

Construção de um Dataset para Análise da Evolução dos bugs de Sistemas

Pedro Araujo Pires

Introdução

- Dataset
 - Conjunto de informações sobre 9 sistemas, geradas para 180 versões (uma a cada 2 semanas).
- Quais informações?
 - Modelo FAMIX do código fonte.
 - Séries temporais de métricas de código fonte (CK + 11).
 - Série temporal de bugs em classes.

Introdução

- Inspiração
 - **An Extensive Comparison of Bug Prediction Approaches**
Marco D'Ambros, Michele Lanza, Romain Robbes In Proceedings of MSR 2010 (7th IEEE Working Conference on Mining Software Repositories). IEEE CS Press, 2010.
 - Dataset com 5 sistemas:
 - Eclipse JDT Core
 - Eclipse PDE UI
 - Equinox Framework
 - Lucene
 - Mylyn
 - Aproximadamente 100 versões

Sistemas

- Eclipse JDT Core
- Eclipse PDE UI
- Equinox Framework
- Lucene
- Geronimo
- Hadoop MapReduce
- TV-Browser
- Hibernate
- Axis 2

Extração dos dados

1 Repositório SCM:

- *Checkout* de 180 versões de cada sistema.
- Download do *log* dos *commits* dos sistemas, e extração das seguintes informações:
 - Data de criação do *commit*.
 - Mensagem do desenvolvedor.

Extração dos dados

1 Código fonte:

- Remoção das classes de teste.
- Geração do modelo FAMIX (arquivo MSE) para cada versão.

Extração dos dados

1 Issue Tracker

- Download das *issues* dos repositórios.
 - *Bugs*
 - *Fixed*
- Extração das informações relevantes:
 - Data de abertura.
 - ID do *bug*.

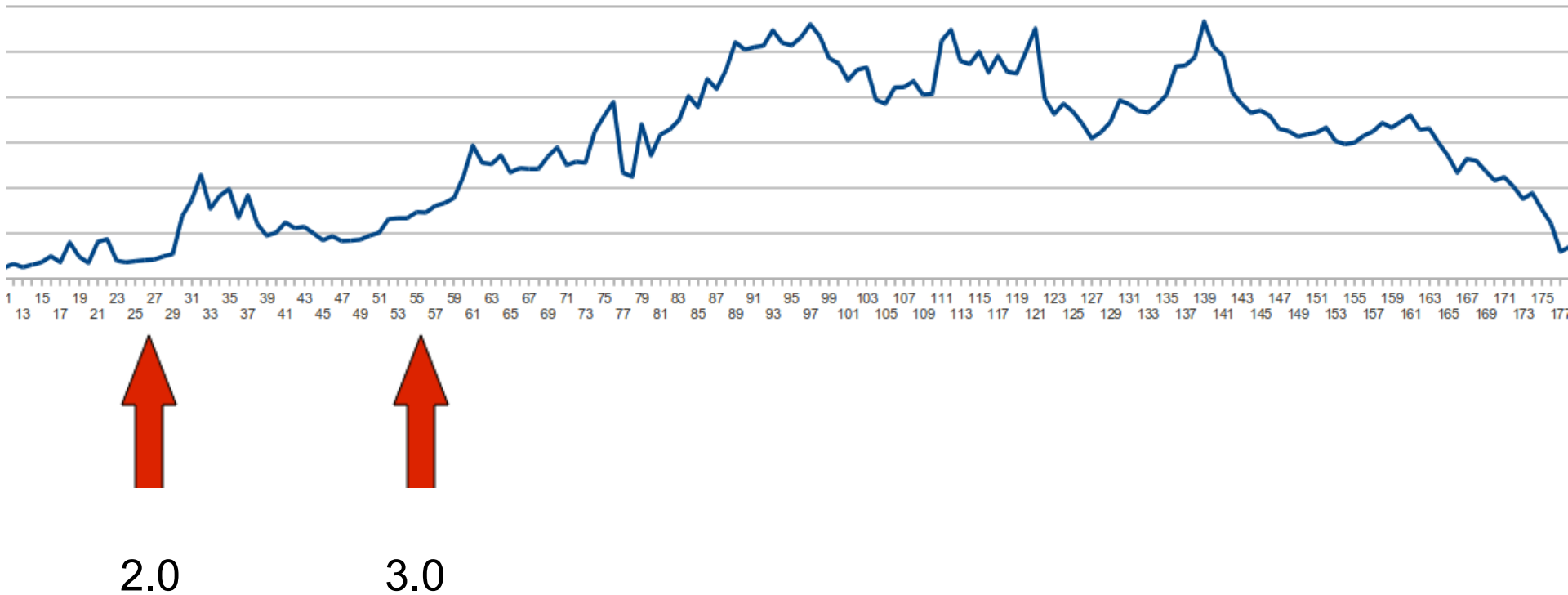
Extração dos dados

1 Extração dos dados:

- Mapeamento de *bugs* para classes.
 - Identificação do ID do *bug* na mensagem de *commit*.
 - Presença do ID na mensagem indica quais arquivos (classes) foram modificados para a resolução do *bug*.
- Cálculo do tempo de vida dos *bugs* em cada sistema.
 - Início: data de criação da *issue* no *issue tracker*.
 - Fim: data do *commit* onde o bug foi corrigido.
- Geração das métricas de código fonte (Plataforma Moose).

Análise dos dados

- Análise dos bugs do Eclipse JDT Core:



Conclusões

- Este trabalho estende o dataset criado por Dambros.
- Reúne várias informações sobre a evolução dos sistemas, e seus bugs.
- Pode ser usado para:
 - validação de técnicas de predição de bugs.
 - Análise de padrões de aparecimento de bugs.