

Grant Proposal

Jincheng He

Department of Computer Science
University of Southern California
jinchenh@usc.edu

I. INTRODUCTION

This research mainly focuses on investigating how different purposes of commits impact on software open source software quality. Using data from open-source software, where we can get enough metadata, and tools where we can acquire existing software quality metrics for analysis, we are to evaluate how different changes, the commits developers contribute to the software repository impact the software quality and how to improve the development process for the sake of better software quality.

Furthermore, this research investigates in a more atomic level of how different commits, with respect to their purposes, differ in code pattern and how can we improve the way people code based on the results, consequently contributing to a better code quality.

II. PREVIOUS WORK

In this direction, research has been conducted to investigate when, where, how and what the developers contribute to the projects and how they may impact software quality. However, there is little work focusing on purpose-related categorizations and how different categories may impact software quality.

III. OUR PLAN

A. Stage One

Research has been done to investigate the impact of some behaviors of developers when they contribute to software, especially open source software. However, in the previous research in this area, they haven't reveal the correlation between the change type and the code as well as their impact on the software quality. In this stage, we start with categorization of commit changes in open source software repositories and show their impact on the software quality and validate the categories by prediction models.

B. Stage Two

In this stage, as we have established a categorization, we will be working on how we can improve the quality of the categorization, which means having a more convincing categorization that represents commit changes more accurately. Once this is done. We will investigate the relation between the categorizations and atomic changes in code to reveal whether they have close connection and how the atomic changes impact software quality.

C. Stage Three

Once the categorization and its relation with atomic changes in code are revealed, we will dive into the next stage which is guidelines for developers about how to better contribute to an open-source software (and write code) and a prediction model for predicting categorization.

This direction could succeed since the current techniques in machine learning, statistics, natural language processing will be sufficient to support this research.

IV. BENEFITS

Once it succeeds, we can provide guidelines about how open-source software developers can better contribute to projects. If we get convincing results in the second stage, we will also be able to provide more reliable coding standards which can lead to an improvement of code quality, thus software quality.

V. TIMELINE AND MILESTONES

This research will take 3 to 5 years for these three stages, although this direction may take more than ten years since changing standards is a gradual process. The success of this direction depends on how the applied techniques evolve in the coming years. The midterm milestone could be a reasonable high prediction accuracy while the final exam milestone could be the completion of the new systematic coding standards.

VI. INTELLECTUAL ADVANCEMENT