Bug Report Classifier

Simon Curty
Haidar Osman
The Problem

It’s Not a Bug, It’s a Feature: How Misclassification Impacts Bug Prediction

Kim Herzig
Saarland University
Saarbrücken, Germany
herzig@cs.uni-saarland.de

Sascha Just
Saarland University
Saarbrücken, Germany
just@st.cs.uni-saarland.de

Andreas Zeller
Saarland University
Saarbrücken, Germany
zeller@cs.uni-saarland.de

Abstract—In a manual examination of more than 7,000 issue reports from the bug databases of five open-source projects, we found 33.8% of all bug reports to be misclassified—that is, rather than referring to a code fix, they resulted in a new feature, an update to documentation, or an internal refactoring. This misclassification introduces bias in bug prediction models, confusing bugs and features: On average, 39% of files marked as defective actually never had a bug. We discuss the impact of this misclassification on earlier studies and recommend manual data validation for future studies.

Index Terms—Mining software repositories, bug reports, data

<table>
<thead>
<tr>
<th>PROJECT DETAILS.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Maintainer</td>
</tr>
<tr>
<td>HTTPClient</td>
</tr>
<tr>
<td>Jackrabbit</td>
</tr>
<tr>
<td>Lucene-Java</td>
</tr>
<tr>
<td>Rhino</td>
</tr>
<tr>
<td>Tomcat5</td>
</tr>
</tbody>
</table>
The Problem

- Issues
- Bug reports

- Correctly classified
- Misclassified
- Actually a bug
- No bug

---

It's Not a Bug, It's a Feature:
How Misclassification Impacts Bug Prediction
Is this a bug?

Categories:
- Bug
- Feature Request
- Improvement
- Documentation
- Refactoring
- Something else?
How can we tell?

Anatomy of an issue report:

- Title/Summary
- Meta information
- Description
- Comments
- Other information
How can we tell?

Categories:
• Bug
• Feature Request
• Improvement
• Documentation
• Refactoring
• Something else?
List with issue IDs

ID, CLASSIFIED, TYPE
HTTPCLIENT-569, BUG, BUG
HTTPCLIENT-366, BUG, BUG
HTTPCLIENT-302, BUG, BUG
HTTPCLIENT-104, IMPROVEMENT, BUG
HTTPCLIENT-85, BUG, BUG
HTTPCLIENT-196, BUG, BUG

Data flow

Importer → Retriever

Retriever → Parser

Parser → Machine Learning Model

Bugzilla

JIRA
Machine Learning Model

• In what category does a given issue belong?
  -> Multiclass Classification Problem

• Set of already classified issues
  -> Supervised Learning
How does it work?

Text representation

```
Text representation
```

Vectorised

```
0.1, 0.239708674365, 0.5, 0.751758983214
0.1, 0.807715642339, 0.77858522143023
1.8, 0.43184023707231, 0.2287794674931
0.8, 0.45127607441732, 0.66957414266103
0.1, 0.525379503934, 0.97595541949688
0.2, 0.6244609896065, -3.5712172457417
-1.9, 0.6798409028372, 0.3227985198413
0.1, 0.76462111109025, -0.59085119089674
-1.9, 0.54075220047797, -0.24549458795326
-1.4, 0.84223027334724, 0.13685386209334
0.2, 1.1268571110583, 1.09465523562788
0.8, 0.663614264813507, 0.401367979403406
0.0, 0.97699736770473, -0.884055062855628
0.9, 0.77224622666051, 0.98564721847961
1.0, 0.27265319806747, 1.24216376252738
0.8, 1.1713169057799, 0.384626892844234
0.0, 0.8256098415523, 2.98009468969564
0.8, 0.42619438668344, -0.24674949684645
1.0, 0.54316914030598, -0.55947977948386
1.0, 0.40519304850988, -0.65785556737486
1.0, 0.19578665680018, 0.25442612919349
-1.8, 0.6778473681861, -3.27270683674903
1.8, 0.47807941278742, 0.96106575204648
0.1, 0.95410191245293, -1.41699629741271
1.1, 0.8505819938087, 0.88052099558288
1.9, 0.85201428822543, -1.98092782566354
-0.2, 0.61930621246295, -0.91735995764943
-1.9, 0.57705211533498, 0.85664787783597
1.0, 1.07213170468168, 0.64735382888642
1.3, 1.2114266188492, 0.56413156226803
1.0, 0.830715752106, 0.4989367949914
0.2, 0.67378400738832, -0.261285121872568
-1.8, 0.33067729685859, -0.57998129578977
```

Neuronal network

```
```

→ output
The angle $\theta$ is the similarity between two vectors.

Vector operations:

$\text{Library} - \text{Books} = \text{Hall}$

Logic analogies:

$\text{Beijing is to China what [Rome, Florence, Venice] is to Italy}$
Accuracy

- Training set
- Test set

Cosine Similarities from a document:

- BUG: 0.6
- IMPROVEMENT: 0.1
- RFE: 0.3
- REFACTORING: 0.5
First Results

Category accuracies:

• Bug: 12%
• Improvement: 21%
• Refactoring: 18%
• Feature Request: 97%
Problems

• Are the classifications in the training set correct/coherent?

• Accuracy and recall of the model

• How should it be usable?
  • Webservice?
  • Browser plugin?