Review report for the dissertation thesis

Author: Ing. Tomáš Řehořek

University: Czech Technical University in Prague, Faculty of Information Technology

Dissertation thesis: Manipulating the Capacity of Recommendation Models in Recall-Coverage Optimization

Supervisor: doc. Ing. Pavel Kordík, Ph.D.

Date: August 2018

The topic of the proposed dissertation thesis is actual. Especially for the e-commerce it is often highly valuable to recommend long-tail items. That's the reason that several major industry players are active in this area. Similarly, one of the views to personalized recommendation suggests that recommending top trending, well known items is useless for the user. This is, hand in hand, connected to qualitative metrics as novelty or serendipity.

The thesis consists of five chapters which cover introduction, analysis, proposed approach, main results and conclusions. The appendices present lists of author's publications (which are quite confusing for the reader). The thesis is well written, despite some language issues, the author clearly expresses his ideas. There are also several format inconsistences.

The structure, which was chosen, however, slightly decreases the readability of the thesis. As the author tries to do not repeat facts presented in the section 2 (analysis), there are plenty of cross references across the thesis. Moreover, from my perspective, the thesis includes enormous amount of equations (121 only in section 2), which are often over-complicated (e.g.. Eq. 2.94, 2.95). Together, this makes the thesis quite challenging to read and understand.

The goals of the dissertation cover all aspects which are required – the state-of-the-art analysis of the research topic, proposing novel approach and its evaluation. Author proves the ability to conduct systematic literature overview, formulate a problem and propose solutions with

Ilkovičova 2
84104 Bratislava
Slovensko
www.fiit.stuba.sk

michal.kompan@stuba.sk
www.fiit.stuba.sk/~kompan

evaluation. Moreover, some of the authors results were accepted by the research community by the means of several papers accepted on international conferences or workshops.

Author used appropriate methods for the dissertation goals fulfillment. This indicates the results presented in the thesis to be valid. From the survey of the state-of-the-art perspective, I would appreciate if the author could describe the methodology which was used to create the systematic literature review.

Author described in total 14 "main contributions" of the thesis. This is in my opinion quite large number and refers to low granularity which is used. In the contrary, the evaluation addresses discussion on model behavior while manipulation with capacity hyperparameters. I personally like the usage of both academic and industry datasets. As the differences between academic and industry recommenders are often addressed by the author, I would appreciate the evaluation of proposed approach in the real-case scenario through various metrics (e.g., revenue, user trust or satisfaction). In this way, not only the correctness of proposed approach would be shown, but also the real contribution in the long-tail recommendations for industry.

I appreciate that proposed dissertation is strictly focused on the problem, author defined in the introduction. However, I personally miss a brother context of concepts and problems, which author described. For example, the conclusion of the thesis consists of: a very short summary, a list of contributions and a short feature work proposal. I would expect some "philosophical" discussion on the proposed approach, the domain of recommender systems and connections between academic and industry goals (based on the introduction of the thesis).

As the results of the thesis refer to various interesting problems, I also miss author's comments on these topics: session-based recommendation and context (which are often used for users with limited history) or user modeling (which is connected to various weighting schemes author uses for predictions).

Questions:

Were proposed approaches used in a real-case scenario? Which pros and cons were identified?

What is the author's opinion on further research directions in recommender systems?

As the author's publication coverage of the thesis is rather small, is there any plan to further publish results?

To sum it up, proposed dissertation addresses an actual research problem. Author proposed approach to improve the long-tail recommendations for several algorithms. Despite some limitations of the thesis and evaluation, I consider the results as a contribution to the state-of-the-art. Moreover, I hope, that the results obtained by the author, will be further published.

The author of the dissertation proved the ability to conduct research and achieve scientific result. In accordance with par. 47, letter 4 of the Law Nr. 111/1998 (The Higher Education Act) I do

recommend the thesis for the presentation and defense with the aim of receiving the Doctor of Philosophy degree.

V Bratislave 07.01.2019

doc. Ing. Michal Kompan, PhD.

Manipulating the Capacity of Recommendation Models in Recall-Coverage Optimization

Doctoral dissertation of Tomáš Řehořek

Faculty of Information Technology, Czech Technical University in Prague

This dissertation tackles the topic of long-tail recommendations, and, specifically, proposes a systematic approach to controlling the number of such recommendations in the recommendation list generated by a recommender system algorithm. The approach is well-grounded in the theoretical aspects of machine learning, and it is also independent of the specific recommender system algorithm used.

