e)}{\sum_{m=1}^M \exp(w_m^e)}$$

[0042] 其中,  $\gamma_m^e$  值表示  $m$  通道  $e$  结点的特征嵌入表示的重要性, 获得学习后的权值  $\gamma_m = [\gamma_m^e] \in R^{n \times 1}$ ,  $n$  表示嵌入特征维度; 将三个通道的特征嵌入表示  $Z_m$  进行融合获得新的特征嵌入表示  $Z$ :

$$[0043] \quad Z = \sum_{m=1}^M \beta_m \cdot Z_m$$

[0044] 其中,  $\beta_m = \text{diag}[\gamma_m]$ 。

[0045] 进一步地, 所述步骤S4具体如下:

[0046] 采用谱聚类算法对特征嵌入表示  $Z$  进行聚类, 使用高斯核定义仿射矩阵  $S$ , 如果  $e \neq o$ , 则

$$[0047] \quad S_{eo} = \exp\left(\frac{\|z_e - z_o\|_2^2}{\Omega}\right), \quad e, o = 1, 2, \dots, N$$

[0048] 否则  $S_{eo} = 0$ ;  $S_{eo}$  代表结点  $e$  和结点  $o$  通过高斯核函数计算得到的相似度,  $N$  表示结点数,  $z_e$  表示结点  $e$  的特征嵌入表示,  $\|\cdot\|$  是向量  $z_e$  和  $z_o$  之间的欧几里得距离,  $\Omega$  是尺度参数,  $S_{eo}$  对称且非负; 然后, 运行谱聚类算法将  $N$  个结点聚类到  $K$  个不同划分中;

[0049] 谱聚类后, 构建簇的  $k$ - $d$  树, 找到与结点  $z_e$  同一个簇中曼哈顿距离最近的邻居结点  $z_o$ , 定义聚类损失函数  $L_{clus}$ :

$$[0050] \quad L_{clus} = \sum_e \min_0 d(z_e, z_o), \quad e, o = 1, 2, \dots, u$$

[0051] 其中,  $d(z_e, z_o)$  表示结点  $e$  和结点  $o$  之间的曼哈顿距离  $|z_e - z_o|$ ;

[0052] 综合  $L_{sa}$  和  $L_{clus}$  形成新的损失函数  $L$ :

$$[0053] \quad L = \sum_{m=1}^M L_{sa}^m + \gamma L_{clus}$$

[0054] 其中,  $\gamma > 0$  是超参数;通过最小化损失函数  $L$  训练并优化图神经网络的参数。

[0055] 进一步地,本发明提出了一种多通道注意力图神经网络聚类的单体程序拆分系统,其特征在于,包括:多通道注意力网络模块、多通道图卷积神经网络模块、注意力机制模块和联合学习模块;

[0056] 所述多通道注意力网络模块对JAVA单体程序中获取的类和类之间的多源关系建模并组合形成多通道信息数据对,基于多通道信息数据对和邻接矩阵构建单体程序多属性图;

[0057] 所述多通道图卷积神经网络模块包括多通道图卷积编码器和多通道图卷积解码器,所述单体程序多属性图输入到多通道图卷积编码器获得对应的特征嵌入表示;所述多通道图卷积解码器重构多通道信息数据对并基于重构误差设计重构损失函数;

[0058] 所述注意力机制模块通过注意力机制计算特征嵌入表示中各结点的重要性,并自适应融合为新的特征嵌入表示;

[0059] 所述联合学习模块通过图神经网络和聚类的联合学习框架对新的特征嵌入表示进行聚类,实现微服务的划分,并基于重构损失函数和聚类过程中产生的聚类损失函数,对神经网络的参数进行优化。

[0060] 进一步地,本发明提出了一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序使计算机执行如上所述的多通道注意力图神经网络聚类的单体程序拆分方法。

[0061] 进一步地,本发明提出了一种电子设备,其特征在于,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行计算机程序时,实现如上所述的多通道注意力图神经网络聚类的单体程序拆分方法。

[0062] 本发明的有益效果是:本发明在Java程序运行时调用图分析,获取类与类之间的多属性信息,构建多属性图结构,设计基于多通道的图卷积神经网络模型对多属性图进行特征嵌入表示学习;然后以优化嵌入表示为目标,采用注意力机制,对多通道卷积网络嵌入表示进行融合,形成新的特征嵌入表示;再通过与聚类信息的联合训练,最后使用谱聚类进行聚类,得到微服务拆分结果。本发明充分利用多属性信息和多通道图注意力神经网络的建模能力,有效提升拆分后微服务的功能性和模块性。

## 附图说明

[0063] 图1为本发明多通道注意力图神经网络聚类的单体程序拆分方法的流程图。

[0064] 图2为本发明实施例提供的Java程序类调用示例图。

[0065] 图3为本发明实施例提供的微服务拆分示意图。

[0066] 图4为本发明多通道注意力图神经网络聚类的单体程序拆分方法的框架图。

[0067] 图5a-5d为本发明实施例提供的不同数据集上三种通道的图结点注意力系数热图,其中图5a是daytrader数据集,图5b是acmeAir数据集,图5c是jpetstore数据集,图5d是plants数据集。

[0068] 图6a-6c为本发明实施例提供的不同数据集嵌入维度对不同指标性能的影响结果图,其中图6a是ICP指标,图6b是SM指标,图6c是IFN指标。

## 具体实施方式

[0069] 现在结合附图对本发明作进一步详细的说明。

[0070] 本发明提出了一种多通道注意力图神经网络聚类的单体程序拆分系统,包括:

[0071] 多通道注意力网络模块,用于对JAVA单体程序中获取的类和类之间的多源关系建模并组合形成多通道信息数据对,基于多通道信息数据对和邻接矩阵构建单体程序多属性图;

[0072] 多通道图卷积神经网络模块,包括多通道图卷积编码器和多通道图卷积解码器,单体程序多属性图输入到多通道图卷积编码器获得对应的特征嵌入表示,多通道图卷积解码器重构多通道信息数据对并基于重构误差设计重构损失函数;

[0073] 注意力机制模块,用于通过注意力机制计算特征嵌入表示中各结点的重要性,并自适应融合为新的特征嵌入表示;

[0074] 联合学习模块,用于通过图神经网络和聚类的联合学习框架对新的特征嵌入表示进行聚类,实现微服务的划分,并基于重构损失函数和聚类过程中产生的聚类损失函数,对神经网络的参数进行优化。

