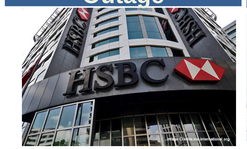
Machine Learning Driven Software Test Case Selection

Victor Cheruiyot
Concordia University of Edmonton

Background

No matter how technology advances, software testing will always be non-negotiable. Every week new stories emerge of software failing across a myriad of industries; sparking chaos, halting business, or even costing lives.

HSBC's Major IT Outage



In January 2016, HSBC suffered a major IT outage,. The bank took almost 2 days to recover and get back to normal functioning.

Airbus Software Bug Alert

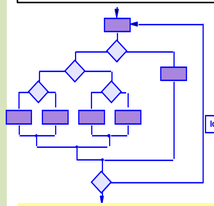
In May 2015, Airbus detected a software bug in its A400M aircraft that had caused a fatal crash in Spain. Four crew members died in a test flight in Seville.

911 Call Outage

In April 2015, Emergency services got stalled for six hours for seven US states. This affected 81 call centers, and 6,000 calls were unattended.

Complexity of Test Case Design and Selection

- Software testing is a quality control activity which focuses on identifying defects (which are then removed).
- After completion of coding software products are subjected to testing with the help of the different test cases.

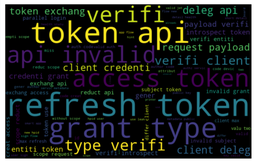


total number of execution paths?

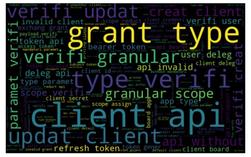
- It is impossible to completely test any nontrivial module or any system.
- Theoretical limitations: Halting problem
- Practical limitations: Prohibitive in time and cost.

Machine learning based test case selection strategy could overcome manual labor and complete software testing successfully with minimum risk.

Datasets and Bag of Words (BoW)

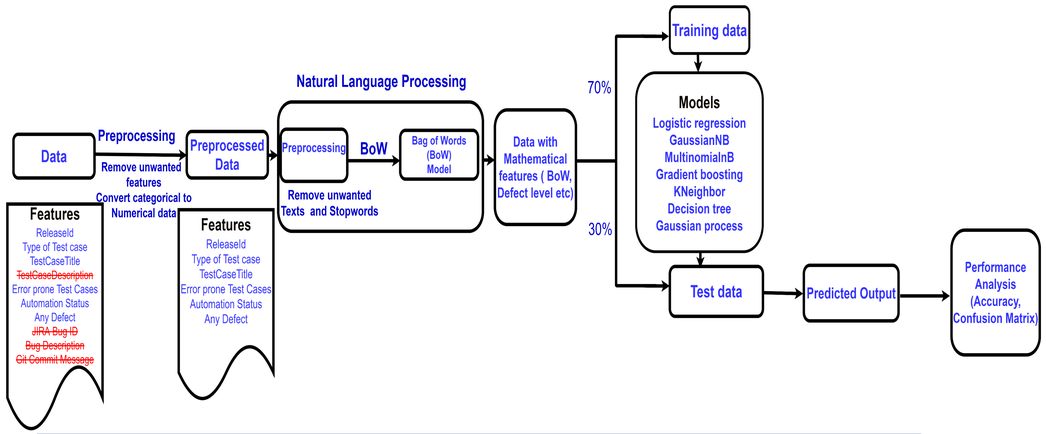


Frequent words in Selected Test Cases



Frequent words in Non-selected Test Cases

ReleaseID	Type of Test Case	TestCaseTitle	Error Prone Test Cases	Automation Status	Any Defect	Target
R20.2.1	Sanity	Get the short and detailed health status APIs	No	Yes	Yes	1
R20.2.1	Sanity	Get the AuthZ service metadata	No	Yes	Yes	1
R20.2.1	Sanity	Get the public keys for validating token	No	Yes	Yes	1
R20.2.1	API/Functionality	Verify Client delegation API: Exchange Access ...	Yes	Yes	No	1
R20.2.1	API/Functionality	Verify Client delegation API:Exchange Access t...	No	Yes	No	1



Flowchart of Proposed ML Driven Software Test Case Selection

Performance Evaluation

Classifier	Train accuracy	Test accuracy	F1 score
Regression	90.2	84.2	68.7
GaussianNB	80.4	78.2	70.1
MuiltnomialNB	87.6	84.2	70.1
Gradient Boosting	100	96.9	94.1
KNeighbor	100	96.9	93.8
Decision Tree	**100**	**99.2**	**98.6**
Gaussian Process	100	98.5	98.6

Confusion Matrix

Decision Tree Classifier		Predicted	
		Positive	Negative
Observed	Positive	98	1
	Negative	0	34

Conclusions

- Regression testing is required to conduct after updating any software components.
- We proposed a machine learning based test case selection strategy for regression testing.
- We implemented natural language processing tools for software test case selection.
- Classification performance of seven different classifiers was presented.
- Experimental results demonstrate that machine learning based approach can reduce the bias and manual labour of domain expert for software regression testing.
- In future, we would investigate the feature selection strategy for natural language processing to automate the whole software test case selection procedure.

Contact

Victor Cheruiyot
MscIT, Concordia University of Edmonton
Email: vcheruiy@student.concordia.ab.ca

References

1.Everton Narciso et al. (2014). Test Case Selection: A Systematic Literature Review. International Journal of Software Engineering and Knowledge Engineering. 24. 653-676.
2.Uma Sharma et al. (2012). Test Case Selection and Prioritization for Regression Testing using Fault Severity. INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY. 9. 1171-1177.
3. Ahlam Ansariet al. (2016), Optimized Regression Test Using Test Case Prioritization, Procedia Computer Science, Volume 79, Pages 152-160

© POSTER TEMPLATE BY GENIGRAPHICS® 1.800.790.4001 WWW.GENIGRAPHICS.COM