Up-to-dateness of the dissertation.

The topic of the dissertation is up to date. The recommender system research community has been long aware of the importance of recommending long-tail items, but there is little research that tackles the topic specifically.

This dissertation represents a welcome breath of fresh air for two reasons: First, the mathematical formulation is clear, comprehensive and systematic, and second, the work focuses on analysis of system behavior and at understanding why certain behaviors arise and how they can be controlled.

Formal structure and organization of the dissertation.

The dissertation is well organized and a pleasure to read. Chapter 2 is a little long compared to the other chapters, but it quickly becomes clear why this needs to be the case. The author should consider extending Chapter 2 into a short textbook on recommender systems: Many students would appreciate the clarity and systematicity of the formulations, already mentioned above.

Completion of the dissertation objectives.

The dissertation completes its objectives. Given that it is motivated by the research-industry gap, it would have been nice to have additional discussion of the larger implications of the findings of the dissertation at the end. However, the amount of work that it takes to create a systematic framework as this dissertation has done should not be underestimated. For this reason, I do not fault the dissertation on its scope.

Methods used in the dissertation.

The dissertation uses an interesting and welcome extension to the usual experimental setup adopted in recommender system research. The dissertation is not attempting to maximize a single value and much recommender system research does, but rather improve systems simultaneously according to multiple objectives, and also explain and illustrate how this improvement happens.

Care was taken to choose the correct evaluation metrics. In this way, the work avoids the trap of proposing new evaluation metrics simultaneously with a new approach that optimizes those evaluation metrics. The careful experimental design carried out in this thesis should be more widely adopted in the recommender system research world.

Results and contributions of the dissertation.

The dissertation makes contributions at two levels. First, it presents a framework to attack an important problem in recommender system. Second, it serves as an example of how

issues with recommender systems can be successfully addressed by a thorough and systematic mathematical formulation of the problem.

Remarks, objections, notes, and questions for the defense.

Here, I mention a couple of the rough points of the dissertation:

- It may not be wise to map ratings on to the interval [-1,1], since information is lost about whether an item was given a low score or was explicitly rated as a dislike; further some algorithms may not take negative numbers as input (Section 2.1.3.2).
- The GDPR currently restricts the use of data by companies in Europe. This impacts the author's assumption that "...interaction data are a side product of normally using the service, and they might not have been originally intended as a data source for Recommender Systems". It would have been nice for the legal aspects of data use to have received mention in the dissertation. Specifically, a couple of words on the legal status of the industry data sets used by the author would have been appropriate (presumably the data was anonymized).
- Chapter 5 "Conclusions" is very brief, and does not close the loop to the introduction: Has the dissertation truly succeeded in narrowing the research-industry gap with respect to long-tailedness?

The version of the dissertation read contained a number of typos in the text, although none were to be discovered in the extensive formulas.

Questions for the defense:

- The work is valuable in a practical setting, because it provides a way to control the proportion of long-tail items in the results list generated by the recommender system. However, in practical situations, control could also be achieved by merging the results of two recommenders, one recommending popular items, and one recommending long-tail items. What would be the advantages/disadvantages of such an approach and why was it not discussed here? (The dissertation mentions the case of news recommendation, and the author seems to be aware of that certain articles must always be in the list because they are relevant to everyone. Combining two different sources of recommendation seems to be the safest way to make sure that people are not lost in a news filter bubble.)
- The dissertation (2.5.5 Model Capacity, Underfitting and Overfitting) mentions that the Vapnik-Chervonenkis dimension as being an important measure of model capacity. Then, however, the discussion goes on to distinguish between the theoretical capacity and the actual learning abilities of the algorithm. The analysis focuses on the actual learning abilities of the algorithm. Is there any sense to attempting to estimate the Vapnik-Chervonenkis dimension for recommender systems? Why or why not? How would we do it?
- The experiments demonstrate that β controls the tradeoff between coverage and recall. In a practical situation, how should the operating point of β be chosen? Is there a dependence on N that is assumed when measuring recall?
- The recommender system community often debates whether a system must necessarily involve personalization in order to be considered a recommender system, and what personalization actually means. Let's assume for this question that personalization means that no two users receive the same recommendation list. Does driving a recommender system into the long-tail necessarily increase

personalization? Why or why not? Should we even be worried about this definition of personalization in an industry setting? (Does it apply?)

