Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward

Gregory M. Kapfhammer
Department of Computer Science
Allegheny College
http://www.cs.allegheny.edu/~gkapfham/

1st International Workshop on Regression Testing
Co-Located with the 4th IEEE International Conference on Software Testing, Verification and Validation
Berlin, Germany, March 2011
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Presentation Overview

Current Trends

Increased Interest

Regression Testing Community
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Model of Regression Testing
Model of Regression Testing

Start ➔ Coverage Report
Model of Regression Testing

Start

Coverage Report

Selection
Reduction
Prioritization
Model of Regression Testing

Start

Coverage Report

Original Test Suite

Selection

Reduction

Prioritization
Model of Regression Testing
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Model of Regression Testing

Start → Coverage Report

Original Test Suite
- Selection
- Reduction
- Prioritization

Modified Test Suite
- Test Suite Execution
- Test Coverage Monitoring

Program Adequacy Criterion

Test Results → End
Model of Regression Testing

Use the Coverage Report During the Next Round of Regression Testing

Start → Coverage Report → Original Test Suite → Selection → Reduction → Prioritization → Modified Test Suite → Test Suite Execution → Test Coverage Monitoring → Test Results → End
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Model of Regression Testing

Make a New Test Suite for the Next Round of Regression Testing
Practitioners are unwilling to use methods for “lack of empirical studies” [12]
Practitioners are unwilling to use methods for “lack of empirical studies” [12]

Testing tools do not produce outputs in the best format (e.g., per-test coverage)
Practitioners are unwilling to use methods for “lack of empirical studies” [12]

Experimental assessment could stagnate due to inaccessibility of artifacts
Model of Experimental Evaluation
Model of Experimental Evaluation

Start

Programs
Test Suites
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward

Model of Experimental Evaluation

Start → Programs → Test Suites → Conduct Experiments
Model of Experimental Evaluation

Start → Programs → Conduct Experiments → Technique Configurations → Regression Testing Techniques
Model of Experimental Evaluation

Start

- Programs
- Test Suites

Conduct Experiments

- Regression Testing Techniques
- Technique Configurations

Data Sets
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Model of Experimental Evaluation

- Start
  - Programs
  - Test Suites
  - Technique Configurations
  - Conduct Experiments
    - Regression Testing Techniques
    - Data Sets
  - Visualization
    - Statistical Analysis
      - Summarized Data Sets
      - Models
      - Graphs
      - Diagrams
Model of Experimental Evaluation

Iteratively Perform Visualization and Statistical Analysis

Start → Programs → Conduct Experiments → Technique Configurations → Data Sets → Visualization → Statistical Analysis → Summarized Data Sets → Models → Repeat

Regression Testing Techniques → Graphs → Diagrams
Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques
Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques

56% of Papers Surveyed by Yoo and Harman Only Used the SIR Programs [3]
Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques

Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques

Authors Do Not Release Tools That Conduct Experiments and Analyze Results
Model of Experimental Evaluation

Conduct Experiments with Additional Programs, Test Suites, and Techniques

Without Using Data Mining Methods, Researchers May Miss Important Trends
Thought Provoking Assertions

Jonathan B. Buckheit
and David L. Donoho
Department of Statistics
Stanford University
For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Jonathan B. Buckheit and David L. Donoho
Department of Statistics
Stanford University
For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Does the Regression Testing Community Want to be Scientific?

Jonathan B. Buckheit and David L. Donoho
Department of Statistics
Stanford University
For a field to qualify as a science, it is important first and foremost that published work be reproducible by others.

Thought Provoking Assertions

Does the Regression Testing Community Want to be Scientific?

What Does it Mean for Research to be Reproducible?

Jonathan B. Buckheit and David L. Donoho
Department of Statistics
Stanford University
Thought Provoking Assertions

Gary King
Department of Government Harvard University
Thought Provoking Assertions

**Reproducible**: sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.

Gary King
Department of Government Harvard University
Thought Provoking Assertions

**Reproducible**: sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author.

Can the Regression Testing Community Adhere to This Standard?

