Skoltech
Skolkovo Institute of Science and Technology

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 (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 20-08-2018**

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 PhD thesis by Evgeny Frolov titled "*Low rank models for recommender systems with limited preference information*" is devoted to addressing the problem of insufficient user preference information. The work particularly focuses on the SVD-based models, where SVD can be used either directly or as an atomic operation within an iterative process. The author proposes three models, one of them relies on the matrix-based formulation and the other two are tensor-based.

The author of the thesis uses the *PureSVD* approach as the base for further improvement. In this approach all unknown entries are simply replaced with zeroes, which allows to omit weighting and to solve the low rank approximation problem directly with the help of truncated SVD. Both matrix-based and tensor based methods, proposed by the author, rely on this zero-imputation scheme.

The topic of *recommender systems*, which can be viewed as a subarea of *machine learning*, *data mining* and *information retrieval* areas. Recommender systems are ubiquitous in the modern world and are widely applied in many different domains from retail and entertainment to education and research. There are three major challenges faced by any recommender system. First of all, due to a rapid growth of information, the systems have to be scalable, allowing to deal with millions or even billions of objects. One of the most widely used approaches that allows to tackle large scale problems and provide decent quality of recommendations is matrix factorization. Secondly, an input data is often extremely incomplete, which requires special treatment as it may significantly affect generalization ability of recommender models, including the models based on a factorization approach. This type of problems is typically solved by introducing additional sources of knowledge about users, items and/or interactions between them to put additional constraints and help find more reliable solutions. Lastly, modern recommender systems must have capability to respond to user actions instantly in order to keep users engaged. Moreover, they have to deal with a constant flow of new users or new items, which were nor a part of the model. Real-world systems are expected to provide intermediate solutions to address this kind of problems without the need for expensive computations. Addressing all these challenges in a single model was a non-trivial task and is a current topic of an extensive research.

Despite a great variety of matrix and factorization methods, the author the thesis was able to demonstrate that even a relative simple SVD-based approach called PureSVD in many cases provides superior quality of recommendations comparing to more sophisticated approaches. Moreover, his approach provides a number of computational benefits related to both model tuning and recommendations generation phases. The author of the thesis takes this approach as a base for his own research and generalizes it in several ways to address specific challenges that are not easily solvable within the standard approach.

One of the generalizations is related to construction of a hybrid model. While there are many hybrid models already developed, the most frequent approach used in many competitions is Factorization Machines, as it provides the most general computational framework for hybridization. The author of the thesis uses this approach for comparison with his methods. The author also outlines the key differences of his most general approach to the related tensor-based methods, such as Multiverse and TAPER and explains the main advantages of his approach over them.

Experiments conducted by the author follow a standard and widely accepted evaluation protocol and are quite well-designed. Comparison with competive methods wre also included. The datasets used in the experiments are well known benchmarks. They are collected from real-life recommendation services and are often used for benchmarking.

Moreover, the author has developed a recommendation framework to automate many parts of research, increase research reproducibility and make experiments easier to conduct. The framework is cross-platform, written in Python and is openly available online via the author's Github repository.

The main part of the thesis encompasses three approaches developed by the author. The first approach introduces a tensor-based formulation to address the problem positivity bias in both model construction and evaluation. Even though the well-known Tucker decomposition is used to build the model, the approach itself presents an original view on the problem. In addition to that the author proposes a new evaluation metric, which helps to perform a more thorough evaluation. The work based on these ideas was reported in the top-ranked specialized conference on recommender systems and was published in the proceedings of these confrences.

The second approach employs the generalized SVD formulation to extend standard PureSVD model with the ability to additionally account for side information. Adapting the generalized SVD for this specific problem represents a fresh idea never challenged before and extends the family of known SVD-based models.

The third approach takes the previous two and combines them within a single unified model. Specifically for this approach, the author derives a variation of the Tucker decomposition method based on a newly proposed hybrid higher order orthogonal iteration algorithm for sparse data. Despite presenting promising results, the last two approaches are only available as preprints on Arxiv and not yet published anywhere else.

Even though the work by the author is novel and can certainly find its application in certain areas (especially where using other approaches would be computationally prohibitive), a comparison with more recent state-of-the-art approaches would help to deliver a more complete picture on the applicability of the proposed methods.

Extending experiments to several more datasets would also be beneficial and helpful in understanding of the method's performance.

Hybrid approach concerns with sparse similarity matrices only. An analysis of the possibility to operate on dense similarity matrices at large scale would be a good complementary part to the research.

Among 5 publications of the author, two are not yet published anywhere and exist only as preprints on except Arxiv.

Presented methods are mostly applicable for the problems of order not higher than 3 or 4. It is still an open problem how to generalize them to higher orders as it would render both SVD and Tucker decomposition inapplicable.

Another direction for further research is unification of data preprocessing which currently requires to construct similarity matrices and depends on the way it is done. More automated approach would help to move towards a more general solution.

In summary, in my opinion the author contributions are as follows

- Three new methods for recommender systems are developed. They deal with several different aspects of insufficient preferences information problem and also allow to improve the quality of recommendations in the most general case as well.
- The proposed methods compare favorably with some of the state-of-the-art results. Moreover, these methods are relatively simple and due to this highly practical due to computational efficiency and the ability to work in highly dynamic environments due to a simple folding-in computation, inherited from the predecessor models.
- A new software was developed to support research. The software has grown into a versatile multipurpose recommendation framework.

.

Provisional Recommendation

x☐ *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*