

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

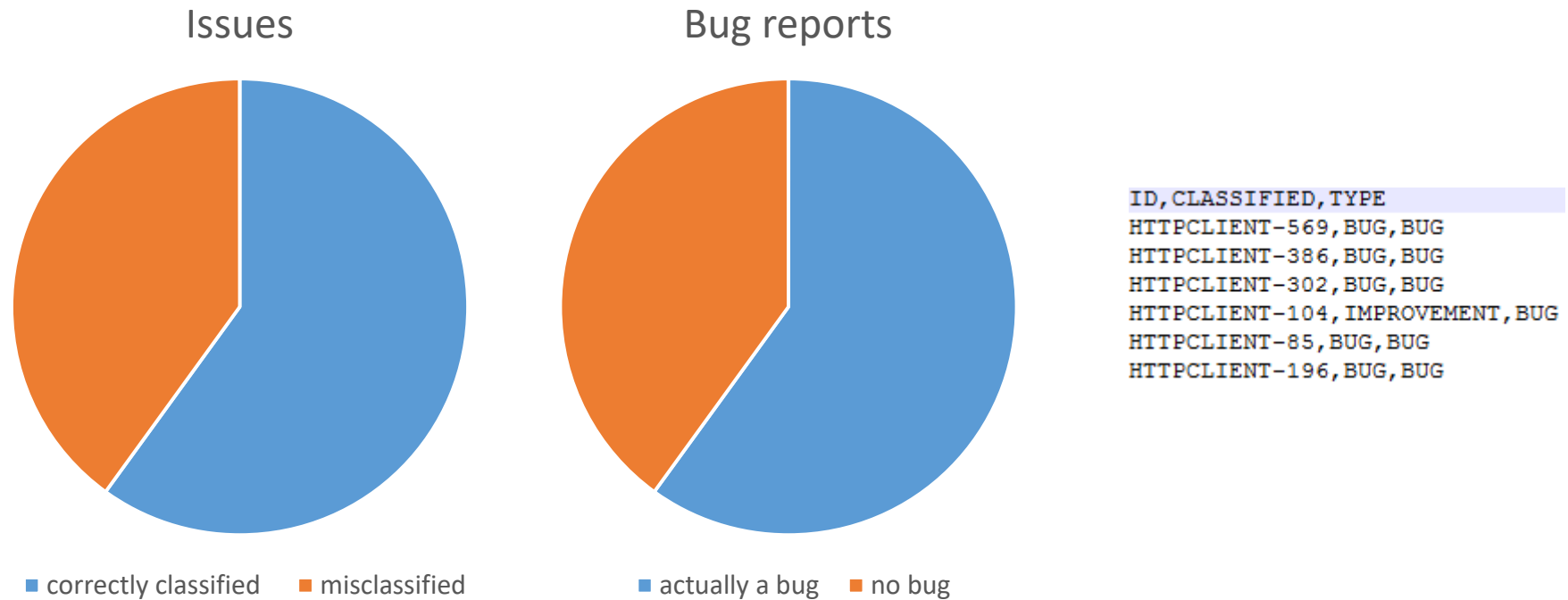
Abstract—In a manual examination of more than 7,000 issue reports from the bug databases of five open-source projects, we found 33.8% of all bug reports to be *misclassified*—that is, rather than referring to a code fix, they resulted in a new feature, an update to documentation, or an internal refactoring. 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 I
PROJECT DETAILS.

	Maintainer	Tracker type	# reports
HTTPClient	APACHE	Jira	746
Jackrabbit	APACHE	Jira	2,402
Lucene-Java	APACHE	Jira	2,443
Rhino	MOZILLA	Bugzilla	1,226
Tomcat5	APACHE	Bugzilla	584

The Problem



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?



Agile Board

Details

Type:	Bug	Status:	CLOSED
Priority:	Major	Resolution:	Fixed
Affects Version/s:	2.0 Alpha 1	Fix Version/s:	2.0 Beta 1
Component/s:	HttpClient (classic)		
Labels:	None		
Environment:	Operating System: All Platform: All		
Bugzilla Id:	https://issues.apache.org/bugzilla/show_bug.cgi?id=16421		

Description

The commons release process (<http://jakarta.apache.org/commons/releases.html>) is a good starting place, but is out of date. When we do our own releases, there are some other steps particular to maven, the test-local and the webapp tests that must be documented.

