Identifying Software Process Management Challenges: Survey of Practitioners in a Large Global IT Company

Monika Gupta, Ashish Sureka, Srinivas Padmanabhuni, Allahbaksh Asadullah

monikag@iiitd.ac.in

PhD Scholar
Introduction

Fast changing, hyper-competitive, interconnected global market place

Process Management Challenges

Intuition Insufficient

Process Mining
Introduction

Fast changing, hyper-competitive, interconnected global market place

Process Management Challenges

“If one cannot measure it, one cannot improve it”
Introduction

Process Mining:

• Extract knowledge from event logs recorded by information systems [1].

• Event logs (e.g. transaction logs) with four fields:

<table>
<thead>
<tr>
<th>CaseID</th>
<th>Event</th>
<th>Timestamp</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

• Tools and framework for process mining:
  ➢ ProM ¹ (Open Source)
  ➢ Disco ² (Commercial)

Software Repositories:

- Vast data generated and archived in multiple information systems. For example: Issue tracking system, peer code review system, version control system

- Uncover interesting and actionable information for process improvement.

- Diverse applications such as:
  - uncovering runtime process model [2, 3],
  - discovering process inconsistencies and inefficiencies [2, 4], and
  - provide CMMI assessors with relevant information to support process assessment and improvement [5].

Research Aim

• Identify the process challenges encountered by managers that can be addressed by novel applications of process mining, by conducting survey and an interview study.

• Investigate importance, benefits and challenges of solving those problems.
• Conduct two-phase online survey and an interview study at Infosys Limited\(^3\), a large, global, software company having more than 170000 software professionals.

• It is a CMM 5 company with well defined processes in place.

• We adopt methodology by Begel et al. with some contextual modifications \([6]\).

Research Methodology

Survey and in-depth interviews
Identify process challenges encountered while managing projects and benefits of solving the problem

Purposive Sampling
Project Management: TL, PM, SPM

46 Participants
130 Items

Generic Problem Formulation
Group similar items and describe each group with brief problem statement
[30 unique problems for 111 items]

classification based on Process Mining Perspective [8 Categories]

Control Analysis
Organizational Analysis
Conformance Analysis
Enhancement Analysis
Preventive Analysis
Comparative Analysis
Performance Analysis
Case Analysis

Importance Analysis Survey
Send 30 brief problem statements to distinct practitioners with same characteristics and ask them to indicate how important is it to have process mining team solve this problem [Essential / Worthwhile / Unimportant / Unwise / I don’t understand]
[1262 ratings by 43 participants]

Net Importance Metric
Categorize problems to different importance level based on the value of metric derived from survey ratings

Commonality and Importance Analysis
Categorize problems based on joint values of frequency in first survey and importance in second survey
Research Methodology: Phase I

- Pilot study and survey design:
  - Definition of process mining along with some examples.

  *Please list up to three problems that you encounter during the software development process management and you would like process mining team to solve for you. Also, mention the benefits you will have if the problem is solved.*

- Participants and data collection

- Generic problem statement formulation:
  - Each statement has first component as task followed by cause and benefit.

  E.g., During issue resolution, detection and analysis of PING-PONG patterns due to bug tossing between developers to reduce resolution time
Research Methodology: Classification

• **Control Analysis**: Discovery of actual process from the execution logs and analyze them to better understand the actual runtime process.

• **Organizational Analysis**: Analyze individuals, team coordination and interactions with the process.

• **Conformance Analysis**: Detect inconsistencies with the defined process and flag anomalies.

• **Enhancement Analysis**: Improve or extend the existing process model using information about actual process recorded in the event log.

• **Preventive Analysis**: Focus is on using insights derived from event log mining for better planning and process design even before the execution happens.

• **Comparative Analysis**: Compare multiple processes from various perspectives along different dimensions such as experience of actors and turnaround time.

• **Performance Analysis**: Identify inefficiencies and imperfections such as most time consuming activities, rework, cause of delay, bottlenecks in the process.

• **Case Analysis**: Focuses on properties of a case such as its path and people working on it.
Net Importance Metric

How important is it to have a process mining team solve this problem?

