Skoltech

Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Alexander Fonarev

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Matrix Factorization Methods for Training Embeddings in Selected Machine Learning Problems

Supervisor: Prof. Ivan Oseledets

Chair of PhD defense Jury Prof. Andrzej Cichocki ***Email***: *a.cichocki@skoltech.ru*

Date of Thesis Defense: 19 September 2018

Name of the Reviewer: Dr. Andre Uschmajew

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 31-08-2018**

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 thesis of Fonarev considers the problem of embeddings in machine learning tasks, that is, the meaningful mapping of given data into real vectors, such that structural features (similarities, semantics, intrinsic dimensionality etc.) are respected. Such embedding of data is obviously an essential step in the mathematical modeling of optimization and learning applications. Fonarev uses a general low-rank matrix factorization framework for this task, introduced in Chapter 3. He shows that many existing methods fit into this framework, and then develops some new methods for certain applications, specifically for embedding of categorical data (Chapter 4), for word embeddings in language processing using Riemannian optimization (Chapter 5), and for embedding of cold users in recommender systems using maximum volume submatrices (Chapter 6). The results of Chapter 5 and 6 have been accepted and published in excellent conference proceedings in the area, with the substantial contribution by Fonarev.

The thesis is generally well written and covers an impressive scope of subjects within machine learning. At some points this comes at a price of a certain brevity when reviewing existing methods or explaining the new approaches in detail. But in turn, Fonarev has succeeded in giving a very broad overview over the subject and making the general ideas very intuitive and understandable. Therefore, I enjoyed reading this thesis. Of course, some improvements should be made, a list of suggestions has been submitted to the candidate

The contributions of this thesis are more on the algorithmic side, but quite substantial and highly relevant to applications. The developed algorithms have been implemented an tested on common big data sets, and seem to outperform certain state-of-the-art methods in the conducted experiments. This is very impressive in such an active and competitive field as machine learning, and strongly advertises the low-rank approach to embeddings to the community. In particular, the algorithm presented for cold users in recommender systems has found industrial applications. The chapter on recommender systems also contains interesting theoretical aspects of the maximum volume concept for submatrices.

In conclusion, Fonarev has submitted a well written and organized thesis, proposing new and competitive algorithms for highly relevant problems in contemporary machine learning tasks. His work supports the idea of matrix factorization for training embeddings, and effectively demonstrates how mathematical concepts like Riemannian optimization and maximum volume submatrices can make an important contribution in this field.

Fonearev 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*