

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:

I confirm the absence of any conflict of interest	Signature:  Date: 10-09-2018
---	---

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 thesis consists of two main parts: (a) an overview of the low rank models in recommender systems and (b) three main approaches to handling the limited preference information problem. The quality of the survey of the low rank models is excellent. I have a couple of minor comments about the survey that I intend to discuss with the candidate in person; but this should not preclude the thesis defense in any way.

The main contribution of the thesis is the introduction of the HybridCoFFee method that combines the CoFFee approach described in Chapter 5 (which is also another contribution of the thesis) and the hybrid factorization model that incorporates "side" information about users and items into the standard SVD method(s). This combined HybridCoFFee approach exhibits the balance between being less affected by and sensitive to the data sparsity problem, maintaining high relevance of recommendations and generating safe predictions that avoid disappointing recommendations. This result is significant because tensors have been extensively used in RSEs over the last 5 – 6 years, while being sensitive to the data sparsity problem (which limits their applicability). Therefore, this work mitigates this problem while providing a novel, reliable and practically significant way to use tensors together with the appropriate "side" information that addresses the known weaknesses listed above. Therefore, I am satisfied with the quality of the thesis.

Relevance of the topic of dissertation work to its actual content

I found it to be relevant

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

The results reported in the thesis are significant. As explained before, the main result about the HybridCoFFee method is novel, relevant, and it provides several significant advantages over the state-of-the-art. The method is well-executed and well-presented. Furthermore, the first result (reported in Chapter 5) has been published in one of the premier conferences in the field (ACM RecSys) where it was well received by the scholars working in the RS area (I attended that RecSys conference and have heard the feedback). The survey part of the thesis has also been published in a reputable outlet. Although the last two results have not been published yet, I think that they are of sufficient quality and should be published by the candidate in one of the premier conferences shortly.

Relevance of the obtained results to applications

Tensors are extensively used in recommender systems. However, they have certain shortcomings, data sparsity being one of the major issues. The proposed method addressed several of these shortcomings and, therefore, should be relevant to the business world.

The quality of publications

ACM RecSys conference is the leading conference in the field of Recommender Systems.

The 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.

One question that the candidate should think about and address at the thesis defense is this. Section 5.1.2 (and several other sections) state that the "Relevance Score denotes the *likeliness* of observing a certain (user, movie, rating) triplet." Is there any possible probabilistic interpretation of the "likeliness" notion for the Relevance Score? The candidate should comment on this probabilistic connection/interpretation and, possibly, explore it further.

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