+ve: [Essential | Worthwhile]
-ve: [Unimportant | Unwise]

\[
NIM = \frac{[(C(E) + 0.75 \times C(W) - C(UW) - 0.75 \times C(UI))]}{\text{No. of responses}}
\]

<table>
<thead>
<tr>
<th>1</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
<th>0.6</th>
<th>0.5</th>
<th>0.4</th>
<th>0.3</th>
<th>0.2</th>
<th>0.1</th>
</tr>
</thead>
</table>

0 to -1

Very important | Important | Less important | Not worth solving
Commonality and Importance

On the basis of *total members in group of each problem statement (Common or Uncommon)* along with the *NIM (Important or Unimportant)*:

1. Common and Important
2. Common and Unimportant
3. Uncommon and Important
4. Uncommon and Unimportant
Commonality and Importance

On the basis of *total members in group of each problem statement (Common or Uncommon)* along with the *NIM (Important or Unimportant)*:

1. **Common and Important**
2. Common and Unimportant
3. Uncommon and Important
4. Uncommon and Unimportant
## Gaps and Challenges

<table>
<thead>
<tr>
<th>Problem</th>
<th>Tool</th>
<th>Research</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify BOTTLENECKS and inefficiencies causing delay to take remedial actions and have better estimation in future.</td>
<td>⭐</td>
<td></td>
<td>[7, 8, 9]</td>
</tr>
<tr>
<td>Enable early detection and PREVENTION OF DEFECTS instead of fixing them during the later stage by understanding patterns of escaped defects</td>
<td></td>
<td>⭐</td>
<td>[10]</td>
</tr>
<tr>
<td>Avoid putting efforts on LESS SIGNIFICANT ACTIVITIES by identifying redundant or unnecessary steps of process</td>
<td>⭐</td>
<td>⭐</td>
<td>[11]</td>
</tr>
<tr>
<td>Inspect REOPENED issues to identify the root cause and recommend verification for future issues based on learning from issues reopened in the past</td>
<td>⭐</td>
<td>⭐</td>
<td>[7,12,13]</td>
</tr>
<tr>
<td>Need for efficient TASK ALLOCATION mechanism by considering individuals’ skills, interests, and expertise as well as team compatibility for better utilization of resources</td>
<td>⭐</td>
<td></td>
<td>[14]</td>
</tr>
</tbody>
</table>

Gaps and Challenges

- Data unavailability
- Noisy data
- Integrating heterogeneous IS
- Different perspectives
- Concept drift
- Operational support
- Corrective actions
Study Limitations

- Participants of same organization are surveyed, organizational culture may create bias.

- Examples given to illustrate process mining applications in the first survey can shape the thoughts of respondents.

- Few problems from requirements gathering phase.

- Most problems are inferred from the long statements received in responses.

- Importance survey responses may be influenced by individual differences among the participants.
Research Contributions

- A list of 30 process management challenges covering diverse perspectives (8 categories).

- We enlist the benefits of solving problems from each category to motivate both, researchers and industry.

- Identify more important problems by calculating net importance metric using responses of second survey. Therefore, focus on more important problems.

- Investigation of most important and frequent problems to identify the gap and challenges involved in solving those problems.
Acknowledgement

- Supported by Prime Minister’s fellowship for PhD Students (awarded to first author) by SERB, CII and industry sponsor, Infosys Limited.

- Thanks to the participants from two survey.

- I acknowledge Microsoft for the travel support.

- Thanks for the SIGSOFT CAPS award.
Surveys, collected responses and consolidated lists for both the phases are available:

