

# 图中的注意力模型

## 摘要：

图结构的数据很自然地出现在许多不同的应用范围里。通过将数据表示为图，我们可以捕捉实体（比如节点）与实体之间彼此的关系。许多有用的见解可以从图形结构数据中获得，正如不断增长的专注于图挖掘的工作所证明的那样。但是，在现实世界中，图形可能很大 - 具有许多复杂模式 - 并且噪声可能会对有效图形挖掘造成问题。处理这个问题的有效方法是将“注意力”纳入图挖掘解决方案。

注意机制允许方法专注于图的任务相关部分，帮助它做出更好的决策。在这项工作中，我们对图形注意力模型的新兴领域的文献进行了全面而有针对性的调查。我们引入了三种直观的分类法来对现有工作进行分组。这些基于问题设置（输入和输出的类型），所使用的关注机制的类型和任务（例如，图分类，链接预测等）。我们通过详细的示例激励我们的分类法，并从独特的角度使用每个分类法来调查竞争方法。

最后，我们强调该领域的一些挑战，并讨论未来工作的有希望的方向。

## 关键字：

计算方法→人工智能; 机器学习; 逻辑和关系学习; 计算数学→图算法; 组合学; 图论; 信息系统→数据挖掘; 计算理论→图算法分析; 流, 次线性 and 近线性时间算法;

## 1 简介

那些可以被自然地建模为图的数据可以在各种各样的领域中被找到，包括全球网络[Albert et al. 1999]，生物信息学[Pei et al. 2005]，神经科学[Lee et al. 2017]，化学信息学[Duvenaud等. 2015]，社交网络[Backstrom和Leskovec 2011]，科学引文和合作[Liu et al. 2017]，城市计算[郑等人. 2014]，推荐系统[Deng et al. 2017]，传感器网络[Aggarwal et al. 2017]，流行病学[Moslonka? Lefebvre et al. 2011]，异常和欺诈分析[Akoglu et al. 2015]和生态学[Allesina et al. 2005]。

例如，可以使用图表来捕获社交网络上的用户之间的交互，其中节点表示用户，并且链接表示用户交互或友谊 [Backstrom和Leskovec 2011]。另一方面，在化学信息学中，我们可以通过将原子视为节点并将原子之间的键视为边来构建分子图[Duvenaud et al. 2015年]。

相关领域已经出现大量相关的工作。我们大致分类为图形挖掘领域[Aggarwal和Wang 2010]，专注于从图形数据中获取洞察力。在这方面已经研究了许多有趣和重要的问题。这些包括图表分类[Duvenaud et al. 2015]，链接预测[Sun et al. 2011]，社区检测[Girvan and Newman 2002]，功能性脑网络发现[Bai et al. 2017]，节点分类和聚类[Perozzi等. 2014年]，影响最大化[He and Kempe 2014]不断出现新问题建议。

图1

如图所示。注意是用来给每种类型的邻居赋予重要性的。链接大小表示数量注意，我们想要适用于每个邻居。在这个例子中，我们通过关注a看到了这一点node的同学们，我们可以更好的预测目标感兴趣的活动类型。用颜色看最好。

然而，在现实世界中，图形在结构上既大又复杂[Ahmed et al. 2014; Leskovec and Faloutsos 2006]以及嘈杂[He and Kempe 2014]。这些对图表挖掘技术构成了巨大的挑战，特别是在性能方面[Ahmed et al. 2017]。

许多技术被提出来来解决这个问题。例如，由[Wu et al. 2015]提出的利用同一个图的多种视图来改善分类性能的方法。而[Zhang et al. 2016a]在类似的动机下利用辅助非图数据作为侧视图。另一种流行的技术涉及识别任务相关或有区别的视图[Shi et al. 2012; Zhu et al. 2012]。

最近，出现了一种解决上述问题的新方法，这是通过将注意力集中到图形挖掘解决方案中来实现的。[Velickovic et al. 2018]。注意机制通过允许模型“专注于输入的最相关部分来做出决策”来帮助模型。最初在深度学习社区引入了注意力，以帮助模型处理数据的重要部分[Bahdanau et al. 2015; Mnih et al. 2014]。

该机制已成功应用于解决各种任务的模型。[Mnih et al. 2014]使用注意力来瞥见输入图像的相关部分以进行图像分类;另一方

面, [Xu et al. 2015]使用注意力集中于图像字幕任务的图像的重要部分。同时[Bahdanau et al. 2015]利用注意力机制, 通过在输出句子中生成相应单词时分配反映输入句子中不同单词重要性的权重, 来帮助改善机器翻译任务。最后注意力机制也都同样被用于图像[Yang et al. 2016]和自然语言问答任务[Kumar et al. 2016]。然而, 涉及注意力的大部分工作已经在计算机视觉和自然语言处理领域中完成了。

最近以来,人们对图表的注意力模型越来越感兴趣, 并且各种各样的技术开始被提出。[Choi et al. 2017; Feng et al. 2016; Han et al. 2018; Lee et al. 2018;Ma et al. 2017; Velickovic et al. 2018].尽管在所有这些论文中注意力的定义略有不同, 但竞争方法确实有共同点, 因为注意力被用于使模型能够专注于图的任务相关部分。

我们在第六节更精确地来讨论这些定义, 这些应用注意力机制到图中的方法的主要策略。图1显示了在图表设置中注意何时有用的激励示例。

特别是, 在图表上使用注意力的主要优点可归纳如下:

- (1) 注意力机制允许模型避免或忽略图形的嘈杂部分, 从而改善信噪比 (SNR) [Lee et al. 2018; Mnih]. 2014。
- (2) 注意允许模型为图中的元素 (例如, 节点邻域中的不同邻居) 分配相关性得分, 以突出显示具有最多任务相关信息的元素, 进一步提高SNR [Velickovic et al 2018]
- (3) 注意力也为我们提供了一种方法, 使模型的结果更具解释性[Choi et al.2017;Velickovic et al. 2018]  
例如, 通过分析模型对不同的注意力在医学本体图中的组件, 我们可以确定导致特定医学状况的主要因素。[Choi et al. 2017]