Activity

All **Comments** Work Log History Activity Transitions

▼ [Jeff Dever](#) added a comment - 26/Jan/03 15:22

The first pass is in there, I'll refine it over time.
<http://jakarta.apache.org/commons/httpclient/releases.html>
Still need to link to it from the nav bar.

▼ [Jeff Dever](#) added a comment - 26/Jan/03 23:11

I'll call this done. The release process is pretty complete now, and can be evolved over time.

How can we tell?

Anatomy of an issue report:

- Title/Summary
- Meta information
- Description
- Comments
- Other information

HttpComponents HttpClient / HTTPCLIENT-144

Add a document describing the HttpClient release process

Agile Board

Details

Type:	<input checked="" type="checkbox"/> Bug	Status:	CLOSED
Priority:	↑ Major	Resolution:	Fixed
Affects Version/s:	2.0 Alpha 1	Fix Version/s:	2.0 Beta 1
Component/s:	HttpClient (classic)		
Labels:	None		
Environment:	Operating System: All Platform: All		
Bugzilla Id:	https://issues.apache.org/bugzilla/show_bug.cgi?id=16421		

Description

The commons release process (<http://jakarta.apache.org/commons/releases.html>) is a good starting place, but is out of date. When we do our own releases, there are some other steps particular to maven, the test-local and the webapp tests that must be documented.

Activity

All **Comments** Work Log History Activity Transitions

▼ [Jeff Dever](#) added a comment - 26/Jan/03 15:22

The first pass is in there, I'll refine it over time.
<http://jakarta.apache.org/commons/httpclient/releases.html>

Still need to link to it from the nav bar.

▼ [Jeff Dever](#) added a comment - 26/Jan/03 23:11

I'll call this done. The release process is pretty complete now, and can be evolved over time.

How can we tell?

Categories:

- Bug
- Feature Request
- Improvement
- Documentation
- Refactoring
- Something else?

Agile Board

Details

Type:	<input checked="" type="checkbox"/> Bug	Status:	CLOSED
Priority:	↑ Major	Resolution:	Fixed
Affects Version/s:	2.0 Alpha 1	Fix Version/s:	2.0 Beta 1
Component/s:	HttpClient (classic)		
Labels:	None		
Environment:	Operating System: All Platform: All		
Bugzilla Id:	https://issues.apache.org/bugzilla/show_bug.cgi?id=16421		

Description

The commons release process (<http://jakarta.apache.org/commons/releases.html>) is a good starting place, but is out of date. When we do our own releases, there are some other steps particular to maven, the test-local and the webapp tests that must be documented.

Activity

All **Comments** Work Log History Activity Transitions

▼ [Jeff Dever](#) added a comment - 26/Jan/03 15:22

The first pass is in there, I'll refine it over time.
<http://jakarta.apache.org/commons/httpclient/releases.html>
Still need to link to it from the nav bar.

