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: Dmitry Yarotsky

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 17-11-2024**

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 topic of the thesis is tensor train (TT) decomposition and its applications to several domains in which a compact data representation is essential. TT is currently one of the leading modern universal numerical technologies for data compression and efficient manipulation. The thesis presents a series of new results further extending the applicability of the technology.

The thesis includes five chapters. The first chapter is a nicely written introduction to the topic of tensor trains. The other four chapters present the research results of the author and correspond to his four publication. Chapter 2 proposes a TT-based method of density estimation for distributions represented by finite samples (TTDE). Chapter 3 proposes a TT-based approach to compression of fully-connected layers of neural networks (TTMLinear). Chapter 4 proposes a quantization-based method of compressing stored nonlinear activations (Few-Bit Backward). Chapter 5 proposes a TT-based method of compressing point clouds and using it as an index structure for vector database (TT Point Cloud).

The thesis is fairly long (150 pages), but well-written and is abundantly illustrated (in particular with Penrose-type diagrams for tensors), which further improves its readability.

On the whole, the thesis is a solid and comprehensive study, reflecting the author's deep and broad involvement in TT and its applications. Georgii was the main (co-)contributor in each of the four publications the thesis is based on. The publications are peer-reviewed and presented at international conferences; two were presented at A/A* conferences. The applications discussed in the thesis involve state-of-the-art dataset and models. Accordingly, the research reported in the thesis is high-quality and advances state of the art in Tensor Trains and their applications.

Summarizing, I have no doubt that the presented thesis is a high-quality work, and its author Georgii Novikov deserves to be awarded the Skoltech PhD degree.

I noticed only a couple minor writing issues:

* the provided link https://github.com/PgLoLo/fewbit-tests does not work

*In definition 1.2.1, probably "THE function T" is meant rather than "A function T"; otherwise the definition looks like an incomplete definition of a hitherto unknown function T

* The captions to figures should be left-justified, not centered. The separation between figure captions and the main text should be increased; currently the captions are hard to distinguish from the main text. For example, it took me some time to realize that Eq. (1.28) is not repeated twice in the main text and that rather the top equation belongs to the figure caption.

* Section 2.7.1 confusingly looks like an empty section. I assume it is meant to only include a few figures, but in a scientific publication all the numbered figures are usually supposed to be referred to in the main text. You could just put a sentence "See the relevant diagrams in Figures …" in this section.

Provisional Recommendation

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

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*