Gary King
Department of Government Harvard University
Reproducible Research Through Sharing

Artifacts Used in Experiments

Test Suite
Reproducible Research Through Sharing

Artifacts Used in Experiments

Test Suite
Reproducible Research Through Sharing

Artifacts Used in Experiments

Test Suite

Data Sets

...
Reproducible Research Through Sharing

Artifacts Used in Experiments

Repository

Test Suite

Data Sets

...
Reproducible Research Through Sharing

Artifacts Used in Experiments

Dataverse

Repository

Test Suite

Data Sets

...
Reproducible Research Through Sharing

Artifacts Used in Experiments

Test Suite

......

Data Sets

Repository

Dataverse

Web Site
Reproducible Research Through Sharing

Artifacts Used in Experiments

- Test Suite
- Data Sets

Repository

Dataverse

Web Site

Version Control
Reproducible Research Through Sharing

Deposit the Artifacts from Experimentation in One or More Repositories

Artifacts Used in Experiments
- Test Suite
- Data Sets

Repositories
- Dataverse
- Web Site
- Version Control
Reproducible Research Through Sharing

External Researchers Access the Repositories

Artifacts Used in Experiments

Dataverse

Test Suite

Web Site

External Researchers

Version Control

Data Sets
Reproducible Research Through Sharing

Internal Researchers Access the Repositories

Artifacts Used in Experiments

Dataverse

Web Site

Version Control

Repository

Test Suite

Data Sets

Internal Researchers

...
Reproducible Research Through Sharing

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Reproducible Research Through Sharing

Internal Researchers Access the Repositories

- Dataverse
- Web Site
- Version Control

Artifacts Used in Experiments

- Test Suite
- Data Sets

Repository

Costs of Reproducible Research?

Benefits of Reproducible Research?

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Introduction

Challenges

Potential Way Forward

Conclusion

Reproducible Research Through Sharing

Internal Researchers Access the Repositories

Dataverse

Web Site

Version Control

Artifacts Used in Experiments

Repository

Test Suite

Data Sets

Internal Researchers

Costs of Reproducible Research?

Benefits of Reproducible Research?

Increase the Citation of Your Paper [17]

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Reproducible Research Through Sharing

Artifacts Used in Experiments

- Test Suite
- Data Sets
- Internal Researchers

Repository

Web Site

Version Control

Dataverse

Internal Researchers Access the Repositories

Allow Others to Replicate Your Experiments [15]

Benefits of Reproducible Research?

Costs of Reproducible Research?
Reproducible Research Through Sharing

Artifacts Used in Experiments → Repository

Dataverse → Repository
Web Site → Repository
Version Control → Repository

Internal Researchers Access the Repositories

Test Suite → Data Sets

Lower the Barriers to Entry for New Researchers [16]

Costs of Reproducible Research?
Benefits of Reproducible Research?

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Reproducible Research Through Sharing

Internal Researchers Access the Repositories

Artifacts Used in Experiments

Test Suite

Data Sets

... 

Repository

Dataverse

Web Site

Version Control

Internal Researchers

Costs of Reproducible Research?

Benefits of Reproducible Research?

Find Your Own Deliverables More Easily [13]
Practical Suggestions: Pick Tools Carefully
Practical Suggestions: Pick Tools Carefully

Use the R Language for Statistical Computation When:

[Image of the R logo]
Practical Suggestions: Pick Tools Carefully

Use the R Language for Statistical Computation When:

- Conducting Experiments
Practical Suggestions: Pick Tools Carefully

Use the R Language for Statistical Computation When:

- Conducting Experiments
- Visualizing Data
Practical Suggestions: Pick Tools Carefully

Use the R Language for Statistical Computation When:

- Conducting Experiments
- Visualizing Data
- Data Mining
Practical Suggestions: Pick Tools Carefully

Simple Shell or Full-Featured IDEs

Use the R Language for Statistical Computation When:
- Conducting Experiments
- Visualizing Data
- Data Mining
Practical Suggestions: Pick Tools Carefully

- Simple Shell or Full-Featured IDEs
- Use the R Language for Statistical Computation When:
  - Conducting Experiments
  - Visualizing Data
  - Data Mining
- Advanced Data Analysis Methods
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward

