

基于图神经网络的分子表示学习方法研究综述

汪浩¹⁾

¹⁾(华中科技大学计算机科学与技术学院 武汉 430037)

摘 要 构建适当的分子表示是材料科学、化学和药物设计等众多任务的核心，机器学习技术最近被用于医学、生物学、化学和材料工程的各种应用，其中一个重要的任务是学习分子表示以预测分子的性质，这是许多下游应用（如虚拟筛选和药物设计）中的主要子任务。最近的研究将分子抽象为图结构数据，并使用图神经网络（GNN）进行分子表示学习。图神经网络（GNN）被认为是一种有吸引力的分子特性预测建模方法，大量研究表明，与传统的基于描述符的方法相比，GNN 在分子图建模方面取得了显著成果。分子数据中的原子、化学键之间有丰富复杂的结构、语义信息，分子结构可以自然地抽象为具有属性节点和边的拓扑图，使用 GNN 进行分子表示学习主要问题是如何处理节点、边缘信息，以及如何处理节点与节点、节点与边的交互信息。基于图神经网络的分子表示学习方法主要经过了四个发展阶段和类型，只考虑节点、同时考虑节点和边缘信息、考虑节点和边缘和两者之间的交互信息、还考虑 3D 图上未成键但彼此接近的原子。本文首先介绍了基于图神经网络的分子表示学习方法的研究背景与意义、存在的挑战，然后介绍基于图神经网络的分子表示学习方法的原理与优势，然后按照发展阶段、侧重方向归纳该方向上的近期主要成果，最后对本文进行总结并对未来研究方向进行展望

关键词 图神经网络；分子表示学习；注意力机制；分子特性预测

A Survey of Research on Molecular Representation Learning Methods Based on Graph Neural Networks

WANG Hao¹⁾

¹⁾(School of Computer Science and Technology, Huazhong University of Science and Technology, 430037, China)

Abstract Constructing appropriate molecular representations is the core of many tasks in materials science, chemistry, and drug design. Machine learning technology has recently been used in various applications in medicine, biology, chemistry, and materials engineering. One of the important tasks is to learn molecular representations. Predicting the properties of molecules is a major subtask in many downstream applications such as virtual screening and drug design. Recent research abstracts molecules into graph structure data and uses graph neural networks (GNN) for molecular representation learning. Graph Neural Network (GNN) is considered to be an attractive predictive modeling method for molecular properties. A large number of studies have shown that, compared with traditional descriptor-based methods, GNN has achieved significant results in molecular map modeling. There are rich and complex structure and semantic information between the atoms and chemical bonds in molecular data. The molecular structure can be naturally and simply abstracted as a topological graph with attribute nodes and edges. The main problem of using GNN for molecular representation learning is how to deal with node and edge information. , And how to deal with node-to-node, node-to-edge interaction information. The molecular representation learning method based on graph neural network has mainly gone through four development stages and types. It only considers nodes, considers node and edge information at the same time, considers nodes and edges and the interactive information between the two, and also considers unbonded 3D graphs. But atoms close to each other. This article first introduces the research background and significance of the molecular representation learning

method based on graph neural network, and the existing challenges, and then introduces the principles and advantages of the molecular representation learning method based on graph neural network, and then summarizes the technical direction according to the development stage and focus direction. The main recent results of the above, and finally the summary of this article and prospects for future research directions.

Key words Graph neural network; molecular representation learning; attention mechanism; molecular property prediction

1 引言

分子特性的准确表征有助于寻找具有所需药理学和吸收、分布、代谢、排泄和毒性特性的化学品,是经典的化学信息学任务之一,但目前在很大程度上仍然是一个开放的挑战,其解决方案可能会开启深度学习在药物发现行业中的广泛应用。

一个经典的例子是药物研发的虚拟筛选。药物研发的传统物理筛选(即根据通过物理实验测试的特性选择分子)通常是准确和有效的,但非常昂贵和缓慢。相比之下,虚拟筛选(即通过机器学习方法根据预测的特性选择分子)可以在几分钟内完成预测数百万个分子。因此,它可以是物理实验之前的一个很好的过滤筛选环节,有助于加速药物研发过程并显著降低资源需求,筛选的成功率则取决于分子表示学习算法的预测性能。

