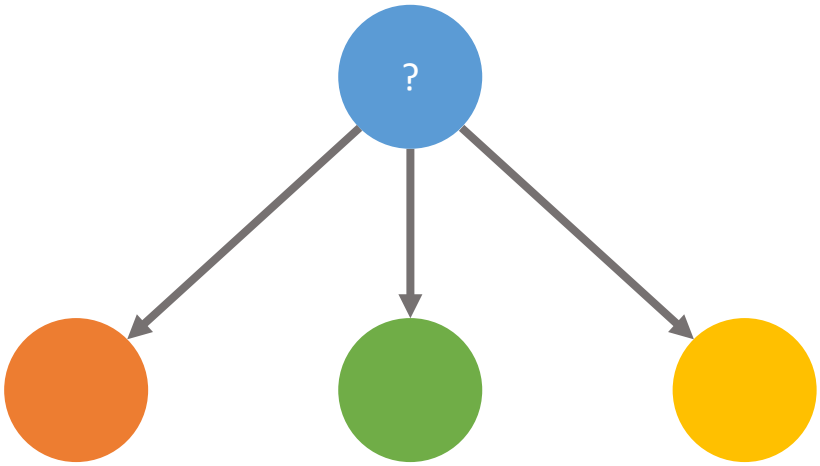
ISSUE REPORT ANALYSIS

Simon Curty

Bachelor Thesis / Seminar Project

Issue Report Classification
(Bachelor thesis)



Bug Report Quality Assessment
(Seminar Project)

Misclassification of Issue Reports

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

HttpComponents HttpClient / HTTPCLIENT-144

Add a document describing the HttpClient release process

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:	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The commons release process (http://jakarta.apache.org/commons/releases.html) is a good starting place, but is out of date.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.

People

Assignee: Jeff Dever

Reporter: Jeff Dever

Votes:

0 Vote for this issue

Watchers:

