

Simple and Deep Graph Convolutional Networks

摘要

图卷积神经网络 (GCN) 是一种强大的深度学习方法, 适用于图结构数据。最近, GCNs 及其后续变体在现实世界数据集上的各种应用领域都表现出了卓越的性能, 尽管取得了成功, 但目前大多数 GCN 模型是浅层的, 原因是存在过度平滑问题。在本文中, 我们研究了深度图卷积网络的设计和分析问题。我们提出了 GCNII, 它是对普通 GCN 模型的扩展。GCNII 是对 vanilla GCN 模型的扩展, 采用了两种简单而有效的技术: 初始残差和恒等映射。我们提供的理论和实证证据表明, 这两种技术能有效缓解过度平滑问题。我们的实验表明, 在各种半监督 and 全监督任务中, 深度 GCNII 模型都优于最先进的方法。

关键词: GCN; GCNII; 深度图神经网络

1 引言

理论上一个神经网络更深的时候它获取的信息会更多, 模型整体表现也会更好, 但是对于图神经网络, 无论是 GCN, 还是 GAT, 它们最好的效果都是由两层获得的, 当给它们提高模型的深度, 模型的表现反而下降了, 这称为过度平滑 (over-smoothing)。通过在基本的图卷积网络 (GCN) 框架上添加初始残差和单位映射, GCNII 模型在增加深度的同时防止模型受到过度平滑问题的影响, 从而提高了在各种任务上的性能表现。

2 相关工作

2.1 图神经网络

图神经网络 (Graph Neural Network, 简称 GNN) 是一类专门用于处理图结构数据的深度学习模型。与传统的深度学习模型专注于处理向量数据不同, GNN 被设计用于学习和推理图数据中节点和边的复杂关系, 在推荐系统、生物信息学、交通规划等领域取得了显著的成功。然而许多图神经网络存在过平滑问题, 无法拓展到深层, 随着网络深度的增加整体性能下降。针对这个问题, Xu 等人提出 JKNet [6], 使用密集跳过连接来组合每一层的输出, 以保留节点表示的局部性以缓解过度平滑; Rong 等人提出 DropEdge [4], 通过从输入图中随机删除一些边, 可以减轻过度平滑的影响; Wu 等人提出 SGC [5], 尝试通过在单个神经网络层中应用图卷积矩阵的 K 次方来捕获图中的高阶信息; Gasteiger 等人提出 PPNP 和 APPNP [1], 将图卷积矩阵的幂替换为 Personalized PageRank 矩阵, 以解决过度平滑问题。

2.2 残差网络

残差网络（Residual Network，简称 ResNet）[2] 是一种深度学习架构，通过引入残差连接（residual connections）解决了训练非常深的神经网络时的梯度消失和梯度爆炸问题。它的核心思想是将输入直接添加到网络中间层的输出，构建了一种跨层的信息流通机制，使得网络在学习过程中更容易捕捉到残差部分的信息。这种结构使得训练非常深的网络变得更加容易和高效，为深度学习模型的发展提供了一个重要的思路和技术基础。ResNet 在图像分类、目标检测等领域取得了显著的性能提升，并成为许多深度学习任务中的基础模型。