在本文中, 我们对图表中的文献进行了全面而有针对性的评论。据我们所知, 这是关于这一主题的第一部著作。我们引入了三种不同的分类法, 将现有工作分组为直观类别。然后, 我们通过详细的示例激励我们的分类法, 并使用每个分类法从特定的角度来调查竞争方法。特别是, 我们通过问题设置 (由输入图的类型和主要问题输出定义), 所使用的注意类型以及任务 (例如, 节点分类, 图表分类等) 对现有工作进行分组)。

在以前的工作中, 已经研究了不同类型的图 (例如, 齐次图, 异构图, 有向非循环图等) 具有不同的属性 (例如, 归因, 加权或未加权, 有向或无向) 和不同的输出 (例如, 节点嵌入, 链接嵌入, 图形嵌入)。第一个分类法允许我们从这个角度来调查这个领域。第二种分类法解决了在图表中应用注意力的主要策略。然后, 我们引入了一个最终分类法, 按应用领域对方法进行分组; 这向读者展示了已经解决了哪些问题, 或许更重要的是, 揭示了尚未应用注意力模型的基于图形的重要问题。图2显示了建议的分类法。

此外, 我们还总结了图表关注领域尚未解决的挑战, 并为未来的工作提供了有希望的方向。

## 1.1 主要贡献

本项工作的主要贡献如下:

- (1) 我们引入了三种直观的分类法来分类各种图形注意模型, 并使用这些分类法调查现有方法。据我们所知, 这是关于图注意重要领域的第一次调查。
- (2) 我们通过从分类学的角度讨论和比较不同的图表注意力模型来明晰每个分类法
- (3) 我们强调在图表关注领域尚未解决的挑战。
- (4) 我们还就这一新兴领域未来工作的潜在领域提出建议

图2

建议的分类法基于 (a) 问题设置, (b) 使用的注意类型和 (c) 任务或问题来对图注意模型进行分组。

## 1.2 本文的范围

在本文中, 我们将重点研究和分类各种适用于图形的技术 (我们在第2节中给出了注意的一般定义)。这些方法中的大多数都将图形作为输入并解决了一些基于图形的问题, 例如链接预测[Sun et al. 2011], 图分类/回归[Duvinaud et al. 2015], 或节点分

类[Kipf和Welling 2017]。但是，我们也考虑了关注图形的方法，尽管图形只是问题的几种输入类型之一。

我们不会尝试调查未明确应用注意力的一般基于图形的方法的广泛领域，已经对此进行了多项工作，每项工作都有一个特定的重点。[Cai et al. 2018; Getoor and Diehl 2015; Jiang et al. 2013]

## 1.3 综述的结构

本次综述的其余部分安排如下。我们首先在第2节中介绍有用的符号和定义。然后我们使用第3节到第5节来讨论使用我们的主要分类法的相关工作（图2a）。我们组织了第3-5节，使得现有工作中的方法按照它们计算的主要嵌入类型进行分组（例如，节点嵌入，边缘嵌入，图形嵌入或混合嵌入）；然后将这些方法进一步划分为它们支持的图形类型。在第6节和第7节中，我们转向不同的视角，并使用其余的分类法（图2b和图2c）来指导讨论。然后，我们将在第8节讨论挑战以及未来工作的有趣机会。最后，我们在第9节中总结了综述结果。

## 2 问题的形式

在本节中，我们定义了讨论中出现的不同类型的图形；我们还介绍了相关的符号。

定义1（同构图）

设 $G = (V, E)$ 是图，其中 $V$ 是节点（顶点）的集合， $E$ 是 $V$ 中节点之间的边集。更进一步，让 $A=[A_{ij}]$ 为 $n \times n$ ，对于 $n=|V|$ ， $G$ 的邻接矩阵，其中 $A_{ij} = 1$ 。存在 $(v_i, v_j) \in E$ 和 $A_{ij} = 0$ 。否则。当 $A$ 是二进制矩阵（即， $A_{ij} \in \{0,1\}$ ）时，我们认为该图是未加权的。注意， $A$ 也可以编码边权重；对于 $(v_i, v_j) \in E$ ， $A_{ij} = w$ ，给定权重 $w \in \mathbb{R}$ 。在这种情况下，图表被称为加权。此外，如果 $A_{ij} = A_{ji}$ ，对于任何 $1 \leq i, j \leq n$ ，则图表是无向的，否则它是有向的。

给定邻接矩阵 $A$ ，我们可以构造一个右随机矩阵 $T$  - 也称为转移矩阵 - 它只是 $A$ ，行标准化为1。另外，给定顶点 $v \in V$ ，让 $\Gamma v$ 是节点 $v$ 邻域中的顶点集合（例如，邻域的流行定义就是 $v$ 的单跳邻域，或者 $\Gamma v = \{w \mid (v, w) \in E\}$ ）

定义2（异构图）。

异构图被定义为 $G = (V, E)$ ，其由一组节点对象 $V$ 和连接 $V$ 中的节点的一组边 $E$ 组成。异构图还具有节点类型映射函数 $\theta: V \rightarrow TV$ 和边界类型映射函数，其定义为 $\xi: E \rightarrow TE$ ，其中 $TV$ 和 $TE$ 分别表示节点类型集和边缘类型。节点 $i$ 的类型表示为 $\theta(i)$ （其可以是作者，论文或异构书目网络中的会议），而边缘的类型 $e = (i, j) \in E$ 表示为 $\xi(i, j) = \xi(e)$ （例如，共同作者关系，或“发表于”关系）