0 Start watching this issue

Dates

Created: 26/Jan/03 00:02

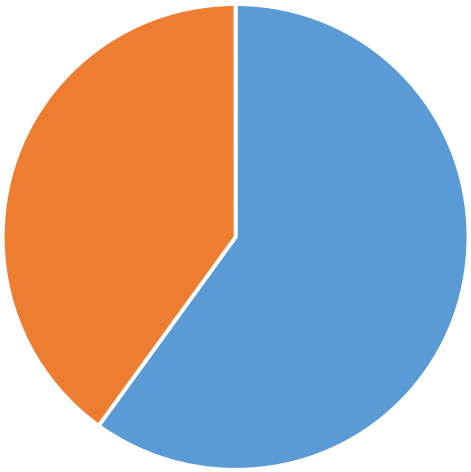
Updated: 22/Apr/07 07:10

Resolved: 22/Apr/07 07:10

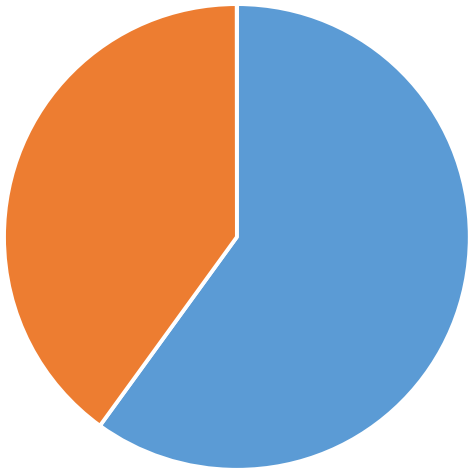
Agile

View on Board

Issues



Bug reports




```
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
```

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

Kim Herzig
Saarland University
Saarbrücken, Germany

Sascha Just
Saarland University
Saarbrücken, Germany

Andreas Zeller
Saarland University
Saarbrücken, Germany

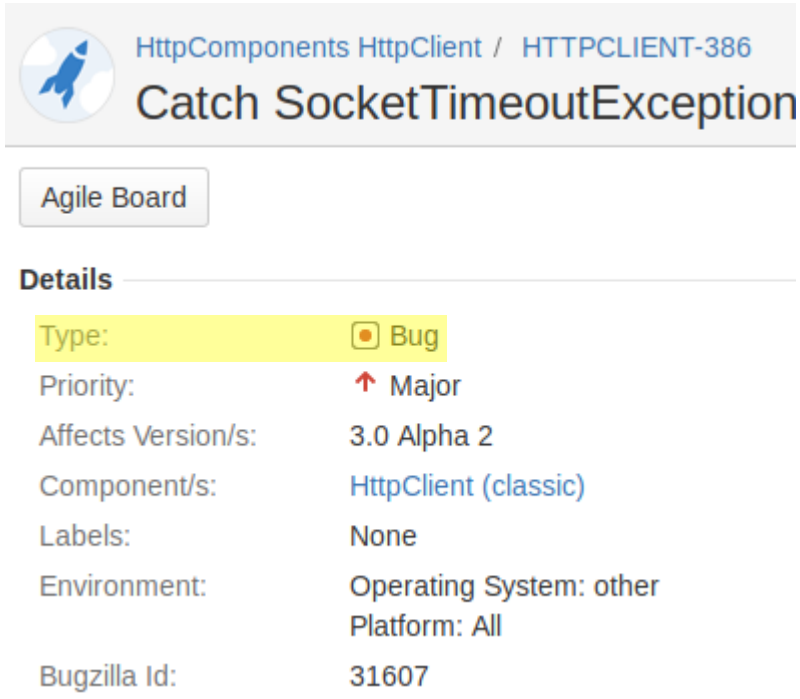
Goal

- Issue Report Classifier: A program that tells you, if an issue report describes a bug or not

- Better precision than reporters (over 66%)

Type: Issue category on tracker

Label: Issue category according to paper



Issue Report Classifier

HttpComponents HttpClient / HTTPCLIENT-390

SSL + proxy + Host auth + Keep Alive off causes

Agile Board

Details

Type:	Bug
Priority:	Blocker
Affects Version/s:	3.0 Alpha 2
Component/s:	HttpClient (classic)
Labels:	None
Environment:	Operating System: other Platform: Other
Bugzilla Id:	31981

- High priority: 1
- Version: 1
- Links 0
- Code sample: 0
- System specifications: 0
- Patch: 1
- Documentation keywords 0

Description

The combination of SSL tunnelling, host authentication, and disabled persistent connection support (HTTPD KeepAlive off) causes an infinite loop in HttpMethodDirector.

The problem has been reported on the httpclient-dev list by Rindress MacDonald <RMacDona at enterasys.com>

Attachments

ASF.LICENSE.NOT.GRANTED--auth.patch

HttpComponents HttpClient / HTTPCLIENT-144

Add a document describing the HttpClient release process

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:	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.

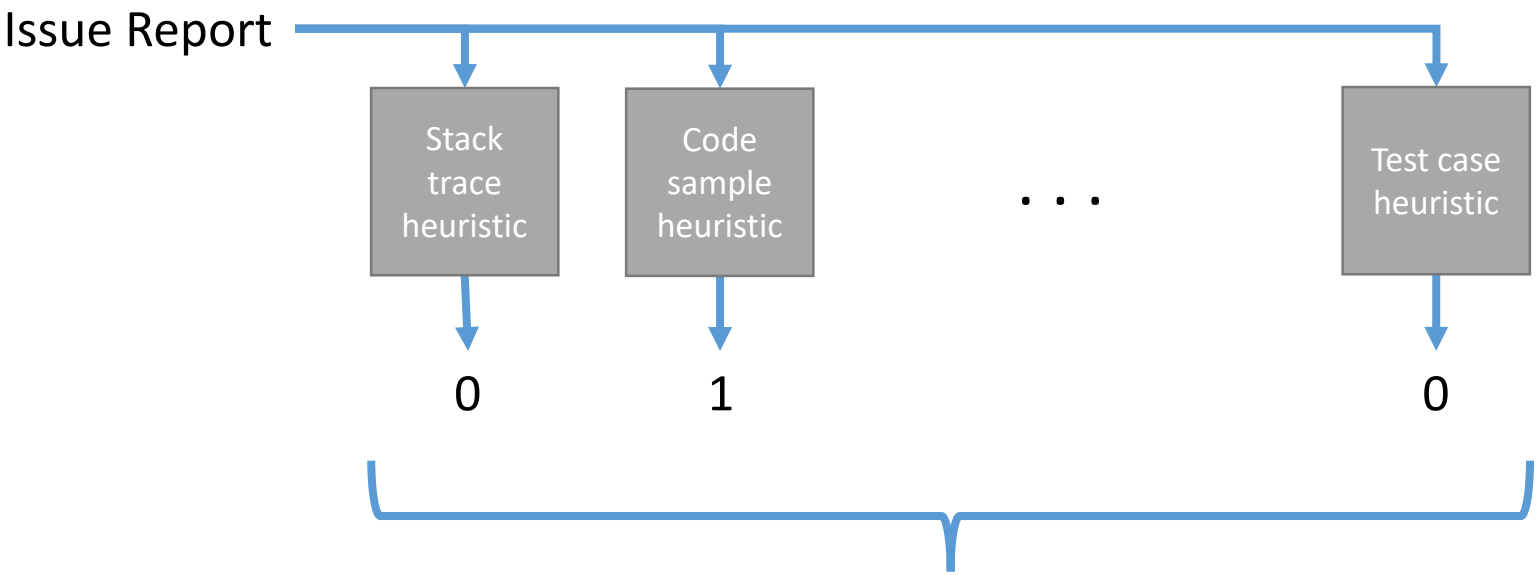
- High priority: 1
- Version: 1
- Links 1
- Code sample: 0
- System specifications: 0
- Patch: 0
- Documentation keywords 1