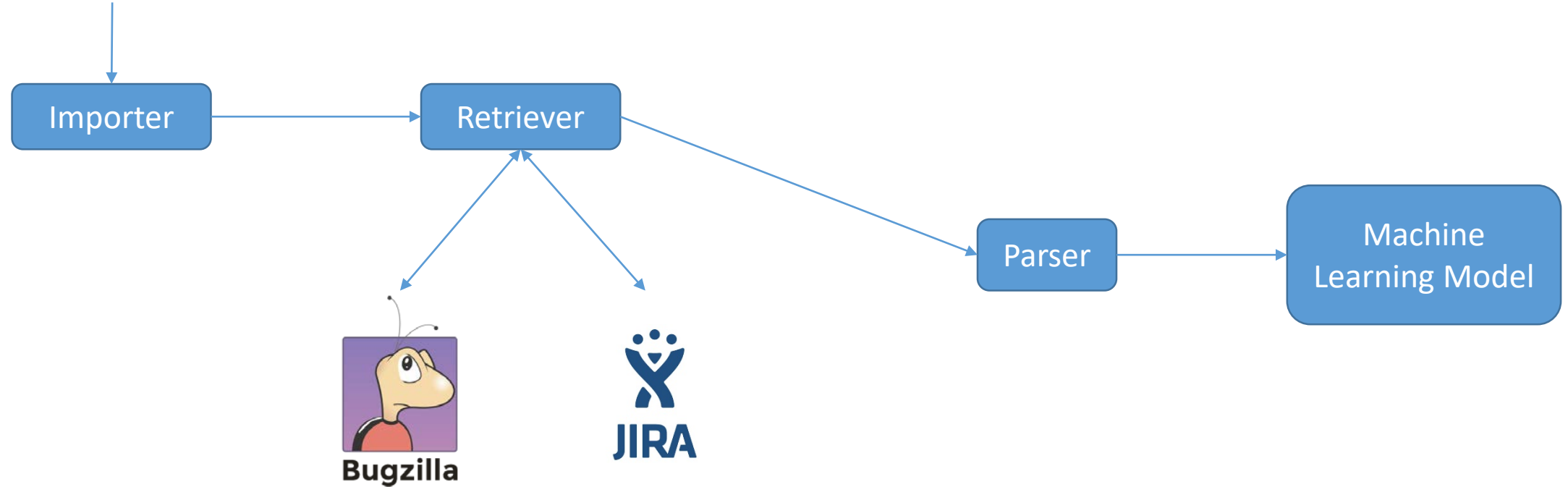
▼ [Jeff Dever](#) added a comment - 26/Jan/03 23:11

I'll call this done. The release process is pretty complete now, and can be evolved over time.

List with issue IDs

```
ID, CLASSIFIED, TYPE  
HTTPCLIENT-569, BUG, BUG  
HTTPCLIENT-386, BUG, BUG  
HTTPCLIENT-302, BUG, BUG  
HTTPCLIENT-104, IMPROVEMENT, BUG  
HTTPCLIENT-85, BUG, BUG  
HTTPCLIENT-196, BUG, BUG
```

Data flow



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

```
INTELLIGENT-MAILING-56
How to commons logging

Commons-logging was derived from
attactant.log, still using the old logging
which should be removed. Some complaints on
mailing list about setting up
commons-logging: should be single and well
documented.

Can I send in a patch?

As this one touches practically every file,
and changes a dependency
relationship and configuration settings,
you'll have to be careful with this
one. You'll likely have to modify build.xml,
build.properties.sample and some
documentation files as well.
I'm working on a few other changes that
should make the build dependencies
single to handle as well, so you might want
to hold off for that. Feel free to
submit that bug to yourself, or I'm sure it
will be a high priority feature.
You may also wish to consider the other
logging bug in align all the usage of
logging levels to be consistent (should be
submitted as a separate patch).
There might also want to get some discussion
going on the mailing list for the
requirements for what the policy should be:
http://nagaya.apache.org/hogzilla/
show_bug.cgi?id=18860

Patch applied to local source tree for
testing. Note called for removal of the
old attactant.log package as it is no longer
used.

Patch provided by DMS applied.

INTELLIGENT-MAILING-56
How to commons logging

Commons-logging was derived from
attactant.log, still using the old logging
which should be removed. Some complaints on
mailing list about setting up
commons-logging: should be single and well
documented.

Can I send in a patch?

As this one touches practically every file,
and changes a dependency
relationship and configuration settings,
you'll have to be careful with this
one. You'll likely have to modify build.xml,
build.properties.sample and some
documentation files as well.
I'm working on a few other changes that
should make the build dependencies
single to handle as well, so you might want
to hold off for that. Feel free to
submit that bug to yourself, or I'm sure it
will be a high priority feature.
You may also wish to consider the other
logging bug in align all the usage of
logging levels to be consistent (should be
submitted as a separate patch).
There might also want to get some discussion
going on the mailing list for the
requirements for what the policy should be:
http://nagaya.apache.org/hogzilla/
show_bug.cgi?id=18860

Patch applied to local source tree for
testing. Note called for removal of the
old attactant.log package as it is no longer
used.

Patch provided by DMS applied.

INTELLIGENT-MAILING-56
How to commons logging

Commons-logging was derived from
attactant.log, still using the old logging
which should be removed. Some complaints on
mailing list about setting up
commons-logging: should be single and well
documented.

Can I send in a patch?

As this one touches practically every file,
and changes a dependency
relationship and configuration settings,
you'll have to be careful with this
one. You'll likely have to modify build.xml,
build.properties.sample and some
documentation files as well.
I'm working on a few other changes that
should make the build dependencies
single to handle as well, so you might want
to hold off for that. Feel free to
submit that bug to yourself, or I'm sure it
will be a high priority feature.
You may also wish to consider the other
logging bug in align all the usage of
logging levels to be consistent (should be
submitted as a separate patch).
There might also want to get some discussion
going on the mailing list for the
requirements for what the policy should be:
http://nagaya.apache.org/hogzilla/
show_bug.cgi?id=18860
```

