

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:

I confirm the absence of any conflict of interest	Signature:  Date: 11-09-2018
---	---

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

Brief evaluation of the thesis quality and overall structure of the dissertation

The topic of the thesis – proposal and study of a general framework to training embeddings using the low-rank factorization approach (as presented in Chapter 3, p. 53 of the dissertation) – is very relevant and interesting. The proposed framework is general enough to work across different domains, ranging from recommender systems to search engines to the NLP problems (word2vec and beyond). Furthermore, the claimed contributions of the thesis include:

- a. Proposal of a general (albeit simple) framework of dealing with embeddings
- b. Formulation and solution of the problem of unsupervised search for embeddings of categorical feature values
- c. Formulation of the generalized Skip-Gram Negative Sampling word embeddings training procedure and presentation of the new embedding training method based on Riemannian optimization
- d. Introduction of a new cold start method in recommender systems that is based on the rectangular generalization of the previously introduced Maxvol method.

Out of these four claimed contributions, I find the first one to be quite basic (although useful since it is the “glue” for the whole thesis and puts different disparate approaches into one common framework). The second contribution is also quite marginal. The last two are significantly more interesting and constitute the main two contributions of the thesis, in my opinion. These last two problems were analyzed fairly well and the research is of high quality, although with some minor lapses that I will discuss with the Candidate after his thesis defense and that should not affect nor delay the thesis defense in any way.

The structure of the dissertation is acceptable, and I am satisfied with the overall quality of the thesis (modulo a few minor comments, some of which can be found below).

Relevance of the topic of dissertation work to its actual content

It is relevant.

Scientific significance of the results obtained and their compliance with the international level and current state of the art

Out of the four contributions that are claimed in the dissertation, the first one (the framework) is very simple but relevant and useful, as explained above. Significance of the second contribution is very marginal. The third contribution constitutes an extension of the previously proposed methods, including those described in [68] and [69]. However, I found this extension, relying on Riemannian optimization (see page 72), to be novel, interesting, and leading to better performance results, as demonstrated in Chapter 5. Therefore, I view this contribution to be significant.

The last contribution is an extension of the previously proposed Maxvol approach [35] where square matrices of Maxvol are replaced with rectangular submatrices. This extension is conceptually quite simple (“from squares to rectangles”), but is well executed, uses non-trivial analysis, and the advantages of the proposed method were empirically demonstrated. Therefore, I also found this contribution to be significant.

By the way, both results were published in reputed conferences (ICDM and ACL), although with many co-authors.

Relevance of the obtained results to applications (if applicable)

Both word2vec types of embeddings and working solutions to the cold start problem are of practical importance to the industry. Moreover, according to the Candidate, the cold start method described in Chapter 6 has been successfully used in the Yandex recommender system, which is impressive.

Quality of publications

Are high: both the IEEE ICDM and ACL are premier conferences in the field.

Summary of issues to be addressed before/during the thesis defense

I have several minor comments about the manuscript that I intent to discuss with the candidate after the thesis defense. These comments are minor and should not prevent nor delay the thesis defense in any way. Although English is OK, it could have been improved (via various editing services and/or tools).

Finally, exposition of the thesis can be tighter. For example, I had to guess and also go many pages back on several occasions while reading the thesis, trying to understand what the Candidate is trying to say. As another example, the Candidate claims on page 100 that "... the corresponding figures are not demonstrated here *due to the lack of space*" [italics are mine]. Surely, the Candidate must be joking!

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