Vector Component Examples

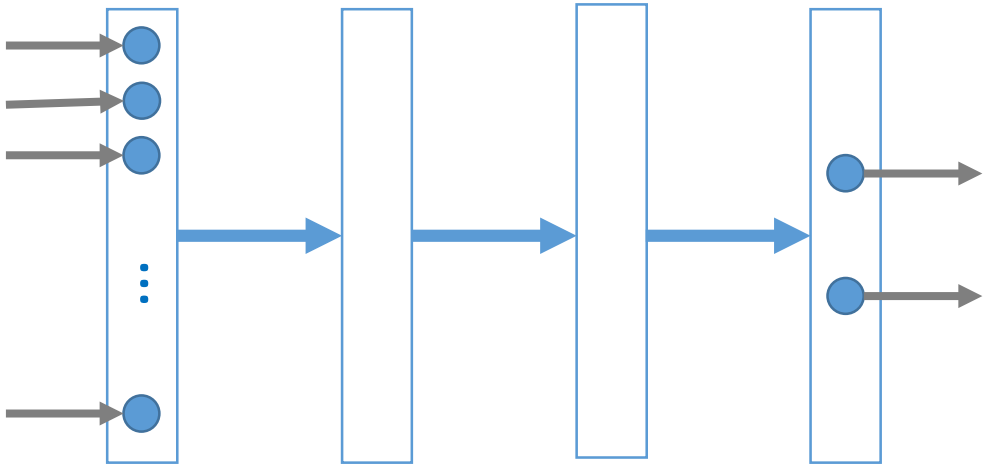
- Stack trace
- Test cases
- Code samples
- Code reference (reference to a specific place in the code)
- Priority/Severity
- Keywords

- 21 in total

Vectorization



FNN



Input layer

Hidden layers
25 hidden nodes

Output layer

Experiment Setup

- 100 independent runs
- Dataset is split into 75% for training and 25% for testing (random sampling)
- Balancing of dataset
- Early stopping (max. 1000 epochs)
- Fixed learning rate of 0.1

Results

	Bug
Precision	0.9
Recall	0.72
F1 Score	**0.8**

Precision, recall and F1 Score of the category **Bug** with binary classification over a dataset with **matching label and type**

Problems and Limitations

- Vectorization is bug centric
- Only for Java
- Some elements hard to detect

Bug Report Quality

What Makes a Good Bug Report?

Thomas Zimmermann, *Member*, *IEEE*, Rahul Premraj, Nicolas Bettenburg, *Member*, *IEEE*, Sascha Just, *Member*, *IEEE*, Adrian Schröter, *Member*, *IEEE*, and Cathrin Weiss

Abstract—In software development, bug reports provide crucial information to developers. However, these reports widely differ in their quality. We conducted a survey among developers and users of APACHE, ECLIPSE, and MOZILLA to find out what makes a good bug report. The analysis of the 466 responses revealed an information mismatch between what developers need and what users supply. Most developers consider steps to reproduce, stack traces, and test cases as helpful, which are, at the same time, most difficult to provide for users. Such insight is helpful for designing new bug tracking tools that guide users at collecting and providing more helpful information. Our CUEZILLA prototype is such a tool and measures the quality of new bug reports; it also recommends which elements should be added to improve the quality. We trained CUEZILLA on a sample of 289 bug reports, rated by developers as part of the survey. The participants of our survey also provided 175 comments on hurdles in reporting and resolving bugs. Based on these comments, we discuss several recommendations for better bug tracking systems, which should focus on engaging bug reporters, better tool support, and improved handling of bug duplicates.

