

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ
БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра теории функций**

Малекин
Александр Игоревич

**ЭФФЕКТИВНЫЕ МЕТОДЫ УМЕНЬШЕНИЯ
РАЗМЕРА ЯЗЫКОВЫХ МОДЕЛЕЙ**

Аннотация к дипломной работе

Научный руководитель:
кандидат физ.-мат. наук,
доцент С.А. Бондарев

Минск, 2025

Реферат

Дипломная работа содержит 28 страниц, 6 иллюстраций, 1 приложение, 8 литературных источников.

Ключевые слова: ЯЗЫКОВАЯ МОДЕЛЬ, ТОКЕНИЗАЦИЯ, n -ГРАММНАЯ ЯЗЫКОВАЯ МОДЕЛЬ, НЕЙРОННАЯ ЯЗЫКОВАЯ МОДЕЛЬ, ПОЛНОСВЯЗНАЯ НЕЙРОННАЯ СЕТЬ, ТРАНСФОРМЕР, СИНГУЛЯРНОЕ РАЗЛОЖЕНИЕ МАТРИЦ, SVD, ОПТИМИЗАЦИЯ.

Цель исследования — методы уменьшения размера языковых моделей.

В дипломной работе изучены виды языковых моделей и их архитектуры. Рассмотрен метод уменьшения размера языковой модели, основанный на механизме сингулярного разложения матриц(SVD). Проведен эксперимент по применению полученных знаний для оптимизации предобученной модели. Представлены результаты эксперимента и его выводы.

Дипломная работа выполнена автором самостоятельно.

Рэферат

Дыпломная праца змяшчае 28 старонак, 6 ілюстрацый, 1 дадатак, 8 літаратурных крыніц.

Ключавыя словы: МОЎНАЯ МАДЭЛЬ, ТАКЕНІЗАЦЫЯ, n -ГРАМНАЯ МОЎНАЯ МАДЭЛЬ, НЕЙРОННАЯ МОВАВАЯ МАДЭЛЬ, ПОЎНАСУВЯЗНАЯ НЕЙРОННАЯ СЕТКА, ТРАНСФАРМЕР, СІНГУЛЯРНАЕ РАЗЛОЖЭННЕ МАТРЫЦ, АПТЫМІЗАЦЫЯ.

Мэта даследавання — метады аптымізацыі моўных мадэляў.

У дыпломнай рабоце вывучаны віды моўных мадэлей і іх архітэктурны. Разгледжаны метады памяншэння памеру моўнай мадэлі, заснаваны на механізме сінгулярнага раскладання матрыц (SVD). Праведзены эксперымент па прымяненні атрыманых ведаў для аптымізацыі прадугледжанай мадэлі. Прадстаўлены вынікі эксперыменту і яго высновы.

Дыпломная праца выканана аўтарам самастойна.

Abstract

The thesis contains 28 pages, 6 illustrations, 1 appendix, 8 literary sources.

Key words: LANGUAGE MODEL, TOKENIZATION, n -GRAMMED LANGUAGE MODEL, NEURON LANGUAGE MODEL, FULLY CONNECTED NEURON NETWORK, TRANSFORMER, SINGULAR VALUE DECOMPOSITION, SVD, OPTIMIZATION.

The aim of the research is methods of optimization of language models.

In this thesis, types of language models and their architectures are studied. A language model size reduction method based on the singular value matrix decomposition(SVD) mechanism is considered. An experiment is conducted to apply the knowledge gained to optimize a pre-trained model. The results of the experiment and its conclusions are presented.

The diploma work was carried out by the author independently.