

引言

作为开篇部分，我们将探讨关于本次图神经网络组队学习的几个问题：

- 首先我们将讨论**为什么我们要关注图深度学习**？具体地，为什么我们要把现实世界的数据表示成图，为什么我们要把深度学习与图连接起来，以及图深度学习的挑战是什么？
- 其次我们将讨论**此次组队学习将涵盖哪些内容**？具体来说，我们将**共同学习哪些话题以及这些话题是如何安排的**。
- 第三，我们将讨论如何更好地参与此组队学习以及**更好地理解图神经网络的理论并掌握其应用**？

为什么要在图上进行深度学习？

在过去的深度学习应用中，我们接触的数据形式主要是这四种：**矩阵、张量、序列（sequence）和时间序列（time series）**。然而来自现实世界应用的数据更多地是图的结构，如社交网络、交通网络、蛋白质与蛋白质相互作用网络、知识图谱和大脑网络等。图提供了一种通用的数据表示方法，众多其他类型的数据也可以转化为图的形式。此外，大量的现实世界的问题可以作为图上的一组小的计算任务来解决。推断节点属性、检测异常节点（如垃圾邮件发送者）、识别与疾病相关的基因、向病人推荐药物等，都可以概括为**节点分类问题**。推荐、药物副作用预测、药物与目标的相互作用识别和知识图谱的完成（knowledge graph completion）等，本质上都是**边预测问题**。

同一图的节点存在连接关系，这表明节点不是独立的。然而，**传统的机器学习技术**假设样本是独立且同分布的，因此传统机器学习方法不适用于图计算任务。**图机器学习**研究如何构建节点表征，节点表征要求同时包含节点自身的信息和节点邻接的信息，从而我们可以在节点表征上应用传统的分类技术实现节点分类。图机器学习成功的关键在于如何为节点构建表征。深度学习已经被证明在表征学习中具有强大的能力，它大大推动了计算机视觉、语音识别和自然语言处理等各个领域的发展。因此，将深度学习与图连接起来，**利用神经网络来学习节点表征**，将带来前所未有的机会。

然而，如何将神经网络应用于图，这一问题面临着巨大的**挑战**。首先，**传统的深度学习是为规则且结构化的数据设计的**，图像、文本、语音和时间序列等都是规则且结构化的数据。但**图是不规则的**，节点是无序的，节点可以有不同的邻居节点。其次，规则数据的结构信息是简单的，而**图的结构信息是复杂的**，特别是在考虑到各种类型的复杂图，它们的节点和边可以关联丰富的信息，这些丰富的信息无法被传统的深度学习方法捕获。

图深度学习是一个新兴的研究领域，它将深度学习技术与图数据连接起来，推动了现实中的图预测应用的发展。然而，此研究领域也面临着前所未有的挑战。

注：以上内容整理自“[Deep Learning on Graphs: An Introduction](#)”!!!

此组队学习涵盖的话题

此组队学习由五个话题组成，每一话题都包含理论部分与实践部分：

- 话题一：
 - 我们将首先学习**简单图论知识**、了解常规的**图预测任务**（见第2节）；
 - 然后学习**基于PyG包的图数据的表示与使用**（见第3节）。
- 话题二：
 - 我们将首先学习**实现图神经网络的通用范式，即消息传递范式**；
 - 其次学习PyG中的**消息传递（MessagePassing）基类的属性、方法和运行流程**；
 - 最后学习**如何自定义一个消息传递图神经网络**（见第4节）。
- 在话题三：
 - **图计算应用中最基础的任务是节点表征（Node Representation）学习**。
 - 我们将以GCN和GAT（两个最为经典的图神经网络）为例，学习**基于图神经网络的节点表征学习的一般过程**；并且通过MLP、GCN和GAT三者节点分类任务中的比较，学习**图神经网络为什么强于普通的MLP神经网络**，以及**GCN和GAT的差别**（见第5节）。
 - 此外，我们还将学习**如何构造一个数据全部存于内存的数据集类**（见第6-1节）；
 - 并学习**基于节点表征学习的图节点预测任务和边预测任务的实践**（见第6-2节）。
- 话题四：
 - 我们将首先分析在超大图上进行节点表征学习面临的挑战**；

- 接着学习**应对挑战的一种解决方案**；
- 最后学习**超大图节点预测任务的实践**（见第7节）。
- 话题五：
 - 我们将首先学习**基于图神经网络的图表征学习的一般过程**（见第8节）；
 - 接着学习**样本按需获取的数据集类的构造方法**（见第9-1节）；
 - 最后学习**基于图表征学习的图预测任务的实践**（见第9-2节）。

除了话题四和话题五都依赖于话题三之外，其余话题都依赖于该话题自身的前一话题。

如何更好地学以及如何学得更好

此次组队学习**要求参与者掌握一定的深度学习知识，并具备使用PyTorch实现神经网络的能力**。学习并掌握此次组队学习涵盖的话题，**需要参与者认真研读理论部分，并动手操作实践部分**。在实践部分，我们将利用PyG库构建图数据集和图神经网络模型。PyG是一个高度封装的库，在方便了我们构建图神经网络的同时，也**隐藏了很多代码运行细节**。要掌握图神经网络的应用，需要大家多多调试代码，观察并分析程序的运行流程。

各位小伙伴在学习过程中，如遇到困难或有疑惑，欢迎在群里向助教们提问。同时我们也呼吁各位小伙伴对其他小伙伴提出的问题积极思考，并踊跃发表自己的看法！

参考资料

- [Deep Learning on Graphs: An Introduction](#)