Index Terms—Testing and debugging, distribution, maintenance, and enhancement, human factors, management, measurement.

SO_TIMEOUT parameter on the method level has no effect

Agile Board

Details

Type:	Bug	Status:	CLOSED
Priority:	Major	Resolution:	Fixed
Affects Version/s:	3.0 Beta 1	Fix Version/s:	3.0 RC1
Component/s:	HttpClient (classic)		
Labels:	None		
Environment:	Operating System: All Platform: PC		
Bugzilla Id:	32742		

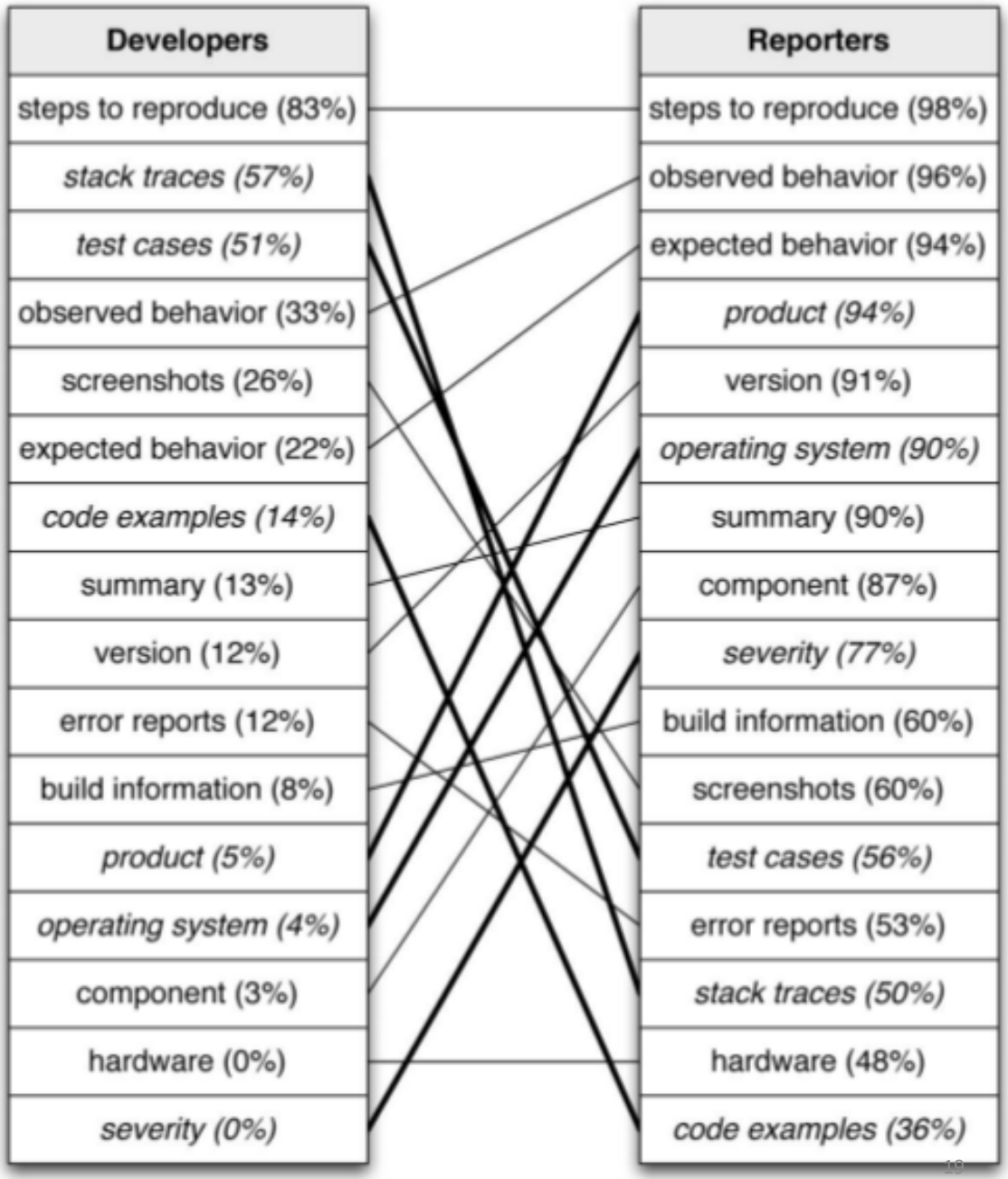
Description

This bug has been reported on the HttpClient user list by Ilya Kharmatsky <ilyak ~~at~~ mainsoft.com>