[0075] 如图1所示,本发明公开了一种多通道注意力图神经网络聚类的单体程序拆分方法,依次包括单体程序多属性图构建、多通道图神经网络特征嵌入表示学习、基于注意力的多通道特征嵌入融合和基于谱聚类的微服务拆分。首先,以Java单体应用为基础,从中获取类和类之间的特征信息构建成多属性图 $G = \{V, E, X_m\}$ ;然后构建基于多通道图卷积神经网络的特征嵌入表示学习模型,每个通道得到特征嵌入表示 $Z_m$ ,每个通道的图卷积神经网络由编码器和解码器组成;再基于结点重要性的注意力融合机制将 $Z_m$ 融合得到新的特征嵌入表示 $Z$ ;最后通过联合学习框架不断优化参数,使用谱聚类算法进行聚类,得到拆分后的微服务。

[0076] 如图2所示,在JAVA应用程序中,程序都是由类组成的,类和类之间的调用关系可以表示为类之间的调用图形式,以图的结构来表示。因此,将程序转换为图结构进行建模是许多微服务提取模型采用的方法,通常通过静态代码分析或者运行时追踪等工具可以获得类和类之间的不同关系,本发明将该问题建模为一个多通道图神经网络和聚类的联合学习框架。将单体应用程序根据代码中类之间的相似性进行划分,例如图3中的6个类经过微服务提取后被划分到3个微服务中。如图4所示,本发明的方法以两个类是否同时出现在同一条调用链中、同时出现在某一条调用链中的次数、是否存在继承依赖关系等多属性出发,以充分利用这些关系构建更加高效的微服务提取方法。

[0077] S1,单体程序多属性图构建。

[0078] 通过对JAVA应用程序动态分析,可获得运行时追踪数据。一条完整的调用链追踪数据信息Trace = (序号,业务用例, $class_i$ ,调用(或返回到), $class_j$ ,调用所在文件和方

法)。每个业务用例包含多条调用链,每条调用链是从业务用例测试的入口点(entrypoint)类开始的,因此根据调用链信息,可获得类与类之间的调用图,通过跟踪在入口点的执行跟踪期间调用的所有类,得到类与某条调用链的关系子图 $X_{class-trace}$ 。如果类 $i$ 出现在入口点 $p$ 所在的执行追踪中,则 $X_{class-trace}(i, p) = 1$ ,否则 $X_{class-trace}(i, p) = 0$ 。进一步计算某个开始点开始的调用链同时包含 $class_i$ 和 $class_j$ 的次数,获得矩阵属性子图 $X_{class-occurrence}$ 。定义继承关系子图 $X_{class-inheri}$ ,如果两个类通过继承关系相关,则 $X_{class-inheri}(i, j) = 1$ ,否则 $X_{class-inheri}(i, j) = 0$ 。为了能够捕捉到应用程序中类和类之间的全面深度关系,对类和类之间的关系进行新的组合形成多通道信息数据对,即 $X_1 = (X_{class-trace}, X_{class-occurrence})$ ,  $X_2 = (X_{class-trace}, X_{class-inheri})$ ,  $X_3 = (X_{class-occurrence}, X_{class-inheri})$ 。使用 $A$ 表示该图的邻接矩阵,定义为 $A_{uv} = \begin{cases} 1, & \text{if } (v_a, v_b) \in E \\ 0 & \text{otherwise} \end{cases}$ 。

[0079] 因为本问题中类之间的调用图固定,因此三个通道属性图具有相同的邻接矩阵 $A$ 。下一步,将 $(X_1, A)$ ,  $(X_2, A)$ 和 $(X_3, A)$ 分别输入到多通道图卷积编码器中。

[0080] S2,多通道图神经网络特征嵌入表示学习。

[0081] 1)多通道图卷积编码器。本发明中,每一个通道都使用了图卷积网络(GCN)。GCN将卷积运算扩展到谱域的图形数据中,其卷积层用于学习图中每个结点的特征嵌入表示和从邻居结点聚合信息,将特征映射到嵌入特征空间。

[0082] 多通道注意力图神经网络聚类的单体程序拆分方法对多通道属性图构建了三个图卷积编码器,通过三通道GCN编码器将多属性矩阵和图结构映射到新的特征嵌入空间中。具体来说,对于 $m(m=1, 2, \dots, M)$ 通道(本实施例取 $M=3$ ),则GCN模型的函数为将图 $G$ 和 $X_m$ 映射到 $d$ 维嵌入特征 $Z_m$ 。则第 $m$ 个图卷积编码器的第 $l$ 层的输出 $Z_m^{(l)}$ 为:

$$[0083] \quad Z_m^{(l)} = \sigma(\tilde{D}_m^{-\frac{1}{2}} \tilde{A} \tilde{D}_m^{-\frac{1}{2}} Z_m^{(l-1)} W_m^l)$$

[0084] 其中, $\tilde{A} = A + I_N$ 是具有自连接的邻接矩阵, $I_N$ 是单位矩阵; $\tilde{D}_m$ 是 $\tilde{A}$ 对应图的度矩阵, $D_{ii} = \sum_j A_{ij}$ , $W_m^l$ 是第 $m$ 个图卷积编码器的权值参数, $\sigma$ 表示非线性激活函数。对于 $Z_m^{(l)}$ ,当 $l=0$ 时, $f_m(G, X_m; \theta) \rightarrow Z_m$ , $Z_m^{(0)} = X_m$ ,当 $l=L$ 时, $H_m^{(l)}$ 是 $H_m$ 的第 $l$ 个隐含层的输入激活矩阵。如选取2层GCN模型,可表示为:

$$[0085] \quad Z_m = f(X_m, A) = \sigma(\hat{A} \text{ReLU}(\hat{A} X_m W_m^0) W_m^1)$$

[0086] 其中 $W_m^{(0)} \in R^{d \times H}$ 是输入层到具有 $H$ 个特征映射的隐含层权值矩阵, $W_m^{(1)} \in R^{H \times F}$ 是隐含层到输出层的权值矩阵。 $X \in R^{N \times d}$ , $N$ 是结点个数, $d$ 是属性特征维度。 $\sigma$ 是激活函数,由此得到不同通道嵌入表示 $Z_m$ 。

[0087] 2)多通道图卷积解码器。为了能够提高每个通道GCN编码器特征表示的有效性,增加了GCN解码器模块。重建多通道输入图的属性图 $X_m$ 和结构图 $A_m$ 。每个解码器模块,与编码器

