PWG 1

Bug prediction models

Mentor
Max Di Penta

Leader
Marco D’Ambros

Emitzá
Guzmán Ortega

Dávid
Tengeri

Amir
Molzam Sharifloo

Zuoning
Yin
Goal

Investigating:

• usage
• barriers for adoption
• threats
• future directions
27 subjects
27 subjects

Industry

Academia
Which models are used in the industry?

- Very simple model: Academia 26%, Industry 17%
- Off-the-shelf tools: Academia 9%, Industry 17%
- Models proposed by researchers: Academia 0%, Industry 0%
- No models, but developers’ skills: Academia 22%, Industry 33%
- Not needed, it is known where bugs are: Academia 4%, Industry 33%
Which models are used in the industry?

- Very simple model: 17% (Academia), 26% (Industry)
- Off-the-shelf tools: 9% (Academia), 17% (Industry)
- Models proposed by researchers: 22% (Academia), 0% (Industry)
- No models, but developers' skills: 22% (Academia), 50% (Industry)
- Not needed, it is known where bugs are: 4% (Academia), 33% (Industry)
Which models are used in the industry?

- Very simple model: 26% Industry, 17% Academia
- Off-the-shelf tools: 17% Industry, 9% Academia
- Models proposed by researchers: 22% Industry, 22% Academia
- No models, but developers' skills: 22% Industry, 9% Academia
- Not needed, it is known where bugs are: 33% Industry, 50% Academia

**Never seen anybody using bug prediction**
Which models are used in the industry?

- Very simple model: 26%
- Off-the-shelf tools: 17% (Academia), 17% (Industry)
- Models proposed by researchers: 9% (Academia), 38% (Industry)
- No models, but developers' skills: 17%
- Not needed, it is known where bugs are: 4%

CRANE @ Microsoft

Never seen anybody using bug prediction
Barriers for the adoption among practitioners?

- Models not available as tools: 52% (Academia), 50% (Industry)
- Accuracy of the prediction: 48% (Academia), 17% (Industry)
- Difficulty in interpretation: 39% (Academia), 17% (Industry)
- QA is not properly modeled: 22% (Academia), 33% (Industry)
Barriers for the adoption among practitioners?

- Models not available as tools: 52% (Academia), 50% (Industry)
- Accuracy of the prediction: 48% (Academia), 39% (Industry)
- Difficulty in interpretation: 33% (Academia), 17% (Industry)
- QA is not properly modeled: 22% (Academia), 17% (Industry)

Developers don't even know that bug prediction models exist.
Barriers for the adoption among practitioners?

Developers don't even know that bug prediction models exist

The models should not only predict the number of bugs but also their characteristics (e.g. severity)
Threats to validity of the models

- Quality of the data: 75%
- Inter-project inapplicability: 54%
- Not meaningful variables: 36%
- Not applicable to new projects: 29%
- Do no explain the phenomena: 25%
Future directions

Models adaptable to different systems
- Models adaptable to different systems
- Improve benchmarks for comparison
- Better model QA activities
- Finer-grained prediction
- Models that deals with incomplete data

Academia: 57% 48% 22% 50% 57%
Industry: 17% 17% 17% 50% 67%

Future directions

Academia

Industry
Future directions

- Models adaptable to different systems: 57% (Academia: 57%, Industry: 67%)
- Improve benchmarks for comparison: 48% (Academia: 17%, Industry: 22%)
- Better model QA activities: 22% (Academia: 17%, Industry: 17%)
- Finer-grained prediction: 50% (Academia: 22%, Industry: 22%)
- Models that deals with incomplete data: 57% (Academia: 57%, Industry: 67%)