

图神经网络 & 半监督学习

本次调研主要动机是由于近期实验实验涉及相关问题而进行的调研，由于实验的设置是半监督学习模式，而传统 GNN 模型或异构信息网络表示模型在应用的时候，大多没有对半监督学习模式下的 GNN 模型做深入的研究。通过调研发现论文主要集中于这两年，分两大类方向，本文讲对这两大类的三篇顶会文章进行介绍。

目 录

图神经网络 & 半监督学习	1
1. 介绍	2
2 相关研究	2
2.1 NeurIPS 2020. GRAND	2
2.1.1 简介	2
2.1.2 模型介绍	2
2.1.3 实验&结论	4
2.2 NeurIPS 2019. GraphMix	4
2.2.1 简介	4
2.1.2 模型介绍	4
2.2.3 实验&结论	6
2.3 NeurIPS 2020. GSNN	6
2.3.1 简介	6
2.3.2 模型介绍	6
2.3.3 实验结果	7
3. 结论	7
参考文献	9

1. 介绍

本次调研主要动机是由于近期实验中遇到相关的问题而进行的相关调研，由于实验的设置是半监督学习模式，而传统 GNN 模型或异构信息网络表示模型在应用的时候，大多没有对半监督学习模式下的 GNN 模型做深入的研究。

通过调研发现，GNN 模型的半监督学习方面的深入研究在近两年才引起学界的深入研究，相关论文较少且基本都是很重量级的论文。通过阅读近两年来发表在 NIPS 顶会上的三篇论文，总体上了解了图上半监督学习理论研究的两个方向，下面主要对这三篇文章进行一个介绍。

2. 相关研究

在该部分主要介绍调研的近两年来发表在 NIPS 顶会上的三篇论文。它们分别从三个角度来设计模型，提升在图神经网络上半监督分类的性能。具体来讲，主要分为：基于数据增强的方法和基于生成式模型的方法两类。前两篇介绍的文章属于基于数据增强的方法，其中，GRAND 模型是通过随机 drop 节点，并通过对比不同 drop 下，模型的训练结果是否一致，来引入一致性正则损失，提升模型鲁棒性；而 GraphMix 是通过引入一个参数共享的 FCN 模型来增强数据，该模型类似于联邦学习，通过交替优化 FCN 和 GNN 网络，使得模型学习到的表示最优，进而提升模型分类效果。第三篇文章 GSNN 通过引入生成式模型架构来提升对稀疏标注和噪声图数据的分类。

2.1 NeurIPS 2020. GRAND

GraphRandom Neural Network for Semi-Supervised Learning on Graphs. [PDF](#)

2.1.1 简介

在 NeurIPS 2020 上，清华大学联合微众银行、微软研究院以及博世人工智能中心提出了 Graph Random Neural Network (GRAND)，一种用于图半监督学习的新型图神经网络框架。在模型架构上，GRAND 提出了一种简单有效的图数据增强方法 Random Propagation，用来增强模型鲁棒性及减轻过平滑。基于 Random Propagation，GRAND 在优化过程中使用一致性正则（Consistency Regularization）来增强模型的泛化性，即除了优化标签节点的 cross-entropy loss 之外，还会优化模型在无标签节点的多次数据增强的预测一致性。GRAND 不仅在理论上有良好的解释，还在三个公开数据集上超越了 14 种不同的 GNN 模型，取得了 SOTA 的效果。

