



(12) 发明专利申请

(10) 申请公布号 CN 117242453 A

(43) 申请公布日 2023. 12. 15

(21) 申请号 202280029598.X

阿尔瓦罗·桑切斯

(22) 申请日 2022.05.30

(74) 专利代理机构 中原信达知识产权代理有限  
责任公司 11219

(30) 优先权数据

63/194,851 2021.05.28 US

专利代理师 周亚荣 李宝泉

(85) PCT国际申请进入国家阶段日

2023.10.19

(51) Int.Cl.

G06N 3/042 (2023.01)

G06N 3/0455 (2023.01)

(86) PCT国际申请的申请数据

PCT/EP2022/064565 2022.05.30

(87) PCT国际申请的公布数据

W02022/248735 EN 2022.12.01

(71) 申请人 渊慧科技有限公司

地址 英国伦敦

(72) 发明人 乔纳森·威廉·戈德温

彼得·威廉·巴塔利亚

凯文·迈克尔·沙尔施密特

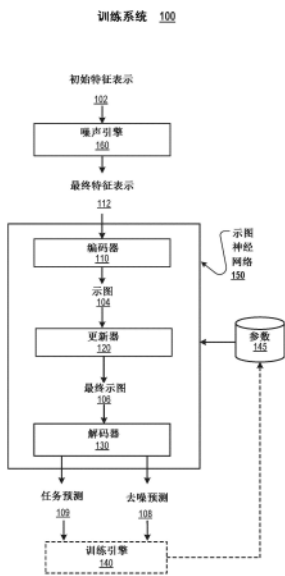
权利要求书3页 说明书18页 附图6页

(54) 发明名称

使用去噪目标训练示意图神经网络

(57) 摘要

用于训练包括一个或多个示意图神经网络层的神经网络的方法、系统和装置,包括在计算机存储介质上编码的计算机程序。在一个方面中,一种方法包括:生成定义示意图的数据,包括:为每个节点生成相应的最终特征表示,其中,针对所述节点中的一个或多个中的每个节点,所述相应的最终特征表示是使用相应噪声从所述节点的相应特征表示生成的修改特征表示;使用所述神经网络的所述示意图神经网络层中的一个或多个来处理定义所述示意图的所述数据,以生成每个节点的相应的更新节点嵌入;以及针对具有修改特征表示的所述节点中的一个或多个中的每个节点,处理所述节点的所述更新节点嵌入以为所述节点生成相应去噪预测。



1. 一种用于训练包括一个或多个示图神经网络层的神经网络的方法,所述方法包括:  
生成定义示图的数据,所述示图包括:(i) 节点集合,(ii) 每个节点的节点嵌入,以及  
(iii) 各自连接相应节点对的边缘集合,包括:

获得每个节点的相应初始特征表示;

为每个节点生成相应的最终特征表示,其中,针对所述节点中的一个或多个中的每个节点,所述相应的最终特征表示是使用相应噪声从所述节点的所述相应特征表示生成的修改特征表示;以及

使用所述节点的所述相应的最终特征表示来生成定义所述示图的所述数据;

使用所述神经网络的所述示图神经网络层中的一个或多个来处理定义所述示图的所述数据,以生成每个节点的相应的更新节点嵌入;

针对具有修改特征表示的所述节点中的一个或多个中的每个节点,处理所述节点的所述更新节点嵌入,以为所述节点生成相应去噪预测,所述相应去噪预测表征所述节点的去噪特征表示,所述去噪特征表示不包括用于生成所述节点的所述修改特征表示的所述噪声;以及

确定对所述神经网络的神经网络参数的当前值的更新,以优化目标函数,所述目标函数测量所述节点的所述去噪预测中的误差。

2. 根据权利要求1所述的方法,其中,针对具有修改特征表示的所述节点中的一个或多个中的每个节点,所述节点的所述相应去噪预测预测用于生成所述节点的所述修改特征表示的所述噪声。

3. 根据权利要求1所述的方法,其中,针对具有修改特征表示的所述节点中的一个或多个中的每个节点,所述节点的所述相应去噪预测预测所述节点的所述相应初始特征表示。

4. 根据权利要求1所述的方法,其中,针对具有修改特征表示的所述节点中的一个或多个中的每个节点,所述节点的所述相应去噪预测表征所述节点的目标特征表示。

5. 根据权利要求4所述的方法,其中,针对具有修改特征表示的所述节点中的一个或多个中的每个节点,所述节点的所述相应去噪预测预测所述节点的增量特征表示,如果所述增量特征表示被添加到所述节点的所述修改特征表示,则产生所述节点的所述目标特征表示。

6. 根据任一前述权利要求所述的方法,进一步包括:处理所述节点的所述更新节点嵌入以生成任务预测,其中,所述目标函数还测量所述任务预测中的误差。

7. 根据权利要求6所述的方法,其中以下两者:(i) 所述节点的所述更新节点嵌入,以及(ii) 在使用所述示图神经网络层来被更新之前的所述节点的原始节点嵌入,都被处理以生成所述任务预测。

8. 根据权利要求6至7中任一项所述的方法,其中,所述示图表示分子,并且所述任务预测是所述分子的平衡能量的预测。

9. 根据任一前述权利要求所述的方法,其中,针对所述神经网络的多个示图神经网络层中的每个示图神经网络层,所述目标函数测量所述节点的去噪预测中的相应误差,所述去噪预测基于由所述示图神经网络层生成的更新节点嵌入。

10. 根据任一前述权利要求所述的方法,其中,针对具有修改特征表示的所述节点中的一个或多个中的每个节点,处理所述节点的所述更新节点嵌入以为所述节点生成所述相应

去噪预测包括：

使用一个或多个神经网络层来处理所述节点的所述更新节点嵌入,以为所述节点生成所述相应去噪预测。

11. 根据任一前述权利要求所述的方法,其中,确定对所述神经网络的所述神经网络参数的所述当前值的所述更新以优化所述目标函数包括：

通过所述示图神经网络层的神经网络参数反向传播所述目标函数的梯度。

12. 根据任一前述权利要求所述的方法,其中,针对所述节点中的一个或多个中的每个节点,所述节点的所述相应的最终特征表示是通过将所述相应噪声添加到所述节点的所述相应特征表示而生成的。

13. 根据任一前述权利要求所述的方法,其中使用所述节点的所述相应的最终特征表示来生成定义所述示图的所述数据包括：

针对包括第一节点和第二节点的每对节点,确定所述第一节点的所述最终特征表示和所述第二节点的所述最终特征表示之间的相应距离;以及

确定与小于预定义阈值的距离相对应的每对节点由所述示图中的边缘连接。

14. 根据任一前述权利要求所述的方法,其中,所述示图进一步包括每个边缘的相应边缘嵌入。

15. 根据权利要求14所述的方法,其中,生成定义所述示图的所述数据包括：

至少部分地基于由各边缘连接的节点的所述相应的最终特征表示之间的差异来为所述示图中的每个边缘生成边缘嵌入。

16. 根据任一前述权利要求所述的方法,其中,所述示图包括分子,所述示图中的每个节点表示所述分子中的相应原子,并且生成定义所述示图的所述数据包括：

基于由所述节点表示的原子类型来为每个节点生成节点嵌入。

17. 根据任一前述权利要求所述的方法,其中,所述神经网络包括至少10个示图神经网络层。

18. 根据任一前述权利要求所述的方法,其中,所述示图神经网络的每个示图神经网络层被配置为：

接收当前示图;以及

根据所述示图神经网络层的当前神经网络参数值来更新所述当前示图,包括：

基于以下来更新所述示图中的一个或多个节点中的每个节点的当前节点嵌入：(i) 所述节点的所述当前节点嵌入,以及(ii) 所述示图中的所述节点的一个或多个邻居中的每个邻居的相应当前节点嵌入。

19. 根据权利要求18所述的方法,其中,所述当前示图包括每个边缘的边缘嵌入,并且其中,更新所述示图中的一个或多个节点中的每个节点的所述当前节点嵌入进一步包括：

至少部分地基于连接至所述节点的一个或多个边缘中的每个边缘的相应边缘嵌入来更新所述节点的所述节点嵌入。

20. 根据权利要求4至16中任一项所述的方法,当从属于权利要求4时,其中,所述示图表示分子,所述示图中的每个节点表示所述分子中的相应原子,每个节点的所述初始特征表示表示所述分子中的对应原子的初始空间位置,并且针对具有修改特征表示的所述节点中的一个或多个中的每个节点,并且所述节点的所述目标特征表示表示原子弛豫之后所述

对应原子的最终空间位置。

21.一种存储指令的一个或多个非暂时性计算机存储介质,所述指令在由一个或多个计算机实行时,使所述一个或多个计算机执行权利要求1至20中任一项所述的相应方法的操作。

22.一种系统,包括:

一个或多个计算机;以及

一个或多个存储设备,所述一个或多个存储设备被通信地耦合至所述一个或多个计算机,其中,所述一个或多个存储设备存储指令,所述指令在由所述一个或多个计算机实行时,使所述一个或多个计算机执行权利要求1至20中任一项所述的相应方法的操作。

## 使用去噪目标训练示图神经网络

[0001] 相关申请的交叉引用

[0002] 本申请要求“TRAINING GRAPH NEURAL NETWORKS USING ADE-NOISING OBJECTIVE (使用去噪目标训练示图神经网络)”的美国临时专利申请序列号63/194,851的提交日期的权益,该申请于2021年5月28日提交并且通过引用全部并入本文。

### 背景技术

[0003] 本说明书涉及使用机器学习模型处理数据。

[0004] 机器学习模型接收输入并且基于接收到的输入生成输出,例如预测输出。一些机器学习模型是参数模型,并且基于接收到的输入和模型的参数的值来生成输出。

[0005] 一些机器学习模型是采用多层模型来为接收到的输入生成输出的深度模型。例如,深度神经网络是包括输出层和一个或多个隐藏层的深度机器学习模型,每个隐藏层对接收到的输入应用非线性变换以生成输出。

### 发明内容

[0006] 本说明书大体上描述了一种训练系统,该训练系统在一个或多个地点中的一个或多个计算机上被实施为训练包括一个或多个示图神经网络层的神经网络的计算机程序。

[0007] 如贯穿本说明书使用的,“示图”(graph)是指至少包括以下的数据结构:(i) 节点集合,以及(ii) 边缘集合。示图中的每个边缘可以连接示图中的相应节点对。示图可以是“有向”示图,即,使得连接一对节点的每个边缘被定义为从第一节点指向第二节点,或者反之亦然,或者可以是“无向”示图,即,使得边缘(或者反向边缘对)与方向不关联。

[0008] 通常,定义示图的数据可以包括定义示图的节点和边缘的数据,并且可以用任何适当的数字格式表示。例如,示图可以由包括元组列表 $\{(i, j)\}$ 的数据定义,其中每个元组 $(i, j)$ 表示连接节点 $i$ 和节点 $j$ 的示图中的边缘。而且,示图中的每个边缘可以与一个或多个边缘特征的集合相关联,并且示图中的每个节点可以与一个或多个节点特征的集合相关联。

[0009] 在一个方面中,描述了一种用于训练包括一个或多个示图神经网络层的神经网络的方法。该方法包括生成定义示图的数据,该示图包括:(i) 节点集合,(ii) 每个节点的节点嵌入,以及(iii) 各自连接相应节点对的边缘集合。在实施方式中,这包括获得每个节点的相应的初始特征表示,并且为每个节点生成相应的最终特征表示,其中,针对一个或多个节点中的每个节点,相应的最终特征表示是使用相应噪声从节点的相应特征表示生成的修改特征表示,并且使用节点的相应的最终特征表示来生成定义示图的数据。在实施方式中,每个节点的节点嵌入是从节点的相应的最终特征表示生成的。

