Relating User Session Clusters to Dynamic Web Application Behavior

Sreedevi Sampath, Sara Sprenkle
University of Delaware

Emily Gibson, Lori Pollock
University of Delaware

Amie Souter
Drexel University

Overview

- Cluster user sessions, us, (a sequence of URLs), by similar URLs using a technique called concept analysis
  - Cluster 1: us1, us2
  - Cluster 2: us4, us5, us6
- Test suite reduction: our heuristic for reduction selects one user session from each cluster such that sessions in the reduced suite cover all URLs of the application

Hypotheses

- A cluster of user sessions exhibits similar program characteristics
  - Program code covered/faults detected by a user session, us, in cluster
  - Measure of similar program characteristics is the common program covered/faults detected by all the sessions in cluster
- Our heuristic selects sessions from clusters that cover different program characteristics
  - Cluster 1
  - Cluster 2

Research Questions

RQ1: How do sessions clustered based on URLs relate to program covered/faults detected by the sessions?
RQ2: Does our heuristic select sessions from clusters that cover different program code and detect different faults?

Subject Web Applications and Results

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Book</th>
<th>CPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>Methods</td>
<td>385</td>
<td>172</td>
</tr>
<tr>
<td>NLOC</td>
<td>7791</td>
<td>9298</td>
</tr>
<tr>
<td>Faults</td>
<td>40</td>
<td>86</td>
</tr>
<tr>
<td>Sessions</td>
<td>125</td>
<td>251</td>
</tr>
</tbody>
</table>

G1: RQ1, program coverage

G2: RQ1, faults detected

G3: RQ2, program coverage

Summary of Results

1. As user sessions become more similar in terms of URLs, similarity in their dynamic program characteristics increases (G1, G2)
2. Our heuristic selects user sessions from clusters that are different in their program coverage and fault detection (G3)