分子表示学习的本质是将分子 m 通过表示函数 $h = g(m)$ 转换为密集特征向量,然后应用各种预测算法根据 $y = f(h)$ 预测目标属性。

其中最关键的挑战是,这些应用程序中的原始数据通常不能直接被现有的学习算法很好地处理,因此需要仔细构建合适的表示。与机器学习(尤其是深度学习)取得重大成就的图像或文本数据不同,分子属性预测问题中最常见的原始输入仅提供化学物质的高度抽象表示(即具有原子属性的原子图)。

为了应对这一挑战,人们提出了各种表示方法,主要分为两类。第一类是早期的预测建模方法,例如定量结构-性质关系(QSPR),是基于固定表示法进行的,例如专家制作的物理化学描述符和分子指纹,原型是摩根指纹。然而,基于描述符的方法假定所有与目标属性相关的信息都被选定的描述符集覆盖,从而限制了模型做出特定问题决策的能力。

第二类是图神经网络(GNN),自然地将分子结构抽象为具有属性节点和边的拓扑图,其中节点特征对应于原子特性,如原子身份和程度,边特征对应于化学键特性,如化学键类型和芳香性,所以可以直观地引入图表示模型,尤其是GNN来学习分子的表示。

大量研究表明,与传统的基于描述符的方法相比,GNN在学习适当的分子表示方面取得了显著成果。

基于图神经网络的分子表示学习方法,以分子图编码的基本化学信息作为输入,通过直接在编码的分子图上进行卷积操作来学习分子表示。图 $G=(V, E)$ 可以定义为一组节点(V)和一组边(E)之间的连接关系。自然地,分子也可以被认为是由一组原子(节点)和一组键(边)组成的图。GNN的一个关键特性是它能够使用图卷积自动学习特定于任务的表示,同时不需要传统的手工描述符和/或指纹。

目前几乎所有容易获得的原子/键级特征都被穷尽,将化学信息综合压缩到基于图的模型的分子图中,其中包括九种原子特征(即原子符号,原子度,形式电荷,自由基电子,杂化,芳香性,氢,手性和手性类型)和四种键特征(即化学键类型,共轭,环和立体)。它们中的大多数以onehot方式编码成分子图,随后将编码的分子图用作输入。

同时在目前的化学分子表示研究中,有两个常用的概念SMILES和RDKit出现在分子初始特征提取过程中。简化分子线性输入规范(英语:Simplified molecular input line entry specification,简称SMILES),是一种用ASCII字符串明确描述分子结构的规范。SMILES用一串字符来描述一个三维化学结构,将化学结构转化成一个生成树,SMILES字符串可以被大多数分子编辑软件导入并转换成二维图形或分子的三维模型。目前化学分子表示研究相关数据集通常提供化学分子的SMILES表示。

RDKit 是一个用于化学信息学的开源工具包，基于对化合物 2D 和 3D 分子操作，利用机器学习方法进行化合物描述符生成，fingerprint 生成，化合物结构相似性计算，2D 和 3D 分子展示等。RDKit 目前常用于提取化学分子的初始特征。

尽管取得了丰硕的进展，但仍有几个问题阻碍了图神经网络对分子表示学习在实际场景中的应用：（1）常见图神经网络对节点、边以及之间交互的信息考虑不是很充分；（2）常见的图卷积操作仅聚合局部信息，并且在堆叠过多 GNN 层时会遇到假死问题，因此这些模型难以学习每个原子的长程依赖性、分子图上的高阶属性；（3）分子表示学习任务的数据标记往往费时费力，可以使用的标记数据不足，模型容易过拟合；

（4）分子在实际中是三维立体的，但分子图只能考虑二维平面上的结构信息。

本文第二节介绍图神经网络和分子表示学习的原理，以及用图神经网络训练学习分子表示的优势；第三节介绍按照发展阶段、侧重方向归纳该技术方向上的近期主要成果，并分析其分别解决了什么问题、思路和优缺点；第四节对本文进行总结并对未来研究方向进行展望。

2 原理和优势

一般来说，GNN 框架的过程可以概括为三个主要步骤：（1）初始化步骤，节点用它们的初始属性或结构特征进行初始化；（2）消息传递步骤，每个节点的特征从它的邻居跨分子图传输到一个消息向量中；（3）消息聚合部分，其中节点消息被聚合或汇集成一个固定长度的特征向量。

