Automated Classification of Static Code Analysis Alerts: A Case Study

29th IEEE International Conference on Software Maintenance
26 September 2013 - Eindhoven, The Netherlands

Ulaş Yüksel[1,2]
Hasan Sözer [2]

[1] Vestel Electronics, Manisa, Turkey
[2] Özyegin University, Istanbul, Turkey

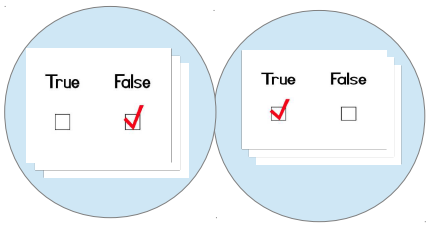
Problem

Static code analysis tools (SCATs) complement manual reviews and testing...

but

if you have 1500 KLOC legacy software ...

Problem (cont.)



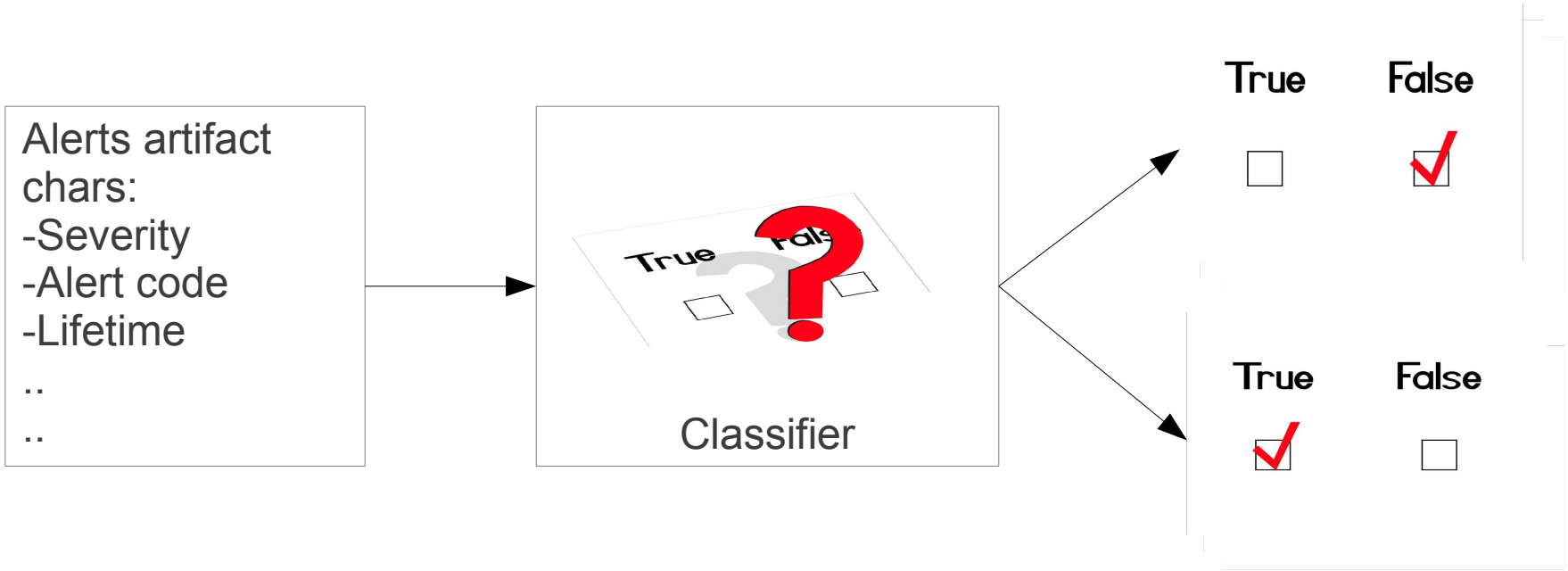
1500 KLOC code...

- May generate up to **3000 alerts** to be inspected.
- **30% - 100%*** false positives.
- May need 250 hours of manual inspection.
- Is automatic classification feasible?

* T. Kremenek and D. Engler, “Z-ranking: using statistical analysis to counter the impact of static analysis approximations,” in Proceedings of the 10th international conference on Static analysis, 2003, pp. 295–315.

Evaluated approach

- Post-process of SCAT alerts to classify with *machine-learning algorithms*
- Based on set of artifact characteristics.

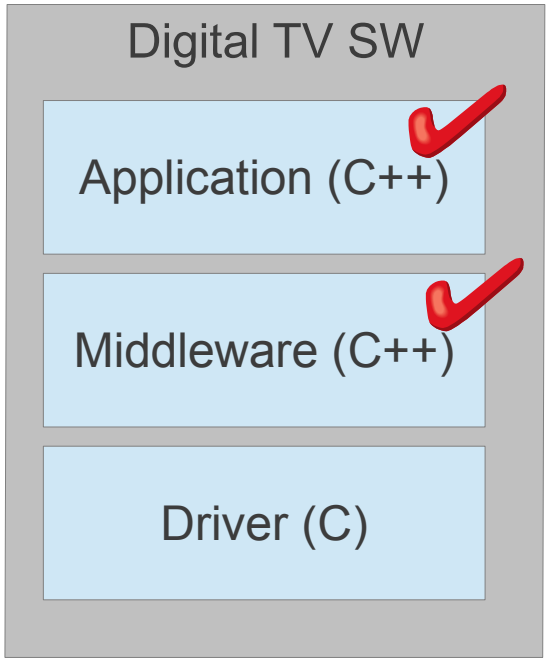


Outline

- Introduction: Problem definition
- Industrial case: Digital TV software
- Data set
- Case study and results
- Future work

