

# 图神经网络

## 1、神经网络基础

### 1.1、图数据的应用场景

重要的四个类别：同构图、异构图、属性图和非显示图

- **同构图**：节点类型和关系类型只有一种。如超链接关系构成的万维网；社交网络
- **异构图**：节点类型和关系类型不止一种。更贴近现实。
- **属性图**：节点和关系都有标签和属性，标签指节点或者关系的类型，属性是节点或关系的附加描述信息。
- **非显示图**：数据之间没有显示的定义出关系，需要依据某种规则或计算方式将数据的关系表达出来。如点云数据。

### 1.2、图神经网络

是一套基于图卷积操作并不断衍生的神经网络理论

在图神经网络出现之前。处理图数据的方法是在数据的预处理阶段将图转化为一组向量表示，问题是可能会丢失图中的结构信息。

近几年，更多的基于空域图卷积的神经网络模型的变体被开发出来，这类方法被统称为 **GNN**

### 1.3、几个定义

- **卷积滤波**：卷积是一种运算方法，滤波操作就是图像对应像素与卷积**滤波器**的乘积之和。卷积滤波器就是利用卷积运算的原理来构成滤波器，来提取特征。卷积运算决定了该滤波器的性能。
- **空域**：空域滤波是指图像平面本身，这类方法直接对图像的像素进行处理。

### 1.4、图数据相关任务的分类

- **节点层面**：主要包括分类任务和回归任务，对节点层面的性质进行预测。
- **边层面任务**：主要包括边的分类和预测（两节点是否构成边）任务，集中应用在推荐业务中。
- **图层面任务**：从图的整体结构出发，实现分类、表示和生成等任务。主要应用在自然科学研究领域，如药物分子的分类。

### 1.5、机器学习是什么？

机器学习理论主要是设计和分析一些让计算机可以自动“学习”的算法。即是让计算机从数据中挖掘有价值的信息。

### 1.6、机器学习的分类

#### 1.6.1、根据训练数据是否有标签

- **监督学习**：训练数据中每个样本都有标签，通过标签可以指导模型进行学习，学到具有判别性的特征，从而对未知样本进行预测。
- **无监督学习**：训练数据完全没有标签，通过算法从数据中发现一些数据之间的约束关系，比如数据之间的关联、距离关系等。如聚类算法。
- **半监督学习**：既有标签数据也有无标签数据，无标签数据中含有一些数据分布相关的信息，可以作为标签数据之外的补充。

## 1.6.2、根据算法输出的形式（属于监督学习的范畴）

- **分类问题**：模型的输出值为**离散值**。例如风控的输出为正常/异常两类结果。
- **回归问题**：模型的输出值为**连续值**。例如用户点击商品的概率，概率越高表示模型认为用户越倾向于点击该商品。

## 1.7、机器学习流程

- **提取图片特征**
- **建立模型**
- **确定损失函数**：用来衡量模型输出值与真实值之间的差异程度。
- **进行优化求解**

### 1.7.1、重要组成部分

- **损失函数**：用来估量模型的预测值和真实值的差异程度，是一个**非负实值函数**，用 $L(y, f(x; \theta))$ 。机器学习中，通过在训练集 $X$ 上最小化损失函数来训练模型，调整 $f$ 的参数 $\theta$ ，使得损失函数值降低。
  - 平方损失函数
  - 交叉熵损失函数
- **优化算法**：
  - **批梯度下降算法**：利用梯度信息，通过不断迭代调整参数来寻找合适的解。
    - 1.通过随机初始化为需要求解的参数赋初始值，作为优化的起点。
    - 2.使用模型对**所有**样本进行预测，计算总体的**损失值**。
    - 3.利用损失值对模型参数求导，得到响应的梯度。
    - 4.基于梯度调整参数，得到迭代之后的参数。
    - 5.重复上述过程，直到停止条件。
  - **随机梯度下降算法**：每次从训练集中随机选择一个样本，计算其对应的损失和梯度，进行参数更新和反复迭代。

