



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Fabio Casati
Dept of Information Engineering and Computer Science
University of Trento
Italy

Trento, Feb 28, 2019

Object: Assessment of Oskar Jan Jarczyk's PhD Thesis

I include below my assessment of Oskar's thesis.

Overall i recommend an accept, with minor edits to clarify the issues related to statistical analysis and to the discussion that at some point it seems to make causal inferences where there seems to be "only" correlational evidence.

I find the topic very interesting and timely and the results quite insightful. Sometimes I missed a deeper analysis that starts from the findings and proceeds to understands the reasons behind them, as this would lead to actionable recommendations for OSS managers and teams. Nevertheless, it is also ok to leave this to future students as the results are sufficient for awarding a thesis.

Detailed comments according to the format are provided below.

1. *What are the research problems and objectives considered in the thesis, and have they been sufficiently clearly described by the author?*

The thesis discusses quality metrics for open source projects, as available on github. It first proposes quality and popularity metrics for such projects and then focuses on specific quality metrics such as speed in responding to issues. Finally, it deals with recommendations, such as how to recommend projects to developers.

I think the metrics and research questions are fairly clear and well articulated.



2. *Does the thesis contain an appropriate analysis of state of the art (based on global scientific literature, current knowledge and applications in industry)? Does the analysis of related work demonstrate sufficient expertise of the author? Have the conclusions of the review of related work been sufficiently clearly stated?*

Yes the thesis contains a proper discussion of the related work, both in the introductory chapter as well as in the individual papers.

3. *Does the research described in the thesis use a correct scientific methodology?*

The methodology is correct for the most part. There are a few point where I think the explanation needs to be extended and some decisions need to be explained better.

One such point is the discussion on significance of certain factors. The thesis would benefit from a discussion of *corrections* when assessing significance, as in many cases you are not starting from an hypothesis but testing many different things, so even by chance some of them will result “significant” – if you consider significant the p value at 0.05.

The second issue is that sometimes the discussion seems to confuse correlation and causality. This happens particularly in the first paper and in the part related to recommendations. I do not think you can derive recommendations from those correlations.

4. *What are the original and innovative contributions of the author, and what is the position of these contributions compared to the state of the art?*
 - a. *How do you evaluate the publication record of the candidate?*

The original contributions are in terms of the quality measures, the predictive model, and the recommendations. The contributions are interesting and innovative, I am not aware of similar predictive models in the literature.

The publication record is sufficient for the awarding of a PhD degree.

5. *Did the author present his results correctly and convincingly? (Please evaluate the clarity, conciseness, correctness of the thesis or presented research articles).*

The presentation is clear and concise, my only observations are related to the points mentioned above in point 3 and then again later in point 6.

6. *What are the weak and strong points of presented research results?*

The thesis is very well written and provide in depth motivations for the identification of metrics and datasets, as well as for the rationale behind the approach. The results are also interesting and in part somewhat surprising in terms of correlations.



I would have expected to see follow up of some aspects, for example in terms of a more in depth analysis of the identified correlations, eg to identify direction of causality.

Another aspect I would have appreciated is a qualitative analysis resulting from interviews with developers to uncover some of the behaviors and motivations behind the behavior. This would have made the analysis much more insightful.

7. *What is the contribution of the thesis to the discipline of information technology?*

Anything that contributes to understanding of open source sw development contributes to IT.

The thesis mentions a set of contributions to IT in a dedicated section, but I think the main contributions to IT coincide with the main contribution of the thesis.

8. Are the presented achievements of the author sufficient to grant him/her a doctoral degree in the field of technical sciences in the discipline of computer science or software engineering?

Yes

Fabio Casati

Fabio Casati