层数相同,结构恰好相反。经过多通道GCN编码器模块后,已获得图嵌入表示 $Z_m$ 。在每个解码器端假定 $\hat{X}_i$ 表示 $m$ 通道的解码器重建属性图, $\hat{A}$ 代表 $m$ 通道的解码器重建结构图。第 $m$ 通道第 $l$ 层的属性特征重建可表示为:

$$[0088] \quad \hat{Z}_m^l = \sigma(\hat{D}^{-\frac{1}{2}} \bar{A} \hat{D}^{-\frac{1}{2}} \hat{Z}_m^{l-1} \hat{W}^{l-1})$$

[0089] 其中 $\bar{A} = 2I_n - A$ , $\hat{D} = 2I_n + D$ ,经过 $L$ 层解码器,多通道结点属性矩阵 $\hat{X}_m = H_m^0$ 。对于视图邻接矩阵的重建:

$$[0090] \quad \hat{A}_m = \sigma(Z_m Z_m^T)$$

[0091] 使用重建图和原始图的误差继续优化网络,因此每个通道GCN网络的重建损失综合属性图重构误差和结构图重构误差,该部分损失函数定义为:

$$[0092] \quad L_{sa} = \min_{\theta_i} \sum_{m=1}^M \|X_m - \hat{X}_m\|_F^2 + \lambda \sum_{m=1}^M \|A - \hat{A}_m\|_F^2$$

[0093] S3,基于注意力的多通道特征嵌入融合。

[0094] 使用注意力机制计算不同通道GCN的嵌入表示 $Z_m$ 中不同结点的重要性 $\beta_m$ ,经过融合得到新的嵌入特征表示 $Z$ 。以结点 $e$ 为例,其在 $m$ 通道的嵌入表示 $Z_m$ ,元素为 $z_m^e \in R^{1 \times h}$ , $m = 1, 2, \dots, M$ 。通过注意力机制先对每个通道的特征嵌入表示进行非线性变换,再用一个共享注意力向量 $p_m \in R^{h' \times 1}$ 得到 $m$ 通道结点 $e$ 的注意力值 $w_m^e$ :

$$[0095] \quad w_m^e = p_m^T \cdot \tanh(W_m \cdot (z_m^e)^T + b_m)$$

[0096] 其中 $W_m \in R^{h' \times h}$ 是权值矩阵, $b_m \in R^{h' \times 1}$ 是偏置向量。对注意力值 $w_m^e$ 进行归一化:

$$[0097] \quad \gamma_m^e = \text{soft max}(w_m^e) = \frac{\exp(w_m^e)}{\sum_{m=1}^M \exp(w_m^e)}$$

[0098] 较大 $\gamma_m^e$ 值表示 $m$ 通道 $e$ 结点的嵌入特征越重要。类似地,对于所有的结点,可获得学习后的权值 $\gamma_m = [\gamma_m^e] \in R^{n \times 1}$ ,进一步地, $\beta_m = \text{diag}[\gamma_m]$ 。将三个通道嵌入表示 $Z_m$ 进行融合可获得最终的嵌入特征表示 $Z$ :

$$[0099] \quad Z = \sum_{m=1}^M \beta_m \cdot Z_m$$

[0100] S4,基于谱聚类的微服务拆分。

[0101] 经过多通道卷积图神经网络和注意力机制融合后,得到了新的嵌入特征表示 $Z$ ,正

常情况下,可采用分步骤的方式,只需要运行聚类方法进行划分即可,但考虑机器学习领域基于多通道注意力图神经网络聚类的单体程序微服务拆分方法的思路,在嵌入表示融合层上执行一个谱聚类层,定义损失函数,再联合优化整个多通道图神经网络模型参数。

[0102] 谱聚类是基于图论的一种典型聚类算法,减少了对样本空间形状的要求,可以处理非凸数据集,同时能有效克服集中经典聚类算法收敛于局部最优的缺点,因此本实施例此处谱聚类算法对嵌入表示 $Z$ 进行聚类。为了运行谱聚类,使用高斯核定义仿射矩阵 $S$ ,如果

$e \neq o$ , 则  $S_{eo} = \exp(\frac{\|z_e - z_o\|_2^2}{\Omega})$ ,  $e, o = 1, 2, \dots, N$ ; 否则  $S_{eo} = 0$ 。其中  $\|\cdot\|$  是向量  $z_e$  和  $z_o$  之间的

的欧几里得距离,  $\Omega$  是尺度参数,  $S_{eo}$  对称且非负; 然后, 运行谱聚类算法将  $N$  个结点聚类到  $K$  个不同划分中。

[0103] 为了能够获得更具有判别性的嵌入表示和最优聚类结果, 谱聚类后, 构建簇的  $k-d$  树, 借助  $k-d$  树实现最近邻的快速查找, 找到与结点  $z_e$  同一个簇中曼哈顿距离最近的邻居结点  $z_o$ 。定义聚类损失函数:

$$[0104] \quad L_{clus} = \sum_e \min_o d(z_e, z_o), \quad e, o = 1, 2, \dots, u$$

[0105] 其中,  $d(z_e, z_o)$  表示结点  $e$  和结点  $o$  之间的曼哈顿距离  $|z_e - z_o|$ 。

[0106] 本发明首先通过多通道图神经网络对多属性图进行建模, 通过注意力机制对特征嵌入表示进行融合, 再使用谱聚类算法进行微服务拆分, 训练过程中首先对损失函数进行了改进。根据网络模型结构, 其损失函数由多通道图卷积网络的重构损失和聚类损失组成。定义为:

$$[0107] \quad L = \sum_{m=1}^M L_{sa}^m + \gamma L_{clus}$$

[0108] 其中,  $\gamma > 0$  是超参数。通过最小化该损失函数训练整个多通道注意力图卷积神经网络的参数, 得到较高质量嵌入质量表示和聚类效果。

[0109] 实验过程由三个步骤组成。第一步是对多通道图卷积神经网络模块网络参数进行预训练。第二步是对整个网络进行正式训练, 得到最终的参数。第三步是使用学习到的嵌入表示  $Z$  进行聚类。

[0110] 1) 预训练优化网络参数。构建图临接矩阵  $A$ , 多源属性图  $X_m$  输入到  $m$  个图卷积编码器中获得嵌入表示  $Z_m$ 。多通道图卷积解码器获得属性重建  $\hat{X}_m$  和结构重建  $\hat{A}_m$ , 使用重建误差优化每个通道图卷积网络参数。将训练得到的卷积自编码器的参数作为网络的初始参数, 计算注意力融合嵌入特征表示  $Z$ , 使用谱聚类方法初始化聚类, 完成初始化。

