

1 Рецензия

1.1 Про что статья и аналоги

Статья «You Can Have Better Graph Neural Networks by Not Training Weights at All: Finding Untrained GNNs Tickets» описывает алгоритм поиска подсетей с хорошим качеством внутри случайно инициализированной графовой нейронной сети. Существует два метода поиска таких подсетей: (1) лотерейные билеты, в которых итеративно уменьшают сети (обучается либо исходная сеть, либо подсети) [4, 8] (2) полностью "сырые" подсети, в которых нет никакого обучения самих нейронных сетей [17, 12, 6]. Рассматриваемая статья принадлежит второму типу и исследует наличие таких подсетей в графовых сетях.

В качестве основы для статьи были взяты методы для сверточных сетей [17, 12]. В первой статье впервые вводится понятие Supermask, говорящее о наличии подсети со случайными весами и хорошим качеством. Во второй статье предлагается алгоритм поиска таких сетей, названный edge-порcup, именно с ним и будут сравниваться на графовых сетях. В [6] предлагалось переинициализировать выкинутые веса, но в рассматриваемой нами статье так не делают. Авторы рассматриваемой статьи предложили новый метод UGT (Untrained GNNs Tickets). Отличается он главным образом следующим: (1) У предыдущих алгоритмов на каждом шаге был зафиксирован уровень разреженности весов, здесь же он итеративно увеличивается (2) удаляем не топ- k в каждом слое, а топ- k по всей сети. Эти отличия позволяют увеличить "область" поиска разреженных подсетей и достигать большего уровня разреженности (99% против 50%, которые были у аналогов).

Авторы провели множество экспериментов с различными популярными GNN архитектурами и сравнили свой алгоритм UGT с edge-порcup и полностью обученными сетями, выявив, что при достаточно большом размере сети UGT достигает качества полностью обученной модели, а edge-порcup проседает. Также UGT работает при очень больших уровнях разреженности, а edge-порcup не особо.

Кроме того, авторы задумались о проблеме сглаживания векторов-представлений вершин при слишком глубоких сетях: вершины становились слабо различимыми, что сильно сказывалось на качестве. UGT работает хорошо даже на глубоких сетях, обходя эту проблему. Также UGT опережает другие методы [5], которые как-то меняли архитектуры GNN, чтобы избежать сглаживания. Также присутствует T-SNE визуализация, наглядно показывающая насколько лучше представления у UGT. Последней рассмотренной проблемой было качества на чуть-чуть испорченных данных, где UGT также сравнимо по качеству с полностью обученными сетями.

1.2 Авторы

Авторов у статьи очень много (11), поэтому здесь будут указаны лишь первые трое, указанных в статье:

1. Tianjin Huang, Eindhoven University of Technology.

Много публикация про anomaly detection и adversarial training, имеющие 5-20 цитирований. В 2022 акцент был смещен на лотерейные билеты и подсети. Было написано 3 статьи, включая рассматриваемую: [13, 9, 14], две из них больше про лотерейные билеты, чем про подсети.

2. Tianlong Chen, University of Texas at Austin.

Много статей про графовые сети и лотерейные билеты с очень большим количеством цитирований (> 100). Можно выделить следующие: [15, 3] (опубликованы на NeurIPS), [2] (CVPR), [16] (ICML). Также он является одним из авторов статьи про лотерейные билеты в GNN: [4].

3. Meng Fang, University of Liverpool

Занимается вообще всяким разным, самая популярная статья (248 цитирований) по RL: [7]. Есть несколько статей про графовые нейронные сети, но не про подсети, имеющие от 0 до 70 цитирований: [1, 10, 11]

1.3 Плюсы/Минусы

Начнем с плюсов:

- Тестирование на многих видах GNN архитектур.
- Много экспериментов для каждой рассматриваемой проблемы.
- Есть псевдокод (минус, что он в апендиксе, он реально бы пригодился для понимания).
- На сайте конференции можно найти весь код, если покопаться.

Теперь минусы:

- Как указали другие ревьюеры, непонятно что на самом деле проще: обучить GNN с нуля или использовать UGT.
- Какую инициализацию изначальную используют авторы непонятно (kaiming, xavier, gloriot, uniform, normal...), влияет ли это на эксперименты? У самих инициализаций также могут быть свои гиперпараметры. Например, инициализация в PyTorch и в TensorFlow одной и той же сети будет разная – будет ли хорошая подсеть при обеих видах инициализации? В коде, насколько я видел, есть параметр отвечающий за метод инициализации, но в статье про это ни слова.
- Что будет если замиксовать архитектуры? Например, два слоя GCN, потом два GAT.
- Все обученные сети имеют фиксированные параметры оптимизаторов и схему обучения в целом в экспериментах. Авторы пишут, что выбрали их на основе нескольких статей, описывающих различные хитрости по обучению GNN. Однако непонятно насколько это оптимальные гиперпараметры и будет ли похожее качество, если взять другой датасет или архитектуру для которой гиперпараметры будут совсем другими.
- В отличие от других ревьюеров мне не понравилось верстка статьи. Картинки не сверху / снизу, а посередине текста. Бывает что в тексте ссылка на картинку, которая находится на 2 страницы назад или вперед. Непоследовательное расположение картинок в тексте: картинка с большим номером в тексте описывается раньше, чем с меньшим.

