

УДК 111.11

## РОЛЬ ФУНКЦИИ ПОТЕРЬ В ОБРАБОТКЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ТЯЖЕЛЫМИ ХВОСТАМИ

Грицких М.А. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Гусарова Н.Ф.  
(ИТМО)

**Введение.** Прогнозирование последовательностей с тяжёлыми хвостами важно для медицины, финансов и других областей. Современные модели (LSTM, трансформеры) подавляют выбросы, что снижает точность предсказаний. В работе исследуется влияние выбора функции потерь на устойчивость моделей к тяжёлым хвостам.

**Основная часть.** Проведена теоретическая оценка градиентов различных функций потерь при переходе от нормального распределения входных последовательностей к тяжёлым хвостам. Используется модель стабильного распределения Леви [1]. Рассматриваются функции потерь: среднеквадратичная ошибка (MSE), средняя абсолютная ошибка (MAE) и функция потерь Коши. Выявлено, что механизм внимания в трансформерах может испытывать трудности [2] при обработке распределений с тяжёлыми хвостами из-за чрезмерного акцентирования на выбросах, однако применение multi-head attention частично нивелирует этот эффект [3]. В режиме авторегрессии градиент функции ошибки Коши уменьшается, MAE остаётся стабильным, а MSE увеличивается, что влияет на устойчивость градиентного спуска при обучении сети.

Экспериментально изучено поведение моделей на синтетических и реальных данных (COVID-19) [4]. Полученные результаты подтвердили теоретические выводы: функции потерь Коши и MAE на 40% превосходят MSE по метрике sMAPE. Анализ результатов также показывает, что модели на основе трансформеров демонстрируют лучшие результаты на шаге предсказания, но уступают LSTM в многократных итерациях предсказания.

**Выводы.** Проведённый анализ продемонстрировал, что выбор функции потерь критически важен для работы трансформеров с последовательностями, имеющими распределения с тяжёлыми хвостами. Использование функций потерь Коши и MAE позволяет повысить точность прогнозирования и устойчивость обучения, в отличие от MSE, которая оказывается наименее эффективной. В дальнейшем целесообразно исследовать комбинированные функции потерь, учитывающие особенности тяжёлых хвостов и внимания в трансформерах.

### Список использованных источников:

1. Gurbuzbalaban M., et al. Heavy Tails in ML: Structure, Stability, Dynamics. <https://neurips.cc/virtual/2023/workshop/66530>
2. Song W., Chen D., Zio E. Heavy Tail and Long-Range Dependence for Skewed Time Series Prediction Based on a Fractional Weibull Process. *Fractal Fract.* 2024, 8(1), 7; <https://doi.org/10.3390/fractalfract8010007>
3. Vaswani A. et al. Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
4. Liu P., Zheng Y. Heavy-tailed distributions of confirmed COVID-19 cases and deaths in spatiotemporal space. *Plos ONE* November 21, 2023 <https://doi.org/10.1371/journal.pone.0294445>