Vincent Guigue
Associate Professor
Tower 26-00, office 519
4 place Jussieu
F-75252 PARIS cedex 05
tel : +33 (0)1 44 27 88 54

To whom it may concern

Paris, June 7^{th} 2022

Object : Report on A. Burashnikova thesis

Aleksandra Burashnikova's thesis proposes to study context-aware recommendation systems based on pairwise learning. The thesis also presents an energy grid analysis system.

Chapter 1 The introduction clearly positions recommender systems within the broader field of information access. A. Burashnikova insists on the importance of taking into account the context, in particular the temporal context, as well as the formatting of the data (taking into account positive and/or negative feedbacks). Finally, the learning criterion is important to finely order the elements to be recommended.
The introduction ends with a quick summary of the different articles written during the thesis.

Chapter 2 This chapter first introduces the usual notations, assumptions and formalisms in supervised learning. It details the notions of empirical risk, overfitting and model selection in the sense of Vapnik's generalization error (Structural Risk Minimization).
The chapter then discusses optimization methods, in particular gradient descent. A. Burashnikova describes and compares in great detail the different formulations of gradient descent with moments.
Classification is then formalized using SVMs, the notion of margin and the hinge cost function.
Neural networks are discussed briefly, with the section briefly describing biomimetic intuition before moving on to perceptron and deep-learning architectures through the various classical activation functions, MLP and cost functions.
Regardless of the framework, the ranking problem corresponds to the definition of a new cost.

This chapter is very clear but rather broad, it is difficult to understand at first reading which elements will be used later.

Chapter 3 This chapter discusses the bibliography of recommender systems, initially distinguishing between content-based systems and collaborative filtering. A. Burashnikova first

describes matrix factorization, emphasizing the shortcomings of the least squares cost function.
While the matrix factorization is very well described, the description of the contributions of CNN & RNN for recommender systems is a bit succinct. In particular, the use of these architectures for modeling user sessions would deserve analyses regarding the representation of items and possible new tasks (such as predicting the continuation of an item sequence). The figures representing the architectures would have benefited from being more focused on the recommendation systems to better highlight the contribution of these approaches to this application. Matrix factorization is an interesting approach to bridge the gap between classical learning techniques and deep learning : the general organization of the section could have benefited from a transition around these aspects.
Similarly, the description of GCNN could have been preceded by an explanation of how to model the recommender system as a bi-partite graph.
The chapter also discusses Transformer architectures for representing sequences of discrete events. A. Burashnikova shows how this very recent family of approaches is very well adapted to user session modeling.
This chapter continues with a section on metrics, which have evolved a lot in the last decade. The description is clear and factual but could have put more emphasis on the properties and prerequisites of the different metrics, for example the focus of MAP and nDCG on the top of the list or the fact that nDCG requires graded interest scores on the responses (those point are mentioned in the next chapter).
Given the interest of this thesis for ranking systems, metrics based on AUC could be mentioned (AUC itself or ATOP for example). In terms of organization, the highlighting of MSE flaws would benefit from being presented in this last section rather than in the matrix factorization.

This section is very complete and educational : it is a critical milestone to understand the next contributions. All the architectures required to understand the suite are well described.

Chapter 4 This chapter presents the first contribution of this thesis around the modeling of user sessions in which some items are validated while others, more numerous, are not.
The SAROS algorithm is based on Bayesian Personalized Ranking (in its exponential form) but distinguishes the cost functions associated with each temporal block. A pair of hyperparameters distinguishes users to be considered from those that are just noise.
A. Burashnikova then demonstrates the convergence of the stochastic gradient descent on her proposed formulation, on this ranking problem, despite the lack of update in some cases.
The chapter continues with a series of experiments where SAROS is put in competition with different state-of-the-art approaches. The fact that SAROS focuses on only some users allows it to be more efficient in learning. The performances are very interesting in MAP and nDCG. Even though these tables are based on a temporal separation of the data and the learning algorithm iterates over temporal blocks, the figures do not illustrate very well the online performance (and its possible evolution over time).

This first contribution is valuable and wide, ranging from theoretical proof of convergence to very impressive application results. As a minor remark, we notice that the performance in this setting is very favorable to BPR and SAROS compared to more recent approaches : a more thorough analysis of this fact would be interesting. This chapter takes up the article published by A. Burashnikova. In terms of form, a standardization of notations with the previous chapter would have benefited both chapters.

Chapter 5 This chapter, still dedicated to contributions, presents variants of SAROS that introduce different memory sizes. This chapter is the occasion to introduce a new dataset and new reference algorithms including the very powerful LightGCN.

This second series of experiments is interesting as it focuses on the memory in time aware RS, which is not studied a lot in the literature. It shows the interest of a deep study of the memory mechanisms around SAROS (and MOSAIC) which can improve at the margin the performances but also the efficiency of learning models.

Chapter 6 This chapter presents a technique for identifying faulty lines in an electrical network. The chapter describes a convolutional architecture for the localization of these faulty lines. The idea is to introduce a composite cost function with a secondary target impacting the target's neighboring nodes, in order to regularize the network and perform the localization better, especially in the case where the measurements on the network are more spaced.

A. Burashnikova distinguishes evaluations on simpler cases where a measurement is available on a neighbor, from the more difficult cases.

Conclusion Aleksandra Burashnikova has presented in this thesis a serious and thorough work around recommender systems. While some parts of the paper could have benefited from a thorough proofreading or a little binding in the notations, the whole presents relevant contributions with theoretical guarantees and supported by extensive experimental campaigns.

These strongs contributions and the associated significant publications allow Aleksandra Burashnikova to defend her thesis.

Vincent Guigue