在上述框架下，已经提出了许多用于有效图表示学习的 GNN 架构。

图卷积包括谱域图卷积和空域图卷积两种：谱域图卷积是根据图谱理论和卷积定理，将数据由空域转换到谱域做处理，有较为坚实的理论基础；空域图卷积则不依靠图谱卷积理论，直接在空间上定义卷积操作，定义直观，灵活性更强。

图卷积神经网络(GCN 即 Graph Convolutional Networks)于谱域图卷积的一种，其核心思想就是在一个 Graph 上定义一个关于 Graph 的卷积操作，实现这一过程需要借助图的拉普拉斯矩阵。

GCN 的主要思想是，对于一个图 $G=(V,E)$ ，输入 X 是一个 $N \times D$ 的矩阵，表示每个节点的特征，

同时有图的邻接矩阵 A 。最后得到一个 $N \times F$ 的特征矩阵 Z ，表示我们学得的每个节点的特征表示，其中 F 是希望得到的表示的维度。

对于 L 层的神经网络来说，可以表示为： $H_{l+1} = f(H_l, A)$ ，其中 $H_1 = X$ ， $H_L = Z$ 。

图注意力网络（GAT 即 Graph Attention Networks）属于空域图卷积的一种，认为卷积可定义为利用注意力机制（attention）对邻域节点进行有区别的聚合。注意力机制模仿了人类观察物品的方式，核心逻辑就是“从关注全部到关注重点”。GAT 模型认为邻域中所有的节点共享相同卷积核参数会限制模型的能力，因为邻域内的每一个节点和中心节点的关联度都是不同的，在卷积聚合邻域节点信息时需要 对邻域中的不同的节点区别对待。

图注意力层的输入是，所有节点特征向量的集合 $h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$ ， $\vec{h}_i \in \mathbb{R}^F$ ，其中 N 为节点集合的里节点的个数， F 为节点特征的维度。矩阵 h 的大小是 $N \times F$ ，代表了所有节点的特征，而 \mathbb{R} 只代表了某一个节点的特征，所以它的大小为 $F \times 1$ 。输出同理。

每一层的输出是：得到更新后的节点特征向量集 $h' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}$ ， $\vec{h}'_i \in \mathbb{R}^{F'}$ 其中 F' 表示新的节点特征向量维度（可以不等于 F ）。

GAT 与 GCN 同样也是一个特征提取器，针对的是 N 个节点，按照其输入的节点特征预测输出新的节点的特征。

为了对输入与输出特征维度进行转换，GAT 需要根据输入的特征至少进行一次线性变换得到输出的特征，所以我们需要一个参数矩阵： $W \in \mathbb{R}^{F' \times F}$ ，这个参数矩阵经过训练可以表示输入特征与输出特征之间的关系。

计算注意力系数： $e_{ij} = a(W\vec{h}_i, W\vec{h}_j)$ ，其中 a 不是一个常数或是矩阵， $a(\cdot)$ 是一个自己定义好的函数。

GAT 引入了 softmax 对所有的 i 的邻居节点 j 的特征向量进行正则化：

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (1)$$

最终结果为

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i \parallel W\vec{h}_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i \parallel W\vec{h}_k]))} \quad (2)$$

通过上述运算得到了中心实体对所有邻居节点的注意力系数,可以用来计算每个节点的更新后的输出特征: $\vec{h}_i = \sigma(\sum_{j \in N_i} \alpha_{ij} W \vec{h}_j)$

图神经网络强大之处在于:它们可以捕获分子的综合信息,包括骨架结构、构象信息和原子特性;他们接受端到端的训练,能够使用图卷积自动学习特定于任务的表示,可能会产生更好的预测表示。

本质上, GNN 旨在通过将来自原子特征向量编码的相邻原子的信息和键特征向量编码的连接键的信息通过递归穿过分子图的消息聚合来学习每个原子的表示,然后是中心原子的状态更新。最后,学习到的原子表示可用于预测分子特性。

