Evaluating the Adaptive Selection of Classifiers for Cross-Project Bug Prediction
Developers usually introduce bugs
Software failures can compromise business

RBS could take until weekend to make 600,000 missing payments after glitch

Delay in payments - including tax credits and disability living allowance - is latest technology failure to affect UK bank’s customers

Software failures can compromise safety

US aviation authority: Boeing 787 bug could cause 'loss of control'

More trouble for Dreamliner as Federal Aviation Administration warns glitch in control unit causes generators to shut down if left powered on for 248 days
Large software systems contain thousands of bugs!
but resources are limited...
Thus, we should predict which components are more prone to be buggy!
Machine Learning features for Bug Prediction

- **Structural Metrics**: CK metrics, LOC, etc...
- **Process Metrics**: # Changes, # Developers, Ownership, etc...
- **Developers Metrics**: Scattering, Micro-Interactions, etc...
Training a Machine Learner for Bug Prediction

Within-Project
Training a Machine Learner for Bug Prediction

Cross-Project
Training a Machine Learner for Bug Prediction

T. Menzies, A. Butcher, D. Cok, A. Marcus, L. Layman, F. Shull, B. Turhan, and T. Zimmermann
"Local versus global lessons for defect prediction and effort estimation"
*IEEE Transactions on Software Engineering*
Training a Machine Learner for Bug Prediction

T. Menzies, A. Butcher, D. Cok, A. Marcus, L. Layman, F. Shull, B. Turhan, and T. Zimmermann

"Local versus global lessons for defect prediction and effort estimation"

*IEEE Transactions on Software Engineering*
Training a Machine Learner for Bug Prediction

T. Menzies, A. Butcher, D. Cok, A. Marcus, L. Layman, F. Shull, B. Turhan, T. Zimmermann
"Local versus global lessons for defect prediction and effort estimation"
IEEE Transactions on software engineering

Local-Project
Selecting a Machine Learner for Bug Prediction
The choice of the machine learner strongly impacts performance!

Performance can increase or decrease up to 30% depending on the used classifier.
Cross-Project Defect Prediction Models: L’Union Fait la Force

Amlan G. Patra, Bharat Odedra, Arvind De Leon
Department of Management & Information Technology, University of Salento, Lecce, Italy
Department of Economics and Computer Science, University of Salerno, Italy

Abstract—Existing defect prediction models use product or process metrics and sometimes also machine learning techniques to identify defective classes. Researchers have proposed and experimentally evaluated different machine learning techniques to improve the detection of defective classes. In this paper, we propose a practical rule-based approach for selecting classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research. We use two different machine learning techniques to learn the behavior of defect prediction models. To evaluate the effectiveness of the proposed approach, we compare the performance of different machine learning techniques and the performance of the proposed approach against the existing approaches. We find that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models.

1. INTRODUCTION

Defect prediction models are used to predict the quality of software components. Defect prediction models are used to identify classes and modules that are more likely to contain defects. These models are used to improve the quality of software components and to reduce the cost of software development. The effectiveness of defect prediction models depends on the accuracy of the models. The accuracy of the models depends on the selection of the model and the selection of the machine learning techniques used in the models.

1.1 Related Work

There have been many studies on defect prediction models. Researchers have proposed and evaluated different machine learning techniques to improve the detection of defective classes. Researchers have also proposed and evaluated different machine learning techniques to improve the detection of defective classes. In this paper, we propose a practical rule-based approach for selecting classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research. We use two different machine learning techniques to learn the behavior of defect prediction models. To evaluate the effectiveness of the proposed approach, we compare the performance of different machine learning techniques and the performance of the proposed approach against the existing approaches. We find that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models.

1.2 Contributions

In this paper, we propose a practical rule-based approach for selecting classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research. We use two different machine learning techniques to learn the behavior of defect prediction models. To evaluate the effectiveness of the proposed approach, we compare the performance of different machine learning techniques and the performance of the proposed approach against the existing approaches. We find that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models.

1.3 Organization

This paper is organized as follows: Section 1 introduces the problem and motivates the need for a practical approach for selecting classifiers and machine learning techniques. Section 2 describes the proposed approach for selecting classifiers and machine learning techniques. Section 3 evaluates the effectiveness of the proposed approach against the existing approaches. Section 4 concludes the paper and suggests future work.

2. Proposed Approach

We propose a practical rule-based approach for selecting classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research. The proposed approach is based on the intuition that classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research can be used to select better machine learning techniques for defect prediction models.

2.1 Rule-Based Approach

The proposed approach is based on the intuition that classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research can be used to select better machine learning techniques for defect prediction models. The proposed approach is based on the intuition that classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research can be used to select better machine learning techniques for defect prediction models.

