

Kolmogorov Arnold Networks as an alternative to MLP in Tabular DL

Г.А. Булгаков¹, Д.В. Руденко¹

Г.Л. Молодцов^{1,2}, Д.О. Медяков^{1,2}

¹Московский физико-технический институт (национальный исследовательский университет)

²Лаборатория фундаментальных исследований Искусственного Интеллекта (BRAIn Lab)

Введение

Методы DL (глубокое обучение) исторически не выигрывали у классических ML-моделей на задачах с табличными данными, состоящими из категориальных, бинарных и числовых признаков. Более того, модели основанные на градиентном бустинге решающих деревьев (как CatBoost, XGBoost) зачастую значительно выигрывали нейросети. Однако современные модели, основанные на полносвязных сетях (MLP) начали превосходить стандартные GBDT-архитектуры. В данной работе исследуется применение в Tabular DL новой архитектуры нейросетей — Kolmogorov-Arnold Networks (KANs) [1]. Эти модели основаны на теореме Колмогорова-Арнольда о представлении многомерной функции в виде суммы и композиции одномерных функций. Но в отличии от MLP, линейные весы и фиксированные активации заменены на обучаемые одномерные функции активации, параметризованные сплайнами. Гипотетически такой подход повышает точность аппроксимации сложных зависимостей и адаптивность модели.

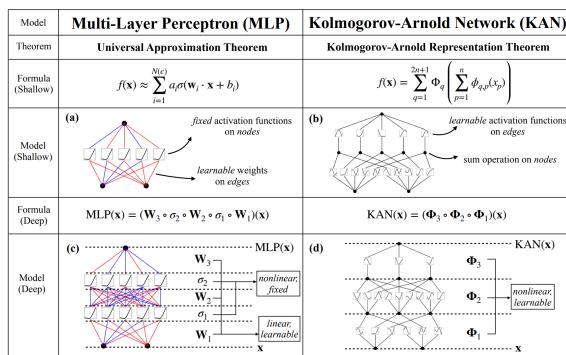


рис 1: Архитектура MLP vs KAN

Методология

В работах [2] и [3] проведены масштабные эксперименты по применению KAN в Tabular DL и были получены многообещающие результаты. Исследования показали, что KAN постоянно превосходит MLP по точности предсказаний на разных датасетах. В статье [3] авторы дополнительно предлагают и исследуют иные виды базовой архитектуры: FastKAN, ChebyshevKAN, предлагающие альтернативные способы параметризации нелинейных обучаемых функций активаций.

Несмотря на положительные результаты касательно применения KAN в Tabular DL, в этих исследованиях не был полностью раскрыт, как потенциал KAN, так и потенциал бейзлайна - MLP.

Для повышения эффективности моделей в настоящей работе предлагается два основных усовершенствования:

1. Применение эмбеддингов для числовых признаков. В частности, используются методы Piecewise Linear Encoding (PLE) и Periodic Linear ReLU

(PLR) [4], которые преобразуют числовые атрибуты в векторные представления. Этот подход позволяет:

- Существенно снизить риск переобучения
- Улучшить обобщающую способность модели

2. Использование продвинутых методов ансамблирования [5] (подобно TabM [5]). Данный подход позволяет моделировать работу классического ансамбля из k моделей, при этом:
 - Усреднение предсказаний обеспечивает более высокую точность
 - Уменьшается дисперсия ошибок
 - Повышается устойчивость к шумам в данных
 - Значительно ускоряется обучение по сравнению с классическими ансамблями

Экспериментальная часть и результаты

Для обеспечения объективного сравнения моделей был проведен тщательный подбор гиперпараметров как для KAN-based, так и для MLP-based архитектур. Тюнинг был проведен с использованием:

- Модуля Optuna для автоматизированного поиска оптимальных параметров
- Фреймворка Weights & Biases (wandb) для трекинга экспериментов и визуализации результатов

Полученные результаты демонстрируют следующие ключевые закономерности:

1. KAN-архитектуры стабильно показывают более высокое качество по сравнению с MLP в задачах Tabular DL, что подтверждает выводы предыдущих исследований [2, 3]
2. KAN-based подходы стабильно превосходят MLP на табличных задачах, что подтверждает выводы [2, 3].
3. Ключевой вклад в улучшение предсказаний вносит не архитектура (KAN/MLP), а наличие эмбеддингов:
 - Модели с PLE и PLR демонстрируют близкую эффективность независимо от базовой архитектуры
 - Качество MLP с продвинутыми эмбеддингами приближается к показателям KAN
4. Анализ различных вариантов KAN выявил:
 - Наилучшую производительность и скорость работы у Fast-KAN и Small-KAN
 - Низкую эффективность и точность Cheby-KAN
 - Оптимальное соотношение точности и вычислительных затрат у Small-KAN
 -

Заключение

Работа демонстрирует, что KANs являются перспективной альтернативой MLP в задачах Tabular DL, особенно в сочетании с современными методами эмбеддингов. Дальнейшие исследования могут быть направлены на оптимизацию и ускорение архитектуры KANs, разработку новых методов параметризации функций и дальнейшую интеграцию KANs в ансамблевые подходы.

Литература

[1] Ziming Liu and Yixuan Wang and Sachin Vaidya and Fabian Ruehle and James Halverson and Marin Soljačić and Thomas Y. Hou and Max Tegmark KAN: Kolmogorov-Arnold Networks 2025

[2]Poeta, Eleonora and Giobergia, Flavio and Pastor, Eliana and Cerquitelli, Tania and Baralis, Elena A Benchmarking Study of Kolmogorov-Arnold Networks on Tabular Data 2024

[3]Ali Eslamian and Alireza Afzal Aghaei and Qiang Cheng TabKAN: Advancing Tabular Data Analysis using Kolmogorov-Arnold Network 2025

[4]Yury Gorishniy and Ivan Rubachev and Artem Babenko On Embeddings for Numerical Features in Tabular Deep Learning 2023

[5]Yury Gorishniy and Akim Kotelnikov and Artem Babenko TabM: Advancing Tabular Deep Learning with Parameter-Efficient Ensembling 2025