目前一些模型尝试集成 transformer 和图神经网络,用 transformer 结构考虑分子图中的高阶属性和替代传统的注意力机制。

Transformer 编码器框架上,其中注意力模块是主要的构建块。注意模块的通常实现是点积自注意,它采用一组查询、键和值 (q, k, v) 的输入,这些查询、键和值 (q, k, v) 从隐藏节点特征 $h(X)$ 投影出来。然后它使用所有键计算查询的点积,并应用 softmax 函数来获得值的权重。通过将 (q, k, v) 的集合堆叠成矩阵 (Q, K, V), 它允许高度优化的矩阵乘法运算。具体来说,输出可以表示为:

$$[Q, K, V] = h(X)[W^Q, W^K, W^V] \quad (3)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (4)$$

公式 (4) 中 d 是 Q 和 K 的维度,而且 transformer 多采用多头注意力机制,其中 1 为总共堆叠的注意力层数,输出可以表示为:

$$Multihead(Q, K, V) = Cat(head_1, \dots, head_l)W^O \quad (5)$$

$$head_i = Attention(h(X)W_i^Q, h(X)W_i^K, h(X)W_i^V) \quad (6)$$

3 研究进展

基于图神经网络的化学分子表示学习方法主要经历了三个阶段:(1)只考虑节点信息;(2)同

时考虑节点和边缘信息;(3)考虑节点、边缘和节点-边缘交互信息。在这个过程中,表示学习还分为无监督和有监督两种,各有优缺点,但都可以为下游任务提供良好的嵌入。

因为化学表达式中原子和化学键数量有限,往往一个原子在不同的分子中携带的信息也有所区别,所以如何提高表示学习方法的泛化能力,就成了很重要的一个问题。接下来介绍三个阶段经典三个基于图神经网络的分子表示学习模型。

3.1 N-gram Graph

N-gram Graph 是一种比较早的无监督的基于图神经网络的分子表示学习方法,它简单、高效,但可以应用于不同学习方法的紧凑表示。

该方法的灵感来自最近的词嵌入方法和自然语言处理中的传统 N-gram 方法,可以表述为一个简单的图神经网络。

在自然语言处理 (NLP) 中, n-gram 指的是连续的单词序列。例如,句子“the dataset is large”的 2-grams 是{“the dataset”, “dataset is”, “is large”}。N-gram 方法为句子构造一个表示向量 $c(n)$, 其坐标是对应的 n-gram 在句子中出现的次数。n-gram 表示已被证明是一个强大的基线模型,但一个明显缺点是它的维度很高,同时很稀疏,可以通过使用词嵌入来缓解。从理论上和经验上都表明,即使使用随机词向量,也可以为下游学习任务保留良好的信息。

N-gram 图方法包括两个步骤:首先嵌入顶点,然后基于顶点嵌入嵌入图。

将分子视为图形,其中每个原子是一个顶点,每个键是一个边。假设图中有 m 个顶点,每个顶点都有有用的属性信息,如分子图中的原子符号和电荷数。这些顶点属性被编码成一个大小为 $m \times s$ 的顶点属性矩阵 V, 其中 s 是属性的数量。

在图中嵌入顶点的典型方法是将每个顶点视为一个标记,并应用类似 CBoW 或其他词嵌入方法。在这里, N-gram Graph 提出了一种变体,它利用了化学原子属性多为离散值的特点。

假如总共有 s 个属性,对于第 j 个属性用一组大小为 K_j 的值表示,则用 h_{ij} 表示编码顶点 i 的第 j 个属性的 one-hot 向量, h_i 由 s 个属性下的向量拼接得到。然后给定一个嵌入维度 r, 学习顶点嵌入矩阵 W 即可。

$$f_i = Wh_i \quad (7)$$

然后给的学习任务：让链接到节点 i 的顶点集可以经过网络后的输出可以匹配节点 i 的嵌入，此过程是无监督的。而且具有一定泛化能力，从一个数据集学习到的 W 可以用于另一个数据集。此外，即使使用随机顶点嵌入也可以提供合理的性能。

N-gram Graph 方法的灵感来自 NLP 中的 N-gram 方法，将其从线性图（句子）扩展到一般图（分子）上。它将图视为“游走袋”：并在其上构建表示。

设 n-gram 指的是图中长度为 n 的游走，n-gram 游走集指图上所有长度为 n 的游走的集合。n-gram p 的嵌入是该游走中顶点嵌入的元素乘积。即，如果 $p = (1, 2, 4)$ ，则 $f_p = f_1 * f_2 * f_4$ 。