2.1.2 模型介绍

如图 2-1，该模型主要由三部分组成：①随机传播；②特征增强；③一致性正则训练。

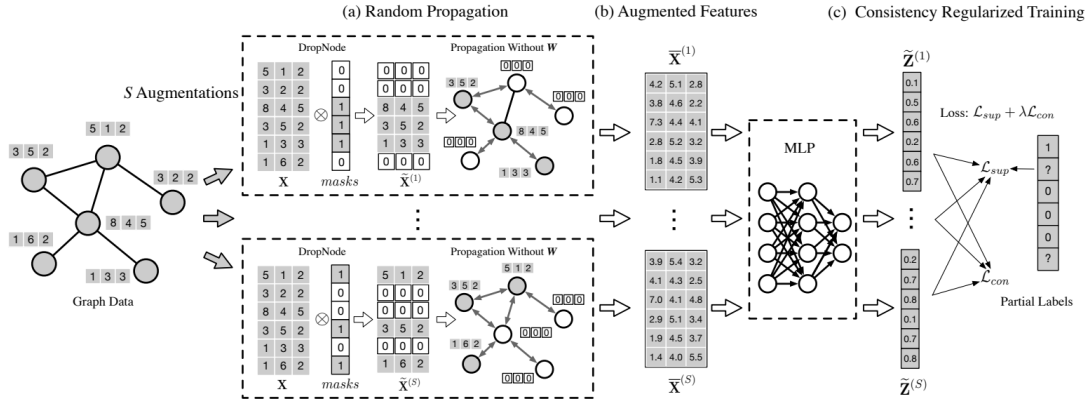


Figure 1: Illustration of GRAND with DropNode as the perturbation method. GRAND designs random propagation (a) to generate multiple graph data augmentations (b), which are further used as consistency regularization (c) for semi-supervised learning.

图 2-1 GRAND 模型架构

2.1.2.1 随机传播

如图 2-1 (a) 所示，该部分引入一个 mask 机制来随机 drop 部分节点的特征，相当于去除部分节点，然后在剩余节点上进行信息的传播(message passing)。引入该机制（作者称之为 DropNode），目的是为了增强鲁棒性。可以看作是一种正则化机制。作者基于同质性假设，认为被 mask 的节点的特征会被其同质的邻居（包含多跳邻居，可通过邻接矩阵阶乘获得第 k-hop 邻接矩阵，再通过加和获取多跳内的邻接矩阵）的特征进行补充。

区别于 dropout，DropNode 是 mask 掉一个节点上的所有特征，而 dropout 是 mask 掉某些特征。

作者运用 s 次随机传播机制，通过 s 次 mask 得到不同的节点表示，后续过程通过正则手段来进一步增强鲁棒性。

2.1.2.2 特征增强

与以往运用 softmax 或 sigmoid 非线性函数来预测概率不同，高部分使用两层 MLP 来学习最终的 logits。

$$\tilde{\mathbf{Z}}^{(s)} = f_{mlp}(\overline{\mathbf{X}}^{(s)}, \Theta),$$

不适用非线性函数的好处是缓解了 GNN 模型 over-smoothing 的通病。

2.1.2.3 一致性正则训练

该部分同样为了增强鲁棒性。

半监督训练的基本损失函数为：

$$\mathcal{L}_{sup} = -\frac{1}{S} \sum_{s=1}^S \sum_{i=0}^{m-1} \mathbf{Y}_i^\top \log \tilde{\mathbf{Z}}_i^{(s)}.$$

该公式表示在 m 个有标注节点上对得到的 s 个预测概率分布（第一步确定有 s 个随机传播生成的节点表示）上进行交叉熵验证。

作者引入另外一个损失函数——一致性正则，该正则针对于所有的节点（既有非标注，也有带标注的节点），主要用于衡量得到的 s 个预测概率分布是否一致。采用以下步骤并分别对应相应公式：

- ① 分别计算各个节点 s 个预测类别概率分布的平均值；

$$\bar{\mathbf{Z}}_i = \frac{1}{S} \sum_{s=1}^S \tilde{\mathbf{Z}}_i^{(s)}.$$

- ② 运用类似 softmax 的方法计算最终各类别的概率值；

$$\bar{\mathbf{Z}}'_{ij} = \bar{\mathbf{Z}}_{ij}^{\frac{1}{C}} / \sum_{c=0}^{C-1} \bar{\mathbf{Z}}_{ic}^{\frac{1}{C}}, (0 \leq j \leq C-1),$$

- ③ 在所有节点上计算一致性正则损失。

$$\mathcal{L}_{con} = \frac{1}{S} \sum_{s=1}^S \sum_{i=0}^{n-1} \|\bar{\mathbf{Z}}'_i - \tilde{\mathbf{Z}}_i^{(s)}\|_2^2.$$

应该认识到，该正则目的是增强鲁棒性，方法是自监督。

2.1.3 实验&结论

作者进行了大量的实验来验证提出的方法的优越性，并验证了各个组件的增益。实验非常全面，几乎涵盖各种 GNN，并在不同大小的数据集上进行验证。

2.2 NeurIPS 2019. GraphMix

GraphMix : Regularized Training of Graph Neural Networks for Semi-Supervised Learning [PDF](#)