[0111] 2) 正式开始训练过程。对图卷积编码器、解码器和注意力层参数进行迭代优化, 得到最终的参数。首先将多通道图形数据和参数输入到卷积自编码器网络。经过模型训练, 获得隐含层嵌入表示和注意力层嵌入表示, 使用注意力层嵌入表示数据谱聚类, 得到划分结果, 计算总体损失函数。通过最小化总体损失函数反向传播训练整体网络。

[0112] 3) 通过学习到的嵌入表示  $Z$  计算非负对称矩阵  $S$ , 执行谱聚类获得最终聚类结果。

[0113] 到此,该发明的训练过程和微服务提取已经计算完成。实施在一台运行Windows 10(64位)、配备NVIDIA GeForce GTX 1660 Ti图形处理单元(GPU)和16gb内存的服务器上进行了所有的实验。使用PyTorch和Python实现,在训练过程中,使用了Adam优化器。为了评估本发明,使用Mono2Mirco提供的4个公开数据集进行测试。daytrader是从在线股票交易系统的运行追踪数据中抽取了含有28个类和103个方法的样本数据。acmeAir是从航空公司订票应用系统的运行追踪数据中抽取了含有73个类、428个方法的样本数据。jpetstore是从宠物上天系统的运行追踪数据中抽取了含有25个类和264个方法的样本数据。该项目是评估微服务提取方法中最常用的基线。如表1所示,plants是从购买植物和宠物的网店运行追踪数据中抽取了含有73个类和428个方法的样本数据。这些应用都是使用各种Java EE框架(如Spring、Apache、Struts等)构建的具有代表性的web应用。

[0114] 表1 各数据集

数据集	#类	#方法	类覆盖率	方法覆盖率	说明
daytrader	33	163	28 (84%)	108 (66%)	在线股票交易系统
plants	109	969	73 (66%)	428 (44%)	购买植物和宠物的网店
acmeAir	66	350	36 (54%)	236 (67%)	航空公司订票应用程序
jpetstore	37	463	25 (67%)	264 (57%)	宠物商店

[0116] 将本发明机基准方法Bunch、MEME、FoSCI和Mono2Micro进行了性能比较。对于所有数据集,参数设置如下:预训练次数都是 $preepochs=350$ ,正式训练次数 $epochs=300$ ,学习率 $\eta=0.01$ ,聚类的数目 $k \in \{2,3,4,5,6\}$ 。每个GCN编码器维度为 $hidden1, hidden2 = \{x | x \in (8,64), x \in N^*\}$ ,  $dropout \in (0.2,0.5)$ ,结构损失权重 $\lambda \in (0,1)$ ,  $\gamma \in (0,1)$ 。

[0117] 性能评价主要从功能独立性、模块性和交互复杂性三个角度进行。

[0118] (1)功能独立性。微服务的功能独立性是指服务提供单一功能职责的程度。一个微服务的功能是单一且一致的,则该服务在应用到新系统上则易于被新系统复用,同时能提高新系统的开发效率。

[0119] (2)模块性。模块性是指微服务能遵循高内聚、低耦合的程度。微服务之间的耦合程度越低、微服务内部聚合程度越高,则说明该微服务的模块性越好。因此,模块化主要用于评估微服务服务内部交互的一致性以及服务之间的松散耦合性(loosely coupled)。

[0120] (3)交互复杂性。服务的交互复杂性是指服务间动态交互行为的复杂程度。服务间的交互行为复杂,则发生故障后,很难有效在源代码中定位故障源头。

[0121] 为了较好地量化评估基于多通道注意力图神经网络聚类的单体程序微服务拆分方法模型的微服务提取性能,本实施例采用了微服务提取中较通用的满足上述要求的评价指标。评价指标主要有3个:

[0122] (1)IFN(independence of functionality)。IFN强调外部依赖,评价微服务是否具有良好定义(well-defined)和独立性(independence)。ifn(interface number)用于量化一个给定服务的接口数目。它基于单一责任原则(single responsibility principle)。IFN表示所有ifn的平均值,可表示为:

$$[0123] \quad IFN = \frac{1}{N} \sum_{i=1}^N ifn_i, \quad ifn_i = |I_i|$$

[0124] 通常IFN值越小,表示给定服务越能承担单一责任。

[0125] (2) SM(structural modularity quality)用于量化微服务内的类间结构一致性,较高的SM值表示比较好的模块化服务,表示为:

$$[0126] \quad SM = \frac{1}{M} \sum_{i=1}^M scoh_i - \frac{1}{(M(M-1))/2} \sum_{i \neq j} scop_{i,j}$$

[0127] 其中  $m_i \in M$ ,  $scoh$ 、 $scop$ 表示量化服务之间的耦合,  $scoh_i = \frac{u_i}{m_i^2}$ ,

$scop_{i,j} = \frac{\sigma_{i,j}}{2 * (m_i * m_j)}$ 。 $u_i$ 表示当M个实体类之间发生结构调用依赖时一个服务内部边的

数目,与之类似, $\sigma_{ij}$ 指的是服务之间的依赖发生时边的数目。

[0128] (3) ICP(inter-partition call percentage)表示服务间调用纯度,用于最小化不同调用的数量和服务间调用的数目,评估微服务之间的交互百分比。可表示为:

$$[0129] \quad icp_{i,j} = \frac{c_{i,j}}{\sum_{i=1, j=1, i \neq j}^M c_{i,j}}$$

[0130] 其中 $c_{i,j}$ 表示不同微服务之间的调用数目。ICP值越小,说明服务之间调用纯度越好。

[0131] 主要目标是验证本发明所提取的微服务是否与功能和模块化的独立性有关,同时,测试结果的评价指标主要是SM、IFN和ICP。与其他方法的实验性能对比,结果如表2到表5所示。性能测试的结果是在参数设置选取的范围内跑了100次,然后取每个性能指标的中位数。其中 $\uparrow$ 表示值越大越好, $\downarrow$ 表示该值越小越好。在所有数据集上,本发明取得了非常具有前景的结果。

[0132] 表2 daytrader数据集上聚类分析结果

daytrader	M2M	FoSCI	MEM	Bunch	CoGCN	MAGEMP
ICP $\downarrow$	0.346	0.748	0.355	0.572	0.455	0.284
SM $\uparrow$	0.078	0.092	0.089	0.269	0.086	0.122
IFN $\downarrow$	1.922	3.489	4.200	-	2.880	4.113