https://github.com/Mining-multiple-repos-data/QualitativeStudy
List of 30 problems
<table>
<thead>
<tr>
<th>ID</th>
<th>Problem Statement</th>
<th>+ve</th>
<th>-ve</th>
<th>C(PI)</th>
<th>NIM</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identify BOTTLENECKS and inefficiencies causing delay to take remedial actions and have better estimation in future.</td>
<td>42</td>
<td>0</td>
<td>7</td>
<td>0.92</td>
<td>Performance</td>
</tr>
<tr>
<td>2</td>
<td>Enable early detection and PREVENTION OF DEFECTS instead of fixing them during the later stage by understanding patterns of escaped defects.</td>
<td>41</td>
<td>0</td>
<td>9</td>
<td>0.91</td>
<td>Case</td>
</tr>
<tr>
<td>3</td>
<td>Avoid putting efforts on LESS SIGNIFICANT ACTIVITIES by identifying redundant or unnecessary steps of process.</td>
<td>41</td>
<td>1</td>
<td>5</td>
<td>0.89</td>
<td>Enhancement</td>
</tr>
<tr>
<td>4</td>
<td>Automatic ADAPTATION OF PROCESS according to different project specifications that is, design process based on knowledge of similar successful projects instead of selecting process only on the basis of experience.</td>
<td>43</td>
<td>0</td>
<td>2</td>
<td>0.85</td>
<td>Preventive</td>
</tr>
<tr>
<td>5</td>
<td>Inspect REOPENED issues to identify the root cause and recommend verification for future issues based on learning from issues reopened in the past.</td>
<td>41</td>
<td>1</td>
<td>6</td>
<td>0.84</td>
<td>Case</td>
</tr>
<tr>
<td>6</td>
<td>Need for efficient TASK ALLOCATION mechanism by considering individuals' skills, interests, and expertise as well as team compatibility for better utilization of resources.</td>
<td>40</td>
<td>2</td>
<td>13</td>
<td>0.83</td>
<td>Organizational</td>
</tr>
<tr>
<td>7</td>
<td>Approvals are part of software development lifecycle (SDLC) and need better management. Design a process for seamless approvals to reduce delays.</td>
<td>40</td>
<td>3</td>
<td>2</td>
<td>0.79</td>
<td>Preventive</td>
</tr>
<tr>
<td>8</td>
<td>Mechanism for CONTINUOUS PROCESS EVOLUTION based on best practices of individuals who exercise the process. Therefore, we improve process by encouraging on-the-job learnings of people rather than dependence on process designers.</td>
<td>39</td>
<td>2</td>
<td>6</td>
<td>0.76</td>
<td>Enhancement</td>
</tr>
<tr>
<td>9</td>
<td>Improve effectiveness of CODE REVIEW PROCESS AND STANDARDIZATION by redesigning check list and updating code analyzers based on the defects reported during testing.</td>
<td>39</td>
<td>2</td>
<td>7</td>
<td>0.74</td>
<td>Enhancement</td>
</tr>
<tr>
<td>10</td>
<td>Facilitate BETTER INTEGRATION between different silos by reconstructing the process thus, reduce rework happening due to differences in understanding.</td>
<td>36</td>
<td>3</td>
<td>6</td>
<td>0.71</td>
<td>Organizational</td>
</tr>
<tr>
<td>11</td>
<td>Handle CHANGING TEAMS seamlessly by analyzing interaction pattern between team members and team dynamics.</td>
<td>38</td>
<td>2</td>
<td>2</td>
<td>0.70</td>
<td>Organizational</td>
</tr>
<tr>
<td>12</td>
<td>Design a technique to TRACE ADHERENCE WITH REQUIREMENTS and adapt process automatically with changing requirements.</td>
<td>36</td>
<td>3</td>
<td>2</td>
<td>0.69</td>
<td>Conformance</td>
</tr>
<tr>
<td>13</td>
<td>PEOPLE VS PROCESS: Identify which factor contributed to what extent towards the success and failure of project.</td>
<td>39</td>
<td>4</td>
<td>2</td>
<td>0.69</td>
<td>Comparative</td>
</tr>
<tr>
<td>14</td>
<td>Simplify tracking of the whole REVIEW PROCESS to identify inefficiencies quickly.</td>
<td>37</td>
<td>4</td>
<td>2</td>
<td>0.67</td>
<td>Control</td>
</tr>
<tr>
<td>15</td>
<td>During issue resolution, detection and analysis of PING-PONG patterns due to bug tossing between developers to reduce resolution time.</td>
<td>36</td>
<td>3</td>
<td>3</td>
<td>0.67</td>
<td>Control</td>
</tr>
<tr>
<td>16</td>
<td>Improve PROJECT PLANNING AND ESTIMATION by complimenting it with the insights derived from event log mining of similar projects done in the past.</td>
<td>38</td>
<td>5</td>
<td>3</td>
<td>0.66</td>
<td>Preventive</td>
</tr>
<tr>
<td>17</td>
