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: Vice-President for Artificial Intelligence and Mathematical Modelling, Professor, Maxim Fedorov

I confirm the absence of any conflict of interest	**Signature:** **Date: DD-MM-YYYY**

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 Ph.D. thesis by Yermek Kapushev proposes several techniques to be applied to build Gaussian Process and kernel-based models in large-scale regression problems and their applications in different fields such as statistics and robotics. The kernel methods are highly applicable in many regression/classification problems on its own as well as part of a larger system. The author not only derived computationally efficient techniques, prove their efficiency theoretically and empirically, but also demonstrate how they can be incorporated into different pipelines. Such demonstrations are important as they open a new way of solving real-world problems that become possible only with the development of large-scale models. All in all, the results are novel, provide a powerful tool to solve many real-world problems and its potential impact is very high.

The thesis consists of 5 chapters. The first chapter contains motivation and necessary background on Gaussian Process models. Chapters 2 and 3 present the developed methods for structured and unstructured data sets while Chapter 4 contains applications of the methods in different areas. The dissertation is well structured and easy to follow. Each technique has a theoretical justification and some of them theoretical guarantees and error bounds followed by an experimental section to support the claims.

The topic of the dissertation represents its actual content. However, the third chapter described the approach for the kernel approximation and contains very brief discussion on how it is connected to Gaussian Processes. Also, in the application section the part that is devoted to density estimate doesn't use Gaussian Processes but just kernel-based approach. The connection to Gaussian Process requires more discussion in these 2 cases.

The quality of the research is justified by high quality publications (1 paper in Q2 journal, 3 papers in Q1 journal, 1 publication in A* conference and 1 publication is submitted to rank B conference, though it is one of the top conferences in robotics).

I highly value the work and believe that the author deserves Ph.D. degree. However, I have several questions and comments for consideration.

- In case of the data set on grid with missing points the proposed approach for tuning Gaussian Process model is efficient in case of small number of missing points. What should we do if the number of missing points is large?
- Why did you choose to develop data-independent approach for kernel approximation? The thesis lacks motivation on this. Where it can be useful? Why did you prefer this approach over data-dependent ones?
- As was mentioned above, Chapter 3 should contain some discussion on how Gaussian Processes and kernel methods are connected and how the developed kernel approximation can be used with Gaussian Processes.
- The kernel approximation technique shows impressive accuracy of the kernel function approximation compared to other techniques. However, on the downstream tasks the improvement is less significant. What is the reason for such phenomenon?
- One of the important parameters is the number of random features. Maybe you could comment on the choice of the number of random features?

There are also a few typos and missing punctuation across the manuscript to be taken care of.

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*