Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

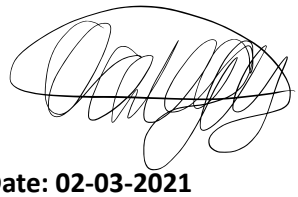
Name of Candidate: Nikita Klyuchnikov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Multi-fidelity classification and active search

Supervisor: Associate Professor Evgeny Burnaev

Name of the Reviewer: Ivan Oseledets

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 02-03-2021**

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

The thesis has 96 pages and 4 chapters. The main topic of the thesis is multi-fidelity, in different tasks. Chapter 2 describes new methods for multi-fidelity classification, Chapter 3 - for active search, and Chapter 4 discusses really numerous applications of the developed methods in different areas. The results are published in reputable venues, are novel and significant. With no doubt, this is a high-level work. There are, however, comments on the presentation of the results. First of all, the term «multi-fidelity» (and fidelity itself) is used straightforwardly, without any introduction and informal description, such as if it well-understandable to all of the readers. The introduction discusses the importance of different data sources, by does not give a simple and intuitive definition of different aspect of fidelity. As far as I can grasp, the synonym to this word is «precision», and that is exactly how the term is used: we have different sources of data, with different costs. I think, just from the start, this should be explicitly described in the introduction, along with the specific examples where and when this problem setting is really a game-changer (there are many!). A formal definition appears in the classification setting only on page 13.

Second comment on the presentation are the references. On the same page 13, the author references «Bayesian approach». However, why not put a reference to some classical textbook? Then, the Gaussian process appears, which could be specified more formally as well.

In Table 2.1 explicit gradients are derived. Although they are important, do we really need them, if we implement the evaluation of the cost functional in some ML framework, the gradients could be computed for «free» ?

The baselines (2.2.4) do not use multi-fidelity; are their any competitive approaches using the same setting?

Table 2.3 does not report standard deviation (although it is standard nowadays).

Section 3.3.3 has «Regret guarantees» but I do not see it there.

Section 4.1 (and Chapter 4) begins without any connection to the previous ones. A small introduction would improve the readability. The «SHiP» acronym should have a full name, and also should have some reference. The Standard Model also should have a reference to it.

The first formula in this chapter are even typeset in a different engine.

It is difficult to understand the meaning of Figure 4.1 in an ML dissertation.

Similar comments apply to Section 4.2: some basic references are needed for a good style.

Page 49: «According there the fundamentals of Machine Learning» - very misleading sentence. There are no such results, and it is not good to use capital letters in the middle of the sentence. The whole motivational paragraph is probably written for a paper for a oil&gas journal, and looks completely alien here.

The whole section is in fact comparison of 3 standard ML methods and looks not very correlated to the topic of the thesis. The conclusion tells that «MF gpc» has been applied to the data-driven lithotype identification at directional oil well drilling, but it is not easy to find in the large section devoted to it. The same applies to the next part.

Finally, the thesis (for the amount of work done) is quite short for my taste. It would benefit from better discussion, and also, probably, some centralized literature review in the beginning.

These comments can be used to improve the thesis. However, it does not change my evaluation as a high-quality work. The strongest result is by far the MS-ASC method. Its presentation stands out from the other parts of the work, and it alone is enough for a PhD thesis. Other chapters are also very interesting. I highly support the thesis defence.

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