3 本文方法

3.1 GCNII 模型概述

此部分对 GCNII 模型进行概述，GCNII 模型如图 1所示：

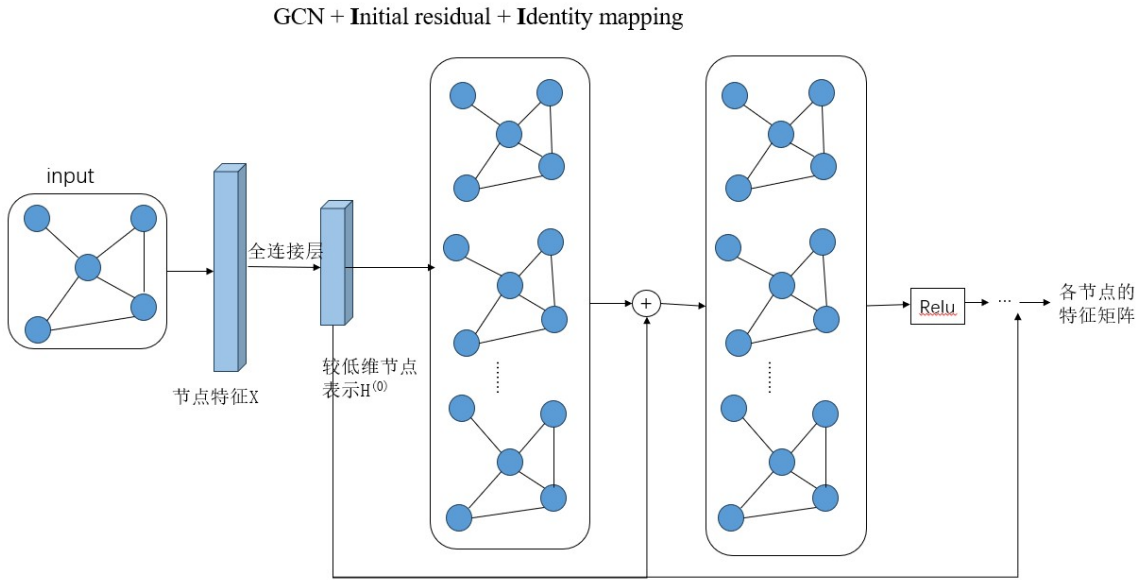


图 1. GCNII 模型

GCNII 模型就是在 GCN 模型的基础上加入了初始残差和单位映射，其公式如下：

$$H^{l+1} = \sigma(((1 - \alpha_l)PH^{(l)} + \alpha_l H^{(0)})((1 - \beta_l)I_n + \beta_l W^{(l)}))$$

在这里，符号的解释如下：

$H^{(l)}$ 是第 l 层的节点表示。 $H^{(l+1)}$ 是第 $l+1$ 层的节点表示，即下一层的节点表示。 σ 是激活函数，通常为 $ReLU$ 。 P 是图卷积矩阵的幂，用于捕捉图的高阶信息。 I_n 是 $n \times n$ 的单位矩阵，其中 n 是节点的数量。 $W^{(l)}$ 是权重矩阵，用于学习每个节点的线性变换。 α_l 和 β_l 是超参数，分别用于控制初始残差和单位映射的权重。

该公式整体描述了 GCNII 模型在每一层的节点更新过程。其中， $(1 - \alpha_l)PH^{(l)} + \alpha_l H^{(l)}$ 部分引入了初始残差机制， $(1 - \beta_l)I_n + \beta_l W^{(l)}$ 部分引入了单位映射机制。这两个机制共同作用，使得模型更好地捕捉图的结构信息，防止过度平滑。

3.2 初始残差模块

为了模拟 ResNet 中的跳跃连接 [2], Kipf 等人 [3] 提出了将平滑表示 $PH^{(l)}$ 与 $H^{(l)}$ 相结合的残差连接。然而, [3] 也表明这种残余连接只能部分缓解过度平滑问题; 当我们堆叠更多层时, 模型的性能仍然会下降。因此作者不使用残差连接来携带来自前一层的消息, 而是构建与初始表示 $H^{(0)}$ 的连接。初始残差连接确保每个节点的最终表示至少保留输入层的 α_l 的一小部分, 即使我们堆叠很多层也是如此。在实践中, 可以简单地设置 $\alpha_l = 0.1$ 或 0.2 , 以便每个节点的最终表示至少包含输入特征的一部分。另外, 作者还提出如果输入特征维度 d 很大, 可以在 X 上应用全连接神经网络来获得较低维的初始表示 $H^{(0)}$ 。

