

## Jury Member Report – Doctor of Philosophy thesis.

**Name of Candidate:** Evgeny Frolov

**PhD Program:** Computational and Data Science and Engineering

**Title of Thesis:** Low-rank models for recommender systems with limited preference information


**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. Michael Thess

<p>I confirm the absence of any conflict of interest</p> <p>(Alternatively, Reviewer can formulate a possible conflict)</p>	<p><b>Signature:</b></p>  <p><b>Date: 19-08-2018</b></p>
---	---

*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 deals with low-rank models for Recommender Systems (RS) when limited preference information is available. This is a highly relevant topic since almost all relevant ecommerce systems like web shops, search engines, and portals today employ recommender systems. At the same time, most of them continuously extend the volume of the content they offered (number of items like products, links, articles). For example, even small online book shops meanwhile offer millions of books and CDs. At the same time, the growth of their transaction volume is usually much slower what leads to a decreasing rate of items to transactions. Thus, the problem to deal efficiently with sparse data is becoming increasingly important.

In the first part of the thesis the author gives an overview of the current state of recommender systems and main approaches. Starting from Content-based Filtering and Collaborative Filtering (CF) he gives a comprehensive introduction to Matrix Factorization (MF) which generalizes CF and allows to handle sparse data more efficiently. He introduced a general formulation of RS problems in terms of a minimization problem. Then he correctly points out that of all the many approaches the ones based on Singular Value Decomposition (SVD) are the most theoretically sound and numerically efficient ones. After that he explains special aspects such as incorporation of bias, incremental updates of models, and solution techniques of the minimization problem. Most important are approaches for generalizations of the SVD, especially by including side information (e.g. item or user attributes). Chapter 3 extends the matrix factorization case to Tensor Factorization (TF) where the emphasis is made on Tucker decomposition, for the same reason as for selecting the SVD in the MF case. In the final part the shortcomings of the current approaches are presented and used to motivate the new work.

In the second part the three novel approaches of the work are developed. The first one aims to improve the quality of recommendations, especially by learning from full feedback information instead of only the positive one (as in most state-of-the-art systems). This is done by introducing the rating as explicit new (ordinal) dimension. Instead of the MF of the user-item rating matrix, we now arrive at a 3D TF delivering a relevance score for each rating. The author uses a Tucker decomposition for factorization and the elegant folding-in method for online recommendations. Experimental results confirm the advantage of the new approach in comparison to current ones. Overall, the idea is simple and very good, although the efficient implementation is more demanding.

The next idea, presented in Chapter 6, is about incorporating side information to improve the handling of sparse data. It is developed for the matrix case by generalizing the SVD approach including similarity-based matrices from side information (between users and between items, respectively). The SVD part seems convincing, especially the nice folding-in technique. In contrast, the information about the construction of the side similarity matrices is a bit scarce. Given that the author acknowledges that this is a difficult topic I would expect to see more descriptions here. Especially, an example would be helpful. At the end of the section the author tests the method on various benchmarks where it shows superior results.

Both novel approaches presented so far, the explicit rating dimension and the incorporation of side information, are based on SVD-type factorizations with folding-in for recommendations. This is very sound and so it is natural to combine them into one model - the higher-order preference model. This is done as third approach in Chapter 7. Even if in principal clear, there are a lot of mathematical and computational issues that the author masters very competent. Final results show the effectiveness of his new approach.

The third final part is devoted to the implementation of the work in a new open-source framework. It is developed in Python and can be used beyond the current RS approaches. I am not a Python expert but

for me the framework seems to be quite sophisticated and robust.

Finally, the author gives the conclusions of his work. They are appropriate and complete. However, I miss a continuative discussion of the work. Especially, how the proposed methods can be utilized for other applications than ratings (i.e. for web shops or search engines). Further, I miss a discussion about possible improvements and extensions.

The work is very structured and written in a clear style. It is highly relevant from both mathematical and practical points of view. The ideas are novel and can be directly used in practical applications. The high value of the work is confirmed by the international publications.

#### **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*