注意，异构图有时被称为类型化网络。此外，二分图是具有两种节点类型和单边类型的简单异构图。同构图只是异构图的一个特例，其中 $|TV| = |TE| = 1$ 。

定义3（归因图）

设 $G = (V, E, X)$ 为属性图，其中 $V$ 是一组节点， $E$ 是一组边， $X$ 是节点输入属性的 $n \times d$ 矩阵，其中每个 $x_i$ （或 $X_i$ ）是节点 $v_i \in V$ 的属性值的 $d$ 维（行）向量， $x_j$ （或 $X_j$ ）是对应于 $X$ 的第 $j$ 个属性（列）的 $n$ 维向量。或者， $X$ 也可以是 $d \times n$ 边缘输入属性矩阵。这些可能是实值的，也可能不是。

定义4（有向无环图（DAG））没有循环的有向图

定义1-3中定义的图形的不同“类别”可以是定向的或非定向的，也可以是加权的或未加权的。其他类别的图表来自于组成不同的“图类”（定义1-3）。例如，定义归因异构网络 $G = (V, E, TV, TE, X)$ 是直截了当的。

定义5（路径）

长度为 $L$ 的路径 $P$ 被定义为唯一索引 $i_1, i_2, \dots$ 的序列。。。 ， $i_L + 1$ 这样 $(v_{i_t}, v_{i_t + 1}) \in E$ 对于所有 $1 \leq t \leq L$ 。路径的长度定义为其包含的边数。

与允许循环的遍历不同，路径不会，因此如果没有循环（所有节点都是唯一的），则walk是路径。请注意，路径是一种特殊的图形，具有非常严格的结构，其中所有节点最多具有2个邻居。

我们现在给出关于图注意概念的一般但正式的定义如下：

定义6（图注意）

给定目标图形对象（例如，节点，边缘，图形等）， $v_0$ 和 $v_0$ 的“邻域”中的一组图形对象 $\{v_1, \dots, v_{|\Gamma_{v_0}|}\} \in \Gamma_{v_0}$ （邻域是特定于任务的，可以表示节点的 $\ell$ -hop邻域或更一般的东西）。注意被定义为函数 $f': \{v_0\} \times \Gamma_{v_0} \rightarrow [0, 1]$ ，它将 $\Gamma_{v_0}$ 中的每个对象映射到相关性分数，该相关性分数告诉我们想要给予特定邻居对象多少关注。此外，通常期望 $\sum_{i=1}^{|\Gamma_{v_0}|} f'(v_0, v_i) = 1$ 。

表1中提供了整个手稿中使用的符号摘要

## 3 基于注意力机制的节点/边嵌入

在本节，以及接下来的两个部分中，我们介绍了各种图注意力模型，并通过问题设置对它们进行分类。为便于参考，我们展示了表2中的所有调查方法，注意在每个提议的分类法下突出它们的属性。我们现在开始定义传统的节点嵌入问题。

定义7（节点嵌入）给定图 $G = (V, E)$ ，其中 $V$ 作为节点集， $E$ 作为边集，节点嵌入的目的是学习函数 $f: V \rightarrow \mathbb{R}^k$ ，使得每个节点 $i \in V$ 映射到 $k$ -维向量 $z_i$ ，其中 $k \ll |V|$ 。节点嵌入矩阵表示为 $Z$ 。

作为输出给出的学习节点嵌入随后可以用作挖掘和机器学习算法的输入，用于各种任务，例如链路预测[Sun et al. 2011]，分类[Velickovic等。2018年]，社区检测[Girvan and Newman 2002]和角色发现[Rossi and Ahmed 2015]。

我们现在定义基于注意力的节点嵌入问题如下：

定义8（基于注意力的节点嵌入）

给定与上面相同的输入，我们学习函数 $f: V \rightarrow \mathbb{R}^k$ ，其将每个节点 $i \in V$ 映射到嵌入向量 $z_i$ 。另外，我们学习第二函数 $f': V \times V \rightarrow [0, 1]$ 以将“注意权重”分配给目标节点 $i$ 的邻域 $\Gamma_i$ 中的元素。函数 $f'$ 定义每个邻居 $j$ ，对于 $j \in \Gamma_i$ ，相对于目标节点 $i$ 的重要性。通常，对于所有 $i$ ， $f'$ 被约束为具有 $\sum_{j \in \Gamma_i} f'(i, j) = 1$ ，对于所有 $k \notin \Gamma_i$ ， $f'(i, k) = 0$ 。基于注意力的节点嵌入的目标是为节点 $i$ 及其更相似或更重要的邻居分配类似的嵌入，即 $\delta(z_i, z_j) > \delta(z_i, z_k)$  iff  $f'(i, j) > f'(i, k)$ ，其中 $\delta$ 是某种相似性度量。

上面，我们定义了基于注意力的节点嵌入；基于注意力的边缘嵌入也可以类似地定义

在这里，我们使用单个部分来讨论节点和边嵌入，因为在基于注意力的边嵌入方面还没有很多工作，并且因为这两个问题非常相似。我们现在根据它们支持的图表类型对基于注意力的节点/边缘嵌入进行计算的不同方法进行分类。

□

### 3.1 同构图

计算基于注意力的节点嵌入的大多数工作都集中在同构图上。此外，所有方法都假设图表被归因，尽管只需要节点标签（在这种情况下，属性大小 $d = 1$ ）。

[Abu-El-Hajja]等人提出的方法。2017年，名为AttentionWalks，最让人想起流行的节点嵌入方法，如DeepWalk和node2vec [Grover和Leskovec 2016; Perozzi等。因此，随机游走用于计算节点上下文。给定具有其对应的转移矩阵 $T$ 的图 $G$ 和上下文窗口大小 $c$ ，我们可以计算在每个节点处开始行走的预期次数，通过以下方式访问其他节点：

$$\mathbb{E}[D[a_1, \dots, a_c]] = \mathbb{I}_n \sum_{i=0}^{c-1} a_i (T)^i$$

这里的 $\mathbb{I}_n$ 是size- $n$ 单位矩阵， $a_i$ ， $1 \leq i \leq c$ ，是可学习的注意权重， $D$ 是游走分布矩阵，其中 $D_{uv}$ 编码节点 $u$ 预期访问节点 $v$ 的次数，假定从每个节点开始步行。在这种情况下，注意力因此被用于引导步行朝向更广泛的邻域或将其限制在更窄的邻域内（例如，当权重是头重脚轻时）。这解决了必须进行网格搜索以识别最佳超参数的问题，因为研究表明这对性能有显著影响[Perozziet al. 2014] - 请注意，对于DeepWalk，权重固定为 $a_i = \frac{1}{c}$ 。