最终将整个分子中所有长度为 n 的 n-gram 得到嵌入 f_p 求和得到图在游走长度 n 上的嵌入。如果 n 最大值为 T ，则将 T 个嵌入拼接得到最后的图嵌入。

$$f_p = \prod_{i \in p} f_i \quad (8)$$

$$f_{(n)} = \sum_{p: n\text{-gram}} f_p \quad (9)$$

$$f_G = [f_{(1)}; \dots; f_{(T)}] \quad (10)$$

通过构造，N-gram 图是置换不变的，即对分子中原子顺序的排列不变。此外，它是无监督的，因此可用于同一数据集上的不同任务，以及不同的机器学习模型。

3.2 GROVER

另一项值得注意的工作是 GROVER，其为分子表征学习提供了一个自监督的、预训练的基于 Transformer 的神经网络，以定制的 GNN 作为自注意力块。其中的 GNN 能够捕获图数据中的结构信息以及节点和边缘消息传递路径上的信息流。

GROVER 认为有两个问题阻碍了 GNN 在实际场景中的使用：(1) 用于监督训练的标记分子不足；(2) 对新合成分子的泛化能力差。

为了解决这两个问题，GROVER 能自监督地通过消息传递 transformer 结构学习分子表示。通过精心设计的节点级、边级和图级的自监督任

务，GROVER 可以从大量未标记的分子数据中学习分子的丰富结构和语义信息。

此外，我们在定制的 GNN 中引入了动态消息传递方案，事实证明该方案可以提高 GROVER 模型的泛化性能。

GROVER 由两个模块组成：节点 GNN 变换器和边缘 GNN 变换器。它们具有类似的结构。

节点 GNN 变换器 GTransformer 的关键组件是图多头注意力，它是针对结构化输入数据量身定制的注意力块。一个普通的注意力块，需要矢量化输入。然而，图输入是未矢量化自然结构数据。设计了一个定制的动态消息传递网络（dyMPN）以从图的节点中提取向量作为查询、键和值，然后将它们输入到注意力块中。

可以利用高表现力的 GNN 模型来更好地模拟分子数据中的结构信息。GTransformer 的高表现力可以归功于它的双级信息提取框架。消息传递过程捕获了图的局部结构信息，因此使用 GNN 模型的输出作为查询、键和值会得到所涉及的局部子图结构，从而构成第一级信息提取。

同时，Transformer 编码器可以看作是节点集合构造的全连接图上 GAT 的变体。因此，在这些查询、键和值之上使用 Transformer 编码器可以提取之间的全局关系节点，这可以实现第二级信息提取。这种双层信息提取策略极大地增强了 GROVER 模型的表示能力。

此外，GTransformer 应用来自输入特征的单个远程残差连接将初始节点/边缘特征信息直接传送到 GTransformer 的最后一层，而不是原始 Transformer 架构中的多个短程残差连接。从这个单一的远程残差连接可以获得两个好处：i) 像普通的残差连接一样，它通过缓解梯度消失问题来改进训练过程，ii) 与 Transformer 中的各种短程残差连接相比，我们的远程残差连接可以缓解消息传递过程中的过度平滑问题。

GROVER 还改进了消息传递机制，设计了一种随机策略的消息传递网络，称为动态消息传递网络（Dynamic Message Passing Network, dyMPN）：一般的消息传递过程，有两个超参数，迭代次数 L 和跳数 K ，跳数与图卷积运算的感受野大小密切相关，会影响消息传递模型的泛化能力。

给定固定的层数 L ，预先指定的跳数可能不适用于不同类型的数据集。于是该论文开发一种随

机策略, 代替预先指定的 K , 实验发现均匀分布和带界的正态分布随机化的效果很好。

dyMPN 使得图卷积运算中的每个节点都具有随机感受野。大量的实验验证表明, dyMPN 比没有随机化策略的普通 MPN 具有更好的泛化性能。

预训练模型的成功关键取决于自监督任务的设计。为了避免下游任务的负迁移, GROVER 在预训练中不使用监督标签, 并在这两个级别上提出新的自监督任务: 上下文属性预测和图级主题预测。

该论文提出一个原则, 一个好的节点级自监督任务应该满足以下性质: 1) 预测目标可靠且容易获得; 2) 预测目标应该反映节点/边的上下文信息。

于是, 该论文没有用常规的像 bert 一样的单节点 mask 预测任务, 而连续 mask 一些节点和边, 然后让模型去预测出来, 这样做增大了语义空间, 使得模型的学习能力更强。

