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: Assistant Professor Alexey Zaytsev, Skoltech

I confirm the absence of any conflict of interest 	**Signature:**  **Date: 28-11-2020**

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 makes a significant contribution to the field of Gaussian Processes and kernel methods in general. The evaluation of the proposed approaches for the several applied problems proves their importance. So, the overall thesis quality is high.

  The structure of the thesis is consistent and is easy to follow. The thesis has five chapters. The first one and the last provide an introduction and a conclusion. Chapters 2 and 3 describe the developed approaches. Chapter 4 considers the application of the proposed methods.

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

  The content of the dissertation is relevant to the declared topic. The thesis develops techniques for the construction of large-scale Gaussian Process models and kernel methods. The author considers two approaches: utilization of the structure in input points to perform computationally efficient inference, usage of quadrature rules to obtain a low-rank approximation of the kernel function. These techniques are of interest on their own for regression tasks. Also, they can serve as building blocks in other applications. The thesis demonstrates the power and usefulness of the developed techniques in diverse problems.

- *The relevance of the methods used in the dissertation*

  The thesis has three main parts.

  The first part considers the problem statement when the input points lie on a multi-dimensional grid. The author considers a vital case where some points in the grid are missing. This case is rarely considered in the literature, so this part is novel and relevant.

  The second part deals with a more general case with no structure in the data. For this case, the author proposes a quadrature-based approximation of the kernel function. This approximation shows improvement over the state-of-the-art approach and has interesting properties. In addition to a good performance, it unifies several random features based kernel approximation techniques.

  Finally, the thesis contains applications that build upon the developed approaches. The author proposes approaches for a tensor completion problem, a density estimate problem, and a simultaneous localization and mapping (SLAM) problem. For all three problems, the proposed methods show state-of-the-art results in numerical experiments.

  All proposed methods lie on top of the current state-of-the-art techniques in Machine learning in terms of novelty and relevance. They are theoretically grounded and supported by results presented in the experimental sections.

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

  The obtained results have a high impact on Machine learning, particularly kernel methods that are now on the rise in ML community. As we see from the applications chapter, robotics and statistics could benefit from the proposed approaches.

  The results were published in top-venue conferences and journals. There is no doubt that this work is on par with the current state of the art. However, the author should consider description of his contribution to each of the publication, as he is the first author only for one accepted publication

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

  The chapter 4 clearly demonstrates the relevance of the results to important applications.

- *The quality of publications*

  The results were published at recognizable (Q1/Q2) journals and were presented at top conferences (one of the works was accepted at NeurIPS with a spotlight talk, the acceptance rate as the spotlight talk is less than 1%). So, the publications are of high quality.

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

  In the case of data on a grid with missing points, the log-likelihood calculation complexity depends on the number of missing points and, in some cases, can be enormous. I suggest devoting more time to the description of the limits of the proposed approaches and how close it is to the theoretical bounds.

  I think that motivation for the random features-based approach lacks details. There are many data-dependent approaches (like Nystrom approximation and follow-ups), variational inference-based approaches. The latter ones should be added to the literature review. Also, it will be interesting to see some kind of discussion of the limitations of the proposed approaches and possible directions of the future research.

  Prior selection in 2.2.2 is not motivated, can we select a better prior? What is the common practice for the selection of prior in this case? Also, the evidence that we need a prior consists only of one figure. Can you elaborate more on this issue and provide more detailed study?

  While the quadrature rules outperform other approaches in terms of the kernel approximation error, on regression/classification problems the improvement is less notable. Could you provide more precise results in this area?

  Minor issues:
  - In some chapters/sections related works are given in separate subsection and in some – in introduction to the section. It is better to choose one scheme and follow it throughout the manuscript.
  - The labels for x axis in Figures 3.1 and 3.2 are not very intuitive, it is better to change it to number of features.

- In some of the figures the font can be enlarged for better readability
- Storage complexity for grid DoE methods is not presented in Chapter 2
- Figure 2.8 – please describe, what is R and N in the caption of the figure
- For chapter 3 the theoretical analysis of the proposed method requires further clarification and some more evident example of what the theorems mean in a more practical spirit, may be a figure on this
- The manuscript is generally well-written but still requires proofreading for typos. E.g. in the abstract:
    - a problem of building large-scale models -> the problem of building large-scale models
    - consider -> consider
    - This very diverse set of problems demonstrate -> This very diverse set of problems demonstrates
    - For the first case we develop technique -> For the first case, we develop a technique
    - Mix of present (develop, show) and past (developed) should be avoided in one paragraph

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*