另一种基于注意力的方法在精神上与DeepWalk，node2vec和LINE等方法非常相似，因为它使用共生信息来学习节点嵌入是GAKE这种方法。重要的是要指出GAKE从知识三元组 $(h, t, r)$ 构建图形，通常称为知识图形，其中 $h$ 和 $t$ 是由关系 $r$ 连接的项。基于三个三元组 $(\text{Jose}, \text{Tagalog}, \text{speaks})$ ， $(\text{Sato}, \text{Nihongo}, \text{speaks})$ 和 $(\text{Jose}, \text{Sinigang}, \text{eat})$ ，我们可以构建一个简单的异构图，其中包含三种类型的节点，即 $\{\text{person}, \text{language}, \text{food}\}$ 和两种类型的关系，即 $\{\text{speaks}, \text{eats}\}$ 。然而，我们将GAKE归类

为同构图方法，因为它似乎没有区分不同类型的关系或节点类型（请参阅metapath2vec [Dong et al.2017]，了解明确模拟不同类型的关系的节点嵌入方法和异构图中的节点）。相反，该方法采用一组知识三元组，并通过取每个三元组（h，t，r）并分别添加h和t作为头部和尾部节点来构建有向图，同时从头部到尾部添加边缘（他们还添加了反向链接）。GAKE与DeepWalk或node2vec等方法之间的主要区别在于它们在计算节点的上下文时包含边。他们定义了三个不同的上下文来获得相关主题（节点或边缘），正式地说它们最大化了对数似然：

$$\sum_{s \in V \cup E} \sum_{c \in \Gamma_s} \log \Pr(s | c)$$

其中 $\Gamma_s$ 是定义 $s$ 的邻域上下文的一组节点和/或边。为了在图中给出给定主题的最终节点嵌入，他们使用注意力机制来获得最终嵌入 $z_s = \sum_{c \in \Gamma_s} \alpha(c) z'_c$ 。其中 $\Gamma'_s$ 是 $\alpha$ 的一些邻域上下文。 $\alpha(c)$ 定义上下文对象 $c$ 的注意权重， $z'_c$ 是 $c$ 的学习嵌入。

另一方面，GAT和AGNN等方法通过结合明确的注意力机制来扩展图卷积网络（GCN）。回想一下，GCN能够使用传播规则通过输入图的结构传播信息： $H^{(l+1)} = \sigma(\frac{1}{2} \check{D}^{-1} \check{A} \check{D}^{-1} H^{(l)} W^{(l)})$  其中 $H^{(l)}$ 表示在具有 $H_0 = X$ 的GCN的层 $l$ 处的学习的嵌入矩阵。此外 $\check{A} = A + I_n$ 是附加自循环的无向图 $A$ 的邻接矩阵。另一方面，矩阵 $\check{D}$ 被定义为 $\check{A}$ 的对角矩阵，换句话说， $\check{D}_{i,i} = \sum_j \check{A}_{i,j}$ 。因此，项 $\frac{1}{2} \check{D}^{-1}$

计算由 $\check{A}$ 定义的图的对称归一化（类似于归一化图拉普拉斯算子）。最后， $W^{(l)}$ 是级别 $l$ 的可训练权重矩阵， $\sigma(\cdot)$ 是非线性，如ReLU，Sigmoid或tanh。

GCN的工作方式类似于Weisfeiler-Lehman算法的端到端可微分版本，其中每个附加层允许我们扩展和集成来自更大邻域的信息。

然而，因为我们在传播中使用术语 $\frac{1}{2} \check{D}^{-1} \check{A} \check{D}^{-1}$ ，所以我们基本上应用由其度数归一化的相邻节点的特征的加权和。GAT和AGNN基本上引入了注意权重值，以确定我们希望从节点 $u$ 的角度给予相邻节点 $v$ 多少。这两种方法的不同之处主要在于注意力的定义方式（我们在第6节中对此进行了阐述）。此外，引入了“多关注”的概念，其基本上定义了图中一对对象之间的多个关注头（例如，权重）。

PRML是另一种学习边嵌入的方法，但它们使用不同的策略来定义注意力。在[赵等人。2017]，定义了路径长度阈值 $L$ 和递归神经网络[Gers et al. 2000]在节点 $u$ 和 $v$ 之间学习长度为1，...， $L$ 的路径的基于路径的嵌入。注意力以两种方式定义。首先，注意用于识别沿路径的重要节点，这有助于计算中间路径嵌入。其次，注意力然后优先考虑更多指示链路形成的路径，并且当这些路径嵌入被集成以形成最终边缘嵌入时，这用于突出重要或任务相关的路径嵌入。我们可以将此方法视为基于注意力的模型，该模型计算节点对的Katz中介性得分

## 3.2 异构图

据我们所知，为异构图计算注意力引导节点嵌入的唯一工作是EAGCN [Shang et al. 2018。非常类似于GAT和AGNN，[Shang et al. 2018]建议关注GCN [Duvenaud et al. 2015; Kipf和Welling 2017]。但是，它们处理的情况是，可以有多种类型的链接连接图形中的节点。因此，他们建议使用“多重注意”，如[Velickovic et al.2018]，并且每个注意机制都考虑仅由特定链接类型定义的邻居。尽管作者使用图形回归任务验证了EAGCN，但是该方法可以在没有对节点级任务进行太多调整的情况下使用，因为EAGCN中的图形嵌入仅仅是学习的注意引导节点嵌入的连接或总和。