## 1.8、神经网络

- **神经元**：是神经网络进行信息处理的基本单元，其主要是模拟生物神经元的结构和特性，接收输入信号并产出输出。
- **神经元基本组成**：输入信号、线性组合、非线性激活函数。
- **多层感知器**：又称**前馈神经网络**，
  - 分为输入层，隐藏层，输出层。隐藏层可以包括一层或者多层；
  - 每一层都由若干神经元组成，每个神经元承担的计算功能包括线性加权和非线性变换（激活函数）；
  - 层与层之间通过权值建立联系，后一层的每个神经元与前一层的每个神经元都产生连接；
  - 输入和输出层的神经元个数基本已知，隐藏层中每层的神经元个数以及使用的层数都是超参数。
- **激活函数**：它的非线性使得神经网络几乎可以任意逼近任何非线性函数。如果不使用激活函数，无论神经网络有多少层，其每一层的输出都是上一层输入的线性组合，这样构成的神经网络仍是一个线性模型，表达能力有限。
  - 选择：连续可导，可以允许在少数点不可导。
  - 常见函数：
    - **S型激活函数**：Sigmoid和Tanh，特点是有界， $x$ 趋于 $-\infty$ / $+\infty$ 有界
    - **ReLU**（Rectified Linear Unit）函数：线性整流函数，是目前深度学习模型中最常用的激活函数。定义为  $x \geq 0$  时，保持 $x$ 不变进行输出； $x < 0$ 时，输出为零，又称**单侧抑制**。

- **训练神经网络**：使用反向传播法来高效训练
  - **神经网络的运行过程**
    - **前向传播**：给定输入和参数，逐层向前进行计算，最后输出预测结果。
    - **反向传播**：基于前向传播得到的预测结果，使用损失函数得到损失值，然后计算相关参数的梯度。
    - **参数更新**：使用梯度下降算法对参数进行更新，重复上述过程，逐步迭代，直到模型收敛。
  - **梯度消失**：原因在于激活函数的饱和性，函数值在趋于上下边界时，梯度通常比较小，在与误差项的乘积更小，多次的反向传播导致梯度值不断下降，最终在靠近输入层的梯度值很低，几乎无法进行参数有效更新，在下次前向传播时，前面层参数无法提供有价值的信息供后面使用，模型也就难以进行有效更新。这就是梯度消失。
  - **局部最优**：深度模型网络通常存在非常多的局部最优解，往往这些局部最优解都能够保证模型的效果。
  - **鞍点**：指在该处梯度为0，但并不是最大或最小值。

---

## 2、卷积神经网络

---

卷积神经网络（Convolutional Neural Network）是一种具有局部连接、权值共享等特点的深层前馈神经网络，在图像和视频分析领域、比如图像分类、目标检测、图像分割等各种视觉任务上取得了显著的效果

### 2.1、卷积与池化

#### 2.1.1、信号处理中的卷积

- **卷积定理**： $(f * g)(t)$  等价于  $F(w)G(w)$ ，可以将时域中的数据转换到频域中的一种方式。
- **时域**：信号的基本性质，描述数学函数或物理信号对时间的关系。
- **频域**：信号的基本性质，描述信号在频率方面特性时用到的一种坐标系。
- **傅里叶变换**：能将满足一定条件的某个函数表示成三角函数（正弦/余弦）或者他们的积分的线性组合。