<td>Investigate the LEAD TIME for issue resolution by analyzing issue resolution process from TIME PERSPECTIVE. Therefore, we can have timely resolutions.</td>
<td>37</td>
<td>5</td>
<td>2</td>
<td>0.64</td>
<td>Performance</td>
</tr>
<tr>
<td>18</td>
<td>Design of more meaningful QUALITY METRICS by understanding runtime process practices to precisely identify the scope of improvement.</td>
<td>35</td>
<td>5</td>
<td>3</td>
<td>0.63</td>
<td>Enhancement</td>
</tr>
<tr>
<td>19</td>
<td>Equip novice with the KNOWLEDGE OF EXPERIENCED PRACTITIONERS by associating efficiency of adopted process with the experience of practitioners.</td>
<td>36</td>
<td>4</td>
<td>4</td>
<td>0.63</td>
<td>Comparative</td>
</tr>
<tr>
<td>20</td>
<td>Facilitate in-depth understanding of point where things went wrong by deriving and understanding actual process at a MORE GRANULAR LEVEL.</td>
<td>36</td>
<td>5</td>
<td>1</td>
<td>0.63</td>
<td>Control</td>
</tr>
<tr>
<td>21</td>
<td>Continuous check on SCHEDULE ADHERENCE is a complex task. Design an automated way to track and preempt if any deviations.</td>
<td>36</td>
<td>6</td>
<td>1</td>
<td>0.60</td>
<td>Conformance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Relate bugs with the <strong>ACTUAL STAGE OF INCEPTION</strong> by understanding issue resolution life cycle along with other relevant attributes.</td>
<td>34</td>
<td>5</td>
<td>3</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Uncover <strong>DEVIATIONS</strong> between the actual process followed by the team and the defined process, their cause, impact on overall outcome and identify the set of people exhibiting more deviations.</td>
<td>36</td>
<td>6</td>
<td>8</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td><strong>INTEGRATE MULTIPLE STANDALONE SYSTEMS</strong> used during SDLC to solve data and process redundancy challenges, and obtain a holistic view.</td>
<td>34</td>
<td>5</td>
<td>5</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Analyze code review life cycle to identify developers who are not reviewing their code properly before they submit it for external review and the deviations from defined checklist. It will help take corrective actions and reduce defects during testing phase.</td>
<td>33</td>
<td>7</td>
<td>2</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Mechanism to manage and keep track of SVN check-ins process that is, activity sequence for merging and branching as it is very important and can help take informed decisions.</td>
<td>27</td>
<td>5</td>
<td>1</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Capture the <strong>ACTUAL STATUS</strong> (reality) of project or any task by discovering runtime process from event logs instead of current manual practice.</td>
<td>31</td>
<td>8</td>
<td>1</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Trace the complete flow and understand <strong>WHICH ISSUE LEADS TO WHICH CODE CHANGE</strong> by analyzing event logs for issue resolution in combination with the code modified in VCS.</td>
<td>31</td>
<td>9</td>
<td>1</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Perform <strong>COMPARATIVE ANALYSIS OF TICKETS</strong> along dimensions such as component, owner (analyst), reporter, type such as performance, regression and security, final resolution such as duplicate, invalid and fixed, and turnaround time to derive useful insights for improvement.</td>
<td>32</td>
<td>10</td>
<td>7</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Identify the group of <strong>ACTIVE VS INACTIVE CONTRIBUTORS, GENERALIST VS SPECIALIST</strong> by analyzing performance of individuals participating in the process.</td>
<td>24</td>
<td>17</td>
<td>2</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

**Control**

**Conformance**

**Preventive**

**Comparative**

**Organizational**
Control Perspective: Benefits

- Minimize unnecessary efforts and delay by detecting the point where process went wrong and its overall impact to take timely corrective actions.
- Reduce rework as the process reality visualized in more reliable way to quickly identify the inefficiencies and phases more prone to error.