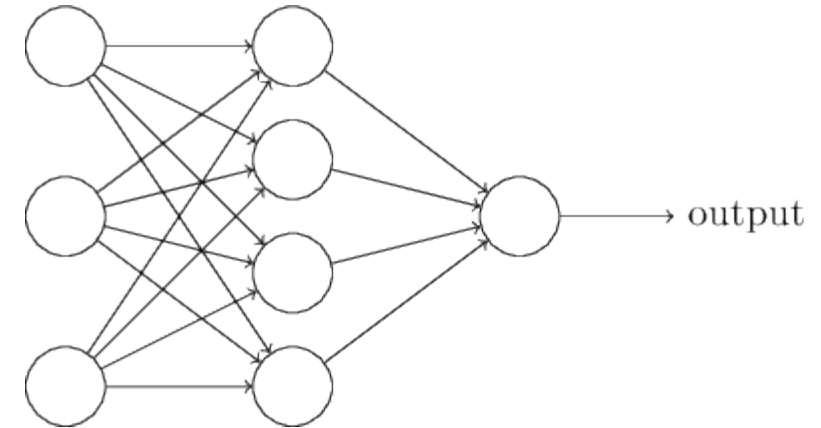


Vectorised

```
1, -7.1239700674365, -5.05175898010314
0, 1.80771566423302, 0.770505522143023
1, 8.43184823707231, -4.2287794074931
0, 0.451276074541732, 0.669574142606103
0, 1.52519959303934, -0.953055551414968
0, 2.6244640096065, -3.57517214574177
1, -9.67984097020372, 2.43273985198413
0, 1.76662111169025, -0.599065119089674
1, -9.54072620047797, -2.44549459765326
1, -4.04223072334724, 9.13685386209343
0, 2.12168571110583, 1.09436523562788
0, -0.683614264815307, 0.401363797043046
0, -0.976997367730473, -0.594055002855628
0, -2.77224622666051, 3.94657421847961
1, -9.272053185066747, 1.24216037652738
0, -1.71173169597796, 0.584266032844234
0, -2.15368098415523, 2.98009468989564
0, 0.426199438668344, -2.046749498884645
1, -1.01334258269519, -10.3345489508607
1, -9.54101540300598, -4.55047974493486
1, 10.19570665680818, 0.254420103019349
1, -8.6778473681861, -3.27270683674903
1, 6.47809481278742, 8.96100657254648
0, -1.95411001422583, -1.14169962947211
1, 11.8505019698607, -0.880526995586208
1, 9.85201428822543, -1.98092728566354
1, -2.61930621246295, -9.91735995764943
1, -9.57702511533498, 0.856047877803597
1, 1.07211379460816, 10.0475383208642
1, 10.2114266188495, 2.56413556126083
1, -1.0837815752106, 10.4989367949914
0, -2.67378400738832, -0.261205121872508
1, -8.53007726968589, -5.67998129587997
```