图级自监督任务也需要可靠廉价的标签。模体 (motif) 是输入图数据中的循环子图, 在分子图数据中普遍存在。分子中一类重要的模体是功能团 (functional groups), 它编码分子丰富的领域知识, 可以很容易地被专业软件检测到, 如 RDKit。

模体预测任务可以表述为一个多标签分类问题, 其中每个模体对应一个标签。对于一个特定的分子 (抽象为图 G), 使用 RDKit 检测每个模体是否出现在图 G 中, 然后将它们用作模体预测任务的目标。

在使用设计的自监督任务在大量未标记数据上预训练 GROVER 模型后, 人们应该获得一个高质量的分子编码器, 它能够输出节点和边的嵌入。这些嵌入可以通过微调过程用于下游任务。各种下游任务可以从预训练的 GROVER 模型中受益。它们大致可以分为三类: 节点级任务, 例如节点分类; 边缘级任务, 例如链接预测; 和图形级任务, 例如分子的属性预测。

3.3 CoMPT

主要思路是引入 Transformer 以学习图结构化数据中的长期依赖关系, 加强节点与边的信息交互, 使用节点边缘消息交互模块计算权重并传递消息, 而不是自注意力机制, 使用具有扩散机制的拓扑连接矩阵来改进消息传递过程, 以减少消息爆炸。

主要包括五点:

CoMPT 需要三个输入: 节点特征 X 、边缘特征 E , 以及使用邻接矩阵 A 。根据节点的原子索引训练一个可学习的位置嵌入向量, 然后将位置向量与原始原子特征得到新的节点特征。边缘特征与源节点特征相加得到带有方向的边特征。

$$h(x_i) = \text{node_embedding}(x_i) + \text{pos}_i \quad (11)$$

$$h(e_{uv}) = \text{edge_embedding}(e_{uv}) + h(x_u) \quad (12)$$

CoMPT 背后的关键思想是我们使用隐藏节点特征 $h(X)$ 和隐藏边缘特征 $h(E)$ 来计算消息交互分数 M , 它取代了原始编码器层中的自注意力分数。具体来说, 隐藏节点特征 $h(X)$ 与查询 Query 投影矩阵、值 Value 投影矩阵相乘, 隐藏边缘特征 $h(E)$ 与键 K 投影矩阵相乘, 分别得到 Query Value Key 矩阵, 消息交互矩阵 T 由 Query Key 矩阵内积生成: 这个张量积意思是: 通过计算每个节点和边之间的标量积, 以生成与分子相关的结构信息, 用于最终预测。

$$[Q, V] = h(X)[W^Q, W^V] \quad (13)$$

$$K = h(E)W^K \quad (14)$$

$$T = \text{matmul}(Q, K.\text{transpose}(2, -1)) \quad (15)$$

为了保持分子图的连通性, 根据节点的邻居选择三种类型的矩阵: 节点与其输出边交互

(Mo)、节点与其输入边交互 (Mi) 和节点与其自环边交互 (Md)。在计算三个矩阵后, 使用 softmax 函数将它们归一化, 并合并计算得到最终消息交互矩阵 M , 这个操作消除了重复计数的自循环消息, 避免了信息爆炸。

用消息交互矩阵 M 更新节点隐藏特征 $h(X)$ 和边缘隐藏特征 $h(E)$, 同时为了使训练步骤更加稳定, 使用残差连接以减少梯度的消失。

$$h(X) = \text{matmul}(M, V) \quad (16)$$

$$h(E) = M \odot K \quad (17)$$

$$h_{k+1}(X) = h_k(X) + \text{Encoder}(h_k(X), h_k(E)) \quad (18)$$

$$h_{k+1}(E) = h_k(E) + \text{Encoder}(h_k(X), h_k(E)) \quad (19)$$

使用高斯核函数加上衰减系数，对消息传递根据节点距离远近进行一定程度衰减。

应用这种机制后，我们可以在一定程度上推迟聚合过程的过度平滑。

$$M(u, v) = M(u, v)e^{-\alpha A(u, v)} \quad (20)$$

对于图级任务，采用门控循环单元 (GRU) 进行全局池化以获得分子级别的固定特征向量

$$z = \sum_{x \in \mathcal{V}} \text{GRU}(h(x)) \quad (21)$$

CoMPT 的本质是利用节点和边之间注意力分数从中心节点周围的节点和边获取不同权重的特征来更新中心节点，在此过程中，还考虑节点之间的距离，如果距离远就进行一定程度的衰减。