[0010] 该方法使用神经网络的示图神经网络层中的一个或多个来处理定义示图的数据,以生成每个节点的相应的更新节点嵌入。针对具有修改特征表示的一个或多个节点中的每个节点,该方法处理节点的更新节点嵌入,以为节点生成相应去噪预测,该相应去噪预测表征节点的去噪特征表示,该去噪特征表示不包括用于生成节点的修改特征表示的噪声。该

方法确定对神经网络的神经网络参数的当前值的更新,以优化目标函数,该目标函数测量节点的去噪预测中的误差,具体地优化节点的相应去噪预测。

[0011] 在一些实施方式中,针对具有修改特征表示的一个或多个节点中的每个节点,节点的相应去噪预测用于生成节点的修改特征表示的噪声。

[0012] 在一些实施方式中,针对具有修改特征表示的一个或多个节点中的每个节点,节点的相应去噪预测预测节点的相应初始特征表示。

[0013] 在一些实施方式中,针对具有修改特征表示的一个或多个节点中的每个节点,节点的相应去噪预测表征节点的目标特征表示。

[0014] 在一些实施方式中,针对具有修改特征表示的一个或多个节点中的每个节点,节点的相应去噪预测预测节点的增量特征表示,如果该增量特征表示被添加到节点的修改特征表示,则产生节点的目标特征表示。

[0015] 在一些实施方式中,该方法进一步包括:处理节点的更新节点嵌入以生成任务预测,其中,目标函数还测量任务预测中的误差。

[0016] 在一些实施方式中,以下两者:(i) 节点的更新节点嵌入,以及(ii) 在使用示图神经网络层来被更新之前的节点的原始节点嵌入,都被处理以生成任务预测。

[0017] 在一些实施方式中,示图表示分子,并且任务预测是对分子的平衡能量的预测。

[0018] 在一些实施方式中,针对神经网络的多个示图神经网络层中的每个示图神经网络层,目标函数测量节点的去噪预测中的相应误差,所述去噪预测基于由示图神经网络层生成的更新节点嵌入。

[0019] 在一些实施方式中,针对具有修改特征表示的一个或多个节点中的每个节点,处理节点的更新节点嵌入以为节点生成相应去噪预测包括:使用一个或多个神经网络层来处理节点的更新节点嵌入,以为节点生成相应去噪预测。

[0020] 在一些实施方式中,确定对神经网络的神经网络参数的当前值的更新以优化目标函数包括:通过示图神经网络层的神经网络参数反向传播目标函数的梯度。

[0021] 在一些实施方式中,针对一个或多个节点中的每个节点,节点的相应的最终特征表示是通过将相应噪声添加到节点的相应特征表示而生成的。

[0022] 在一些实施方式中,使用节点的相应的最终特征表示来生成定义示图的数据包括:针对包括第一节点和第二节点的每对节点,确定第一节点的最终特征表示和第二节点的最终特征表示之间的相应距离;以及确定与小于预定义阈值的距离相对应的每对节点由示图中的边缘连接。

[0023] 在一些实施方式中,该示图进一步包括每个边缘的相应边缘嵌入。

[0024] 在一些实施方式中,生成定义示图的数据包括:至少部分地基于由边缘连接的节点的相应的最终特征表示之间的差异来为示图中的每个边缘生成边缘嵌入。

[0025] 在一些实施方式中,示图表示分子,该示图中的每个节点表示分子中的相应原子,并且生成定义示图的数据包括:基于由节点表示的原子类型来为每个节点生成节点嵌入。

[0026] 在一些实施方式中,神经网络包括至少10个示图神经网络层。

[0027] 在一些实施方式中,示图神经网络的每个示图神经网络层被配置为:接收当前示图;以及根据示图神经网络层的当前神经网络参数值来更新当前示图,包括:基于以下来更新示图中的一个或多个节点中的每个节点的当前节点嵌入:(i) 节点的当前节点嵌入,以及

(ii) 示图中的节点的一个或多个邻居中的每个邻居的相应当前节点嵌入。

[0028] 在一些实施方式中,当前示图包括每个边缘的边缘嵌入,并且更新示图中的一个或多个节点中的每个节点的当前节点嵌入进一步包括:至少部分地基于连接至节点的一个或多个边缘中的每个边缘的相应边缘嵌入来更新节点的节点嵌入。

[0029] 在一些实施方式中,示图表示分子,该示图中的每个节点表示分子中的相应原子,每个节点的初始特征表示表示分子中的对应原子的初始空间位置,并且针对具有修改特征表示的一个或多个节点中的每个节点,并且节点的目标特征表示表示原子弛豫之后对应原子的最终空间位置。

[0030] 根据另一方面,提供了一种存储指令的一个或多个非暂时性计算机存储介质,在所述指令由一个或多个计算机实行时,使所述一个或多个计算机执行本文描述的方法的操作。

[0031] 根据另一方面,提供了一种系统,包括:一个或多个计算机;以及一个或多个存储设备,被通信地耦合至一个或多个计算机,其中,所述一个或多个存储设备存储指令,在所述指令由一个或多个计算机实行时,使所述一个或多个计算机执行本文描述的方法的操作。

[0032] 如贯穿本说明书使用的,“嵌入”是指数值的有序集合,例如数值的向量、矩阵或其他张量。

[0033] 本说明书中描述的主题可以被实施在特定实施例中,以便实现以下优点中的一个或多个。

[0034] 本说明书中描述的系统可以训练示图神经网络(即,包括一个或多个示图神经网络层的神经网络)来生成示图中的节点的去噪预测。具体地,在示图上训练示图神经网络之前,系统可以使用噪声来修改示图中的节点的特征表示,例如通过将噪声添加到节点的特征表示。去噪预测可以例如预测修改节点的特征表示的噪声的值,或者预测示图中的节点的原始特征表示的重构(即,在使用噪声修改特征表示之前)。

[0035] 训练示图神经网络以生成去噪预测可以使示图神经网络的训练正规化,并且具体地,可以实现对具有大量示图神经网络层——例如多于100个示图神经网络层——的示图神经网络进行有效训练。相反,在示图神经网络的性能随着更多的示图神经网络层的添加而饱和或者甚至降低之前,许多常规系统被限制为训练具有少得多的示图神经网络层(例如少于10层)的示图神经网络。当由本说明书中描述的系统训练时,较深的示图神经网络(即,具有更多的示图神经网络层)可以为更复杂的预测任务实现与使用较浅的示图神经网络(即,具有更少的示图神经网络层)可实现的相比更高的预测准确性。

[0036] 生成去噪预测需要每个节点嵌入对唯一信息进行编码,以便对节点的特征表示进行去噪,这可以减轻“过度平滑”的影响,例如节点嵌入在通过多个示图神经网络层处理后变得几乎相同。而且,训练示图神经网络生成去噪预测可以降低“过度拟合”的似然性,例如因为添加到示图中的节点的特征表示的噪声防止示图神经网络记忆原始节点特征表示。训练示图神经网络生成去噪预测也鼓励示图神经网络隐式学习“真实”示图的分布,即,具有未修改的节点特征表示,并且示图神经网络可以充分利用该隐式知识来实现“任务”预测的更高准确性。因为所描述的技术与涉及丢弃节点或边缘特征的其他技术不同地工作,所以它们可以与这些其他技术组合。因此,本说明书中描述的系统通过使得能够有效训练较深

的示图神经网络来实现更高的准确性,同时减轻过度平滑和过度拟合的影响,从而能够更有效地使用计算资源(例如存储器和计算功率)。

[0037] 本说明书的主题的一个或多个实施例的细节在下面的附图和描述中陈述。主题的其他特征、方面和优点将通过描述、附图和权利要求而变得显而易见。

## 附图说明

[0038] 图1是用于训练示图神经网络的示例训练过程的框图。

[0039] 图2图示了可以由训练系统执行的操作的示例。

[0040] 图3是用于使用训练系统来训练示图神经网络的示例过程的流程图。

[0041] 图4、图5和图6图示了示例实验结果。

[0042] 在各个附图中,相同的附图标记和名称指示相同的元件。

## 具体实施方式

[0043] 图1是用于训练示图神经网络150——例如包括一个或多个示图神经网络层的神经网络——的示例训练系统100的框图。系统110是在一个或多个地点中的一个或多个计算机上被实施为计算机程序的系统的示例,下面描述的系统、组件和技术被实施在其中。

[0044] 系统100可以通过使用示图神经网络150生成示图104中的节点的去噪预测108来训练示图神经网络150。通常,“示图”是指至少包括以下的数据结构:(i) 节点集合,以及(ii) 边缘集合。例如,示图可以被表示为例如 $G = (V, E)$ ,其中 $V$ 是节点集合,并且 $E$ 是边缘集合。示图中的每个边缘可以连接示图中的一对节点。在一些实施方式中,示图 $G$ 可以附加地使用示图级特性 $g$ 来表示,例如 $G = (V, E, g)$ ,其中,示图级特性 $g$ 可以包括由示图104表示的整个系统的任何适当方面。作为特定示例,示图104可以表示物理系统,例如分子,示图104中的每个节点可以表示例如分子中的原子,并且示图级特性可以包括例如分子的能量。通常,示图104可以表示任何适当类型的系统,例如粒子的集合、点云或社交网络。下面更详细地描述可以由示图104表示的系统的示例。

[0045] 训练系统100可以包括噪声引擎160,该噪声引擎160可以被配置为处理示图104中的每个节点的初始特征表示102,并且为示图104中的每个节点生成相应的最终特征表示112。通常,节点的“特征表示”可以表征由该节点表示的元素的任何适当方面。例如,针对表示分子的示图,节点的初始特征表示102可以包括例如由该节点表示的原子在分子中的空间位置(例如 $x$ 、 $y$ 和 $z$ 坐标)。在一些实施方式中,节点的初始特征表示102可以包括由该节点表示的元素的特征的有序集合。例如,节点的初始特征表示可以包括由该节点表示的原子在分子中的空间位置(例如表示为 $\mathbb{R}^3$ 中的向量)以及来自可能的原子类型集合的由该节点表示的原子类型(例如包括碳、氧、氮等中的一个或多个)。

[0046] 针对一个或多个节点,噪声引擎160可以通过使用例如高斯噪声的相应的噪声修改初始特征表示102(的至少一部分)来生成最终特征表示112。在一些实施方式中,噪声引擎160可以从例如高斯分布的分布中对一个或多个节点中的每个节点的相应噪声值进行随机地采样,并且将相应噪声值添加到该节点的相应初始特征表示102以生成最终特征表示112。针对示图104中的剩余节点,最终特征表示可以与初始特征表示相同。作为特定示例,噪声引擎160可以如下修改示图102中的一些或全部节点的初始特征表示102:

$$[0047] \quad \tilde{v}_i = v_i + \sigma_i \quad (1)$$

[0048] 其中,  $v_i$  是节点  $i$  的初始特征表示,  $\sigma_i$  是节点  $i$  的噪声, 并且  $\tilde{v}_i$  是节点  $i$  的最终特征表示。作为特定示例, 如果示图104表示分子, 示图104中的每个节点表示分子中的原子, 并且示图104中的节点的初始特征表示102包括: (i) 由该节点表示的原子的空间位置, 以及 (ii) 由该节点表示的原子类型, 那么噪声引擎160可以通过将噪声与表示由节点表示的原子的空间位置的特征相加 (或以其他方式组合) 来生成该节点的最终特征表示112。即, 针对示图104中的每个节点, 噪声引擎160可以使用噪声来扰动表示对应原子的空间位置的特征, 使得该节点的最终特征表示定义对应原子的扰动空间位置。每个节点的相应初始特征表示可以进一步包括例如定义由该节点表示的原子类型的特征, 并且噪声引擎160可以可选地避免将噪声组合 (例如添加) 到表示原子类型的特征。