Attachments

ASF.LICENSE.NOT.GRANTED--so-timeout.patch	1 kB	17/Dec/04 06:19

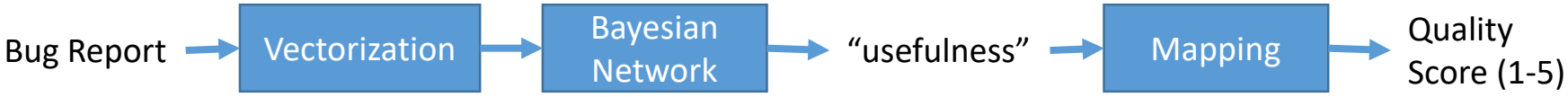
Mismatch between of what developers deemed important and what was provided by reporters



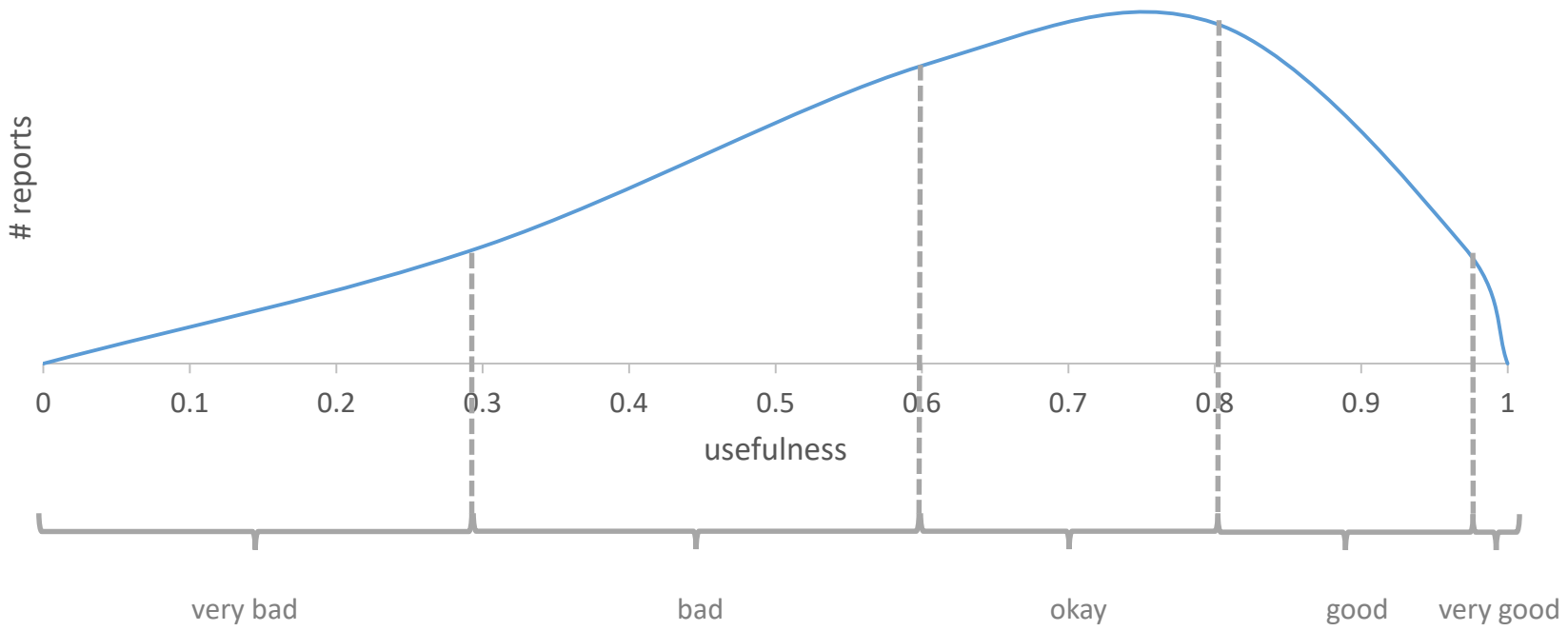
Goal

- Bug Report Quality Estimator: A program that informs the reporter about the usefulness of the report
- Intuitive score as output
- Why would/should a reporter care?

Bug Report Quality Estimation



Mapping


reports
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
usefulness
very bad
bad
okay
good
very good

Problems and Limitations

- Dataset from the paper incomplete -> missing ground truth
- Quite a bit of speculation involved