## 时域卷积定理

若函数  $f(x)$  以及  $g(x)$  都在  $\mathbb{R}$  上绝对可积，则卷积函数

$$f * g(x) = \int_{-\infty}^{+\infty} f(x-t)g(t)dt$$

的傅里叶变换存在，且

$$\mathcal{F}[f * g] = \mathcal{F}[f] \cdot \mathcal{F}[g]$$

## 频域卷积定理

若  $f(x)$  的傅里叶变换为  $F(\omega)$ ， $g(x)$  的傅里叶变换为  $G(\omega)$ ，则有

$$\mathcal{F}[f(x)g(x)] = \frac{1}{2\pi}[F(\omega) * G(\omega)]$$

## 时域卷积定理

若函数  $f(x)$  以及  $g(x)$  都在  $\mathbb{R}$  上绝对可积，则卷积函数

$$f * g(x) = \int_{-\infty}^{+\infty} f(x-t)g(t)dt$$

的傅里叶变换存在，且

$$\mathcal{F}[f * g] = \mathcal{F}[f] \cdot \mathcal{F}[g]$$

## 频域卷积定理

若  $f(x)$  的傅里叶变换为  $F(\omega)$ ， $g(x)$  的傅里叶变换为  $G(\omega)$ ，则有

$$\mathcal{F}[f(x)g(x)] = \frac{1}{2\pi}[F(\omega) * G(\omega)]$$

- **卷积的典型应用：**针对某个线性时不变的系统，给定输入信号  $f(t)$  和系统响应  $g(t)$ ，求系统的输出。

### 2.1.2、深度学习中的卷积操作

- **单通道卷积：**卷积核不需要进行显示翻转，卷积网络中的卷积核是通过自动学习得到的。

导致的问题：

- 进行多次卷积运算后，输出的尺寸会越来越小
- 越是边缘的像素点，对于输出的影响越小，因为卷积运算在移动到边缘的时候就结束了。中间的像素点会参与多次计算，但是边缘的像素点可能只会参与一次运算，这就导致边缘信息的丢失。

解决的办法：

- 在输入矩阵的边缘使用0进行填充，使得边缘处的像素值也能进行计算。输入维度由  $M*N$  变成  $M+2*N+2$
- **多通道卷积：**在单通道卷积的基础上添加了通道数  $C$ ，多通道卷积使用多个卷积核来提取更丰富的特征。

计算过程：

- 输入向量  $X$ ， $H*W$  维， $C$  个通道；与卷积核  $R$ ， $K*k$  维， $C$  个通道， $C'$  个卷积核个数进行卷积运算
- 卷积完成后，为每个特征图加一个偏置
- 得到输出向量  $Y$ ， $H'*W'$  维， $C'$  个通道
- 过程中的参数为卷积核与偏置，参数总量为  $K^2*C*C'+C'$
- 池化

- 目的：降维，降低计算量，并在训练初期提供一些平移不变性。
- 具体：使用一个固定大小的滑窗在输入上滑动，每次将滑窗内的元素聚合为一个值作为输出。
- 分类：平均池化，最大支持化。

## 2.2、卷积神经网络

通过将卷积层与池化层进行堆叠得到的

### 2.2.1、卷积神经网络的结构

- 一部分由**卷积层和池化层交替堆叠**构成的骨干网络，主要用于从输入中提取丰富的特征。
- 一部分是全连接层，将卷积得到特征图展平，主要作用是聚合全局信息并将其映射到输出空间。

### 2.2.2、卷积神经网络的特点

- **局部连接**：图像通常具有局部连接性，卷积计算每次只在与卷积核大小对应的区域进行。输入和输出是局部连接的。
- **权值共享**：不同的区域使用相同的卷积核参数，减少了参数量，带来了平移不变性。
- **层次化表达**：低层次的卷积一般是提取一些简单的特征，如颜色、边缘、角，对应局部性特征。  
中间层次的卷积得到的特征开始变得抽象，比如纹理结构。  
高层次的卷积得到的特征更加抽象，与图像的语义、具体包含的目标。