4 总结和展望

图神经网络 (GNN) 被认为是一种有吸引力的分子特性预测建模方法，大量研究表明，与传统的基于描述符的方法相比，GNN 在分子图建模方面取得了显著成果。

GNN 的一个关键特性是它能够使用图卷积自动学习特定于任务的表示，同时不需要传统的手工描述符和/或指纹。

N-gram Graph 可以简单高效地学习分子表示，该方法的灵感来自最近的词嵌入方法和自然语言处理中的传统 N-gram 方法，但它只考虑了的节点信息，且对于所有节点给与相同的权重。

GROVER 探索了大规模预训练 GNN 模型的潜力，凭借精心设计的自监督任务和大量表达架构，可以从巨大的未标记图中学习丰富的隐式信息。GROVER 验证了自监督预训练方法在图学习领域的能力。尽管取得了成功，但 GNN 预训练在以下方面仍有探索的空间：更多的自监督任务，更多的下游任务，训练超大型 GNN 模型时可能存在梯度消失和过度平滑。

CoMPT 通过基于 Transformer 模型加强节点和边之间的消息交互来改善分子表示，还引入了一种消息扩散机制来衰减消息丰富爆炸以及消息

传递过程中的过度平滑，同时考虑了节点位置和边的方向。

分子在实际中是三维立体的，但目前基于图神经网络的分子表示学习方法只能考虑二维平面上的结构信息，未来可以考虑在 3D 空间中彼此接近但没有化学键的原子对，模拟它们的相互作用对于准确确定分子构象至关重要，尤其是对于大分子和多分子复合物。

参考文献

- [1] Shengchao Liu, Mehmet F Demirel, and Yingyu Liang. N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. In *Advances in Neural Information Processing Systems*, pages 8464–8476, 2019.
- [2] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 2020.
- [3] Jianwen Chen, Shuangjia Zheng, Ying Song, Jiahua Rao, and Yuedong Yang. Learning Attributed Graph Representations with Communicative Message Passing Transformer. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*
- [4] Shitong Luo, Chence Shi, Minkai Xu, Jian Tang. Predicting Molecular Conformation via Dynamic Graph Score Matching. *Advances in Neural Information Processing Systems*, 2020.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [6] T.N.Kipf and M.Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” in *Proc. ICLR*, 2017.
- [7] P. Velićković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” in *Proc. ICLR, VAN, CAN*, Apr. 30-May. 3, 2018.
- [8] Youngchun Kwon, Dongseon Lee, Youn-Suk Choi, Myeonginn Kang, and Seokho Kang. Neural message passing for nmr chemical shift prediction. *Journal of Chemical Information and Modeling*, 60(4):2024–2030, 2020.
- [9] Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [10] Shengchao Liu, Mehmet F Demirel, and Yingyu Liang. N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. In *Advances in Neural Information Processing Systems*, pages 8464–8476, 2019.
- [11] Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin He. Molecular property prediction: A multilevel quantum

- interactions modeling perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 1052–1060, 2019.
- [12] Łukasz Maziarka, Tomasz Danel, Sławomir Mucha, Krzysztof Rataj, Jacek Tabor, and Stanisław Jastrzebski. Molecule attention transformer. arXiv preprint arXiv:2002.08264, 2020.
- [13] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 33, 2020.
- [14] Ying Song, Shuangjia Zheng, Zhangming Niu, Zhang-Hua Fu, Yutong Lu, and Yuedong Yang. Communicative representation learning on attributed molecular graphs. In Proceedings of the TwentyNinth International Joint Conference on Artificial Intelligence, (IJCAI 2020), pages 2831–2838, 2020.
- [15] M Withnall, E Lindelöf, O Engkvist, and H Chen. Building attention and edge message passing neural networks for bioactivity and physical-chemical property prediction. *Journal of Cheminformatics*, 12(1):1, 2020.
- [16] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.
- [17] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 2019.
- [18] Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, and Tie-Yan Liu. On layer normalization in the transformer architecture. arXiv preprint arXiv:2002.04745, 2020.