IncBL: Incremental Bug Localization
Zhou Yang, Jieke Shi, Shaowei Wang and David Lo

Overview

Motivation:
- Software repositories evolve, and new bug reports emerge.
- Current information retrieval-based bug localization (IRBL) tools usually need to re-construct the model to adapt to changes in bug reports and codebases.
- Some simple IR models (e.g., VSM [1]) can be updated incrementally while no such work is about advance models (e.g. rVSM in BugLocator [2]).
- No open-source bug localization tool adopts incremental update strategies.

Contribution:
We present IncBL, the first open-source bug localization tool updating models incrementally. It reduces running time by 77.79% on average while maintaining a competitive level of accuracy. IncBL has been integrated as a GitHub App and can also be deployed locally.

Incremental Updates

The VSMs can be incrementally updated by deletion/update/addition operations in term-document matrix and document frequency matrix. The new df and idf can be computed by:

\[ d_{\text{new}}^f(w) = d_{\text{old}}^f(w) + \lfloor \text{sign}(A_{m}^{\text{new}}(w)) - \text{sign}(A_{m}^{\text{old}}(w)) \rfloor \]

\[ idf_{\text{new}}^d(w) = idf_{\text{old}}^d(w) + \log \left( \frac{M + \Delta M}{M} \right) \]

This figure illustrates how term-document matrix (A), document frequency (df) and document representation (D) can be updated incrementally.

Pipeline and method:
IncBL is based on BugLocator [2]. The workflow is as follows:
- **Step 1.** Processing code files: pre-processing and vectorizing texts using Vector Space Mode (VSM).
- **Step 2.** Processing bug reports: pre-processing texts and computing the similarity (SimiScore) between new and past fixed bug reports.
- **Step 3.** Localizing buggy files: computing the similarity between code files and bug reports, then combining it with SimiScore to find relevant buggy files.


<table>
<thead>
<tr>
<th></th>
<th>BugLocator</th>
<th>IncBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running Time</td>
<td>1x</td>
<td>4.50x (on average)</td>
</tr>
<tr>
<td>MRR</td>
<td>0.328</td>
<td>0.331</td>
</tr>
</tbody>
</table>

Experiment Results

This box-plot shows the ratio of IncBL running time over the running time of BugLocator.

Usage Instruction

**Step 1.** Install GitHub App.
**Step 2.** Issue a bug report.
**Step 3.** Get lists of potentially buggy files.

Reference