[0134] 表3 acmeAir数据集上聚类分析结果

acmeAir	M2M	FoSCI	MEM	Bunch	CoGCN	MAGEMP
ICP $\downarrow$	0.527	0.706	0.589	0.55	0.444	0.301
SM $\uparrow$	0.072	0.095	0.097	0.177	0.038	0.062
IFN $\downarrow$	3.375	4.375	4.333	3.875	2.846	2.805



[0136] 表4 jpetstore数据集上聚类分析结果

	jpetstore	M2M	FoSCI	MEM	Bunch	CoGCN	MAGEMP
[0137]	ICP↓	0.333	0.478	0.434	0.477	0.582	0.263
	SM↑	0.054	0.044	0.124	-	0.091	0.116
	IFN↓	1.857	3.750	3.429	7.948	2.533	2.148

[0138] 表5 plants数据集上聚类分析结果

	plants	M2M	FoSCI	MEM	Bunch	CoGCN	MAGEMP
[0139]	ICP↓	0.381	0.682	0.320	0.501	0.571	0.608
	SM↑	0.078	0.135	0.210	0.155	0.133	0.138
	IFN↓	6.000	4.875	4.750	6.357	4.875	4.576

[0140] 从总体性能上比较,表2到表5列举了不同方法在四个数据集上的SM、IFN和ICP值。从表中可以看出在与同类方法CoGCN相比,在daytrader数据集上,ICP提升了37.6%,SM提升了41.86%。在acmeAir数据集上,ICP提升了32.21%,SM提升了63.16%,IFN提升了1.44%。在jpetstore数据集上,ICP提升了54.81%,SM提升了27.47%,IFN提升了15.2%。在plants数据集上,ICP提升了6.48%,SM提升了3.76%,IFN提升了6.13%。这表明基于多通道注意力图神经网络聚类的单体程序微服务拆分方法的多通道信息的充分利用和有效融合显著提升了提取微服务在功能性和模块性上的效果。

[0141] 值得提及的是,与当前最具优势的其他非深度学习模型比较,基于多通道注意力图神经网络聚类的单体程序微服务拆分方法在daytrader数据集的ICP指标、在acmeAir数据集上的ICP指标和IFN指标、在jpetstore数据集上的ICP指标、在plants数据集上的IFN指标上都取得了最好的实验效果。表明了基于多通道深度学习的注意力卷积网络具有更强的嵌入表示学习能力,进一步提升了聚类的效果。

[0142] 图5a到图5d给出了不同数据集上三种通道的图结点注意力系数热图。注意力融合机制是本模型的一个关键组成部分,它可以衡量和量化不同通道的重要性和不同通道中不同结点的重要性。图5a到图5d中可视化了在四个不同数据集上三个通道中每个图结点的注意力系数。从图中可以看出,三个通道具有不同的重要性,在不同数据集上,通道1和通道3的权值比通道2相对较高。实验结果表明基于多通道注意力图神经网络聚类的单体程序微服务拆分方法可以很好地利用不同通道的不同特征进行聚类。

[0143] 图6a到图6c给出了不同数据集嵌入维度对不同指标性能的影响结果图。总体趋势,随着特征嵌入维度增大,ICP先增大后减小,SM先减小后增大,IFN受特征嵌入维度变化的影响较小。对于jpetstore和daytrader数据集,特征嵌入维度为8或64时,总体性能较好,对于acmeAir数据集,特征嵌入维度16时,总体性能较好。对于plants数据集,特征嵌入维度可在32与64之间平衡选择。

[0144] 表6给出了本发明在jpegstore中提取的微服务结果,示出了基于Java Spring应用框架的jpetstore数据集在本发明基于多通道注意力图神经网络聚类的单体程序微服务拆分方法上的微服务提取结果。jpetstore是一宠物销售的购物应用程序,数据集中主要包含37个实体类,针对数据集中获取的多通道信息,经过基于多通道注意力图神经网络聚类

的单体程序微服务拆分方法之后,可以得到四个微服务。每个微服务中包含的实体类如表6所示。

[0145] 表6 jpegstore中提取的微服务结果

划分	微服务提取
[0146] SC <sub>0</sub>	jpetstore.domain.Account
	jpetstore.domain.LineItem
	jpetstore.domain.Sequence
	jpetstore.web.spring.AccountForm
	jpetstore.domain.logic.AccountValidator
	jpetstore.domain.logic.PetStoreImpl
	jpetstore.web.spring.UserSession
	jpetstore.dao.ibatis.maps.SqlMapAccountDao
	jpetstore.dao.ibatis.maps.SqlMapSequenceDao
	jpetstore.web.spring.AccountFormController
	jpetstore.web.spring.SignonController
	jpetstore.web.spring.SignonInterceptor
	jpetstore.web.spring.SignoffController
SC <sub>1</sub>	jpetstore.Domain.Category
	jpetstore.dao.ibatis.maps.SqlMapCategoryDao
	jpetstore.web.spring.ViewCategoryController
SC <sub>2</sub>	jpetstore.Domain.Product
	jpetstore.web.spring.ViewProductController
	jpetstore.web.spring.SearchProductsController
	jpetstore.dao.ibatis.maps.SqlMapProductDao jpetstore.dao.ibatis.maps.SqlMapProductDao::ProductSearch
SC <sub>3</sub>	jpetstore.Domain.Order
	jpetstore.web.spring.OrderForm
	jpetstore.domain.logic.OrderValidator
	jpetstore.dao.ibatis.maps.SqlMapOrderDao
	jpetstore.web.spring.OrderFormController
	jpetstore.web.spring.ListOrdersController jpetstore.web.spring.ViewOrderController
SC <sub>4</sub>	jpetstore.Domain.Cart
	jpetstore.Domain.CartItem
	jpetstore.Domain.Item
	jpetstore.dao.ibatis.maps.SqlMapItemDao
	jpetstore.web.spring.UpdateCartQuantitiesController
	jpetstore.web.spring.ViewItemController
	jpetstore.web.spring.ViewCartController
	jpetstore.web.spring.AddItemToCartController
jpetstore.web.spring.RemoveItemFromCartController	

[0147] 每行代表某个聚类中包含的类。在这个示例中可以看到,每个微服务中内类之间具有较高的相似性,聚类结果较均匀。最终结果中包含了用户、目录、产品、订单和购物车等5个微服务。