2.2.1 简介

GraphMix 是一种基于图神经网络的半监督对象分类的正则化技术，它利用了经典深度神经网络正则化的最新进展。具体来说，本文提出了一种统一的方法，其中通过参数共享，基于插值的正则化和自我预测的目标，与图神经网络一起训练全连接的网络。提出的方法与结构无关，因此它可以应用于图神经网络的任何变体。尽管 GraphMix 机制很简单，但通过它，我们甚至可以使用更简单的架构（例如 GCN）来持续改善结果并达到或接近最先进的结果。

2.1.2 模型介绍

GraphMix 的模型架构如图 2-2 所示，图中不同颜色标识节点是否带标签。

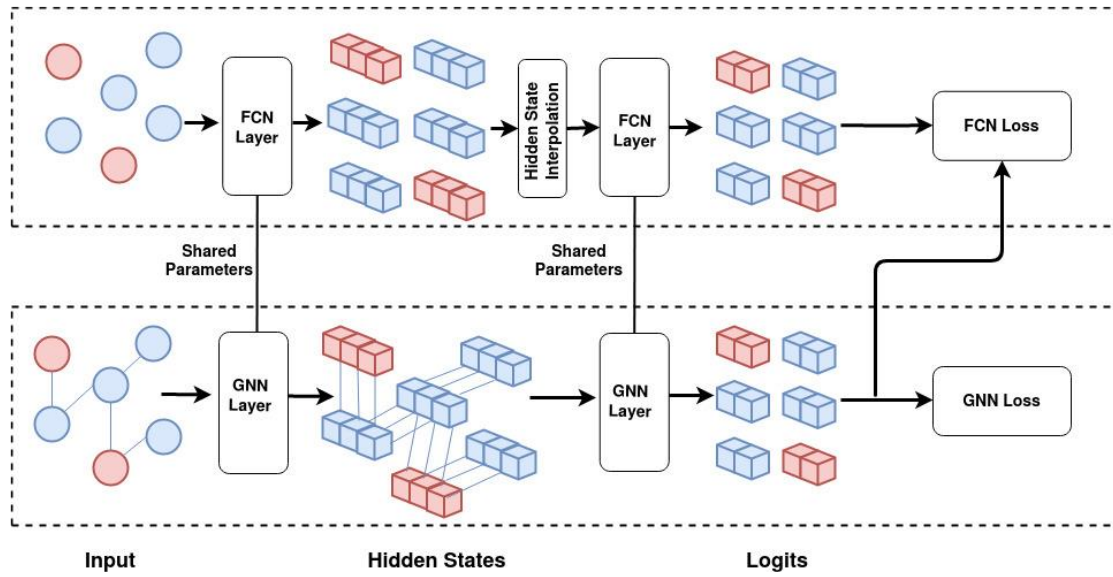


图 2-2 GraphMix 模型结构

从上图可知，GraphMix 通过使用一个全连接网络（FCN）来增强图神经网络的学习，在这两个网络中，它们在节点特征变换的阶段共享参数。全连接网络通过集成节点的隐状态以及标注样本进行训练，因此可以使得更好的节点表示迁移到 GNN 模型中；另一方面，GNN 通过利用图结构对未知标签节点的预测可被用于增强 FCN 模型的输入。最终，FCN 损失和 GNN 损失交替优化。

图 2-3 显示的是 GraphMix 的训练过程。其中，第 6 行是使用 *Manifold Mixup* 进行 FCN 监督损失的计算；第 10 行通过使用 GNN 模型对未标注节点进行预测；12 行通过 FCN 模型进行未标注节点的预测；13 行计算 FCN 的带权损失；15 行使用传统方法计算 GNN 模型的损失。