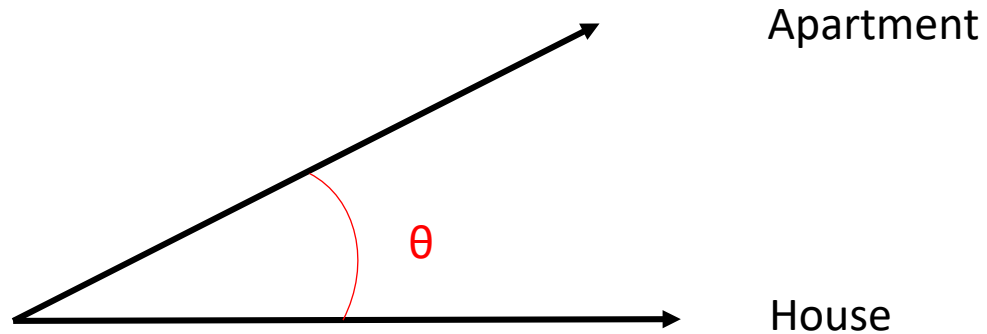


Neuronal network



Similarity

Similarity of apartment and house



The angle θ is the similarity between two vectors.

Vector operations:

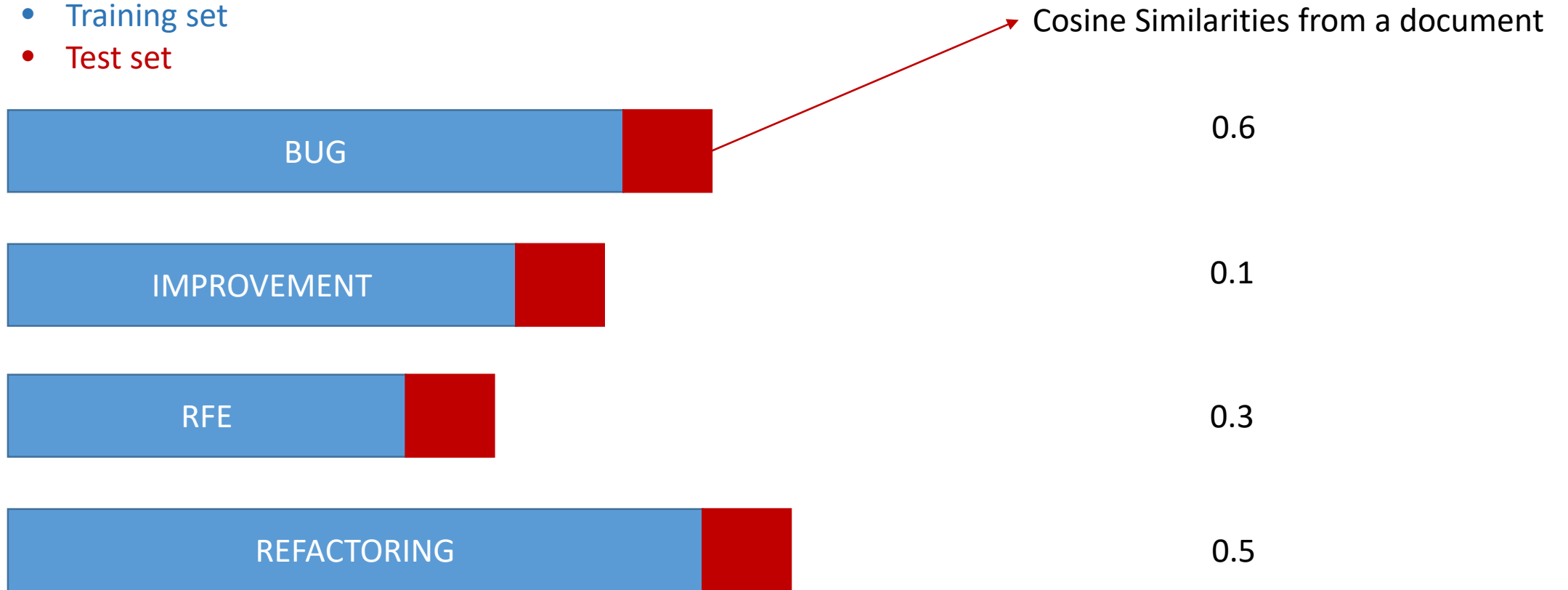
Library - Books = Hall

Logic analogies:

*Beijing is to China what
[Rome, Florence, Venice] is to Italy*

Accuracy

- Training set
- Test set



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?