[0148] 综上,本发明针对单体应用程序微服务拆分问题,对面向对象程序设计中的类实体关系特征表示进行建模,提出了一种基于多通道注意力深度图卷积网络聚类的微服务拆分方法,考虑了不同类结点特征和拓扑结构信息的优势,达到了提高了模型性能的效果,再使用多通道图卷积神经网络对特征嵌入进行表示学习,通过注意力机制对嵌入特征的融合得到了更好的特征表示方式,最后通过谱聚类算法进行聚类得到微服务拆分结果。在4个通用数据集上进行了较广泛的验证实验,实验结果表明,多通道注意力图卷积神经网络聚类方法提升了微服务拆分的模块性和功能性的性能。

[0149] 此外,本发明还提出了一种计算机可读存储介质,存储有计算机程序,计算机程序使计算机执行如上所述的多通道注意力图神经网络聚类的单体程序拆分方法。

[0150] 此外,本发明还提出了一种电子设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行计算机程序时,实现如上所述的多通道注意力图神经网络聚类的单体程序拆分方法。

[0151] 以上仅是本发明的优选实施方式,本发明的保护范围并不仅限于上述实施例,凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理前提下的若干改进和润饰,应视为本发明的保护范围。

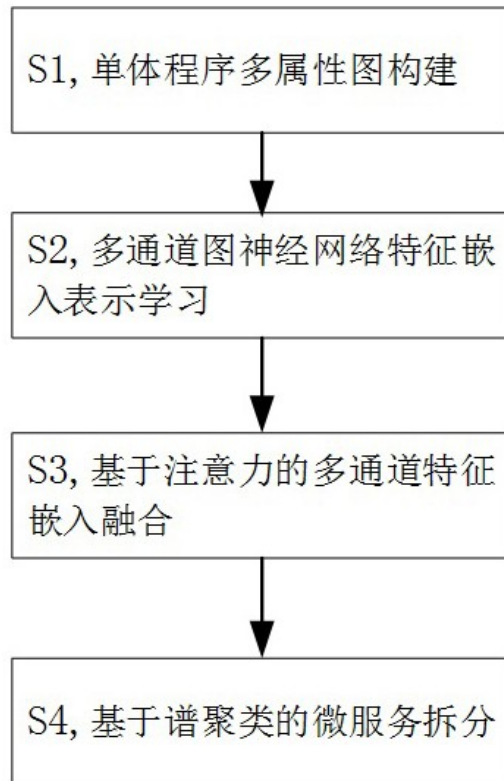


图1

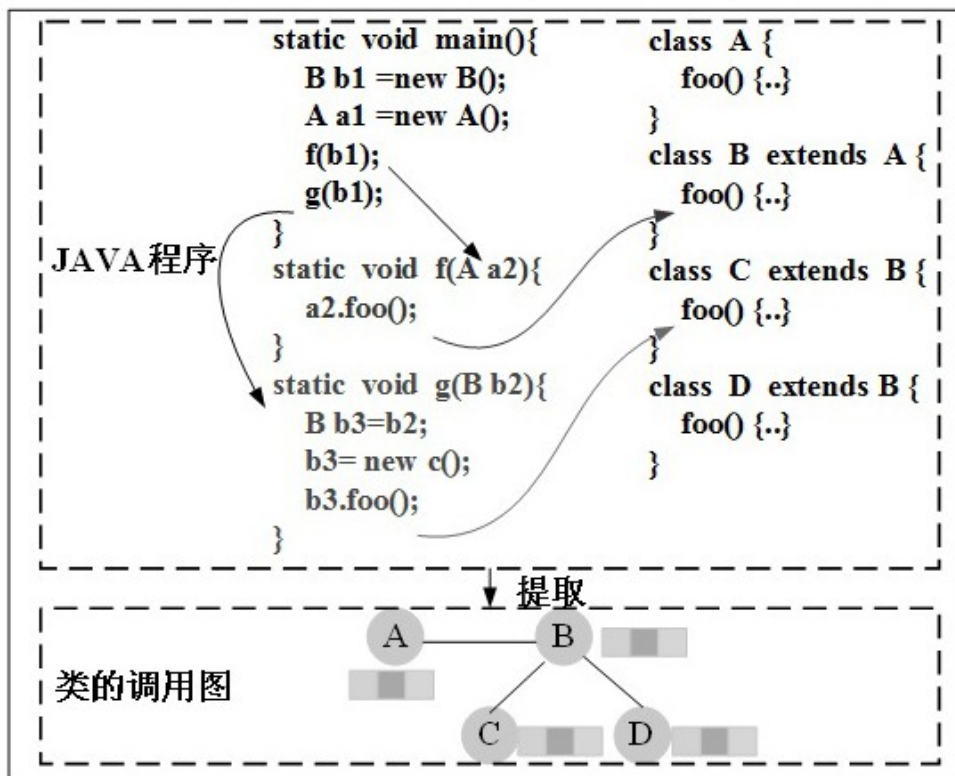


图2

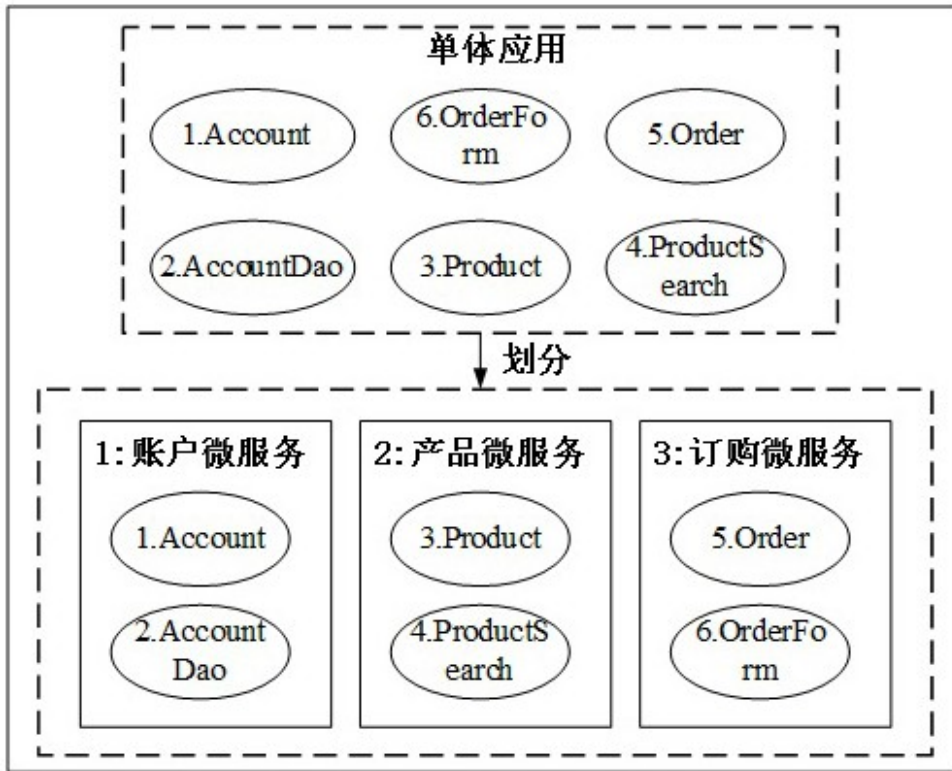


图3