3.3 单位映射模块

但是, 仅仅加入初始残差模块是不够的, APPNP [1] 在个性化 PageRank 的背景下采用了类似的方法来建立初始剩余连接。然而通过实验发现对特征矩阵进行多次非线性操作会导致过度拟合, 从而导致性能下降。因此, APPNP 在不同层之间应用线性组合, 仍然是一个浅层模型, 这表明仅初始残差的想法不足以将 GCN 扩展到深度模型。在此基础上, 作者又引入了单位映射模块。在每一层的权重矩阵中添加一个单位矩阵, 与 ResNet 的动机类似, 单位映射可以确保一个深层的 GCNII 模型至少达到与其浅层版本相同的性能。另外, 有研究指出特征矩阵不同维度之间的频繁交互会降低模型在半监督任务中的性能, 而将平滑表示 $PH^{(l)}$ 直接映射到输出可减少这些相互作用。

4 复现细节

4.1 与已有开源代码对比

论文作者提供了开源代码, 但是在 inductive learning 任务中, 作者未提供数据集, 且代码存在错误, 无法成功运行。因此, 我通过网上资源获取了 PPI 数据集, 并对源代码进行了适当的修改, 达到了预期的效果。

4.2 实验内容

实验内容包括三个部分: GCNII 模型在半监督节点分类任务上的表现、GCNII 模型在完全监督节点分类任务上的表现、GCNII 模型在归纳学习上的表现。

实验中用到了八个数据集, 如图 2 所示:

Dataset	Classes	Nodes	Edges	Features
Cora	7	2,708	5,429	1,433
Citeseer	6	3,327	4,732	3,703
Pubmed	3	19,717	44,338	500
Chameleon	4	2,277	36,101	2,325
Cornell	5	183	295	1,703
Texas	5	183	309	1,703
Wisconsin	5	251	499	1,703
PPI	121	56,944	818,716	50

图 2. 数据集

其中 Cora、Citeseer、Pubmed 三个数据集为论文引用数据集，每个节点代表一篇论文，边代表两个论文之间存在引用关系；Chameleon、Cornell、Texas、Wisconsin 四个数据集为网页引用数据集，每个节点代表一个网页，边同样代表两个网页间存在引用或推荐关系；PPI 数据集为蛋白质相互作用数据集，每个节点代表一个特定的蛋白质分子，而边表示两个蛋白质之间存在相互作用。在实验中，作者还引入了 GCNII 的变体，其公式为：

$$H^{l+1} = \sigma((1 - \alpha_l)PH^{(l)}((1 - \beta_l)I_n + \beta_l W_1^{(l)}) + \alpha_l H^{(0)}((1 - \beta_l)I_n + \beta_l W_2^{(l)}))$$

即对平滑表征和初始表征使用不同的权重参数。

4.3 创新点

无

5 实验结果分析

在半监督节点分类任务上：

Method	Cora	Citeseer	Pubmed
GCN	81.5	71.1	79.0
GAT	83.1	70.8	78.5
APNP	83.3	71.8	80.1
JKNet	81.1 (4)	69.8 (16)	78.1 (32)
JKNet (Drop)	83.3 (4)	72.6 (16)	79.2 (32)
Incep (Drop)	83.5 (64)	72.7 (4)	79.5 (4)
GCNII	85.3 (64)	74.3 (32)	80.2 (16)
GCNII*	85.4 (64)	74.2 (32)	79.2 (16)

图 3. GCNII 在半监督节点分类任务上的表现 (复现结果)

在完全监督节点分类任务上：

Method	Cora	Cite.	Pumb.	Cham.	Corn.	Texa.	Wisc.
GCN	85.77	73.68	88.13	28.18	52.70	52.16	45.88
GAT	86.37	74.32	87.62	42.93	54.32	58.38	49.41
Geom-GCN-I	85.19	77.99	90.05	60.31	56.76	57.58	58.24
Geom-GCN-P	84.93	75.14	88.09	60.90	60.81	67.57	64.12
Geom-GCN-S	85.27	74.71	84.75	59.96	55.68	59.73	56.67
APNP	87.87	76.53	89.40	54.30	73.51	65.41	69.02
JKNet	85.25(16)	75.85(8)	88.94(64)	60.07(32)	57.30(4)	56.49(32)	48.42(8)
JKNet (Drop)	87.46(16)	75.96(8)	89.45(64)	62.08(32)	61.08(4)	57.30(32)	50.59(8)
Incep (Drop)	86.86(8)	76.83(8)	89.18(4)	61.71(8)	61.62(16)	57.84(8)	50.20(8)
GCNII	88.45(64)	76.99(64)	89.51(64)	60.09(8)	74.59(16)	69.46(32)	74.12(16)
GCNII*	88.27(64)	76.97(64)	90.14(64)	62.46(8)	78.11(16)	77.30(32)	80.78(16)