然而，上述工作假定给定的异构网络仅具有一种类型的节点，即 $|\Gamma_V| = 1$

## 3.3其他特殊情况

与基于注意力的图形嵌入领域不同，对于出现在某些领域中的特殊类型的图形（例如，表示为DAG的医学本体，或表示为某些星图的某些异构网络），似乎没有计算注意力引导的节点嵌入的工作。

由于上述方法适用于一般图形，因此它们应适用于更具体类型的图形，并且应该能够将它们直接应用于这些情况。

## 4 基于注意力的图嵌入

同样，我们通过定义传统的图嵌入问题开始讨论。

定义9（图嵌入）：给定一组图形，图嵌入的目的是学习一个函数 $f: \mathbf{G} \rightarrow \mathbb{R}^k$  其将每个输入图 $G \in \mathbf{G}$ 映射到长度为 $k$ 的嵌入向量 $z$ 。此外，我们学习第二函数 $f'$ ，其将“注意权重”分配给不同的 给定图的子图允许模型在计算其嵌入时优先考虑图的更重要部分（即子图）。

## 4.1 同构图

[Lee et al. 2018; Ryu et al. 2018; Xu et al. 2018]已经提出了几种用于在均匀图上学习图嵌入的方法。

[Ryu et al. 2018]只需使用GAT的表述并对计算注意力的方式做一些调整。然而，除了如何计算注意力量的变化之外，GAT与[Ryu et al. 2018]之间没有太大区别。2018。为了在分子图上获得图形回归任务的最终图形嵌入，所提出的方法在最终GAT层之后添加完全连接的层以平坦化每个节点的输出。

另一方面，[Xu et al. 2018]提出了解决自然语言问题回答任务的graph2seq。给定一组事实，以句子的形式，他们的方法非常独特，因为他们将输入转换为属性有向同质图。为了获得用于序列生成任务的图嵌入，他们首先学习图中每个节点的节点嵌入。通过连接每个节点的前向和后向表示来形成节点嵌入，这些表示是通过聚合来自每个节点的前向（使用前向链路遍历的邻居）和向后（类似地，使用反向链路遍历的邻居）邻域的信息而得到的表示。最终的图嵌入是通过汇集单个节点嵌入或简单地聚合它们来获得的。然而，在graph2seq的情况下，通过在序列生成期间也参与单个节点嵌入来应用注意。由于嵌入 $z_i$ 的每个节点在节点 $i$ 周围的区域中捕获信息（我们可以将其视为聚焦在 $i$ 周围的子图），因此注意允许在输出句子（序列）中生成单词时优先考虑图的特定部分。这捕获了直观，即图形的不同部分可以主要与不同的概念或单词相关联。

最后，在之前的工作中[Lee et al. 2018]，我们提出了GAM，它使用两种类型的注意来学习图形嵌入。主要思想是使用注意力引导的步骤来对图形的相关部分进行采样，以形成子图嵌入。在GAM中，我们走了一段长度 $L$ 。让 $1 \leq i \leq L$ 是在walk中发现的第 $i$ 个节点， $x_i$ 是相应的节点属性向量，RNN用于集成节点属性 $\{x_1, \dots, x_L\}$ 形成子图嵌入（在walk期间覆盖的子图或区域）。在每个步骤期间，使用注意机制来确定哪个相邻节点更相关以允许步行覆盖更多与任务相关的图形部分。为了获得最终的图形嵌入，我们将 $z$ “代理”部署到图的各个部分进行采样，从而产生嵌入 $\{s_1, \dots, s_z\}$ 。然后，在将这些子图嵌入组合以形成图嵌入之前，使用第二关注机制来确定各子图嵌入的相关性。换句话说，注意机制被定义为函数 $\alpha: R_k \rightarrow [0, 1]$ ，其将给定子图嵌入映射到相关性得分。因此，图形嵌入被定义为 $\sum_i z_i \alpha(s_i)$

## 4.2 异构图

回想一下我们之前的讨论，即EAGCN [Shang et al. 2018]用于研究图回归的任务。EAGCN使用类似于GAT的方法来生成图嵌入。由于该方法使用与GAT类似的注意机制，因此注意力集中在确定在计算图的节点嵌入时要注意的重要邻居。然后，最终的图嵌入是注意引导的节点嵌入的连接/总和。应用第二种注意机制并不困难，类似于[Lee et al. 2018]，以加权不同节点嵌入的重要性。

## 4.3 其他特殊情况

注意力最初是在计算机视觉和NLP领域进行研究的，并且提出了各种注重基于序列的任务的RNN模型 [Bahdanau et al. 2015; Luong et al. 2015]。由于序列在技术上与路径没有区别，因此我们在此设置下引入了一些值得注意的关注模型。由于序列在技术上与路径没有区别，因此我们在此设置下引入了一些值得注意的关注模型。长度为 $L$ 的序列例如，句子，可以表示为长度为 $L$ 的有向属性图 - 节点 $i$ 的第 $i$ 个属性向量 $x_i$ 则是序列中第 $i$ 个分量的表示（一个 - 例如，热词嵌入或word2vec嵌入）。在提出的方法 [Bahdanau et al. 2015; Luong et al. 2015; Ma et al. 2017]，将节点属性向量 $\{x_1, \dots, x_L\}$ 一个接一个地馈入RNN。回想一下，在最简单的情况下，递归神经网络通过规则为每个输入 $i$ 计算隐藏的嵌入 $h_i$ 。

$$h_i = \sigma(W_h x_i + U_h h_{i-1} + b_h)$$

其中 $W_h$ 和 $U_h$ 分别是输入 $x_i$ 和先前隐藏嵌入 $h_{i-1}$ 的可训练权重矩阵； $b_h$ 是偏置向量， $h_0$ 通常初始化为零向量。