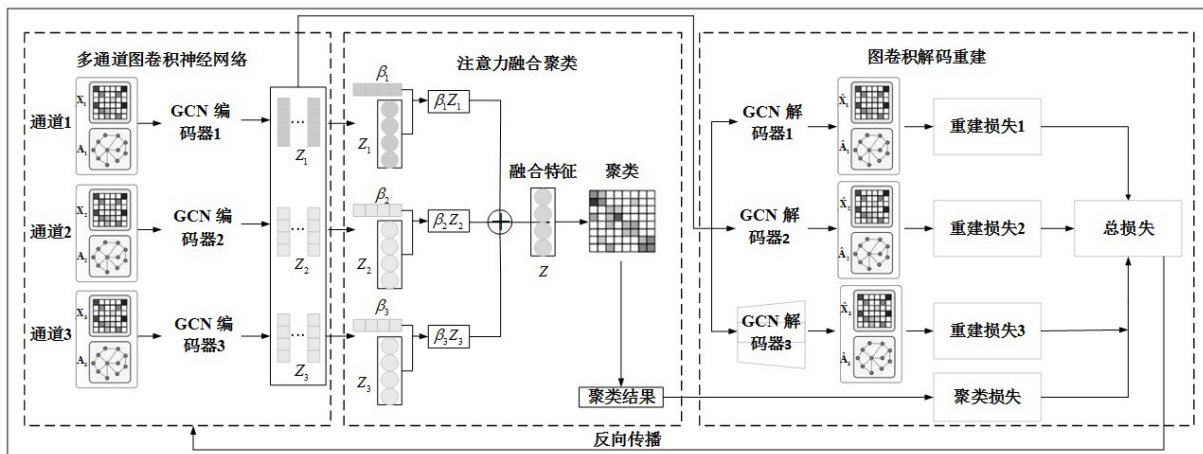


图4

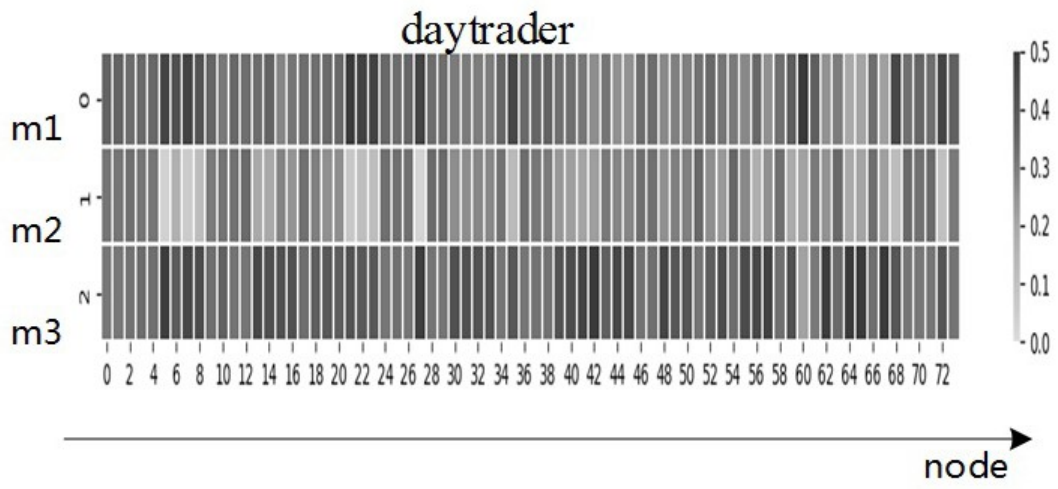


图5a

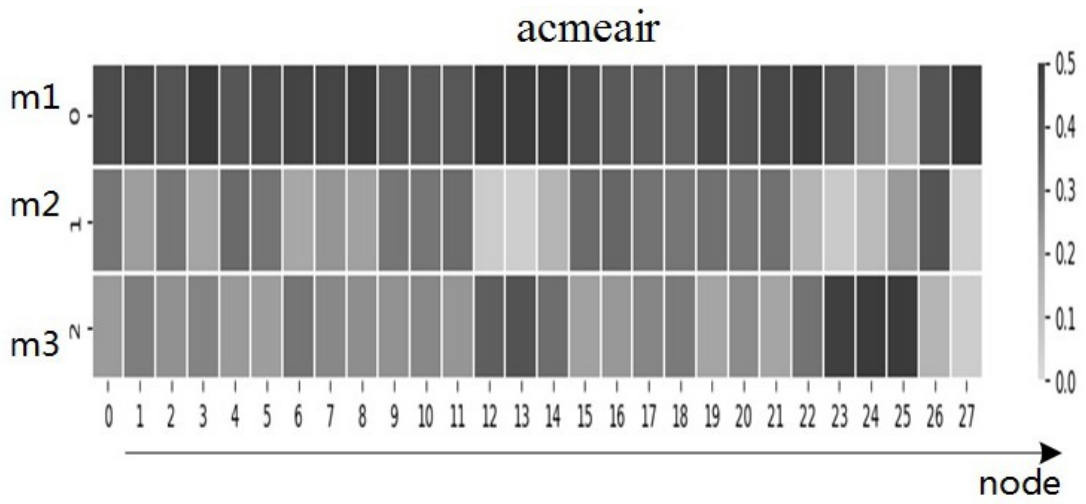


图5b

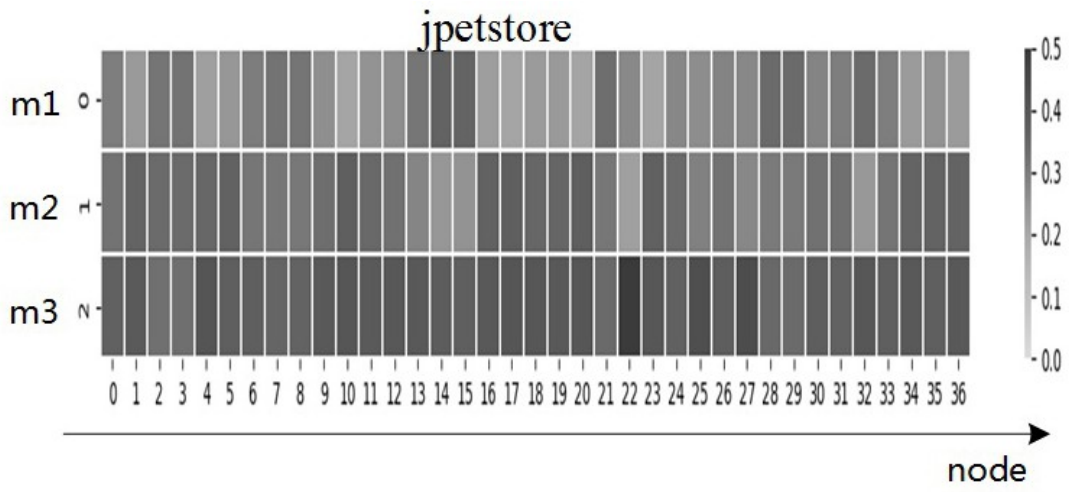


图5c

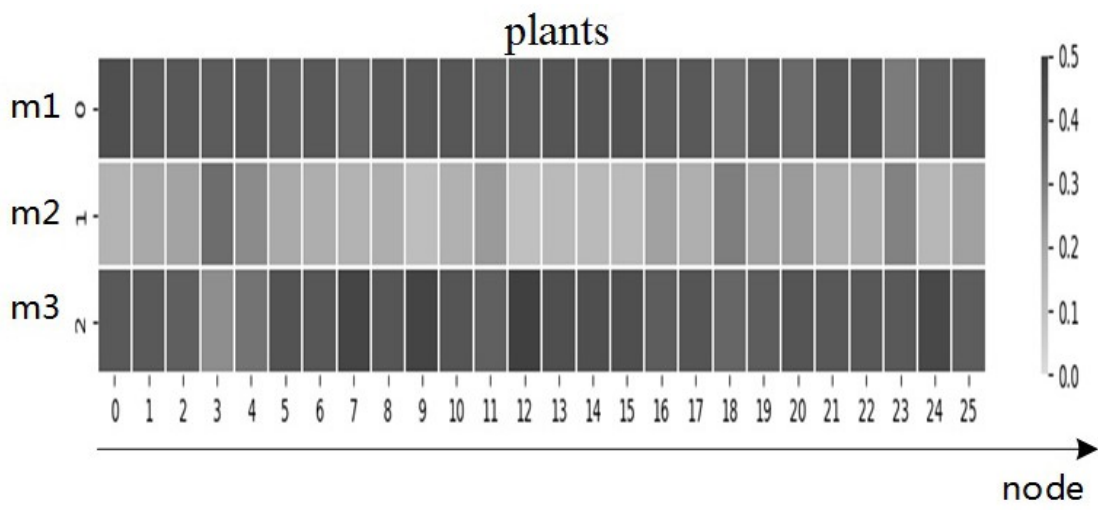


图5d

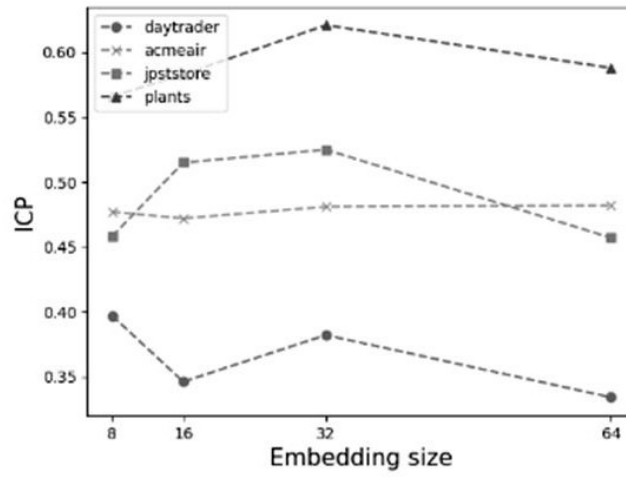


图6a

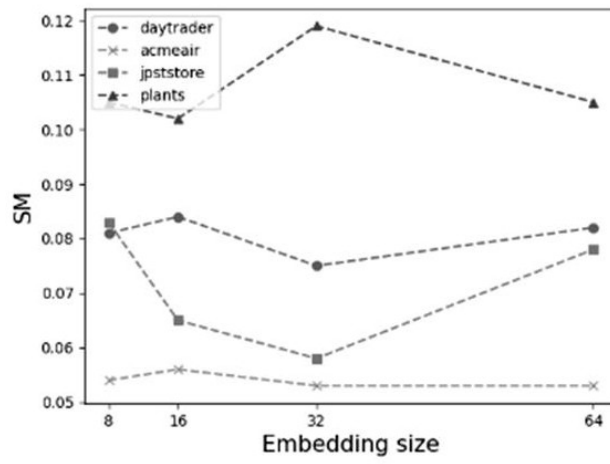


图6b



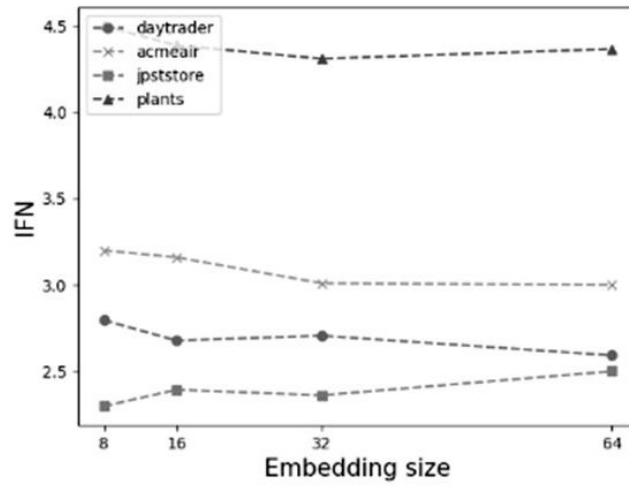


图6c