MIMANSA : Process Mining Software Repositories from Student Projects in an Undergraduate Software Engineering Course

Megha Mittal(megha1243@iiitd.ac.in), Ashish Sureka(ashish@iiitd.ac.in)

Abstract

- Software Engineering is a practice oriented and applied discipline with objectives like:
  - Learn fundamental technical concepts in SE courses.
  - Learn software development processes, teamwork and project management.
  - Exposure to popular SE tools.
- Artefacts like SRS, project plans, test plans, source code are produced at various milestones.
- We present an application of mining three software repositories: team wiki, version control system and issue tracking system in the context of an undergraduate SE course.
- We propose visualizations, metrics and algorithms to provide an insight into practices and procedures followed during project development and to provide a multi-faceted view to the instructor serving as a feedback tool on development process and quality by students.

Research Motivation & Aim

**MOTIVATION**

1. Effective mechanism is required for SE course instructors to:
   - Gain visibility and insights on software development processes followed by student teams.
   - Provide appropriate feedback on process improvement.
2. Our motivation is to develop tools and techniques for solving problems encountered by SE course instructors.

**AIM**

1. To investigate the application of process mining event log data generated from Wiki, VCS and ITS.
2. To define new process quality metrics and visualization.
3. To examine the application of existing process quality metrics.

Research Contributions & Experimental Dataset

**CONTRIBUTIONS**

1. Analyzing Wiki, VCS and ITS activity event logs from a process mining perspective.
2. Visualizations and metrics providing feedback on various aspects such as:
   - Workload distribution
   - Consistency in contributions
   - Quality of commit messages
   - Efficiency of bug fixing process
3. Dataset

**DATASET**

<table>
<thead>
<tr>
<th>No. of Students</th>
<th>Total Commits</th>
<th>ITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>Wiki = 1167</td>
<td>Bugs Reported = 482</td>
</tr>
<tr>
<td>99</td>
<td>VCS = 925</td>
<td>Bugs Resolved = 404</td>
</tr>
</tbody>
</table>

Results

**Team Wiki**

- Figure 1a shows consistency in development activity of various teams across timeline.
- Figure 1b represents behavior of teams near to the deadline.
- Figure 1c helps in determining the distribution of workload across members of a team.

**VCS**

- Figure 2a shows the distribution of workload across members of a team.
- Figure 2b reflects intra-team behavior near to the deadline.
- 2 metrics are proposed: "Developer-Component Entropy" and "Component-Developer Entropy" to quantify developer wise component and component wise developer contribution.
- It serves as an indicator of extent of co-development at component level.

**ITS**

- Figure 3b portrays the related results for 2 algorithms implemented to detect the compliance of observed process with defined design-time process.
- Time taken to repair bugs is depicted in Figure 3c.
- Figure 3d is a histogram which shows the distribution of bugs with different priority in various components of a project.
- Bug Fixing Score represented in Figure 3e & Figure 3f is proposed to indicate the efficiency in bug fixing process.
- Figure 3f reveals the bugs activity process of the student projects.

Conclusion

- Mining activity logs generated by Wiki, VCS and ITS presents useful insights to the course instructor.
- Visualizations and metrics obtained through activity log helps in characterizing:
  - Variation in quality of commit messages & consistency in activity.
  - Variations in degree of distribution of workload & process compliance & bug fixing quality.
- Proposed framework provides effective visibility to the instructor for improving academic outcome and learning methodology.

References