[0049] 在一些实施方式中, 噪声引擎160可以通过使用例如从高斯分布采样的噪声值的相应的噪声缩放节点的初始特征表示102来生成节点的最终特征表示112。在一些实施方式中, 噪声引擎160可以使用具有与初始特征表示102相同维度的噪声来修改节点的初始特征表示102。例如, 如果初始特征表示是  $N$  维向量, 那么噪声也可以是  $N$  维向量。通常, 噪声引擎160可以以任何适当的方式使用相应噪声来生成示图104中的节点的最终特征表示112。

[0050] 在仅针对示图104中的一些节点 (例如不是示图104中的所有节点) 修改初始特征表示102的实施方式中, 噪声引擎160可以随机选择示图104中的初始特征表示102被修改的节点。

[0051] 示图神经网络150可以包括: (i) 编码器110, (ii) 更新器120, 以及 (iii) 解码器130, 下面将更详细地描述每一个。

[0052] 噪声引擎160可以将示图104中的节点的最终特征表示112提供给编码器110。编码器110可以被配置为使用节点的相应最终特征表示112来生成定义示图104的数据。例如, 编码器110可以为由图示104所表示的系统中的每个元素指派示图104中的相应节点。然后, 编码器110可以实例化示图104中的节点对之间的边缘。通常, 编码器110可以以任何适当的方式实例化示图104中的节点对之间的边缘。

[0053] 在一些实施方式中, 编码器110可以通过为每对节点确定这些节点的最终特征表示之间的相应距离来实例化示图104中的节点对之间的边缘。然后, 编码器110可以确定与小于预定义阈值的距离相对应的每对节点通过示图104中的边缘连接。阈值距离可以是任何适当的数值。在一些实施方式中, 编码器110可以基于由示图104表示的系统的类型来实例化示图104中的节点对之间的边缘。

[0054] 作为特定示例, 如果示图104表示分子, 并且示图104中的每个节点表示分子中的原子, 那么编码器110可以在示图104中在一对节点之间指派边缘, 该边缘与由该对节点表示的分子中的原子之间的键相对应。在一些实施方式中, 最终特征表示之间的距离可以表征由节点表示的原子之间的局部相互作用。例如, 阈值距离可以表示连接性半径 ( $R$ ), 使得连接连接性半径内的节点对的边缘表示分子中的邻居原子的局部相互作用。对示图104中的邻居节点的搜索可以经由例如kd树算法的任何适当的搜索算法来执行。

[0055] 除了指派节点并且实例化边缘之外, 编码器110还可以为示图104中的每个节点生成相应的节点嵌入。通常, 实体的“嵌入”可以将实体表示指代为数值的有序集合, 例如数值

的向量或矩阵。

[0056] 编码器110可以通过使用节点嵌入子网络来为每个节点生成节点嵌入。编码器110的节点嵌入子网络可以处理示图104中的每个节点的最终特征表示112,并且生成示图102中的每个节点的节点嵌入。作为特定示例,如果示图104表示分子,则节点嵌入子网络可以基于例如由节点表示的原子的类型来生成节点的节点嵌入。作为另一特定示例,节点嵌入子网络可以基于原子是吸附质还是催化剂的一部分来生成节点嵌入,例如节点嵌入可以包括吸附质的1和催化剂的0。

[0057] 在一些实施方式中,除了为示图104中的每个节点生成节点嵌入之外,编码器110还可以使用编码器110的边缘嵌入子网络为示图104的每个边缘生成边缘嵌入。例如,编码器110的边缘嵌入子网络可以处理示图104中的节点的最终特征表示112,并且至少部分地基于由该边缘连接的节点的相应最终特征表示112之间的差异来生成示图104中的每个边缘的边缘嵌入。作为特定示例,连接一对节点的边缘k的嵌入 $e_k$ 可以被表示如下:

$$[0058] \quad e_k = \text{concat}(e'_{RBF,1}(|d|), \dots, e'_{RBF,c}(|d|), \frac{d}{|d|}) \quad (2)$$

[0059] 其中,d是连接节点对的边缘的向量位移, $|d|$ 是距离, $e'_{RBF,c}(|d|)$ 是下面由等式(3)定义的径向贝塞尔基函数,并且concat表示连结操作。

$$[0060] \quad e'_{RBF,c}(|d|) = \sqrt{\frac{2}{r}} \frac{\sin(\frac{c\pi}{R}d)}{d} \quad (3)$$

[0061] 通过该方式,编码器110可以生成定义示图104的数据,该示图104包括:(i) 节点集合,(ii) 各自连接相应节点对的边缘集合,(iii) 每个节点的节点嵌入,以及可选地,(iv) 每个边缘的边缘嵌入。

[0062] 在生成定义示图104的数据之后,编码器110可以将数据提供给更新器120。更新器120可以在多个内部更新迭代上更新示图104以生成最终示图106。“更新”示图是指通过例如基于示图中的邻居节点的节点和/或边缘嵌入更新示图中的一些或所有节点和边缘的节点和/或边缘嵌入来在示图中所包括的节点和边缘之间执行消息传递步骤(例如信息传播步骤)。更新器120可以包括一个或多个示图神经网络层,并且每个示图神经网络层可以被配置为接收当前示图并且根据示图神经网络层的当前参数更新当前示图。更新器120可以包括任何数量的示图神经网络层,例如1、10、100或者任何其他适当数量的示图神经网络层。在一些实施方式中,更新器120包括至少10个示图神经网络层。

[0063] 具体地,每个示图神经网络层可以被配置为基于:(i) 该节点的当前节点嵌入,以及(ii) 该节点在示图104中的一个或多个邻居中的每个邻居的相应当前节点嵌入,来更新示图104中的每个节点的当前节点嵌入。如果示图104中的一对节点通过边缘彼此连接,则它们是“邻居”节点。在示图104附加地包括每个边缘的边缘嵌入的实施方式中,每个示图神经网络层也可以基于连接至示图104中的节点的一个或多个边缘中的每个边缘的相应边缘嵌入来更新节点的节点嵌入。

[0064] 作为特定示例,在每次更新迭代时,每个示图神经网络层可以被配置为如下确定将节点u连接至节点v的边缘的当前消息向量 $m_{uv}^{(t+1)}$ :

$$[0065] \quad m_{uv}^{(t+1)} = \psi_{t+1}(h_u^t, h_v^t, m_{uv}^{(t)} + m_{uv}^{(t-1)}) \quad (4)$$

[0066] 其中,  $h_u^t$  是节点u在前一更新迭代中的节点嵌入,  $h_v^t$  是节点v在前一更新迭代中的节点嵌入,  $m_{uv}^{(t)}$  和  $m_{uv}^{(t-1)}$  是各自在前一相应更新迭代中确定的边缘的消息向量, 并且  $\psi_{t+1}$  是由示图神经网络层实施为例如全连接神经网络层 (例如针对每个边缘都相同) 的消息函数。在确定消息向量后, 在每次更新迭代中, 示图神经网络层可以更新通过边缘连接至节点v的节点u的当前节点嵌入  $h_u^t$ , 如下:

$$[0067] \quad h_u^{(t+1)} = \varphi_{t+1}(h_u^t, \sum_{N_v} m_{vu}^{(t+1)}, \sum_{N_u} m_{uv}^{(t+1)}) + h_u^t \quad (5)$$

[0068] 其中  $h_u^{(t+1)}$  是更新迭代的更新节点嵌入, 更新函数  $\varphi_{t+1}$  由示图神经网络层实施为例如全连接神经网络层 (例如针对每个节点都相同), 第一求和是对节点v的邻居节点  $N_v$  的总数进行的, 并且第二求和是对节点u的邻居节点  $N_u$  的总数进行的。

[0069] 更新器120的最终更新迭代生成定义最终示图106的数据。最终示图106可以具有与初始示图104相同的结构 (例如最终示图106可以具有与初始示图104相同数量的节点和相同数量的边缘), 但是具有不同的节点嵌入。在一些实施方式中, 最终示图104可以附加地包括不同的边缘嵌入。

[0070] 更新器120可以向解码器130提供定义最终示图106的数据。解码器130可以被配置为处理定义最终示图106的数据, 以针对具有修改特征表示的一个或多个节点中的每个节点生成去噪预测108。具体地, 针对具有修改特征表示的一个或多个节点中的每个节点, 解码器130可以使用一个或多个神经网络层来处理该节点的更新节点嵌入, 以生成该节点的相应去噪预测108。去噪预测108可以表征该节点的去噪特征表示, 它不包括用于生成该节点的修改特征表示的噪声。

[0071] 在一些实施方式中, 节点的去噪预测108可以预测用于生成节点的修改特征表示的噪声。

[0072] 在一些实施方式中, 节点的去噪预测108可以预测节点的初始特征表示。例如, 如果示图104表示分子, 那么节点的去噪预测108可以在通过使用噪声修改初始空间位置之前预测原子在由节点表示的分子中的初始空间位置。

[0073] 在一些实施方式中, 节点的去噪预测108可以表征节点的目标特征表示。节点的“目标特征表示”可以表征由示图104中的节点表示的元素的任何适当方面。作为特定示例, 如果节点的初始特征表示102包括原子在分子中的空间位置, 那么节点的目标特征表示可以包括原子在该分子中的不同空间位置, 例如在分子的原子弛豫之后原子的空间位置。作为另一示例, 如果示图表示社交网络并且示图中的每个节点表示社交网络中的相应用户, 那么每个节点的目标嵌入可以表征例如对应用户在指定时间段 (例如一天) 内与社交网络交互的时间量 (例如以分钟为单位)。在一个示例中, 每个节点的去噪预测108可以包括输出特征表示, 该输出特征表示是该节点的目标特征表示的估计。在另一示例中, 每个节点的去噪预测108可以包括该节点的增量特征表示的预测, 如果该增量特征表示被添加到该节点的修改特征表示, 则产生该节点的目标特征表示。

[0074] 由于生成去噪预测108需要每个节点嵌入对唯一信息进行编码, 以便对节点的特征表示进行去噪, 这可以减轻“过度平滑”的影响, 例如节点嵌入在通过多个示图神经网络层处理后变得几乎相同。

[0075] 在一些实施方式中,解码器130可以处理节点的更新节点嵌入以生成任务预测109。任务预测可以是例如输入示图104的单个输出或者输入示图104中的每个节点的相应输出。通常,任务预测109可以是表征由示图104中的节点表示的元素中的一个或多个的任何适当的预测。任务预测109可以是例如分类预测或回归预测。分类预测可以包括可能的类集合中的每个类的相应分数,其中类的分数可以定义由示图104表示的元素集合被包括在该类中的似然性。回归预测可以包括一个或多个数值,每个数值都是从连续的值范围中提取的,它表征由示图104表示的元素集合。

[0076] 在一个示例中,为了生成任务预测,解码器130可以处理 (i) 节点的更新节点嵌入,以及 (ii) 在使用示图神经网络层来被更新之前的节点的原始节点嵌入。作为特定示例,解码器130可以生成任务预测 $y$ 如下:

$$[0077] \quad y = W^{Update} \sum_{i=1}^{|V|} MLP_{Update}(a_i^{Update}) + b^{Update} + W^{Enc} \sum_{i=1}^{|V|} MLP_{Enc}(a_i^{Enc}) + b^{Enc} \quad (6)$$

[0078] 其中 $a_i^{Update}$ 是节点 $i$ 的更新节点嵌入, $a_i^{Enc}$ 是节点 $i$ 的原始节点嵌入, $|V|$ 是示图中的节点总数, $MLP_{Update}$ 和 $MLP_{Enc}$ 例如分别是更新器和编码器的全连接神经网络层, $b^{Update}$ 是更新器的偏置项, $b^{Enc}$ 是编码器的偏置项, $W^{Update}$ 是更新器的线性神经网络层,并且 $W^{Enc}$ 是编码器的线性神经网络层。

