Practitioners' Expectations on Automated Code Comment Generation
HU-ICSE'22

Summary of Paper

- This paper had a wide-ranging focus - from motivating what an automated code comment generation tool is useful for, to presenting a holistic overview of expectations and characteristics for such a tool from stakeholders and practitioners who write software. It also surveyed the domain of automated comment generation and summarized works published in the top research venues in SWE. The literature survey results provide a detailed inspection of where the current focus of researchers in this area is, and the interviews and online survey results conducted pinpoint specific areas where the expectations of practitioners lie with respect to ongoing and previous work in the research area.

Critical Evaluation Criteria

- **Soundness of Approach**
  In my opinion, the paper does a decent job expanding upon all its stated focus areas. The interviews and global online survey are done in a systematic manner which is clearly explained and well defined. While appreciating the overall approach, there were a few design choices that are worth analyzing critically.
  Firstly, prima facie, the interviews conducted were somewhat limited in scope in the sense that the number and quality of participants selected for interviews are not explained well enough. Moreover, according to the authors, the survey was designed and refined based on the responses from the interviews. So, if the interview participants were not representative of a larger community of software engineer practitioners from across the globe, then the conclusions drawn from those interviews and the design of the survey itself could be impacted. The authors partially mention this issue in the threats to validity section but deem it to be a mild threat. However, in my understanding, 16 interviews from software practitioners is not a large enough sample size that is representative of the opinion of millions of developers worldwide, and could potentially introduce selection bias into the study. To mitigate this, more information about who the interviewees were and how they were selected would have been helpful, as well as incorporating some sort of diversity criteria in the selection process for the interviewees. To summarize, there needs to be more proof that the selected participants for the interviews are representative enough such that their opinions about current code commenting practices and code generation tools hold true across the larger population.
  Apart from this, I believe the online survey covers most of the important topics that would be helpful in drawing conclusions from (in this particular domain). The survey participant selection, however, was done in a relatively informal way, such that most of the survey participants were from the wider social network of the authors. Objectively speaking, this could have the risk of survey participants falling prey to confirmation bias, since they could be aware of the authors' affiliations and their current and previous works and skew the results in favor of automatic code generation tools (which is a research area that the authors have been actively working in). By my own admission, this is likely a mild threat to the validity of the authors' results, but a threat, nonetheless.
  Lastly, the literature review was also systematically done and explained concisely, and the overall approach for that was sound.

- **Novelty**
  The novelty of the paper is not in any major technical contributions but in the extent the authors go to get an understanding of the needs and expectations of software practitioners

regarding what are the desirable and undesirable characteristics of a code comment generation tool. The paper lays the groundwork in terms of defining the necessary characteristics for automated code comment generation tools. While other works exist that have studied code documentation practices extensively, this work is larger in scope and more focused on answering questions regarding very specific issues in code commenting practices. The overall methodology in gathering information is relatively sound, and the results and comparison with the existing literature presented are novel in the sense that they qualitatively evaluate their online survey results with current and previous work in the domain.

- **Clarity of Relation with Related Work**
  The authors compare their work to two domains which they consider the most similar, specifically, work on automated code comment generation models, and work on code documentation practices in general. The works on automated code comment generation are mostly cited to provide context on what the current state of the research is in the domain, while the documentation practices works are most related in my opinion. Table 4 in the paper presents a neat summary of related works in the documentation domain. Overall, the authors do a good job of comparing their work with related work in the two domains and distinguish their contributions well.

- **Quality of Evaluation and Results**
  Barring the issues with participant selection for the interviews, the authors do an excellent job of evaluating participant responses. The results are presented well, with clean, well-designed figures and tables displaying the results. The findings they presented are backed up with significant survey data and intuitively make sense as well. I found it most useful how the authors individually evaluated and presented results for the different expectations of practitioners for statement, method, and class level comments. This benefits future work in the area, especially those who are developing automated code comment generation models, as it neatly characterizes the desirable and undesirable properties of such a model from the perspective of the end-users. The paper also summarizes very well what existing work has been focused on in terms of model development and how it differed from practitioners' expectations.

- **Ability to Replicate**
  The authors release a replication package that has the interview guide as well as the survey questionnaire which was used. The authors also release the raw survey data and analytical results which were derived from the survey data and presented in the paper. Hence, the same interview guide and online survey questionnaire can definitely be used to replicate this work. However, without more details on the selection criteria for the interviewees and participants and online survey, results may differ.

- **Quality of Presentation**
  The research questions were phrased well, the figures were informative and concise. The paper is organized in a methodological manner and each research question was motivated clearly. The results discussed were clear, coherent, and logical, and backed up with consistent figures. There was not too much technical jargon in the paper which made it very readable and understandable.