在这种情况下，我们可以将 $h_i$ 视为节点 $i$ 的节点嵌入，其中集成了来自包含节点 $1, \dots, i$ 的子路径的信息。然后可以将注意力应用于节点嵌入以生成注意力引导的图形嵌入。特别地，在被 [Bahdanau et al. 2015]称为RNNSearch中提出的方法将图形嵌入定义为 $z = \sum_{i=1}^L \alpha_i h_i$ ，其中根据其任务的相关性将注意力分配给每个隐藏嵌入 $h_i$ 。

然而，当路径 $L$ 的长度很大时，参加RNNSearch中的所有隐藏嵌入可能不会产生最佳结果。[Luong et al. 2015]建议使用两种注意机制，第一种在精神上与 [Bahdanau et al. 2015]相似。但是，第二种注意力允许模型在输入中选择一个局部点并集中注意力。更具体地说，根据任务的需要，第二关注机制输出位置 $1 \leq p \leq L$ ，并且图形嵌入由隐藏节点嵌入 $\{h_{p-D}, \dots, h_{p+D}\}$ 构成，其中 $D$ 是经验选择的关注窗口大小。

最后，Dipole [Ma et al. 2017]将类似于 [Bahdanau et al. 2015]的注意力模型应用于顺序医学数据。他们用它来诊断或预测序列中的医疗状况。在实践中，两者都是 [Ma et al. 2017]和 [Bahdanau et al. 2015]使用双向模型，其使用两个RNN处理输入序列，一个以原始顺序获取输入序列而另一个以相反顺序获取输入序列。

## 5 基于注意的混合嵌入

在本节中，我们将讨论关注图数据的方法。然而，计算的嵌入是“混合的”，因为这里的方法也采用其他模态（例如，文本）的数据，并且学习的嵌入是所有输入的组合。

### 5.1 同构图

像GAKE一样[Feng et al. 2016]，CCM[Zhou et al. 2018]处理知识图。但是，在[Zhou et al. 2018]，研究了序列到序列生成问题。CCM使用编码器-解码器模型[Sutskever et al. 2014]编码句子并输出句子（例如，用于句子翻译或问答）。但是，CCM中使用的设置假定模型引用了问答任务的知识图。他们的知识图由多个连接组件组成，每个组件由多个知识三元组组成。为每个组件*i*学习嵌入 $z_i$ 的子图：

$$z_i = \sum_{n=1}^N \alpha_n^{(i)} [h_n^{(i)}; t_n^{(i)}]$$

其中  $T_i = \{(h_1^{(i)}, t_1^{(i)}, r_1^{(i)}), \dots, (h_N^{(i)}, t_N^{(i)}, r_N^{(i)})\}$  由相应知识三元组从*i*的嵌入组成，而  $\alpha_n^{(i)}$  用于指定不同三元组的重要性（由术语嵌入的串联表示），同时考虑术语及其关系。

当模型处于解码过程中（为输出生成单词）时，它通过参与嵌入 $z_i$ 来选择性地识别重要的子图。此外，注意机制还用于识别所选知识图内的三元组，以识别用于单词生成的重要三元组。与GAKE一样，我们认为CCM是一种同构图方法，因为它似乎没有明确区分不同类型的节点和关系。

JointD / E + SATT [Han et al. 2018]是另一项将注意力应用于知识图的工作，类似于[Feng et al. 2016年 周等人。2018。然而，他们还使用了一个大的文本语料库，其中可能包含对知识图中暗示某种关系的不同术语的引用。他们引入了一种方法，该方法使用注意来学习图形和文本数据的联合嵌入，用于知识图链接预测的任务 在某些设置（例如脑网络构建）下，数据集往往非常小而且噪声很大[Zhang et al. 2016B。诸如[Zhang et al. 2016b]建议使用辅助信息来规范图形构建过程以突出更多的判别模式 - 这在输出图形用于图形分类时很有用。探索在这种环境中增加注意力的可能性很有趣。

### 5.2 异构图

虽然有人提出像[Han et al. 2018; Zhou et al. 2018] 将注意力应用于知识图 - 这被认为是异构图 - 这些模型没有明确区分不同类型的链接和节点。据我们所知，目前还没有任何考虑过此设置的工作。一种可能性是扩展EAGCN提出的观点[Shang et al. 2018]到更一般的异构图。

### 5.3 其他特殊情况

GRAM[Choi et al. 2017] 是用于对临床数据进行分类/回归的基于图的注意力模型。临床记录通常可以通过临床代码  $c_1, \dots, c_{|C|}$  在词汇表中C.GRAM构造一个DAG，其叶子是代码  $c_1, \dots, c_{|C|}$  而祖先节点是更一般的医学概念。[Choi et al. 2017]使用注意机制通过以下方式学习每个叶节点*i*（医学概念）的*k*维最终嵌入：

$$g_i = \sum_{j \in A(i)} \alpha_{ij} e_j$$

其中 $A(i)$ 表示由*i*及其所有祖先和 $e_j$ 组成的集合是节点*j*的*k*维基本嵌入。当C中的概念对医学诊断不太有帮助时（例如，这是一种罕见的概念），使用注意力允许模型参考更具信息性或相关的一般概念。由于患者的临床访视记录被表示为二进制向量  $x \in \{0,1\}^{|C|}$ ，其指示针对特定访问存在哪些临床代码，所以可以堆叠学习的嵌入  $g_1, \dots, g_{|C|}$  以形成 $k \times |C|$ 嵌入矩阵嵌入 $x$ ；然后可以将其输入到用于医学诊断的预测模型中。注意，医学诊断的问题是经典的监督学习任务，其采用临床代码特征向量，但GRAM将注意力应用于医学概念DAG以便嵌入给定的特征向量。