Algorithm 1 GraphMix : A procedure for improved training of Graph Neural Networks (GNN)

```

1: Input: A GCN:  $g(X, A, \theta)$ , a FCN:  $f(X, \theta, \lambda)$  which shares parameters with the GCN. Beta distribution
   parameter  $\alpha$  for Manifold Mixup. Number of random perturbations  $K$ , Sharpening temperature  $T$ .
   Consistency parameter  $\gamma$ . Number of epochs  $N$ .  $\gamma(t)$ : rampup function for increasing the importance of
   consistency regularization.
2: for  $t = 1$  to  $N$  do
3:    $i = \text{random}(0,1)$  // generate randomly 0 or 1
4:   if  $i=0$  then
5:      $\lambda \sim \text{Beta}(\alpha, \alpha)$  // Sample a mixing coefficient from Beta distribution
6:      $\mathcal{L}_{sup} = \mathcal{L}(f(X_L, \theta, \lambda), Y_L)$  // supervised loss from FCN using the Manifold Mixup
7:     for  $k = 1$  to  $K$  do
8:        $\hat{X}_{U,k} = \text{RandomPerturbations}(X_U)$  // Apply  $k^{\text{th}}$  round of random perturbation to  $X_U$ 
9:     end for
10:     $\bar{Y}_U = \frac{1}{K} \sum_k g(Y | \hat{X}_{U,k}; \theta, A)$  // Compute average predictions across  $K$  perturbations of  $X_U$ 
        using the GCN
11:     $Y_U = \text{Sharpen}(\bar{Y}_U, T)$  // Apply temperature sharpening to the average prediction
12:     $\mathcal{L}_{usup} = \mathcal{L}(f(X_U, \theta, \lambda), Y_U)$  // unsupervised loss from FCN using the Manifold Mixup
13:     $\mathcal{L} = \mathcal{L}_{sup} + \gamma(t) * \mathcal{L}_{usup}$  // Total loss is the weighted sum of supervised and unsupervised FCN
        loss
14:   else
15:      $\mathcal{L} = \mathcal{L}(g(X_L, \theta, A), Y_L)$  // Loss using the vanilla GCN
16:   end if
17: end for
18: return  $\mathcal{L}$ 

```

在本文方法中，主要引入了 *Manifold Mixup* 机制，并通过利用参数共享机制将其融入到 GNN 模型框架中，以实现在 GNN 框架下的数据增强，获得更好的节点表示及预测结果。

2.2.3 实验&结论

本文使用经典的 GNN 模型（如 GCN、GAT）作为基础，引入 FCN 网络来实现数据增强，最终结果显示，GraphMix 机制相较于基础 GNN 模型有非常大的提升。

2.3 NeurIPS 2020. GSNN

Graph Stochastic Neural Networks for Semi-supervised Learning

2.3.1 简介

图神经网络（GNN）在半监督节点分类任务中取得了卓越的性能。大多数现有模型都学习判别式分类方法，但是，在存在各种不完美的观测数据（例如标记节点稀少和图结构带有噪声）的情况下，传统的判别式模型缺乏足够的灵活性来探索更好的节点表示。为了提高判别式分类方法的性能，本文提出了一种新的框架，称为图随机神经网络（GSNN），旨在通过同时学习一系列随机函数来建模分类函数的不确定性。具体来说，本文引入了一个可学习的图神经网络，结合高维隐变量来对分类函数的分布进行建模，并进一步采用 *amortised* 变分推论来近似标签和隐变量缺失的联合后验。通过最大化标注样本的似然下限，以端到端的方式，可以有效地训练模型。