- Are there other important problems in recommender systems that, like long-tail recommendations, do not get enough attention in research, and which could benefit from the systematic approach applied here?

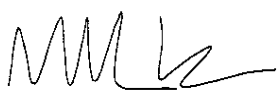
Final evaluation

In sum, the thesis proposes a novel approach to an important problem of recommender systems. The unified system of formalization it introduces is also an important contribution to the recommender system research community. The structure of the thesis is well thought out, and it is dense, but easy to read. The methodology chosen is solid, and has been well applied, and the dissertation succeeds at making its stated contributions.

The author of the dissertation proved the ability to conduct research and achieve scientific results. In accordance with par. 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act) I do recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

Respectfully submitted,

Prof. Martha Larson
Delft University of Technology
Radboud University

Signature of the reviewer: 

Chicago, Illinois, 3 February 2019

CHARLES UNIVERSITY PRAGUE

faculty of mathematics and physics

P.T.

To Whom It May Concern

Faculty of Information Technology

Czech Technical University in Prague

Report on PhD thesis

Manipulating the Capacity of Recommendation Models in Recall-Coverage Optimization

By Tomas Rehorek

A main motivation of this thesis and an important real world challenge is recommendation of items to user based on his/her previous activities especially for non popular items but of possible interest for users. The problem is *up-to date*, important, difficult and widely studied.

Thesis is *formally structured and organized* in 5 chapters, bibliography and a copy of a paper from last RecSys conference. Chapter 1 gives motivation, problem statement, listing of main previous results on which the thesis is built and goals of dissertation. Chapter 2 "Background and State-of-the-Art" introduces the reader to the necessary theoretical background and surveys the current state-of-the-art in a unified common framework. Chapter 3 "Overview of Our Approach" presents approach to controlling and manipulating model capacity for most important algorithms in recall-coverage plane. Chapter 4" Main Results" demonstrates that theoretical model corresponds to empiric behavior on several datasets. Chapter 5 concludes the thesis.

Methods of thesis appropriately combine theoretical models, their implementation and experiments.

Prof. RNDr. Peter Vojtáš, DrSc.
Mobil: +420 739 822 406
vojtas@ksi.mff.cuni.cz
http://www.ksi.mff.cuni.cz/~vojtas/

Department of Software Engineering
Malostranské nám. 25, 118 00 Praha 1
Phone: +420 22191 4239
fax: +420 22191 4323

Results and contributions of this thesis are diverse, interesting and valuable. Very valuable is already Chapter 2 with a very good overview of recommendation models, algorithms and practical insight. One of main theoretical results is a generalized definition of validation reward function capturing also catalog and user coverage. Most important result, from my point of view, is a new way of extending popularity-based regularization to larger spectrum of algorithms (originally another way was proposed in [101] for one specific MF algorithm). This enables to control magnitude of long-tail recommendations. Very nice is an exhaustive set of experiments on a mixture of 7 public and industrial datasets with algorithm-dependent capacity parameters and the response in the recall-coverage plane.

My main *question* is whether author's way of manipulating capacity is a generalization of the approach from [101]. I would appreciate if a comparison of results of [101] and of author for MF algorithm could be presented at defense (e.g. for Netflix data where power-law behavior is present – do all considered data behave by power-law?).

Comparison of experiments with various algorithms on same data could be also interesting – e.g. for MovieLens 1M data there are different scaling of recall (ranging from 0.1 to 0.55) –catalogCoverage (ranging from 0.16 to 0.3) in respective plane. This could be used also to decision for which data which algorithm and which parameters (especially \beta) are giving best result. Hence a further discussion can be interesting: e.g. if (Pareto-)best results are for \beta close to 0, 1 and/or -1 respectively.

Isn't (3.1) just reranking? Could \beta act linearly on score?

Few technical comments: In algorithm 1 while-do should be used instead repeat-until, because C^{k+1} can be nonempty and S^{k+1} empty. (2.6) and (2.7) are too general and never really practically used, (2.8) is better and used.

Finally I *evaluate* of this thesis very high. I think that it will stimulate many further theses.

The author of dissertation proved the ability to conduct research and achieve scientific results. I do recommend the thesis for the presentation and defense with the aim of receiving the PhD degree.

Prague, February 08th 2019

Peter Vojtáš