

Обзор на статью “On Embeddings for Numerical Features in Tabular Deep Learning”

выполнила:
Шешукова Марина



HSE
University

Суть работы

- ★ Piecewise Linear Encoding
- ★ Periodic activation functions

Предшественники и конкуренты

- ★ “Revisiting Deep Learning Models for Tabular Data”
- ★ “Fourier features let networks learn high frequency functions in low dimensional domains”
- ★ “SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training”



Что было дальше?

- ★ “Revisiting Pretraining Objectives for Tabular Deep Learning”
- ★ “ExcelFormer: A Neural Network Surpassing GBDTs on Tabular Data”



Бустинг или Глубинное обучение?

★ “Why do tree-based models still outperform deep learning on typical tabular data?”

Преимущества и недостатки

Сильные стороны работы:

- ★ В статье представлены две простые схемы получения эмбедингов, позволяющие нейронным сетям работать на табличных данных лучше.
- ★ Представлены результаты демонстрирующие, что эмбединги для числовых признаков действительно полезны и эффективны для моделей.

Слабые стороны работы:

- ★ Большинство данных, на которых авторы производят сравнения небольшие, и не до конца понятно: будет ли такой же хороший результат на больших данных.
- ★ Методов представлено два, но не пояснено какой из них эффективнее использовать в реальных приложениях и как выбирать между ними.
- ★ Авторы не дают никакого теоретического обоснования, поэтому не до конца понятно, почему эти методы не зависят от проведенных экспериментов и действительно помогают оптимизации.



Дальнейшее направление работы.

- ★ Можно углубиться в теорию и объяснить, как именно обсуждаемые методы помогают оптимизации на теоретическом уровне.
- ★ Можно применять ко всем признакам не одно и то же преобразование, а разные, что возможно приведет к лучшим результатам.