2.2 Experimental Setup

We evaluate the proposed approach by applying it to a set of defect prediction models. The proposed approach is based on the intuition that classifiers and machine learning techniques that are more complementary and less redundant than those used in existing research can be used to select better machine learning techniques for defect prediction models.

2.3 Results

We evaluate the effectiveness of the proposed approach by comparing the performance of the proposed approach against the existing approaches. We find that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models.

3. Conclusion

We conclude that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models. We also conclude that the proposed approach is more effective than the existing approaches and can be used to select better machine learning techniques for defect prediction models.

References


Acknowledgments

This research is supported by the National Science Foundation under grant number 1703293.

Thanks to the reviewers for their valuable feedback.

Copyright © 2018 Amlan G. Patra, Bharat Odedra, Arvind De Leon.

This article has been accepted for publication in IEEE Transactions on Software Engineering, but has not yet been fully edited by the TSE editors. It can be cited as an IEEE Xplore Press Release.

Electronic version of this article is available at: http://example.com/article.pdf

Classifiers are highly complementary!

Different classifiers capture different sets of buggy components!
Ensemble of classifiers. Learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.

Thomas G. Dietterich
Ensemble of Classifiers

- Validation & Voting
- Bagging
- Boosting
- Stacking
- Random Forest

Which Technique?
ASCI: Adaptive Selection of Classifiers

D. Di Nucci, F. Palomba, R. Oliveto, and A. De Lucia
“Dynamic selection of classifiers in bug prediction: An adaptive method”
IEEE Transactions on Emerging Topics in Computational Intelligence
ASCI: Adaptive Selection of Classifiers

D. Di Nucci, F. Palomba, R. Oliveto, and A. De Lucia

"Dynamic selection of classifiers in bug prediction: An adaptive method
IEEE Transactions on Emerging Topics in Computational Intelligence
Evaluating the Adaptive Selection of Classifiers for Cross-Project Bug Prediction
Experimented Systems

10 Software Systems

i) Less than 50% of Buggy Classes

ii) Randomly Selected

C. Tantithamthavorn, S. McIntosh, A. E. Hassan, and K. Matsumoto
“Automated parameter optimization of classification techniques for defect prediction models,”
*International Conference on Software Engineering 2016*
How does ASCI work in the context of cross-project bug prediction when compared to existing ensemble techniques?
To what extent can local learning improve the performance of ASCI?
Pre-processing steps

Data Cleaning
Pre-processing steps

- Data Cleaning
- Data Normalization
Pre-processing steps

- Data Cleaning
- Data Normalization
- Feature Selection
  Correlation-based Feature Selection
Pre-processing steps

- Data Cleaning
- Data Normalization
- Feature Selection
  - Correlation-based Feature Selection
- Data Balancing
  - Synthetic Minority Over-sampling Technique
Experimented Models

- Naive Bayes (NB)
- ASCII
- Validation & Voting (VV)

Sub-models:
- Naive Bayes
- Multi-Layer Perceptron
- Logistic Regression
- Radial Basis Function
- C4.5
- Decision Table
- Support Vector Machine
Validation Metrics

Accuracy

<table>
<thead>
<tr>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>
Validation Metrics

- Accuracy
- F-Measure

<table>
<thead>
<tr>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>
Validation Metrics

- Accuracy
- F-Measure
- Area Under Curve

<table>
<thead>
<tr>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>
Results

Accuracy  F-Measure  AUC-ROC
Thus, we should predict which components are more prone to be subject to bugs!
Conclusions

Thus, we should predict which components are more prone to be subject to bugs!
Conclusions

Thus, we should predict which components are more prone to be subject to bugs!
Conclusions

Thus, we should predict which components are more prone to be subject to bugs!

Ensemble of Classifiers

Ensemble of classifiers: Learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.

Thomas G. Dietterich

ASCI: Adaptive Selection of Classifiers
Conclusions

Thus, we should predict which components are more prone to be subject to bugs!

Experimented Systems

- 10 Software Systems
  - i) Randomly Selected
  - ii) Less than 50% of Buggy Classes

Experimented Models

- Naive Bayes
- ASCI
- Voted by Voting

Ensemble of Classifiers

Ensemble of classifiers. Learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.

Thomas G. Dietterich

ASCI: Adaptive Selection of Classifiers

ASCI: Adaptive Selection of Classifiers: A method for selecting a subset of classifiers to use for a new prediction.

E. Bodshick, J. Kramer, D. Page, and K. Messinger
Conclusions

Thus, we should predict which components are more prone to be subject to bugs!

Ensemble of Classifiers

Ensemble of classifiers. Learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.

Thomas D. Alterskirk

Experimented Systems

10 Software Systems

1) Randomly Selected
2) Less than 50% of Buggy Classes

Experimented Models

ASCI: Adaptive Selection of Classifiers

Models:
- Naive Bayes
- ASCI
- Voting
- Select

Results

Accuracy
F-Measure
AUC-ROC