Skoltech

Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Georgii Novikov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Tensor-train decomposition for efficient data representation and learning

Supervisor: Professor Ivan Oseledets
Co-supervisor: Assistant Professor Maxim Panov

Name of the Reviewer: Sergey Nikolenko

I confirm the absence of any conflict of interest	**Date: 04-12-2024**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications
- The summary of issues to be addressed before/during the thesis defense

The PhD thesis "Tensor-train decomposition for efficient data representation and learning" presents a comprehensive and well-structured study of tensor-train decomposition applied to deep learning and data representation problems. It integrates theoretical foundations, method development, and practical applications. The main contributions of the thesis include:

- *Tensor-Train Density Estimation* (TTDE), a novel method for estimating probability densities using the tensor-train representation, which enables analytical integration and efficient probabilistic queries;
- *TTMLinear layers*, where Georgii introduced a tensor-train-based compression for fully connected neural network layers, focusing on Transformer-based architectures, and developed an optimized backpropagation scheme, making it practical for large language models such as GPT;
- *few-bit backward method* for compressing activations in neural networks with minimal performance trade-offs, significantly reducing memory requirements during training;
- *tensor-train point cloud compression*, an efficient representation for multidimensional point clouds via TTD that supports tasks such as nearest-neighbor search and anomaly detection.

The main topics and contributions of the thesis are highly relevant, directly addressing challenges such as density estimation, memory optimization, and point cloud compression, and the chosen methods are well suited for these tasks.

All results demonstrate significant scientific novelty, introducing TTDE for density estimation, TTMLinear for neural network compression, few-bit backward method for memory optimization, and tensor-train point cloud for efficient representation learning. The thesis also contains extensive experiments, benchmarks, and comparisons that demonstrate the effectiveness of the developed methods. Their practical relevance is evident in applications such as anomaly detection, approximate nearest-neighbor search, and efficient neural network training. Moreover, the methods have already been implemented in open source Python libraries, such as TTDE and Few-Bit Backward. Publications in top tier conferences such as ICML and UAI confirm the high quality and impact of the research.

Minor improvements could include a deeper discussion of limitations and a clearer outline of future directions. For example, it would be interesting to know whether the TTMLinear layer approach be extended to other types of neural networks, such as convolutional or recurrent architectures.

Overall, the thesis makes a valuable contribution to the field and is ready for defense.

Provisional Recommendation

☒ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*