[0079] 作为特定示例,如果示图104表示分子,那么任务预测109可以是由示图104表示的分子的平衡能量、内能或最高占据分子轨道(HOMO)能量中的一个或多个的预测。在一些实施方式中,解码器130可以处理少于所有节点的更新节点嵌入以生成任务预测109,例如在一些情况下,解码器130可以通过处理示图104中的单个节点的更新节点嵌入来生成任务预测109。下面将更详细地描述任务预测的示例。

[0080] 编码器110、更新器120和解码器130可以具有使它们能够执行其规定功能的任何适当的神经网络架构。例如,编码器110、更新器120和解码器130可以具有任何适当数量(例如2层、5层或10层)的任何适当的神经网络层(例如卷积层、全连接层、递归层、注意力层、示图神经网络层等),并且以任何适当的配置(例如作为层的线性序列)连接。

[0081] 系统100可以进一步包括训练引擎140,该训练引擎140可以使用去噪预测108来训练神经网络150。训练引擎150可以评估目标函数,该目标函数针对神经网络150的示图神经网络层中的一个或多个示图神经网络层来测量基于由示图神经网络层生成的更新节点嵌入的节点的去噪预测180中的相应误差。更具体地,示图神经网络150可以在一个或多个示图神经网络层中的每个示图神经网络层处——即,基于由示图神经网络层生成的更新节点嵌入——为示图104中的节点生成相应的去噪预测。在一些实施方式中,目标函数可以附加地测量任务预测109中的误差,例如使用交叉熵误差测量、平方误差测量或任何其他适当的误差测量。作为特定示例,目标函数 $\mathcal{L}$ 可以被表示如下:

$$[0082] \quad \mathcal{L} = \lambda \mathcal{L}_{de-noising} + \mathcal{L}_{task} \quad (7) \text{ 其中 } \mathcal{L}_{de-noising} \text{ 测量节点的去噪预}$$

测中的相应误差, $\mathcal{L}_{task}$ 测量任务预测中的误差,并且 $\lambda$ 是权重因子。权重因子和噪声将根据应用而变化,并且可以利用超参数扫描来被优化;仅作为示例,权重因子可以是0.1级并且噪声标准偏差可以是0.01级。

[0083] 为了优化目标函数,训练引擎140可以例如使用反向传播技术来确定目标函数相

对于神经网络参数的当前值的梯度。训练引擎140然后可以使用梯度来更新神经网络参数的当前值,例如使用任何适当的梯度下降优化技术,例如RMSprop或Adam梯度下降优化技术。具体地,训练引擎140可以通过示图神经网络层的神经网络参数反向传播目标函数的梯度。

[0084] 在一些实施方式中,训练引擎140可以首先预训练神经网络150以仅基于去噪预测来优化目标函数(例如 $\mathcal{L}_{de-noising}$ ),然后训练神经网络150以基于去噪预测和任务预测(例如等式(7)中定义的 $\mathcal{L}$ )来优化目标函数。训练神经网络150生成去噪预测108可以降低“过度拟合”的似然性,例如因为添加到示图104中的节点的特征表示的噪声防止神经网络150记忆初始节点特征表示。

[0085] 在一些实施方式中,训练引擎140可以预训练神经网络150以仅基于去噪预测来优化目标函数,然后训练神经网络150以仅基于任务预测来优化目标函数。

[0086] 在一些实施方式中,训练引擎140可以预训练神经网络150以基于去噪预测和第一任务预测来优化目标函数,然后训练神经网络150以仅基于第二任务预测来优化目标函数。第二任务预测可以不同于第一任务预测。例如,第一任务预测可以包括预测分子的HOMO能量,而第二任务预测可以包括预测分子的平衡能量。

[0087] 可选地,在预训练神经网络150以基于去噪预测来优化目标函数之后,训练引擎140可以“冻结”神经网络的一些参数,然后基于任务预测在目标函数上仅训练神经网络的未冻结参数。(冻结神经网络的参数可以指将参数的当前值指定为固定的静态值,它在训练期间不会被进一步修改)。例如,训练引擎140可以基于去噪预测预训练神经网络以优化目标函数,冻结编码器110和更新器120的参数,然后基于任务预测在目标函数上仅训练解码器的参数。

[0088] 接下来将更详细地描述系统100和经训练的示图神经网络150的示例应用。通常,使用经训练的示图神经网络150涉及获得节点的特征表示;使用节点的特征表示来生成定义示图104的数据;以及使用示图神经网络150来处理定义示图104的数据,以生成每个节点的相应的更新节点嵌入。然后,取决于应用,来自示图神经网络的输出可以包括以下中的一个或多个:从示图104的更新节点嵌入解码的特征;去噪预测108;以及任务预测109。

[0089] 在一些实施方式中,示图104可以表示一个或多个分子;此处的“分子”包括例如一大块原子,诸如催化剂的表面。然后,示图104中的每个节点可以表示分子中的相应原子。通常,节点的特征表示定义了由节点表示的原子的类型。它可以包括其他特征,诸如原子序数、原子是电子还是质子供体或受体、原子是否是芳香族的一部分、例如碳原子的杂化程度、以及在氢原子未被明确表示的情况下附接至原子的氢的数量。在一些实施方式中,节点的特征表示可以包括原子在分子中的(3D)空间位置。在一些实施方式中,节点的特征表示不包括原子在分子中的(3D)空间位置,即,分子可以由键和原子类型定义。在表示不包括空间位置噪声的情况下,噪声可以通过随机改变一个或多个特征或节点(以及可选地还改变边缘的一个或多个特征)来被添加;去噪预测然后可以包括特征的重构。在一些实施方式中,例如在对两个或更多个交互实体建模的情况下,节点特征可以指示原子属于哪个实体。因此,通常节点的特征表示定义了分子的结构和性质(例如原子的类型)。

[0090] 在一些实施方式中,训练神经网络以从分子中的原子的初始结构识别原子的所得结构。该结构可以从节点嵌入中被解码,例如作为从节点嵌入被解码的原子的(3D)空间位

置,或者作为键和原子类型;例如在一些实施方式中,它可以从节点的相应去噪预测中导出。而且或者代替地,神经网络被训练以生成任务预测,其中,任务预测可以表征分子的一个或多个预测特性,例如分子的平衡能量、分解分子所需的能量或分子的电荷。例如,任务是预测分子的一个或多个特性,诸如:结合态预测,例如原子结合紧密程度的测量,诸如分裂一个或多个分子所需能量的测量,或者键角或长度的测量;或者一个或多个分子的HOMO或LUMO能量;或者分子中的电子分布的特性,诸如大小、电荷、偶极矩或静态极化率。

[0091] 一个或多个分子的几何结构在局部能量最小值下的局部随机畸变几乎肯定是更高的能量配置。因此,在节点的去噪预测在通过使用噪声修改原子之前预测原子的初始空间位置的实施方式中,隐式地训练神经网络以通过初始结构来确定平衡或弛豫结构。

[0092] 如先前提及的,取决于应用,经训练的示图神经网络150可以通过使用从示图104的节点嵌入被解码的特征或者去噪预测或者任务预测来使用。

[0093] 一个示例应用涉及使用经训练的示图神经网络150来获得催化剂分子或与催化剂相互作用的分子。在该应用中,节点的特征表示包括定义催化剂分子或与催化剂相互作用的分子的结构和性质的特征。然后,来自示图神经网络的输出可以包括例如从示图的更新节点嵌入或从表示分子相互作用时的所得结构的去噪预测被解码的特征;和/或表征分子的所得状态的任务预测,例如分子的平衡能量,或由相互作用引起的能量改变,或者分解分子所需的能量。所得结构或表征分子所得状态的预测可以被用于获得催化剂分子或与催化剂相互作用的分子,例如通过筛选多个候选分子。筛选可以是识别以期望的方式——例如特别强烈地——相互作用的那些分子;或者筛选出不合适的分子;或者识别与多个其他分子相互作用的催化剂分子或与多个不同的催化剂分子相互作用的(可能是有用的或者是不想要的)分子。

[0094] 筛选过程可以例如涉及使用来自示图神经网络的输出来确定多个候选催化剂分子和/或与催化剂相互作用的候选分子中的每一个的分数;以及使用该分数来选择一个或多个候选。该方法可以进一步涉及制备催化剂分子或与通过该方法获得的催化剂相互作用的分子;以及可选地在真实世界中测试相互作用。

[0095] 在该应用的特定实例中,催化剂分子包括酶或酶的受体部分,并且与催化剂相互作用的分子是酶的配体,例如受体或酶的激动剂或拮抗剂。配体可以是例如药物或工业酶的配体。分子中的一个或两者可以包括蛋白质分子。

[0096] 进一步的相关应用涉及使用经训练的示图神经网络150来识别抑制病原体复制的药物分子,即,获得与病原体分子相互作用的药物分子。病原体分子是与病原体相关联的分子,当药物分子与病原体分子相互作用时,病原体的复制受到抑制。因此,在上述方法中,使用病原体分子代替催化剂分子。因此,节点的特征表示然后可以包括定义分子的特征,并且来自示图神经网络的输出被用于筛选候选药物分子和/或病原体分子以获得药物分子。该方法还可能涉及在真实世界中制造并且可选地测试针对病原体的药物分子。

[0097] 另一示例应用涉及使用经训练的示图神经网络150来确定化学反应的反应机制,以制备涉及两个或更多个分子相互作用的产物。然后一个或多个分子可以被修改以修改反应机制,例如提高反应速度或产物产率。然后可以使用反应机制来制备产物。节点的特征表示然后可以包括定义分子的特征。来自示图神经网络的输出可以包括例如从示图的更新节点嵌入或从表示分子相互作用时的所得结构的去噪预测被解码的特征;和/或表征分子的

所得状态的任务预测。例如,输出可以预测分子沿着反应坐标的一种或多种过渡态的能态、结合态和构象中的一个或多个。

[0098] 在以上应用中,节点的特征表示然后可以包括通过对真实世界分子进行的一次或多次度量确定的特征,例如使用电子显微镜来表征分子的结构或性质。以该方式获得的特征然后可以由经训练的示图神经网络处理以获得示图神经网络输出,例如表征分子的一个或多个特性的任务预测输出,例如平衡能量、结合态、键角或长度的测量;HOMO或LUMO能量或者大小、电荷、偶极矩或静态极化率。

[0099] 可以被用于训练示图神经网络150以执行以上任务的一些示例训练数据集是:The OC20 dataset (Chanussot et al., “The Open Catalyst2020 (OC20) Dataset and Community Challenges (开放催化剂2020 (OC20) 数据集和社区挑战)”, ACS Catalysis, 6059-6072, 2020, arXiv:2010.09990); the QM9 dataset, Ramakrishnan et al., “Quantumchemistry structures and properties of 134kilo molecules (134千分子的量子化学结构和特性)”, Sci Data 1, 140022 (2014); 来自Open Graph Benchmark的the OGBG-PCQM4M dataset, Hu et al., “Open Graph Benchmark: Datasets for Machine Learning on Graphs (开放示图基准: 示图上的机器学习数据集)”, arXiv:2005.00687; 以及同样来自the Open Graph Benchmark的OGBG-MOLPCBA。在一些实施方式中,示图104可以表示物理系统,示图104中的每个节点可以表示该物理系统中的相应对象,并且任务预测可以表征物理系统中的一个或多个对象的相应的预测未来状态,例如物理系统中的一个或多个对象中的每个对象在未来时间点的相应位置和/或速度。