К плюсам самого метода UGT можно отнести тот факт, что он позволяет без изменений сразу решать несколько проблем: поиск подсети, проблема сглаживания, проблема измененного распределения. Однако метод работает хорошо только при достаточно больших сетях. Это неудивительно, поскольку в целом гипотеза о существовании подобных подсетей изначально строилась на том, что должно быть довольно много весов.

1.4 Публикация

Статья была опубликована на Learning on Graphs Conference в качестве устного выступления. Это совсем новая конференция, созданная специально для исследований в области графов, чтобы продвигать статьи, которым не хватает места на крупных конференциях по типу NeurIPS, ICML. Цитирований у статьи на данный момент нет, но и вышла она не так давно: 1я версия 10го сентября, изменена (публикация в LoG) 9го декабря. Тема исследования довольно узкая, поэтому статью не быстро заметят.

1.5 Дальнейшая работа

Основным минусом, на мой взгляд, является непонятная инициализация. Эксперименты и рассуждения с этим связанные были бы полезны. Также, как указывали другие ревьюеры, хотелось бы увидеть не только эмпирические, но и теоретические рассуждения по поводу превосходства UGT в задаче сглаживания, а также как UGT работает не на графовых сетях. По сути в статье не особо использовалось, что у нас именно графовая сеть.

Список литературы

- [1] Yu Cao, Meng Fang, and Dacheng Tao. Bag: Bi-directional attention entity graph convolutional network for multi-hop reasoning question answering. *arXiv preprint arXiv:1904.04969*, 2019.
- [2] Tianlong Chen, Jonathan Frankle, Shiyu Chang, Sijia Liu, Yang Zhang, Michael Carbin, and Zhangyang Wang. The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16306–16316, 2021.
- [3] Tianlong Chen, Jonathan Frankle, Shiyu Chang, Sijia Liu, Yang Zhang, Zhangyang Wang, and Michael Carbin. The lottery ticket hypothesis for pre-trained bert networks. *Advances in neural information processing systems*, 33:15834–15846, 2020.
- [4] Tianlong Chen, Yongduo Sui, Xuxi Chen, Aston Zhang, and Zhangyang Wang. A unified lottery ticket hypothesis for graph neural networks. In *International Conference on Machine Learning*, pages 1695–1706. PMLR, 2021.
- [5] Tianlong Chen, Kaixiong Zhou, Keyu Duan, Wenqing Zheng, Peihao Wang, Xia Hu, and Zhangyang Wang. Bag of tricks for training deeper graph neural networks: A comprehensive benchmark study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [6] Daiki Chijiwa, Shin’ya Yamaguchi, Yasutoshi Ida, Kenji Umakoshi, and Tomohiro Inoue. Pruning randomly initialized neural networks with iterative randomization. *Advances in Neural Information Processing Systems*, 34:4503–4513, 2021.
- [7] Meng Fang, Yuan Li, and Trevor Cohn. Learning how to active learn: A deep reinforcement learning approach. *arXiv preprint arXiv:1708.02383*, 2017.
- [8] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.
- [9] Tianjin Huang, Tianlong Chen, Meng Fang, Vlado Menkovski, Jiaxu Zhao, Lu Yin, Yulong Pei, Decebal Constantin Mocanu, Zhangyang Wang, Mykola Pechenizkiy, et al. You can have better graph neural networks by not training weights at all: Finding untrained gnns tickets. *arXiv preprint arXiv:2211.15335*, 2022.

- [10] Yinya Huang, Meng Fang, Yu Cao, Liwei Wang, and Xiaodan Liang. Dagn: Discourse-aware graph network for logical reasoning. *arXiv preprint arXiv:2103.14349*, 2021.
- [11] Yinya Huang, Lemao Liu, Kun Xu, Meng Fang, Liang Lin, and Xiaodan Liang. Discourse-aware graph networks for textual logical reasoning. *arXiv preprint arXiv:2207.01450*, 2022.
- [12] Vivek Ramanujan, Mitchell Wortsman, Aniruddha Kembhavi, Ali Farhadi, and Mohammad Rastegari. What’s hidden in a randomly weighted neural network? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11893–11902, 2020.
- [13] Lu Yin, Shiwei Liu, Fang Meng, Tianjin Huang, Vlado Menkovski, and Mykola Pechenizkiy. Lottery pools: Winning more by interpolating tickets without increasing training or inference cost. *arXiv preprint arXiv:2208.10842*, 2022.
- [14] Lu Yin, Vlado Menkovski, Meng Fang, Tianjin Huang, Yulong Pei, and Mykola Pechenizkiy. Superposing many tickets into one: A performance booster for sparse neural network training. In *Uncertainty in Artificial Intelligence*, pages 2267–2277. PMLR, 2022.
- [15] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.
- [16] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. When does self-supervision help graph convolutional networks? In *international conference on machine learning*, pages 10871–10880. PMLR, 2020.
- [17] Hattie Zhou, Janice Lan, Rosanne Liu, and Jason Yosinski. Deconstructing lottery tickets: Zeros, signs, and the supermask. *Advances in neural information processing systems*, 32, 2019.