Chapter 1

Znalezione

- http://www.slideshare.net/marcodambros/bug-prediction-and-analysis
- "Predicting defects for Eclipse" - Thomas Zimmerman. Rahul Premraj, Andreas Zeller.
  - McCabe complexity
  - Method LOC
  - Total LOC
  - Total LOC
  - Linear regression model
  - Pre-release defects
- "Mining metrics to predict component failures" - Nachiappan Nagappan
- "Improving defect prediction Using temporal deatures and Non Linear Models" - Abraham Bernstein
- "Predicting Faults Using the Complexity of Code Changes" - Ahmed E. Hassan
- "A Bug's Life" Visualizing a Bug Database - Marco D'Ambros
- Jak to si robi w google:
  - http://google-engtools.blogspot.com/2011/12/bug-prediction-at-google.html
  - BugCache for Inspections : Hit or Miss? - Rahman

Chapter 2

Klasteryzacja

Pomys polega na tym, aby uywajc istniejcych miar w repozytoriach projektu stworzy model predykcji za pomoc klasteryzacji oraz zbiorw przyblionych (Rough Sets).

2.1 Badanie

W tym rozdziale prezentujemy krtkie podsumowanie prowadzonych bada.

Dane

Dane, za pomoc ktrych przeprowadzamy wstpne badania, pochodz ze strony [3]. Jest to repozytorium uyte w artykule [2], dla porwnania wielu modeli predykcji defektw. Dlatego, one jak i sam artyku stanowi wspaniay wzorzec i porwnianie dla naszych wasnych bada. Coprawda repozytorium to zawiera drobny bd, ale przetworzenie ich daje nam 697 klas z opisujcymi ich metrykami oraz iloci defektw przed i po dacie "release".

Modele

Do powyszych danych zastosowano 4 modele okrelajce grupy klass.

Klasteryzacja

Uywajc poczonych metryk CK oraz OO wylistowanych klas, poddano w tren sposb otrzymany zbir klasteryzacji. W tym celu uyto narzdzia "Rattle" i algorytmu Klasteryzacji Hierarchialnej. Wynikiem byo 20 klastrw, ktrym nastpnie okrelono poziom defektywnoci wedug iloci defektw wykrytych przed zamkniciem projektu.

Ilo defektw

Tutaj take podizielono klasy na 20 grup ale cakiem innym sposobem. Podziau dokokonano dokadnie na podstawie iloci defektw odnalezionych w klasie przed zamkniciem projektu. Przy czym biorc pod uwag fakt, e klass, w ktrych nie wykryto adnych defektw, jest znacznie wicej, podzielono dodatkowo t grup na 12 grup. Podobnie jak poprzedni model, i ten poddano treningowi, okrelajc poziom defektywnoci dziki redniej iloci defektw w klasie danej grupy.

Rwne grupy

Wszystkie klasy podzielono rwno na 20 grup. Podziau dokonano wedug kolejnoci alfabetyczne. Czyli mniej wicej losowo. Podobnie jak poprzednie modele, i ten poddano treningowi za pomoc redniej iloci bugw wykrytych przed zamkniciem projektu wykrytych w klasach grupy.

Rwne grupy - defekty

W tym modelu, klasy take podzielono na 20 rwnych grup. Ale tym razem wedug kolejnoci wzgldem iloci defektw wykrytych przed zamkniciem projektu.

Wyniki

Jak to zostao ju odkryte w artykule [1] Zimmermana. Ilo defektw, odkrytych przed zamkniciem projektu, mocno wpywa na ich ilo znalezionych po. Dlatego widoczna jest wysoka jako predykcji niezalenie jaki model podziau na grupy zastosowalimy
TODO: wklei wyniki stworzonych modeli
TODO: porwna ich wyniki w miar sensowny sposb, tak aby dao to si jako porwna z Dambrosem.

Pewne symbole: DMC, LZ77, LZ78.

Bibliography

[1]

[2] M. D'Ambros, M. Lanza, and R. Robbes. An extensive comparison of bug prediction approaches. In *Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on*, pages 31 –41, may 2010.

[3] Marco D'Ambros, Michele Lanza, and Romain Robbes. Bug prediction dataset.