[0100] 一个示例应用涉及使用经训练的示图神经网络150来预测物理系统的状态或控制物理系统。节点的特征表示可以包括通过对象确定的特征。这样的特征可以包括对象的质量或者惯性矩、位置、定向、线速度或角速度或加速度;边缘可以表示连接的或交互的对象,例如由关节连接的对象。从示图神经网络的输出——例如从更新节点嵌入、去噪预测108或任务预测被解码的特征——可以定义例如单个时间步或多个时间步上的滚动的物理系统中的对象的未来状态的预测。输出可以被用于提供动作控制信号,用于取决于未来状态来控制对象。例如,经训练的示图神经网络150可以被包括在模型预测控制(MPC)系统中,以预测物理系统的状态或轨迹,供控制算法在控制物理系统时使用,例如以最大化回报或最小化从未来状态预测的成本。

[0101] 在一些实施方式中,示图104可以表示点云(例如由激光雷达或雷达传感器生成),示图104中的每个节点可以表示点云中的相应点,并且任务预测可以预测由点云表示的对象类。

[0102] 在一些实施方式中,示图104可以表示文本的一部分,示图104中的每个节点可以表示文本部分中的相应词语,并且任务预测可以预测例如在文本部分中表达的情绪,例如积极的、消极的或中性的。

[0103] 在一些实施方式中,示图104可以表示图像,示图104中的每个节点可以表示图像的相应部分(例如图像的像素或区域),并且任务预测可以表征例如图像中描绘的对象的类。

[0104] 在一些实施方式中,示图104可以表示部分或完全自动驾驶车辆附近的环境,示图104中的每个节点可以表示环境中的相应代理(例如行人、骑自行车的人、车辆等)或环境的元

素(例如交通灯、交通标志、车道等),并且任务预测可以预测例如由示图中的节点表示的一个或多个代理的相应未来轨迹。例如,预测输出可以表征由示图中的节点表示的车辆代理将做出一个或多个可能的驾驶决策的相应似然性,例如直行、变道、左转或右转。在该示例中,为了预测由示图中的节点表示的代理的未来轨迹,系统可以仅针对表示该代理的节点处理更新节点嵌入,即,不针对示图中的其他节点处理更新节点嵌入。示图的边缘可以表示例如代理或元素的物理接近度或连通性;连通性可以被定义为存在连接代理或元素的路线,诸如道路或通路。例如,经训练的示图神经网络可以被用于控制真实世界环境中的机械代理。经训练的示图神经网络可以处理节点的特征表示,该特征表示包括例如针对每个代理或元素来表示环境的其他代理或元素的特征、其他代理或元素的类型以及生成用于控制代理的示图神经网络输出的另一代理或元素的位置、配置、定向、线速度或角速度或者加速度。

[0105] 在一些实施方式中,示图104可以表示社交网络(例如在社交媒体平台上),示图中的每个节点可以表示该社交网络中的相应的人,示图中的每个边缘可以表示例如社交网络中两个对应的人之间的关系(例如“追随者”或“朋友”关系),并且任务预测可以预测例如社交网络中的哪些人可能在未来执行某个动作(例如购买产品或参加活动)。

[0106] 在一些实施方式中,示图104可以表示道路网络,示图中的每个节点可以表示道路网络中的路线段,示图中的每个边缘可以表示两个对应的路线段在道路网络中被连接,并且任务预测可以预测例如穿过道路网络的指定路径所需的时间或者通过道路网络的指定路径上的交通量。

[0107] 在一些实施方式中,示图104可以是表示例如由神经网络模型执行的计算操作的计算示图,示图中的每个节点可以表示一组一个或多个相关计算(例如由一组一个或多个神经网络层执行的操作),并且示图中的每个边缘可以表示一组计算的输出被提供为另一组计算的输入。在这些实施方式中,任务预测可以预测例如应该执行与示图中的每个节点相对应的操作的相应计算单元(即,来自可用计算单元的集合),例如以最小化执行由示图定义的操作所需的时间。每个计算单元可以是例如相应的线程、中央处理单元(CPU)或图形处理单元(GPU)。因此,经训练的示图神经网络可以被用于执行将计算操作指派给物理或逻辑计算单元的任务。经训练的示图神经网络可以处理包括表示计算组的特征的节点的特征表示,以生成示图神经网络输出(例如从更新节点嵌入、去噪预测或任务预测被解码的特征),从而识别应该执行与示图中的每个节点相对应的操作的相应计算单元。

[0108] 在一些实施方式中,如上所述,示图104可以表示蛋白质,示图中的每个节点可以表示蛋白质的氨基酸序列中的相应氨基酸,并且示图中的每个边缘可以表示蛋白质中的两个对应氨基酸在蛋白质的结构中分离小于阈值距离(例如8埃(Angstrom))。在这些实施方式中,任务预测可以预测例如蛋白质的稳定性或蛋白质的功能。

[0109] 在一些实施方式中,示图104可以表示知识库,示图中的每个节点可以表示知识库中的相应实体,并且示图中的每个边缘可以表示知识库中的两个对应实体之间的关系。在这些实施方式中,任务预测可以预测例如与知识库中的一个或多个实体相关联的缺失特征。

[0110] 图2图示了可以由图1中的训练系统100执行的操作的示例。系统100可以通过使用神经网络150来训练示图神经网络150以生成去噪预测234。在一些实施方式中,系统100可

以通过使用神经网络150生成任务预测232来附加地训练神经网络150。

[0111] 如上面参照图1描述的,系统100可以通过使用等式(7)中定义的目标函数来训练神经网络。在任何训练迭代中,目标函数可以包括测量以下中的误差的项:(i)去噪预测(例如 $\mathcal{L}_{de-noising}$ ), (ii)任务预测(例如 $\mathcal{L}_{task}$ ),或(iii)两者。在一些实施方式中,系统100可以首先预训练神经网络150以仅基于去噪预测(例如 $\mathcal{L}_{de-noising}$ )来优化目标函数,然后训练神经网络150以基于去噪预测和任务预测来优化目标函数。

[0112] 该系统可以包括噪声引擎,该噪声引擎可以被配置为生成示图中的节点的最终特征表示。在一些实施方式中,该示图可以表示分子202,并且该示图中的每个节点可以表示分子202中的相应原子。在这样的情况下,噪声引擎可以处理示图中的每个节点的初始特征表示,其中节点的初始特征表示表示分子202中的对应原子的初始空间位置。噪声引擎可以生成节点的最终特征表示,其中节点中的一些或全部的最终特征表示是使用相应噪声生成的修改特征表示。例如,噪声引擎可以通过将噪声添加到由节点表示的分子202中的原子的初始空间位置来生成节点的最终特征表示。

[0113] 神经网络150可以包括:(i)编码器210,(ii)更新器220,以及(iii)解码器230。

[0114] 噪声引擎可以向编码器210提供示图中的节点的最终特征表示。编码器210可以被配置为生成定义示图的数据。例如,编码器210包括节点嵌入子网络,它可以基于例如每个节点的最终特征表示来生成每个节点的节点嵌入。

[0115] 在生成定义示图的数据之后,编码器210可以向更新器220提供定义该示图的数据。更新器可以包括一个或多个示图神经网络层,例如N个示图神经网络层。每个示图神经网络层可以被配置为基于节点的当前节点嵌入和示图中的节点的一个或多个邻居中的每个邻居的相应当前节点嵌入来更新示图中的一个或多个节点的当前节点嵌入。例如,如图2所图示的,示图神经网络层215可以基于邻居节点的当前节点嵌入来更新由实心圆示出的节点的节点嵌入。

[0116] 解码器230可以被配置为针对具有修改特征表示的一个或多个节点中的每个节点,处理节点的更新节点嵌入以为节点生成相应去噪预测234。去噪预测234可以表征该节点的去噪特征表示,它不包括用于生成该节点的修改特征表示的噪声。如上面参照图1描述的,去噪预测234可以预测用于生成节点的修改特征表示的噪声。在一些实施方式中,去噪预测234可以预测该节点的初始特征表示,例如在该节点被噪声破坏之前由该节点表示的原子在分子202中的初始空间位置。

[0117] 在一些实施方式中,节点的去噪预测234可以表征节点的目标特征表示。例如,如图2所图示的,目标特征表示可以指定原子弛豫之后原子在由节点表示的分子202中的最终空间位置。以该方式,系统100可以将分子202中的原子的(例如由节点的初始特征表示指定的)初始空间位置映射到分子202中的原子的最终空间位置。

[0118] 在一些实施方式中,解码器230可以生成任务预测232。通常,任务预测232可以是表征由示图中的节点表示的元素中的一个或多个元素的任何适当的预测。如图2所图示的,任务预测232可以是例如原子弛豫后分子202的平衡能量。

[0119] 训练神经网络150以生成去噪预测234可以鼓励神经网络150隐式学习“真实”示图的分布,即,具有未修改的节点特征表示,并且神经网络150可以充分利用该隐式知识来实现任务预测232的更高准确性。

[0120] 接下来更详细地描述用于使用训练系统100来训练神经网络150的示例过程。

[0121] 图3是用于使用训练系统来训练示图神经网络的示例过程300的流程图。为了方便起见,过程300将被描述为由位于一个或多个地点中的一个或多个计算机的系统执行。例如,根据本说明书适当编程的训练系统——例如图1的训练系统100——可以执行过程300。

[0122] 该系统生成定义示图的数据,该示图包括:(i) 节点集合,(ii) 每个节点的节点嵌入,以及(iii) 各自连接相应节点对的边缘集合(302)。如上面参照图1描述的,系统可以基于每个节点的特征表示来生成定义示图的数据。例如,该系统可以为每个节点获得相应的初始特征表示,并且为每个节点生成相应的最终特征表示。针对一个或多个节点中的每个节点,相应的最终特征表示可以是修改特征表示,它使用相应的噪声从节点的相应特征表示被生成,例如通过将相应的噪声添加到节点的相应特征表示来被生成,如等式(1)定义的。

[0123] 系统可以使用节点的相应的最终特征表示来生成定义示图的数据。例如,针对包括第一节点和第二节点的每对节点,系统可以确定第一节点的最终特征表示和第二节点的最终特征表示之间的相应距离。然后,系统可以确定与小于预定义阈值的距离相对应的每对节点由示图中的边缘连接。如图2所图示的,示图可以表示分子,并且示图中的每个节点可以表示分子中的相应原子。在该情况下,每个节点的初始特征表示可以表示例如对应原子在分子中的初始空间位置。该系统可以通过基于由节点表示的原子类型为每个节点生成节点嵌入来生成定义示图的数据。通常,示图可以表示任何适当的物理系统。

[0124] 在一些实施方式中,该示图可以进一步包括每个边缘的相应边缘嵌入。在这样的情况下,系统可以通过至少部分地基于由边缘连接的节点的相应最终特征表示之间的差异为示图中的每个边缘生成边缘嵌入来生成示图,例如如上面的等式(2)和等式(3)定义的。

[0125] 该系统使用神经网络的一个或多个示图神经网络层来处理定义示图的数据,以生成每个节点的相应的更新节点嵌入(304)。在一些实施方式中,神经网络包括至少10个示图神经网络层。每个示图神经网络层可以被配置为更新当前示图。

[0126] 例如,如上面参照图1描述的,每个神经网络层可以接收当前示图,并且根据示图神经网络层的当前神经网络参数值来更新当前示图。这可以包括例如基于以下来更新示图中的一个或多个节点中的每个节点的当前节点嵌入:(i) 节点的当前节点嵌入,以及(ii) 示图中的节点的一个或多个邻居中的每个邻居的相应当前节点嵌入,例如如上面的等式(4)和等式(5)定义的。在一些实施方式中,当前示图可以进一步包括每个边缘的边缘嵌入。在这样的情况下,每个神经网络层可以至少部分地基于连接至节点的一个或多个边缘中的每个边缘的相应边缘嵌入来更新节点的节点嵌入。