2.3.2 模型介绍

如 2-4 展示的是 GSNN 的模型架构。

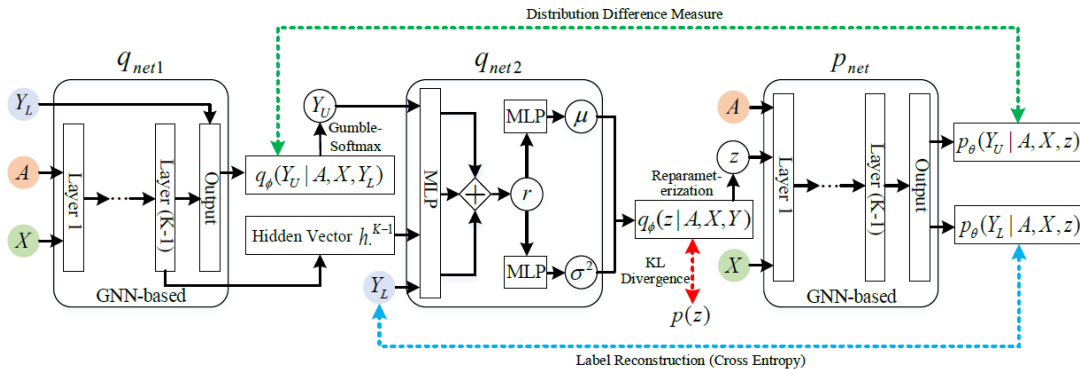


图 2-4 GSNN 模型架构

应该认识到 GSNN 模型是利用生成式方法来学习分类函数分布，以获取更好的分类结果。该模型主要包含三个网络：

1) q_{net1} :该网络就是传统的 GNN 模型。以图结构、节点初始表示、及相关标签数据作为输入，学习得到节点的引状态，以及分类函数 q_ϕ 。

$$h_v^k = \rho^{k-1} \left(\sum_{u \in Ne\{v\} \cup \{v\}} a_{v,u}^{k-1} h_u^{k-1} W_{q_{net1}}^{k-1} \right), \quad k = 1, \dots, K$$

$$q_\phi(y_v | A, X, Y_L) = Cat(y_v | h_v^K), \quad v \in V_U$$

2) q_{net2} :这一网络以及后面的网络就属于经典的变分自编码器的框架了。该网络基于节点进行，以 q_{net1} 的节点分类结果，以及学习到的该节点隐状态和真实的标签作为输入，利用三个 MLP 学习隐变量的分布。

$$r_v = \text{MLP}([h_v^{K-1} || y_v]), \quad v \in V$$

$$r = \text{Readout}(\{r_v\}_{v \in V})$$

$$q_\phi(z | A, X, Y) = \mathcal{N}(z; \mu(r), \sigma^2(r)I)$$

3) q_{net3} : 通过采样隐变量，结合图结构以及节点初始表示作为输入，使用 GNN 模型学习到最终的节点分类函数。

$$p_\theta(y_v | A, X, z) = Cat(y_v | e_v^K), \quad v \in V$$

模型训练阶段利用了诸多手段，比如利用蒙特卡洛估计来估计期望，利用重参数化方法计算方差，利用 Gumbel-Softmax 重参数化方法进行梯度回传。

总的来说，本文基于变分自编码(VAE)的相关理论，设计了一个融合 GNN 模型的分函数生成模型，想要理解这一部分需要对 VAE 的理论知识进行深入了解。

2.3.3 实验结果

本文比较的对象主要分为两部分，一部分是传统的 GNN 模型（如 GCN、GAT），另一部分是与本文类似，对图结构进行不确定性建模。实验结果表明 GSNN 在大多数 setting 上都优于 baseline 模型。

3. 结论

本文介绍了三篇关于图神经网络半监督学习方面的研究。可以说在图上进行节点分类本质上就是一种半监督学习的模式，但是由于面临稀疏标签、对抗数据、噪声数据等问题，怎样提升深入研究 GNN 半监督分类性能成为这两年来学界的一个研究方向。

通过上面方法介绍，可以看到本文这些新的理论产出基本都是基于传统深度学习中的相关研究结果，比如一致性损失、联邦学习、生成式模型等等，他们通过引入这些可能在图像分类、文本分类等非结构数据领域中的技术，以增强在图

结构数据下模型能力，这种移植性非常值得借鉴。此外，由于本文介绍的相关技术都只是在 GNN 的某几个机制上做些改动，因此通用性比较强，怎样将其应用到异质图神经网络中去是下一个研究目标。

参考文献

- [1] Feng, Wenzheng, Jie Zhang, Yuxiao Dong, Yu Han, Huanbo Luan, Qian Xu, Qiang Yang, Evgeny Kharlamov, and Jie Tang. "Graph Random Neural Networks for Semi-Supervised Learning on Graphs." *Advances in Neural Information Processing Systems* 33 (2020).
- [2] Verma, Vikas, Meng Qu, Alex Lamb, Yoshua Bengio, Juho Kannala, and Jian Tang. "Graphmix: Regularized training of graph neural networks for semi-supervised learning." *arXiv preprint arXiv:1909.11715* (2019).
- [3] Wang, Haibo, Chuan Zhou, Xin Chen, Jia Wu, Shirui Pan, and Jilong Wang. "Graph Stochastic Neural Networks for Semi-supervised Learning." *Advances in Neural Information Processing Systems* 33 (2020).
- [4] Verma, Vikas, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. "Manifold mixup: Better representations by interpolating hidden states." In *International Conference on Machine Learning*, pp. 6438-6447. PMLR, 2019.