### 2.2.3、特殊的卷积形式

- **1\*1 卷积**：用于信息聚合，增加非线性和通道数的变化。
- **转置卷积**：语义分割任务中不可分割的模块，对对象在像素级别上进行分类。
- **空洞矩阵**：超参数空洞率 $r$ 控制卷积核（ $K*K$ ）的扩张程度，使用0填充，扩张后的卷积核大小变为 $k+(k-1)(r-1)$ 。
- **分组卷积**：把特征图分给了多个GPU来处理，最后再把多个GPU的处理结果融合。
- **深度可分离卷积**：一部分沿着深度的逐层卷积，另一部分是1\*1卷积。

---

## 3、表示学习

表示学习是一类可以自动地从数据中去学习有用的特征，并可以直接用于后续的具体任务的方法的统称。

### 3.1.1、表示学习的意义

- 从数据中得到具有判别性特征的方法，减少机器学习对特征工程（人工标注的特征）的依赖。

### 3.1.2、端到端的学习

端到端的学习是一种强大的表示学习方法

- **是什么**：以原始图像作为输入，而不是特征工程得到的特征，输出直接是预测的类别，这中学习方式就是端到端学习。去掉怀疑
- **深度学习模型的优势**：
  - 反向传播算法将误差从输出层向前传递直到输入层，优化算法动态的调节模型参数，使得模型可以自动提取到与任务相关的判别性特征。
  - 能够学习到数据的层次化表达，位于低层的变换得到基础的特征，是构成高层抽象特征的基础。

- **表示学习的任务**：通常是学习这样一个映射： $f: X \rightarrow R$ ，即将输入映射到一个稠密的低维向量空间中。

### 3.1.3、典型表示学习的方法：

- **基于重构损失的方法**-----自编码器：
- **基于对比损失的方法**-----Word2vec：

## 4、图信号处理与图卷积神经网络

图信号处理是离散信号处理理论在图信号领域的应用

### 4.1、图信号与图的拉普拉斯矩阵

- **图信号**：图信号是定义在图节点上的信号，性质包括图信号的强度和图的拓扑结构。
- **拉普拉斯矩阵**：用来研究图的结构性质的核心对象，定义为  $L = D - A$ ， $D$ 是图的度矩阵， $A$ 是图的邻接矩阵。拉普拉斯矩阵是一个反映图信号局部平滑度的算子。

### 4.2、图傅里叶变换

- **是什么**：傅里叶变换是数字信号处理的基石，傅里叶变换将信号从时域空间转换到频域空间，频域视角给信号的处理带来了极大的便利。
- **图位移算子**：算子就是把一个  $R^2$  上的函数变化到另一个  $R^2$  上的函数的一个变换。原图每个点到另一幅图的映射，是通过矩阵（图像处理时也称为掩膜）用来与原图像做运算而实现的。
- **定义**：对于任意一个在图  $G$  上的信号，其图傅里叶变换为： $x_k = \langle v_k, x \rangle$  矩阵点积的和，特征向量称为傅里叶基， $x_k$  是  $x$  在第  $k$  个傅里叶基上的傅里叶系数，傅里叶系数的本质上是图信号在傅里叶基上的投影。
- **总变差**：总变差是图的所有特征值的一个线性组合，权重是图信号相对应的傅里叶系数的平方。代表着图信号整体平滑度的实际意义。
- **图信号的能量**： $E(x) = ||x||^2 = x^T \cdot x$  ( $x_k = \langle v_k, x \rangle$ ) 图信号的能量可以同时从空域和时域进行等价定义。单位向量的图信号能量为1。有了频率的定义，傅里叶系数就可以等价成为图信号在对应频率分量上的幅值，反映了图信号在该频率分量上的强度。
- **图滤波器**：对给定信号的频谱中各个频率分量的强度进行增强或衰减的操作。

性质：

- 线性： $H(x+y) = Hx + Hy$
- 滤波操作是顺序无关的： $H_1(H_2x) = H_2(H_1x)$
- 如果  $H(\lambda) \neq 0$ ，则该滤波操作是可逆的