[0127] 针对具有修改特征表示的一个或多个节点中的每个节点,系统处理节点的更新节点嵌入以为节点生成相应去噪预测(306)。例如,系统可以使用一个或多个神经网络层来处理节点的更新节点嵌入,以为节点生成相应去噪预测。去噪预测可以表征该节点的去噪特征表示,它不包括用于生成该节点的修改特征表示的噪声。

[0128] 在一个示例中,去噪预测可以预测用于生成节点的修改特征表示的噪声。在另一示例中,去噪预测可以预测节点的初始特征表示,例如在通过使用噪声对原子进行修改之前原子在分子中的初始空间位置。在又一示例中,去噪预测可以表征节点的目标特征表示,例如原子弛豫后原子在分子中的新位置。在一些实施方式中,节点的目标特征表示可以是

节点的增量特征表示,如果该增量特征表示被添加到节点的修改特征表示,则产生节点的目标特征表示。

[0129] 系统确定对神经网络的神经网络参数的当前值的更新,以优化目标函数,该目标函数测量节点的去噪预测中的误差 (308)。例如,系统可以通过示图神经网络层的神经网络参数反向传播目标函数的梯度。针对神经网络的多个示图神经网络层中的每个示图神经网络层,目标函数可以测量节点的去噪预测中的相应误差,该去噪预测基于由示图神经网络层生成的更新节点嵌入。

[0130] 在一些实施方式中,该系统可以处理节点的更新节点嵌入以生成任务预测,其中目标函数还测量任务预测中的误差,例如如等式 (7) 定义的。通常,任务预测中的误差可以使用适合于任务的训练数据集来确定,即,可以使用监督学习来训练系统。在这样的情况下,系统可以处理以下两者: (i) 节点的更新节点嵌入,以及 (ii) 在使用示图神经网络层来被更新之前的节点的原始节点嵌入,以生成任务预测。作为特定示例,该示图可以表示分子,并且任务预测可以是分子的平衡能量的预测,例如如图2所图示的。

[0131] 接下来将更详细地描述使用用于训练神经网络的系统实现的示例实验结果。

[0132] 图4图示了使用上面参照图1和图2描述的用于训练神经网络的系统100实现的示例实验结果400。

[0133] 系统100可以通过使用神经网络来训练神经网络以生成去噪预测。如上所述,生成去噪预测需要每个节点嵌入对唯一信息进行编码,以便对节点的特征表示进行去噪,这可以减轻“过度平滑”的影响,例如节点嵌入在通过多个示图神经网络层处理后变得几乎相同。在图4中,“MAD”是可以量化“过度平滑”的节点嵌入的分集的测量,其中,数字越高指示节点嵌入的分集水平越高。如图4所图示的,与其他可用技术(例如“DropEdge”和“DropNode”)相比,本说明书中描述的系统能够贯穿神经网络中维持更高水平的节点嵌入分集。这在神经网络层15处尤其明显,其中,本说明书中描述的系统的节点嵌入的分集的测量高得多。因此,本说明书中描述的系统在减轻“过度平滑”的影响方面可以优于其他可用系统。

[0134] 图5图示了使用上面参照图1和图2描述的用于训练神经网络的系统100所实现的示例实验结果500。在图5中,“先前SOTA”是指使用其他可用系统实现的最先进的性能,并且“ev MAE”是指任务预测的预测误差。

[0135] 左手边示图示出,即使在3个消息传递步之后(例如在神经网络中具有3个示图神经网络层),本说明书中描述的系统也超过了使用其他可用系统实现的最先进的性能。右手边示图示出,即使利用较少数量的神经网络参数(例如在示图神经网络层之间具有共享权重),也可以由本说明书中描述的系统超过最先进的性能。

[0136] 图6图示了比较各种神经网络在预测分子的HOMO能量的任务上的性能的示例实验结果。具体地,示图600的横轴表示用于训练每个神经网络的梯度步的数量,并且示图600的纵轴表示每个神经网络的预测准确性。表现最好的神经网络(在图6中标记为“预训练的GNS-TAT”)是在被训练以执行HOMO能量预测任务之前被预训练以基于去噪预测(如本说明书中描述的)优化目标函数的示图神经网络。因此,这些实验结果图示了通过预训练示图神经网络来基于去噪预测优化目标函数可以实现的优点。

[0137] 本说明书使用与系统和计算机程序组件相关的术语“被配置”。一个或多个计算机

的系统被配置为执行特定操作或动作意味着该系统已经在其上安装了软件、固件、硬件或它们的组合,它们在操作中使该系统执行操作或动作。一个或多个计算机程序被配置为执行特定操作或动作意味着一个或多个程序包括在由数据处理装置实行时使所述装置执行操作或动作的指令。

[0138] 本说明书中描述的主题和功能操作的实施例可以被实施在数字电子电路系统、有形实施的计算机软件或固件、计算机硬件——包括在本说明书中公开的结构及其结构等效物——或者它们中的一个或多个的组合中。本说明书中描述的主题的实施例可以被实施为一个或多个计算机程序,即,在有形的非暂时性存储介质上编码以由数据处理装置实行或者控制数据处理装置的操作的计算机程序指令的一个或多个模块。计算机存储介质可以是机器可读存储设备、机器可读存储衬底、随机或串行存取存储器设备或者它们中的一个或多个的组合。可替代地或者另外,程序指令可以在人工生成的传播信号——例如机器生成的电信号、光信号或者电磁信号——上编码,生成该传播信号是为了对用于传输到合适的接收器装置以供数据处理装置实行的信息进行编码。

[0139] 术语“数据处理装置”是指数据处理硬件,并且涵盖了用于处理数据的所有种类的装置、设备和机器,通过示例包括可编程处理器、计算机或者多个处理器或计算机。该装置还可以是或者进一步包括专用逻辑电路系统,例如FPGA(现场可编程门阵列)或者ASIC(专用集成电路)。除了硬件之外,该装置还可以可选地包括为计算机程序创建实行环境的代码,例如构成处理器固件、协议栈、数据库管理系统、操作系统或者它们中的一个或多个的组合的代码。

[0140] 计算机程序(这也可以被称为或者描述为程序、软件、软件应用、应用、模块、软件模块、脚本或代码)可以用包括编译语言或解释语言或者声明性语言或过程语言的任何形式的编程语言来编写;并且它可以按照包括作为独立式程序或者作为模块、组件、子例程或者适合用于计算环境的其他单元的任何形式来部署。程序可以但不需要与文件系统中的文件相对应。程序可以被存储在文件的保持其他程序或数据——例如被存储在标记语言文档中的一个或多个脚本——的一部分中,被存储在专用于所探讨的程序的单个文件中,或者被存储在多个协作文件——例如存储一个或多个模块、子程序或者部分代码的文件——中。计算机程序可以被部署为在一个计算机上实行或者在位于一个站点处或跨多个站点分布并且通过数据通信网络互连的多个计算机上实行。

[0141] 在本说明书中,术语“引擎”被广泛地用于指代被编程为执行一个或多个具体功能的基于软件的系统、子系统或过程。通常,引擎将被实施为安装在一个或多个地点中的一个或多个计算机上的一个或多个软件模块或组件。在一些情况下,一个或多个计算机将专用于特定引擎;在其他情况下,多个引擎可以在相同的一个或多个计算机上安装和运行。

[0142] 本说明书中描述的过程和逻辑流程可以由一个或多个可编程计算机执行,所述一个或多个可编程计算机实行一个或多个计算机程序以通过操作输入数据并且生成输出来执行功能。过程和逻辑流程也可以由例如FPGA或ASIC的专用逻辑电路系统或者由专用逻辑电路系统和一个或多个编程计算机的组合执行。

[0143] 适合于实行计算机程序的计算机可以基于通用或专用的微处理器或两者或者任何其他种类的中央处理单元。通常,中央处理单元将从只读存储器或者随机存取存储器或者两者接收指令和数据。计算机的必要元件是用于执行或实行指令的中央处理单元以及用

于存储指令和数据的一个或多个存储器设备。中央处理单元和存储器可以由专用逻辑电路系统补充或者被并入到该专用逻辑电路系统中。通常,计算机还将包括用于存储数据的一个或多个大容量存储设备,例如磁盘、磁光盘或者光盘,或者计算机被可操作地耦合以从一个或多个大容量存储设备接收数据或者将数据传送给一个或多个大容量存储设备或者两者都有。然而,计算机不需要具有这样的设备。而且,计算机可以被嵌入到另一设备中,例如仅举数例,移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(GPS)接收器或者便携式存储设备,例如通用串行总线(USB)闪存驱动器。

[0144] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器设备,作为示例包括半导体存储器设备,例如EPROM、EEPROM和闪存设备;磁盘,例如内部硬盘或者可移除盘;磁光盘;以及CD-ROM盘和DVD-ROM盘。

[0145] 为了提供与用户的交互,本说明书中描述的主题的实施例可以被实施在计算机上,该计算机具有:用于向用户显示信息的显示设备,例如CRT(阴极射线管)或者LCD(液晶显示器)监测器;以及键盘和指向设备,例如鼠标或者轨迹球,用户可以通过该键盘和该指向设备来将输入提供给计算机。其他种类的设备也可以被用于提供与用户的交互;例如提供给用户的反馈可以是任何形式的传感反馈,例如视觉反馈、听觉反馈或者触觉反馈;并且来自用户的输入可以以任何形式——包括声学输入、语音输入或者触觉输入——来被接收。另外,计算机可以通过以下方式来与用户交互:将文档发送给由用户所使用的设备并且从由用户所使用的设备接收文档;例如通过响应于从web浏览器接收的请求来将网页发送给用户的设备上的web浏览器。而且,计算机可以通过将文本消息或其他形式的消息发送给个人设备——例如正在运行消息收发应用的智能手机——并且从用户接收响应消息作为回应来与用户交互。

[0146] 用于实施机器学习模型的数据处理装置还可以包括例如专用硬件加速器单元,用于处理机器学习训练或生产——即推理、工作负载——的常见的计算密集型部分。

[0147] 机器学习模型可以使用机器学习框架来被实施和部署,例如TensorFlow框架。

[0148] 本说明书中描述的主题的实施例可以被实施在计算系统中,该计算系统包括:后端组件,例如作为数据服务器,或者包括中间件组件,例如应用服务器,或者包括前端组件,例如具有图形用户界面、web浏览器或应用的客户端计算机,用户可以通过该图形用户界面、该web浏览器或该应用来与本说明书中描述的主题的实施方式交互,或者一个或多个这样的后端组件、中间件组件或前端组件的任何组合。系统的组件可以通过任何形式或介质的数字数据通信——例如通信网络——来互连。通信网络的示例包括局域网(LAN)和广域网(WAN),例如互联网。

[0149] 计算系统可以包括客户端和服务端。客户端和服务端通常远离彼此,并且典型地通过通信网络进行交互。客户端和服务端的关系借助于在相应计算机上运行并且彼此具有客户端-服务端关系的计算机程序来产生。在一些实施例中,服务端将例如HTML页面的数据传输给用户设备,例如用于向与设备交互的充当客户端的用户显示数据并且从该用户接收用户输入的目的。在用户设备处生成的数据,例如用户交互的结果,可以在服务端处从该设备接收。

[0150] 虽然本说明书包含了许多具体实施方式细节,但是这些细节不应该被解释为对任何发明的范围或者可能要求保护的内容的范围的限制,而是作为可以特定于特定发明的特

定实施例的特征的描述。在本说明书中在分开的实施例的上下文中描述的某些特征还可以以组合实施在单个实施例中。相反,在单个实施例的上下文中描述的各种特征也可以分开地或者按照任何合适的子组合被实施在多个实施例中。而且,虽然特征在上面可以被描述为以某些组合起作用,甚至描述为最初像这样要求保护,但是来自要求保护的组合的一个或多个特征在一些情况下可以从组合中切除,并且要求保护的组合可以涉及子组合或者子组合的变化。