图 4. GCNII 在完全监督节点分类任务上的表现 (复现结果)

在半监督节点分类任务上，GCNII 或 GCNII 变体在三个数据集中都实现了最佳性能，在完全监督节点分类任务中，GCNII 或 GCNII 变体在七个数据集中的六个上取得了最好的表现，尤其是在 WISC 数据集上，预测准确率比之前的模型高出了 11%，证明了 GCNII 模型的优越性。另外，将 GCNII 和 GCNII 变体比较，发现 GCNII 在半监督任务上表现得更好，而 GCNII 变体在完全监督任务上表现得更好，这是由于 GCNII 变体模型更为复杂，在标注数据多的情况下能更好的学习图的结构，而在半监督任务中，标注数据少，容易导致过拟合。

在归纳学习任务上：

Method	PPI
GraphSAGE (Hamilton et al., 2017)	61.2
VR-GCN (Chen et al., 2018b)	97.8
GaAN (Zhang et al., 2018)	98.71
GAT (Veličković et al., 2018)	97.3
JKNet (Xu et al., 2018)	97.6
GeniePath (Liu et al., 2019)	98.5
Cluster-GCN (Chiang et al., 2019)	99.36
GCNII	99.53 \pm 0.01
GCNII*	99.56 \pm 0.02

图 5. GCNII 在归纳学习任务上的表现 (作者的结果)


```
Epoch:7997 train loss:0.007 | val loss:0.039 f1:99.078
Epoch:7998 train loss:0.007 | val loss:0.041 f1:99.086
Epoch:7999 train loss:0.007 | val loss:0.041 f1:99.085
Epoch:8000 train loss:0.007 | val loss:0.040 f1:99.059
Train cost: 50899.2485s
Load 7770th epoch
Test f1.:99.53
```

图 6. GCNII 在归纳学习任务上的表现 (复现结果)

```
Epoch:7999 train loss:0.004 | val loss:0.038 f1:99.145
Epoch:8000 train loss:0.004 | val loss:0.036 f1:99.143
Train cost: 50836.8335s
Load 7622th epoch
Test f1.:99.55
(cby_gcni) wangchengyu@gpu01:~/cby/GCNII-master$
```

图 7. $GCNII^*$ 在归纳学习任务上的表现 (复现结果)

整体而言, 复现结果与作者在文中给出的结果几乎一致, 验证了 GCNII 的可行性。

6 总结与展望

GCNII 通过引入初始残差和单位映射的概念, 解决了过度平滑问题, 使得深度图神经网络能够更有效地学习图数据的表示。初始残差通过在每一层构建一个跳跃连接, 而单位映射则通过在权重矩阵中添加一个单位矩阵, 共同作用于防止过度平滑。总体而言, GCNII 的优点在于简单而有效, 通过对传统的 GCN 模型进行简单的修改, 成功地克服了深度图神经网络面临的过度平滑问题。实验证明, GCNII 在半监督、全监督和归纳学习任务上取得了新的 state-of-the-art 成果。展望方面, GCNII 为设计新的图神经网络提供了新的思路, 使得深度神经网络在图数据上的应用更加可行。未来的研究可能会进一步拓展 GCNII 的应用领域, 深入研究其在不同类型的图数据和任务上的性能, 以及如何进一步优化其模型结构。

参考文献

- [1] Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [3] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [4] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. Dropedge: Towards deep graph convolutional networks on node classification. *arXiv preprint arXiv:1907.10903*, 2019.

- [5] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pages 6861–6871. PMLR, 2019.
- [6] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In *International conference on machine learning*, pages 5453–5462. PMLR, 2018.