

ГРАФОВЫЕ ПОДХОДЫ К ОБУЧЕНИЮ НА ТАБЛИЧНЫХ ДАННЫХ

Медведев Алексей Владимирович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: fortunato.mav@gmail.com

Научный руководитель — Дьяконов Александр Геннадьевич

Под влиянием успеха глубоких нейронных сетей на данных со структурой в виде сетки(изображения), стали появляться работы, посвященные поиску операции свертки для данных с графовой структурой. Такие новые типы сверток были успешно применены к задачам распознавания и предсказания позы человека [1], задачам классификации вершин графов цитирований и социальных сетей [2]. Они имеют потенциал в обработке категориальных признаков [3] и в задаче многоклассовой классификации [4].

Но не так много работ посвящено применению графовых сетей на табличных данных, где признаки могут иметь самую разную природу и масштабы. С такими данными отлично работают методы основанные на решающих деревьях. В последние годы появились алгоритмы [5,6], позволяющие обучать ансамбли и бустинги решающих деревьев с помощью алгоритма обратного распространения ошибки. Такие алгоритмы позволяют учить многослойные представления табличных данных, которые по нашему предположению могут быть эффективно использованы при обучении графовых сверточных сетей. Более того, это обучение может проходить совместно, что повышает эффективность алгоритмов.

Первая работа, посвященная данной проблеме [7], использует связку обычного градиентного бустинга и графовой сети, такая модель обучается совместно с помощью алгоритма описанного здесь [5]. Однако вышеупомянутая модель имеет ряд ограничений. Например, авторы не изучают вопрос обучения с помощью градиентного бустинга многослойных скрытых представлений. Скрытый слой такой модели представлен или в виде конкатенации изначальных признаков с выходом бустинга или только выходом бустинга, размерность которого совпадает с размерностью целевой переменной(на первой итерации бустинг учится предсказывать ее). Один из экспериментов показал (Таблицы 1, 2, 3), что такой подход можно легко модифицировать, алгоритм гораздо лучше справляется, когда на вход графовой сети подаются только выходы бустинга. Это значит, что можно

отказаться от конкатенации первоначальных признаков и подавать графовой сети только скрытые представления без привязки к целевой переменной. Еще одно ограничение это невозможность стохастического обучения, которое очень важно для решения проблемы индуктивного обучения на графах.

В нашей работе изучаются все эти аспекты, которым не уделялось внимания в прошлых работах.

gbdt-only	bgnn-agnn	bgnn-appnp	bgnn-gat	bgnn-gcn
True	0.49 ± 0.01	0.59 ± 0.01	0.49 ± 0.01	0.54 ± 0.01
False	0.53 ± 0.01	0.60 ± 0.01	0.55 ± 0.02	0.57 ± 0.01

Таблица 1: Значение метрики RMSE (чем меньше тем лучше) для различных архитектур графовых сетей на наборе данных [8]. Значение в первой колонке указывает на то, использовались ли изначально признаки в качестве входа для графовой сети (False) или только выходы градиентного бустинга (True).

gbdt-only	bgnn-agnn	bgnn-appnp	bgnn-gat	bgnn-gcn
True	6.87 ± 0.2	12.36 ± 0.15	7.00 ± 0.2	7.02 ± 0.2
False	6.96 ± 0.2	13.41 ± 0.16	6.95 ± 0.2	8.32 ± 0.43

Таблица 2: Значение метрики RMSE (чем меньше тем лучше) для различных архитектур графовых сетей на наборе данных [9].

gbdt-only	bgnn-agnn	bgnn-appnp	bgnn-gat	bgnn-gcn
True	1.27 ± 0.08	1.35 ± 0.13	1.26 ± 0.09	1.31 ± 0.11
False	1.29 ± 0.11	1.38 ± 0.14	1.33 ± 0.13	1.37 ± 0.07

Таблица 3: Значение метрики RMSE (чем меньше тем лучше) для различных архитектур графовых сетей на наборе данных [10].

Литература

1. Li M., Chen S., Chen X., Zhang Y., Wang Y., Tian Q. Actional-Structural Graph Convolutional Networks for Skeleton-Based Action Recognition. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 3595-3603.
2. Hamilton W.L., Ying R., Leskovec J. Inductive Representation Learning on Large Graphs. CoRR. abs/1706.02216, 2017.

3. Li Z., Cui Z., Wu S., Zhang X., Wang L. Fi-GNN: Modeling Feature Interactions via Graph Neural Networks for CTR Prediction. CoRR. abs/1910.05552, 2019.
4. Chen Z., Wei X., Wang P., Guo Y. Multi-Label Image Recognition with Graph Convolutional Networks. CoRR. abs/1904.03582, 2019.
5. Feng J., Yu Y., Zhou Z. Multi-Layered Gradient Boosting Decision Trees. CoRR. abs/1806.00007, 2018.
6. Popov S., Morozov S., Babenko A. Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data. CoRR. abs/1909.06312, 2019.
7. Ivanov S., Prokhorenkova L. Boost then Convolve: Gradient Boosting Meets Graph Neural Networks. CoRR. abs/2101.08543, 2021.
8. Pace R., Barry R. Sparse spatial autoregressions. *Statistics and Probability Letters*, 33 (3):291–297, 1997.
9. Tsitsulin A., Mottin D., Karras P., Muller E. Verse: Versatile graph embeddings from similarity measures. In *Proceedings of the 2018 World Wide Web Conference*, pp. 539–548, 2018.
10. Jia J., Benson A. Outcome correlation in graph neural network regression. CoRR. abs/2002.08274, 2020.