[0151] 类似地,虽然操作按照特定次序在附图中描绘并且在权利要求中叙述,但是这不应该被理解为需要这样的操作按照所示的特定次序或者按照相继次序来执行,或者所有图示的操作被执行以实现期望的结果。在某些情况下,多任务处理和并行处理可能是有利的。而且,在上述实施例中的各种系统模块和组件的分离不应该被理解为在所有实施例中都需要这样的分离,并且应该理解,所描述的程序组件和系统通常可以被一起集成在单个软件产品中或者被封装到多个软件产品中。

[0152] 本主题的特定实施例已被描述。其他实施例在以下权利要求的范围内。例如,在权利要求中叙述的动作可以按照不同的次序来执行,并且仍然实现期望的结果。作为一个示例,附图中描绘的过程不一定需要所示的特定次序或相继次序来实现期望的结果。在一些情况下,多任务处理和并行处理可能是有利的。

## 训练系统 100

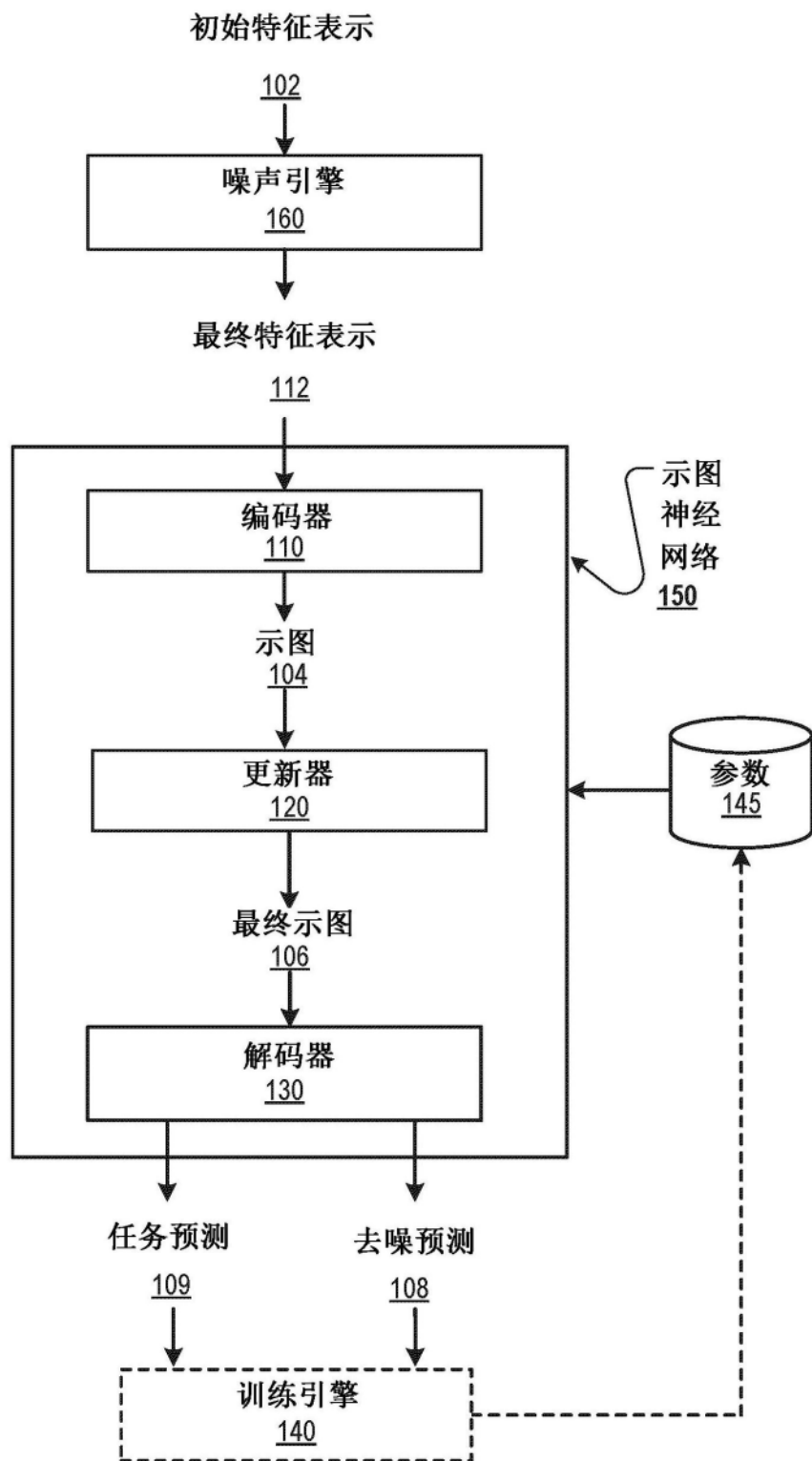


图1

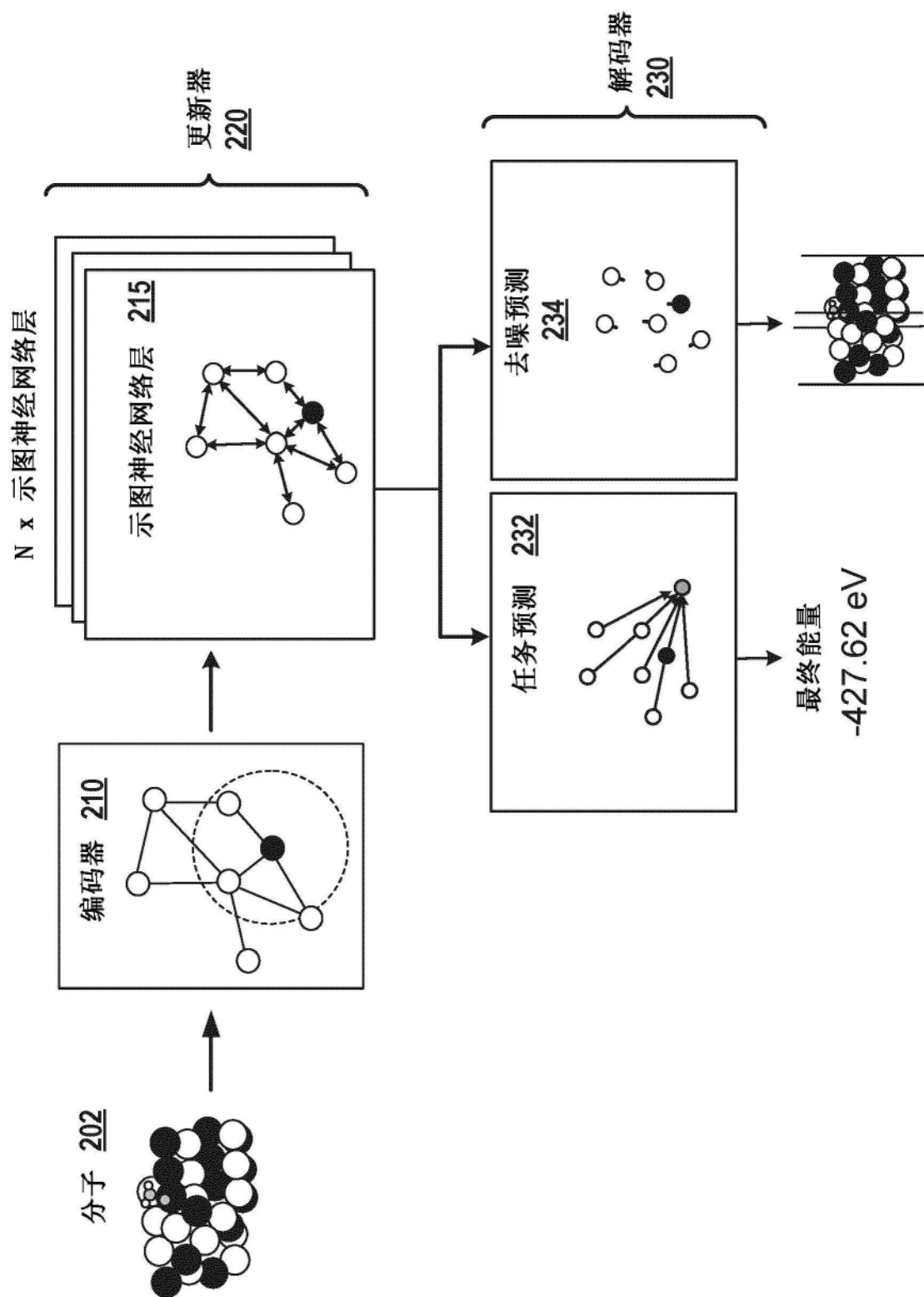


图2

## 训练示图神经网络

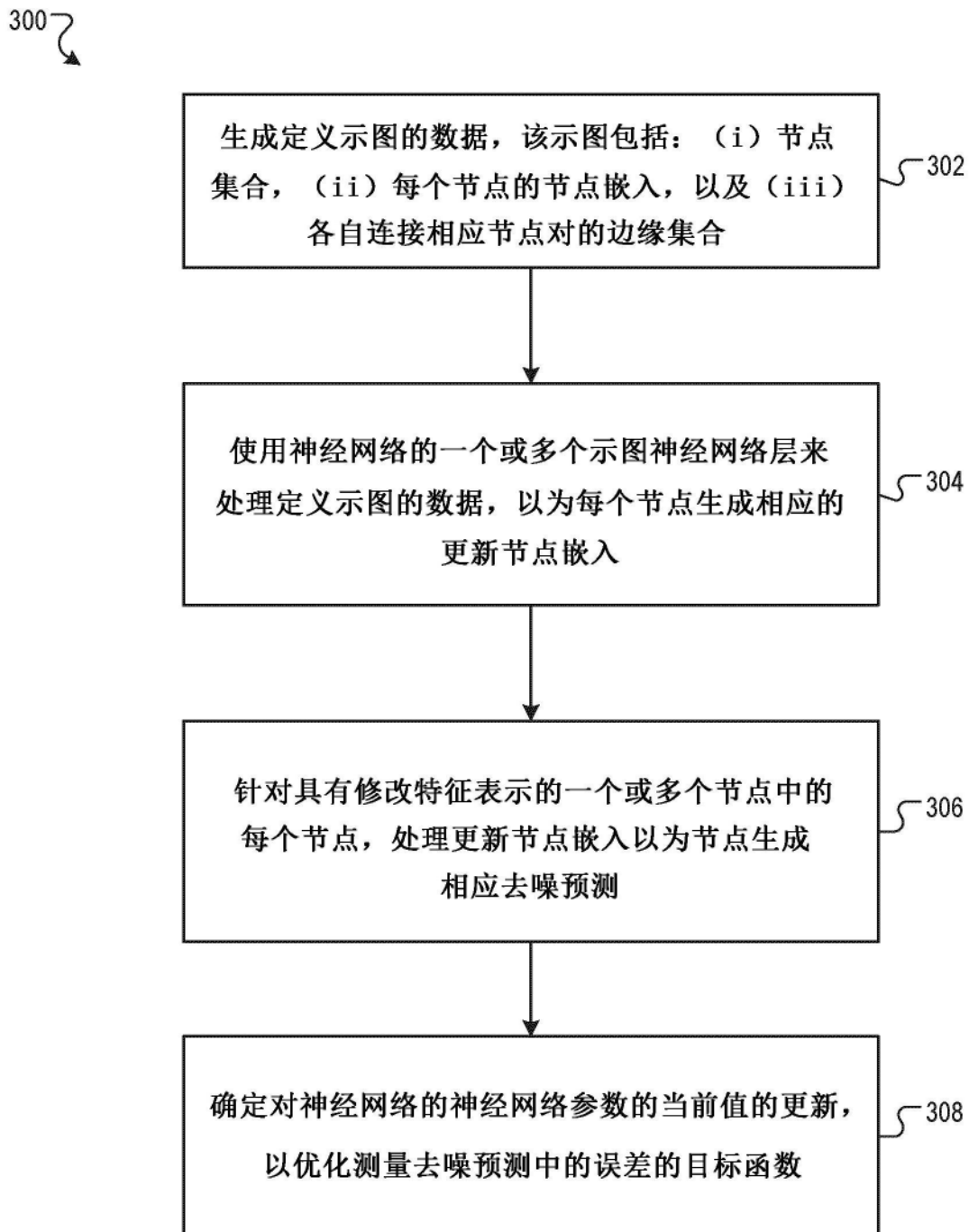


图3

示例实验结果 400

每层节点潜在MAD

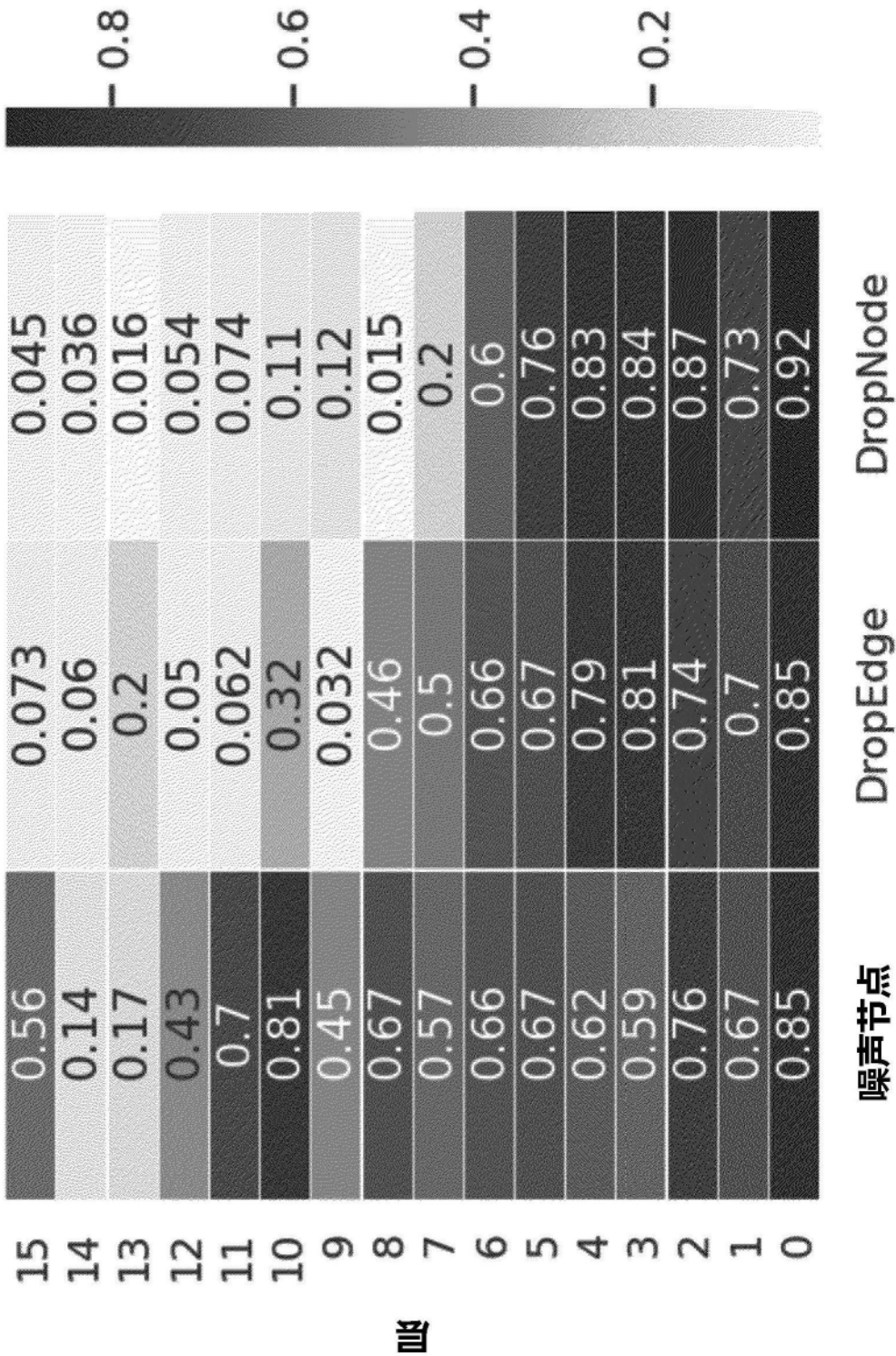


图4

示例实验结果 500

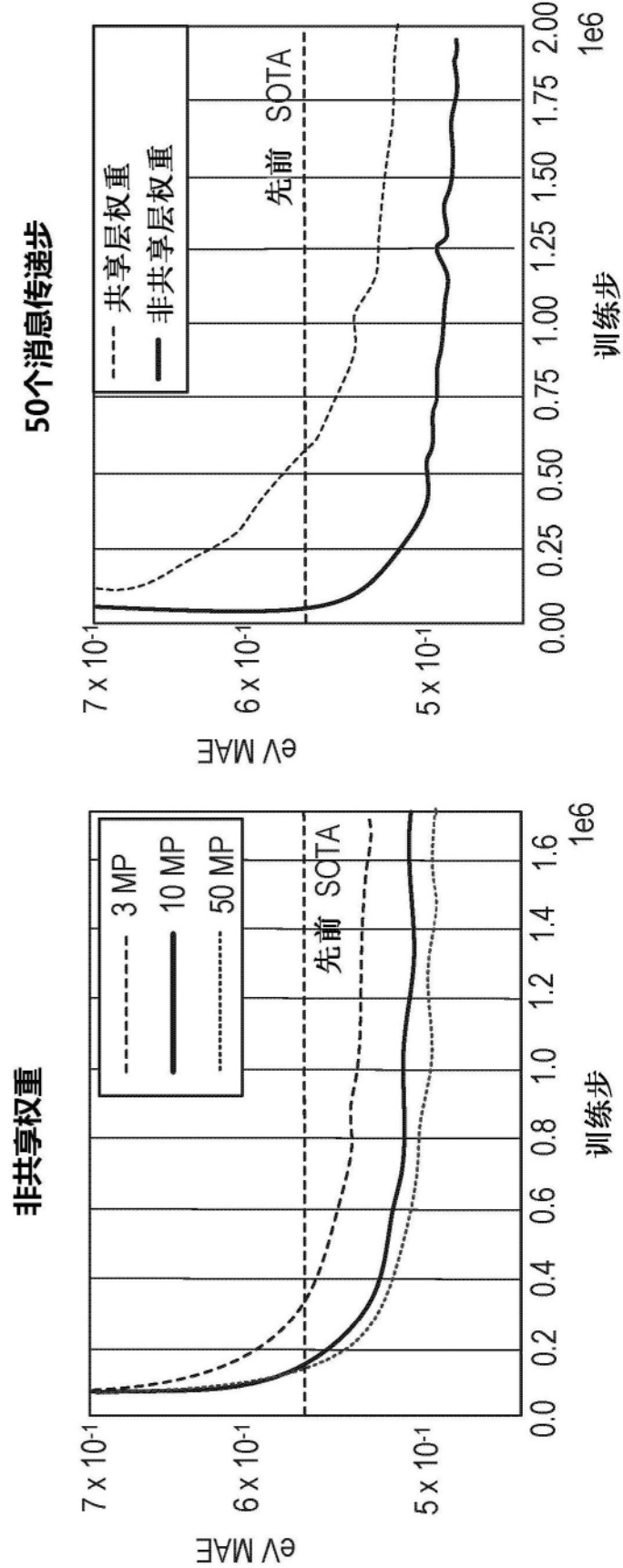


图5

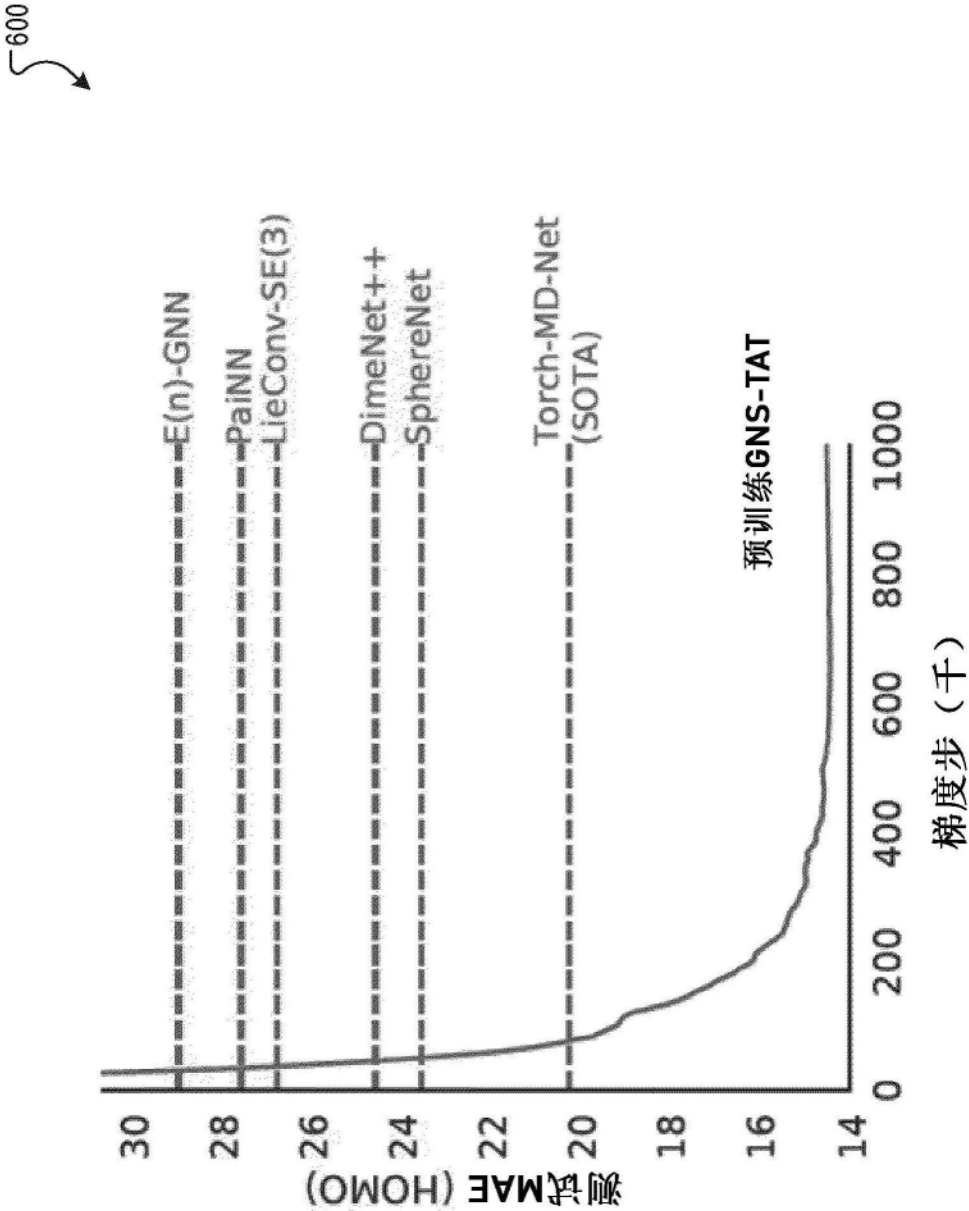


图6