Industrial Case: Digital TV Software

- **VESTEL** Electronics: Leading company in digital TV market
- Supports: Digital video broadcast (DVB), recording, media and web browsing , and so on...
- Embedded
- Over 1500KLOC C/C++ distributed over 8000 files



Data set and characteristics

- 19 weekly run of SCAT
- 1147 Alerts with unique ID and set of artifact characteristics:
  - *Severity, Alert code, Life time, Developer idea, File name, Module name, Open alerts, Total alerts, Total alerts in module, Total alerts in file*
- Manually inspected alert distribution:

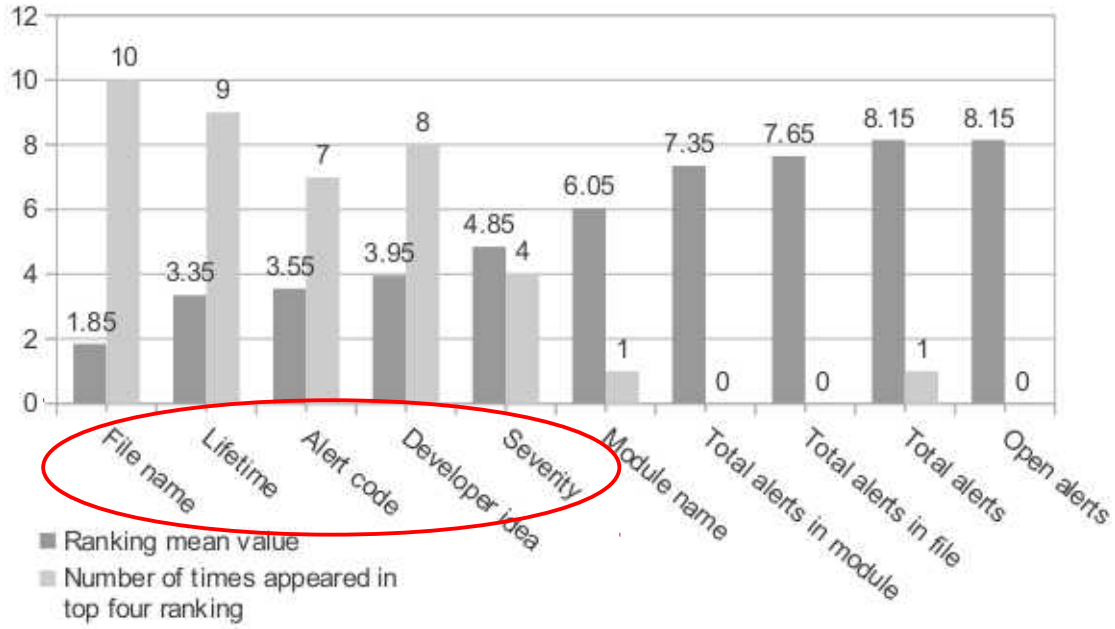
		Manual Inspection Result		
		True Positive	False Positive	Total
Resolution	Resolved	203	62	265
	Not resolved	295	587	882
	Total	498	649	1147

Case Studies on Data Set

- (a) Attribute Evaluation
- (b) Classification with 10-folds cross validation
- (c) Classification during the development life-cycle

Case Study- (a): Attribute Evaluation

- Aimed to see which of the characteristics are relevant for classification.
- With 10 attribute selection evaluator tools of Weka*



* WEKA: Data mining software http://www.cs.waikato.ac.nz/ml/weka/

Case Study- (b): Classification with 10-folds cross validation

- 34 machine learning algorithms available in Weka tested.
- Top classifiers with respect to accuracy:

Classifier	Accuracy	Precision[a]	Recall[a]
Random forest	86.1%	86.1% [0.84,0.88]	86.1% [0.84,0.88]
Random committee	86.4%	86.4% [0.85,0.88]	86.4% [0.84,0.88]
DTNB *	83.6%	83.8% [0.80,0.87]	83.6% [0.84,0.84]

[a] Values in brackets are particular calculations for true positive and false positive classifications respectively.

* DTNB: M. A. Hall and E. Frank, "Combining naive bayes and decision tables,"

Case Study- (c): Classification during the development life-cycle

- First 5 run of SCAT for training set (92% of whole).
- 3 groups of test sets created with the rest:

Group	SCAT run #	True Positive	False Positive	Total
1^{st}	[6,10]	2	13	15
2^{nd}	[11,16]	9	11	20
3^{rd}	[17,19]	34	12	46

- Developer idea and lifetime are modified to reflect the real development life cycle.
- 34 machine learning algorithms available in Weka tested..

Case Study- (c): Classification during the development life-cycle (cont.)

- Best classifier with respect to accuracy: DTNB

Group	Accuracy	Precision[a]	Recall[a]
1^{st}	66.7%	80.7% [0.20,0.90]	66.7% [0.5,0.69]
2^{nd}	80.0%	86.2% [0.69,1.00]	80.0% [1.00,0.64]
3^{rd}	89.1%	90.7% [0.97,0.73]	89.1% [0.88,0.92]

[a] Values in brackets are particular calculations for true positive and false positive classifications respectively.

Conclusion & Future work

- We conclude:
  - Machine learning techniques can be a viable approach for automated classification of SCAT alerts...
- In future:
  - Extend data set with the consequent SCAT runs...
  - Incorporate additional artifact characteristics...

Thank you for your attention

Any Questions?

ÖZYEĞİN ÜNİVERSİTESİ

VESTEL

Contact:

ulas.yuksel@ozu.edu.tr

hasan.sozer@ozu.edu.tr