Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Yermek Kapushev

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Gaussian Process Models for Large-Scale Problems

Supervisor: Associate Professor Evgeny Burnaev, Skoltech

Name of the Reviewer:

I confirm the absence of any conflict of interest	**Signature:** **Date: 26-11-2020**

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 is generally well written and easy to follow. The structure into 5 chapters is appropriate; chapters 1 and 5 contain introduction and conclusions, while chapters 2-4 contain the contributions of the thesis.

The relevance of the topic of dissertation work to its actual content

The thesis investigates a very important problem in Machine Learning and Statistics, which is the problem of scalability of Gaussian processes. These models are widely used in many applications, and scalability is a truly desirable feature. This is especially important for applications where quantification of uncertainty matters. The thesis focuses on the use of structured inputs to achieve scalability and numerical quadrature to improve random feature approximations, which are interesting research areas spanning Machine Learning/Statistics and numerical methods for linear algebra. The thesis applies these ideas to three tasks, namely tensor completion, density estimation, and simultaneous localization and mapping (SLAM).

My understanding is that the thesis makes the most original contributions for (i) the setup of Gaussian processes with inputs on a grid with missing points, which a case that is rarely considered in the literature and (ii) the casting of existing approximation for Gaussian processes as a special case of a proposed quadrature for kernel function.

The thesis makes significant advancements in these direction. In addition, the thesis considers some interesting applications demonstrating the usefulness of the proposed methods.

The relevance of the methods used in the dissertation

Overall, the thesis builds upon recent advancements in the Machine Learning literature on Gaussian processes with structure inputs, and it is therefore of high relevance for this literature. Furthermore, the application on SLAM shows compelling results, indicating that these approaches might have a good impact on applied disciplines too. Here is a detailed breakdown of the relevance of the contributions for each chapter.

Chapter 2 elaborates on the topic of Gaussian Process Models on Multi-dimensional Grids. These have found a reasonable degree of success in the literature of scalable Gaussian processes. This thesis extends this literature by considering the problem of missing inputs, which makes the treatment of these approaches a bit more involved.

Chapter 3 considers the setup where inputs are not necessarily on a grid. This chapter focuses on random feature approximations for Gaussian processes, and quadrature methods to approximate kernels. The chapter shows connections between quadrature methods and known random feature approximations, while deriving error bounds of these approximations.

Chapter 4 discusses the applications of the concepts developed in the thesis to a number of tasks, such as tensor completion, density estimation and simultaneous localization and mapping. The results show improvements over the state-of-the-art.

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

The results obtained in the thesis are significant. In particular, they indicate that it is possible to enjoy scalable inference for Gaussian processes on grid-structured inputs with some missing inputs. Furthermore, the thesis shows some interesting results on the application of quadrature methods to random features. The methodological works appear in selective venues in Machine Learning, suggesting that these are of high international standards. Of particular notice, is the spotlight at NeurIPS, which is arguably one of the top conferences in Machine Learning these days.

The relevance of the obtained results to applications (if applicable)

The results on aircraft engine design and in turbomachinery are interesting and offer some improvements for these manufacturing problems. In addition, the thesis considers three tasks, which are tensor completion problem, density estimation and simultatineous localization and mapping, spanning three areas such as computer vision, Machine Learning/Statistics, and robotics, offering state-of-the-art performance.

The quality of publications

As already mentioned, the methodological contributions appear in a number of selected venues in Machine Learning, such as NeurIPS. In addition, some contributions appear in selective venues in numerical analysis. This suggests that the advancements proposed in the thesis are of international standards.

The summary of issues to be addressed before/during the thesis defense

The thesis develops a number of interesting ideas to scale Gaussian processes to large data by building on the literature of structured inputs. A natural question is the positioning of these approaches with respect to the literature on scalable Gaussian processes based on approximations. For example, the variational sparse GP of Titsias, AISTATS 2009 and followup works, or the random feature approximation in Cutajar et al., ICML 2017 combined with variational inference. The introduction points to the limitations of Gaussian processes, but I would argue that there is a considerable literature on ways around these issues. For instance, kernel design can be bypassed by using deep Gaussian processes or a combination between deep networks and Gaussian processes; scalable approximations, which can work with any likelihood, exists for these approaches. What are the advantages/disadvantages of the proposed formulation? I support the idea that exact Gaussian processes should perform better than approximate methods, and there is literature showing this (Cutajar et al., ICML 2016, Wang et al., NeurIPS, 2019), but I believe it could be interesting to elaborate a bit on these aspects in the introduction and/or in the conclusions to improve the overall vision of the thesis. And some comparisons with these methods in various experiments in the thesis would have been really interesting to see. A further questions to elaborate on could be: "how important is it to carry out exact inference for these models if we introduce constraints on the design of the experiment?", given that current approximate scalable approaches to Gaussian process are truly general.

I think each chapter would benefit from a short introduction indicating the main contribution. For example, it is difficult to extrapolate what is the novel contribution of Chapter 3.

Finally, I'd suggest extending the conclusions by adding some comments on future perspectives for the developments of this thesis in the context of current trends in Machine Learning.

Minor comments:

- While the thesis is generally well-written, I found that the writing could be improved and that a lot of typos could be corrected. Perhaps, a quick proofreading from a native English speaker would help to smooth some of these things out.

- Add some references to the introduction to support some of the statements and acknowledge some previous works. I don't think this needs to be exhaustive, but it should serve as a way to give a bit of context of this work within the literature.
- Maybe switching to a natbib reference style would help, so that author names appear in the references within the text.

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