空域角度：

- 具有局部性，每个节点的输出信号值只需要考虑其  $K$  阶子图
- 可通过  $K$  步迭代式的矩阵向量乘法来完成滤波操作

频域角度：

- 从频域视角能够更加清晰的完成对图信号的特定滤波操作。
- 图滤波器如何设计具有显式的公式指导
- 对矩阵进行特征分解是一个非常耗时的操作，具有  $O(n^3)$  的时间复杂度，相比空域视角中的矩阵向量乘法而言，有工程上的局限性。

- **滤波器种类**：

- **低通滤波器**：只保留信号中的低频部分，更加关注信号中平滑的部分。
- **高通滤波器**：只保留信号中的高频部分，更加关注信号中快速变化的部分。



- **带通滤波器**：只保留信号特定频段的成分。

---

## 5、GCN图神经网络的性质

---

### 5.1、GCN与CNN的关系

- **图象是一种特殊的图数据**，CNN中的卷积运算相较于GCN中的卷积运算，最大的区别是没有显式的表达出邻接矩阵。**GCN中的卷积计算是用来处理更普遍的非结构化的图数据的。**
- **从网络连接方式来看，二者都是局部连接。**GCN的计算作用在其一阶子图上；CNN的计算作用在中心像素的 $N \times N$ 像素栅格内；节点下一层的特征计算只依赖于自身邻域的方式，在网络连接上表现为局部连接的结构。
- **二者卷积核的权重式处处共享的。**都作用于全图的所有节点。参数共享会大大减少每一层网络的数量，可以有效地避免过拟合现象的出现。
- **从模型的层面看，感受域随着卷积层的增大而变大。**每多一层卷积运算，中心节点就能多融合进更外一圈的信息。