Practical Suggestions: Use Tree Models

RM

\text{sel\_method: TOU3, TOU4, TOU5}

\text{child\_density < 0.875}

0.9674
Practical Suggestions: Use Tree Models

RM

sel_method: TOU3, TOU4, TOU5

child_density < 0.875

0.9674

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward

Practical Suggestions: Use Tree Models

**RM**

```
sel_method: TOU3, TOU4, TOU5
child_density < 0.875
0.9674
```

**GB**

```
sel_method: TOU3, TOU4, TOU5
mutation_rate < 0.17
cross_operator: OX1, VR
```

**Practical Suggestions:**
- Use tree models for regression testing.
- **RM:**
  - Selection method (TOU3, TOU4, TOU5)
  - Child density: $< 0.875$
- **GB:**
  - Selection method (TOU3, TOU4, TOU5)
  - Mutation rate: $< 0.17$
  - Cross operator: OX1, VR
Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Practical Suggestions: Use Tree Models

The sel_method variable is always the most important parameter.
Practical Suggestions: Use Tree Models

RM

**sel_method**: TOU3, TOU4, TOU5

- **child_density**: < 0.875
- **0.9674**

GB

**sel_method**: TOU3, TOU4, TOU5

- **mutation_rate**: < 0.17
- **cross_operator**: OX1, VR

Importance of **sel_method** holds for all case study applications
Practical Suggestions: Use Tree Models

**RM**

- sel_method: TOU3, TOU4, TOU5
- child_density < 0.875
- 0.9674

**GB**

- sel_method: TOU3, TOU4, TOU5
- mutation_rate < 0.17
- cross_operator: OX1, VR

**How does the selection method impact efficiency and effectiveness?**
Practical Suggestions: Use Tree Models

RM

sel_method: TOU3, TOU4, TOU5

child_density < 0.875

0.9674

GB

sel_method: TOU3, TOU4, TOU5

mutation_rate < 0.17

cross_operator: OX1, VR

For More Details, Please See Conrad et al. [22]
Practical Suggestions: Use Open Source

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total...
Practical Suggestions: Use Open Source

Gelations is a Genetic Algorithm-Based Test Suite Prioritizer

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total
Practical Suggestions: Use Open Source

Practical Suggestions: Use Open Source

Gelations is a research prototype system for regression test suite prioritization using genetic algorithms. This system is written entirely in version 1.6 of the Java SE programming language, and is accompanied by its own regression test suite written using the JUnit unit testing framework.

Software testing is a crucial part of the software development lifecycle. Regression testing is a form of testing in which all of the old test cases written to cover different parts of a program are combined into a single test suite and executed. This form of testing helps to reveal regressions, or instances in which code that had formerly functioned correctly is broken by later changes to the system. For real-world applications, however, regression test suites can take days or even weeks to execute. One solution to this problem of execution time overhead is to reduce the suite, removing test cases that are redundant or unlikely to detect faults. This approach, however, can compromise the ability of a suite to detect faults. Another approach to this problem is test suite prioritization. Prioritization does not reduce the total

Visit http://raise.googlecode.com/ for Greedy Algorithms
Conclusions and Future Work

The Potential Way Forward May Seem Daunting
Conclusions and Future Work

The Potential Way Forward May Seem *Daunting*

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming
Conclusions and Future Work

Share *One* Item With Your Next Paper

The Potential Way Forward May Seem *Daunting*

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming
Conclusions and Future Work

Once you get your courage up and believe that you can do important problems, then you can. - Richard Hamming

Your Help is Welcomed and Appreciated!

Leader of the Regression Testing Community

Richard Hamming
Conclusions and Future Work

Once you get your courage up and believe that you can do important problems, then you can.

Richard Hamming
Conclusions and Future Work

Future Work:
Complete Case Study of Reproducible Research in Regression Testing

- Data Sets
- Tools
- Visualizations

Leader of the Regression Testing Community

Read the Paper and Contact Me with Comments

One Item With Your Next Paper

Empirically Evaluating Regression Testing Techniques: Challenges, Solutions, and a Potential Way Forward
Conclusions and Future Work

Leader of the Regression Testing Community

Read the Paper and Contact Me with Comments

Acknowledgments:
- Neil Elliott
- Jonathan Kauffman
- Molly Mattis
- Zachary Williams
  provided feedback on drafts of the paper