

ГЕНЕРАТОР СИНТЕТИЧЕСКИХ ГРАФОВЫХ ДАННЫХ С КОНТРОЛИРУЕМЫМИ ТОПОЛОГИЧЕСКИМИ ХАРАКТЕРИСТИКАМИ ДЛЯ ТЕСТИРОВАНИЯ ГРАФОВЫХ НЕЙРОННЫХ СЕТЕЙ

Андреева П.О. (Университет ИТМО), Стебеньков А.С. (Университет ИТМО)

Научный руководитель – кандидат технических наук, Шиков Е.Н.
(Университет ИТМО)

Введение. Традиционный подход к оценке качества графовых нейронных сетей заключается в решении задач на ограниченном наборе реальных графов [1, 2, 3]. Однако, такой подход имеет ограничения, так как диапазон топологических характеристик и ассортативности по атрибутам/меткам таких графов является намного более ограниченным по сравнению с возможными соответствующими характеристиками в графах [4]. Оценка результатов новых архитектур графовых нейронных сетей на ограниченном бенчмарке неизбежно может привести к проблеме переобучения – каждая новая архитектура может показывать высокие точности на OGB и совсем низкие на других датасетах, поскольку будет учитывать только конкретные значения некоторых характеристик бенчмарка. В свою очередь, подход, основанный на генерации синтетических графов, позволяет генерировать наборы графов с заранее заданным распределением параметров. Однако, существующие генераторы позволяют настраивать лишь небольшое количество параметров. В связи с этим, возникает необходимость в разработке генератора графов с контролируемыми топологическими характеристиками.

Основная часть. Авторы статьи [5] показали, что некоторые графовые свертки не работают на наборах данных с низкой ассортативностью по меткам из-за механизма агрегации соседей. С тех пор было разработано несколько методов [6, 7, 8] для решения проблемы дисассортативных графов. Однако ассортативность по меткам не единственный параметр, влияющий на результат графовых сверточных сетей. Некоторые другие свойства графа также могут иметь решающее значение. Например, для алгоритма SuperGAT [8] показано, что результат зависит одновременно от ассортативности и средней степени графа. Авторы статьи [9] пришли к выводу, что длина среднего кратчайшего пути графа может влиять на производительность так же, как и ассортативность – если средний кратчайший путь длинный, то несколько слоев графовой сверточной сети не смогут охватить все узлы, в том числе и те, которые могут оказаться важными. Ассортативность по атрибутам также следует учитывать, поскольку атрибуты во многих методах играют роль входных данных в процедуре передачи сообщений, используемой во всех графовых нейронных сетях. Кроме того, следует рассмотреть коэффициент кластеризации, так как некоторые меры сходства функций потерь из методов обучения без учителя основаны на нем. Итак, свойства, которые следует учитывать при сравнительном анализе графовых сверточных сетей: средняя степень, средний коэффициент кластеризации, средняя длина кратчайшего пути, ассортативность по меткам, ассортативность по атрибутам.

Поскольку не существует методов, позволяющих генерировать графы с заранее заданными желаемыми топологическими характеристиками, предлагается решение, которое может быть использовано в связке с любым другим параметрическим генератором графов, как например моделью Барабаши-Альберта [10], LFR [11], ADC-SBM [12], VTER [13] и другими. Это решение представляет собой использование оптимизационного фреймворка для подбора входных параметров генератора, которые необходимо задать, чтобы сгенерированный граф имел желаемые топологические характеристики (конкретные значения коэффициента кластеризации, ассортативности по меткам, средней степени и средней длины кратчайших путей, а также отдельно корреляция между атрибутами и метками вершин). Кроме того, предлагается способ задания атрибутов, позволяющий контролировать ассортативность по атрибутам состоящий из двух фаз: кластеризация графа

и добавление шума к векторам атрибутов кластеров.

Выводы. Проведен анализ по топологическим характеристикам, влияющим на результат предсказания на графах. Предложен вариант генератора для контролирования выбранных топологических характеристик. Построен фреймворк оптимизации входных параметров для получения графа с желаемыми топологическими характеристиками. Также предложен новый метод генерации атрибутов, позволяющий задавать значение ассортативности по атрибутам независимо от ассортативности по меткам.

Список использованных источников:

1. Giles C.L., Bollacker K.D., Lawrence S. CiteSeer: an automatic citation indexing system // Proc. ACM Int. Conf. Digit. Libr. 1998. P. 89–98.
2. Hu W. et al. Open graph benchmark: Datasets for machine learning on graphs // Adv. Neural Inf. Process. Syst. 2020. Vol. 33. P. 22118–22133.
3. Yang J., Leskovec J. Defining and evaluating network communities based on ground-truth // Knowl. Inf. Syst. 2015. Vol. 42, № 1. P. 181–213.
4. Palowitch J. et al. GraphWorld: Fake Graphs Bring Real Insights for GNNs // Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), August 14–18, 2022, Washington, DC, USA. Association for Computing Machinery, 2022. Vol. 1, № 1.
5. Pei H. et al. GEOM-GCN: Geometric graph convolutional networks // arXiv. 2020. P. 1–12.
6. Bo D. et al. Beyond Low-frequency Information in Graph Convolutional Networks. 2021.
7. Jin W. et al. Node similarity preserving graph convolutional networks // arXiv. 2020.
8. Kim D., Oh A. HOW TO FIND YOUR FRIENDLY NEIGHBORHOOD: GRAPH ATTENTION DESIGN WITH SELF-SUPERVISION. 2021. P. 1–25.
9. Klicpera J., Bojchevski A., Unnemann S.G. PREDICT THEN PROPAGATE: GRAPH NEURAL NETWORKS MEET PERSONALIZED PAGERANK.
10. Albert-László Barabási. 2013. Network science. Chapter 5. The Barabási-Albert Model. Network Science (2013), 1–45
11. Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. 2008. Benchmark graphs for testing community detection algorithms. Physical review E 78, 4 (2008),
12. Anton Tsitsulin, Benedek Rozemberczki, John Palowitch, and Bryan Perozzi. 2021. Synthetic Graph Generation to Benchmark Graph Learning. (2021)
13. Comandur Seshadhri, Tamara G Kolda, and Ali Pinar. 2012. Community structure and scale-free collections of Erdős-Rényi graphs. Physical Review E 85, 5 (2012), 056109

Андреева П.О. (автор)

Стебеньков А.С. (автор)

Шиков Е.Н. (научный руководитель)