### 5.2、GCN能够对图数据进行端到端的学习（一端是数据，一端是任务）

端到端的学习实现了一种自动化地从数据种进行高效学习的机制，但是离不开背后大量的针对特定类型数据的学习任务的适配工作

#### 5.2.1、图数据包含的信息

- **属性信息**：描述了图中对象的固有属性。
- **结构信息**：描述了对象之间的关联性质。

#### 5.2.2、GCN计算过程

首先对属性信息进行仿射变换，学习了属性特征之间的交互模式，然后迭代式的聚合邻居节点的特征，从而更新当前节点的特征。

#### 5.2.3、GCN的优势

- GCN对表示学习和任务学习一起进行端到端的优化。
- GCN对结构信息和属性信息的学习是同步进行的（在同一个网络层里同时学习）。

### 5.3、GCN是一个低通滤波器

### 5.4、GCN的问题-过于平滑

在使用多层GCN之后，节点的区分性变得越来越差，节点的代表向量趋于一致，这使得相应的学习任务变得更加困难，我们将这个问题称为多层GCN的过平滑问题

---

## 6、GNN的通用框架

---

### 6.1、MPNN（Message Passing Neural Network）消息传播神经网络

基本思路：节点的代表向量通过消息函数M（Message）和更新函数U（Update）进行K轮消息传播机制的迭代后得到的。

## 6.2、NLNN (Non-Local Neural Network) 非局部神经网络

非局部神经网络是对注意力机制的一般化总结。

基本思路：NLNN是通过non-local操作将任意位置的输出响应计算为所有位置特征的加权和。

## 6.3、GN (Graph Network) 图网络

基本计算单元包含三个要素：节点的状态、边的状态、图的状态

基本思路：由点更新边，边聚合更新点，点聚合与边聚合更新图。每个元素在更新的时候都需要考虑自身上一轮的状态。

# 7、图分类

图分类任务中实现层次化池化的机制，是GNN需要解决的基础问题

## 7.1、基于全局池化的图分类

**读出机制：**对经过K轮迭代的所有节点进行一次性聚合操作，从而输出图的全局性表示。

## 7.2、基于层次化池化的图分类

- **基于图坍缩 (Graph Coarsening) 的池化机制：**将图划分成不同的子图，然后将子图视为超级节点，从而形成一个坍缩的图。实现对图的全局信息的层次化学习。**实质是节点不断聚合成簇的过程。**
  - **DIFFPOOL算法：**通过学习出一个簇分配矩阵进行子图的划分，得到簇连接强度，确定邻接矩阵，对节点特征进行加和（这样加和损失了子图本身的结构信息）。
  - **EigenPooling算法：**通过图分区的算法（如谱聚类算法）实现对图的划分，得到簇连接强度，之后用子图上的信号在该子图上的图傅里叶变换来代表结构信息于属性信息的整合输出。
- **基于TopK的池化机制：**对图中每个节点学习出一个分数（重要度），基于这个分数的排序丢掉一些底分数的节点，借鉴最大池化，将更重要的信息筛选出来。**实质是一个不断丢弃节点的过程。**
- **基于边收缩 (Edge Contraction) 的池化机制 (Edge Pool)：**边收缩是指并行的将图中的边移除，并将被移除边的两个节点合并，同时保持被移除节点的连接关系。通过归并操作实现图信息的层次化学习。**融合边收缩变换操作和端到端的学习机制。**

# 8、基于GNN的图表示学习

基于GNN图表示学习的优势：

- 非常自然的**融合**了图的属性信息进行学习，而之前的方法大多把图里面的结构信息与属性信息单独进行处理。
- GNN本身作为一个可导的模块，可以任意嵌入到一个支持端到端学习的系统中去，这种特性使得能够与各个层面的有监督学习任务进行有机结合，学习出更加适应该任务的数据表示。
- GNN的很多模型支持归纳学习，多数情况下对新数据的表示学习可以进行直接预测，而不必重新训练。
- 相较于分解类（转换到低维矩阵向量）的方法只能适应小图的学习，GNN保证了算法在工程上的可行性，也能适应大规模的学习任务。

## 8.1、图表示学习



- **主要目标：**将图数据转化成低维稠密的向量化表示。
- **学习对象：**节点的表示学习一直是图表示学习的主要对象。
- **核心：**图表示学习的核心也是研究数据的表示。
- **非线性结构作用下，图表示学习的作用：**
  - 将图数据表示成线性空间中的向量。向量化的表示为擅长处理线性结构数据的计算机提供了便利。
  - 为之后的学习任务奠定基础。图数据的学习任务种类繁多，有节点层面的，有边层面的，还有全图层面的，一个好的表示学习方法可以统一高效地辅助完成这些工作。

## 8.2、基于GNN的图表示学习（无监督的学习方式）

无监督学习的主体在于损失函数的设计，以下是两类损失函数

- **基于重构损失的GNN：**核心在于将节点之间的邻接关系进行重构学习，以及加噪降低GNN学习过平滑的问题。

加噪的常见做法：

- 对原图数据的特征矩阵 $X$ 适当增加随机噪声或进行置零处理；
- 对原图数据的邻接矩阵 $A$ 删除适当比例的边，或者修改边上的权重值。
- **基于对比损失的GNN：**通常对比损失会设置一个评分函数 $D$ ，该得分函数会提高“真实/正”样本的，降低“假/负”样本的得分。

类比词向量，将对比损失的落脚点放在词与上下文中。词就是节点，上下文可以是邻居节点，节点所在的子图、全图。

- **邻居作为上下文：**对比损失就在建模节点与邻居节点的共现关系。
- **子图作为上下文：**强调节点之间的共现关系，更多反映图中节点间的距离关系，缺乏对节点结构相似性的捕捉。
- **全图作为上下文：**对图数据进行加噪，得到负样本数据，将这两组图数据送到同一个GNN中进行学习。为了得到图的全局表示，我们使用读出机制对局部节点的信息进行聚合。在最后的损失函数中，固定全图表示。