## 6 图注意机制的类型

我们现在描述已应用于图形数据的三种主要关注类型。虽然所有三种类型的注意力都有相同的目的或意图，但它们在如何定义或实施注意力方面存在差异。在本节中，我们提供了文献中的示例，以说明每种类型的实现方式。

回想一下我们在Def.6中对注意力的一般定义，我们给出了一个目标图形对象（例如，节点，边缘，图形等） $v_0$ 和一组图形对象在 $v_0$ 的“邻域”  $\{v_1, \dots, v_{|\Gamma\{v_0\}|}\} \in \Gamma\{v_0\}$ 。注意被定义为函数  $f: \Gamma\{v_0\} \times \Gamma\{v_0\} \rightarrow [0,1]$ ，它将  $\Gamma\{v_0\}$  中的每个对象映射到相关性得分。在实践中，这通常以我们在下面介绍的三种方式之一完成。

### 6.1 学习注意权重

给定相应的属性/特征  $x_0, x_1, \dots, x_{|\Gamma\{o\}|}$  对于  $v_0, v_1, \dots, v_{|\Gamma\{v_0\}|}$ , 可以通过以下方式学习注意力权重: 
$$\alpha_{0,j} = \frac{e_{0,j}}{\sum_{k \in \Gamma\{v_0\}} e_{0,k}}$$
 其中  $e_{0,j}$  是  $v_j$  与  $v_0$  的相关性。在实践中, 这通常使用具有可训练函数的softmax来实现, 该函数通过考虑其属性来学习  $v_j$  与  $v_0$  的相关性。在GAT中的实施[Velickovic et al. 2018]说明了这一点: 
$$\alpha_{0,j} = \frac{\exp(\text{LeakyReLU}(a[W_{x_0} \| W_{x_j}]))}{\sum_{k \in \Gamma\{v_0\}} \exp(\text{LeakyReLU}(a[W_{x_0} \| W_{x_k}]))}$$
 其中  $a$  是可训练的注意向量,  $W$  是可训练的权重矩阵, 将输入要素映射到隐藏空间, 并且  $\|$  代表串联。其示意图如图3a所示。

## 6.2 基于相似性的关注

同样, 给定相应的属性或特征, 除了关键差异之外, 可以类似地如上所述地学习第二类型的关注。我们将这种方法称为基于相似性的注意, 因为更多的注意力被分配给具有更多相似的隐藏表征或特征的对象, 这在文献中也经常被称为对齐[Bahdanau et al 2015]。为了说明这一点, 我们使用AGNN中给出的定义[Thekumparampil et al 2018]:

$$\alpha_{0,j} = \frac{\exp(\beta \cdot \cos(W_{x_0}, W_{x_j}))}{\sum_{k \in \Gamma\{v_0\}} \exp(\beta \cdot \cos(W_{x_0}, W_{x_k}))}$$

其中  $\beta$  是可训练的偏差, “cos”代表余弦相似性; 和以前一样,  $W$  是一个可训练的权重矩阵, 用于将输入映射到隐藏空间。请注意, 这与上面的定义非常相似。不同之处在于模型明确地为彼此相关的对象学习类似的隐藏嵌入, 因为注意力基于相似性或对齐。

## 6.3 注意引导walk

虽然前两种注意力集中在选择相关信息以整合到目标对象的隐藏表示中, 但第三种注意力的目的略有不同。我们使用GAM[Lee et al. 2018]来说明这个想法。GAM在输入图上采取一系列步骤, 并使用RNN对来自受访节点的信息进行编码, 以构建子图嵌入。RNN在时间  $t$  的隐藏状态,  $h_t \in \mathbb{R}^h$  对来自步骤1, ...,  $t$  的步行先前访问过的节点的信息进行编码。然后将注意力定义为函数  $f: \mathbb{R}^h \rightarrow \mathbb{R}^k$  将输入隐藏向量  $f(h_t) = r_t + 1$  映射到  $k$  维秩向量, 该向量告诉我们应该为下一步优先考虑哪些类型的节点。然后, 该模型将优先考虑特定类型的相邻节点以用于下一步骤。我们在图3b中说明了这一点。

# 7 图注意力的任务

不同的基于注意力的图形方法可以根据它们解决的问题类型进行广泛划分: 节点级或图级。

## 7.1 节点级

已经提出了许多工作来研究节点级任务的图注意, 其中最值得注意的是节点分类和链路预测[Abu-Ei-Haija等。2017年; 冯等人2016; 韩等人。2018; Thekumparampil等。2018; Velickovic等人。2018; 赵等人。2017年]。尽管每种方法的方法和假设各不相同, 但它们共享一种共同的技术, 即学习注意力引导的节点或边缘嵌入, 然后可以将其用于训练分类器以进行分类或链接预测。不难看出这些方法是针对节点聚类的相关任务实现的。尚未提出基于注意力的图表方法的一些值得注意的节点级任务包括节点/边缘角色发现[Rossi和Ahmed 2015], 以及节点排名[Sun et al. 2009a]的。

## 7.2 图级

多个作品还研究了图表级别的任务, 如图分类和图形回归。在该设置中, 通过关注图的相关部分来构建注意引导图嵌入。EAGCN, Modified-GAT, GAM和偶极子等方法[Lee et al. 2018; Ma et al.2017; Ryu等人。2018; 尚等人。2018]学习用于更标准的基于图形的分类和回归任务的图形嵌入。不难看出这些方法如何应用于图相似性搜索的相关问题[Zheng et al. 2013]

另一方面, 像[Bahdanau et al. 2015年; Luong等。2015年; 徐等人。2018; Zhou等人[201]从输入图数据生成序列。值得注意的是, graph2seq提出了一种方法, 该方法输出给定输入图的序列, 而不是输出序列到序列的更多方法。

最后, GRAM[Choi et al. 2017]将注意力应用于医学本体图, 以帮助学习基于注意力的医疗代码嵌入。虽然他们研究的问题是对患者记录进行分类的问题(由某些医疗代码描述), 但他们工作的新颖性在于将注意力集中在本体图上以提高模型性能。

# 8 讨论和挑战

在本节中, 我们将讨论其他问题, 并强调未来工作的重要挑战。



## 8.1 异构图的基于注意的方法

异构图的研究，也称为异构信息网络，近年来已经变得非常流行，在该领域发表了许多论文[Dong et al. 2017年; Lee和Adorna 2012; 孙等人。 2011年, 2009a, b]。

虽然GAKE, CCM, JointD / E + SATT等方法都认为知识图是一种异构图，但它们并不区分不同类型的链接和节点。虽然EAGCN等方法处理多种类型的链接，但是 不考虑可能存在多种节点的一般情况。 GAM是另一种在不同类型的节点之间区分的方法，因为注意引导的步行基于节点类型优先考虑节点访问。

需要基于注意力的方法研究如何将注意力应用于一般的heterogeneous网络，特别是考虑到不同类型的元路径[Dong et al. 2017]可以影响学习的嵌入。 这一点非常重要，因为基于异构图的方法已被证明优于可以假设图只有单一类型的链接/边的方法。可以参考[Shi et al. 2017]用于基于异构图的方法的调查。

## 8.2 图注意模型的可扩展性

大多数方法[Bahdanau et al. 2015年; Ma等人。 2017年;周等人。 2018]为相对较小的图计算基于注意力的图嵌入工作，并且可能难以有效地缩放到更大的图。

方法如[Lee et al. 2018; Ryu等人。 2018;尚等人。 2018]可能能够处理较大的图形，但它们仍有缺点。例如，GAM[Lee et al. 2018]使用walk来对图表进行采样。对于大型图表，出现的自然问题包括：（1）我们是否需要更长的步行时间，RNN可以处理这些问题吗？（2）散步可以有效地捕获所有相关和有用的结构，特别是如果它们更复杂吗？

像[Ryu et al. 2018;尚等人。 2018]注意到GCN架构似乎是朝着正确方向迈出的一步，因为这可以被描述为Weisfeiler-Lehman算法的基于注意的端到端可微分版本[Duvenaud et al. 2015年; Shervashidze等[2011]特别是如果我们随机训练[Chen et al. 2018。然而，需要在各种大型真实世界图上评估图注意模型以测试其有效性。探索其他方法来提高注意力以提高绩效也是有用的。

## 8.3 归纳图学习

最近，人们对探索基于图形的问题的归纳学习感兴趣[Guo et al. 2018;汉密尔顿等人。 2017]与转导学习不同。前者比后者更有用，因为它可以推广到看不见的数据。

此设置很重要，因为它允许基于图形的方法处理许多真实情况，其中图形中的节点和边缘可以动态显示。它还允许我们学习一个图形数据集中的一般模式，这些模式可以在另一个数据集中使用，就像使用传递学习如何在一个文本语料库上训练模型以在另一个文本数据集上应用一样[Dai et al. 2007年]或训练模型以识别大图像数据集中的一般形状，以应用于较稀疏的数据集[Oquab et al. 2014。转移学习的另一个有趣的例子是[Zhu et al. 2011年]其中跨不同域的数据（例如，文本到图像）进行转移。

未来研究的一个有趣方向是研究基于注意力的归纳学习技术，这些技术可用于识别可推广到其他图形数据集的相关图形模式。展望未来，我们还可以探索跨域或异构转移学习的基于注意力的技术[Zhu et al. 虽然像GAT这样的方法的作者[Velickovic et al. 2018]已经对小图数据集进行了归纳学习的初步研究，我们认为应该在基于图的归纳学习上进行更集中的实验，同时考虑到大量的数据集和设置。

## 8.4 注意力引导的归因walk

最近，艾哈迈德等人。[艾哈迈德等人。 2018]提出了归因步行的概念，并表明它可以用于推广基于图的深度学习和嵌入方法，使它们更强大并且能够学习更合适的嵌入。这是通过将基于图的深度学习和嵌入方法中使用的随机游走（基于节点ID）的概念替换为更合适和更强大的属性步行概念来实现的。更正式，

定义11（归因于步行）。令 $x_i$ 为节点 $v_i$ 的 $k$ 维向量。长度为 $L$ 的属性步行 $S$ 被定义为相邻节点类型的序列：

(1)

$$\{\phi(x_{i_1}), \phi(x_{i_2}), \dots, \phi(x_{i_{L+1}})\}$$

与索引序列  $\{i_1, i_2, \dots, i_{L+1}\}$  相关联,使得  $(v_{i_t}, v_{i_{t+1}}) \in E$  对于所有  $1 \leq t \leq L$  并且  $\phi: x \rightarrow y$  是将节点的输入矢量 $x$ 映射到对应类型 $\phi(x)$ 的函数。

类型序列 $\phi(x_{i_1}), \phi(x_{i_2}), \dots, \phi(x_{i_{L+1}})$ 是步行期间发生的节点类型（与节点ID相对）。结果表明，当独特类型的数量为 $t \rightarrow n$ 时，使用传统随机游走的原始深度图学习模型被恢复为属性步行框架的特例[Ahmed et al. 2018。应当注意，这里的节点类型不一定是指导构图中的节点类型，而是可以根据其局部结构为同构图中的节点计算。

利用随机游走（基于节点ID）的基于注意的方法也可以受益于Ahmed等人[Ahmed等人提出的]归因步行（类型游走）的概念。2018。

## 9 结论

---

在这项工作中，我们对图注意模型的重要领域的文献进行了全面而有针对性的调查。据我们所知，这是第一部此类工作。我们引入了三种直观的分类法来对现有工作进行分组。这些是基于问题设置，使用的注意机制的类型和任务。我们通过详细的例子激励我们的分类法，并用它们从分类学的独特立场来调查竞争方法。我们还强调了该领域的一些挑战，并讨论了未来工作的可能方向。