Predicting Change Impact from Logical Models

Sunny Wong and Yuanfang Cai
Drexel University
Philadelphia, PA USA
Motivation

• Traditional impact scope prediction algorithms
  – Generally only rely on syntactic dependencies
• Data mining (e.g. [Ying 04], [Zimmermann 04])
  – Identifies semantic dependencies
  – May be ineffective when
    • History is limited
    • Architecture is refactored
Approach

• Convert prevailing models into augmented constraint network (ACN) model [Cai 05]
  – Formalizes concept of design rules [Baldwin 00]
  – Assumption relation of ACN captures more than syntactic dependencies, but less than transitive closure [Huynh 08]

• Predict impact scope from ACN model
  – Purely ACN-based
  – Hybrid
Model Conversion
Model Conversion

MapSite_interface: {orig, other};
Room_interface: {orig, other};
EnchantedRoom_interface: {orig, other};
EnchantedRoom_impl: {orig, other};
...

Room_interface = orig ⇒ MapSite_interface = orig;
EnchantedRoom_interface = orig ⇒ Room_interface = orig;
EnchantedRoom_impl = orig ⇒ Room_interface = orig;
EnchantedRoom_impl = orig ⇒ Room_impl = orig;
...

(Room_interface, MapSite_interface);
(EnchantedRoom_interface, Room_interface);
(EnchantedRoom_impl, EnchantedRoom_interface);
Predict Impact Scope

A

B

D

Start

E

A

B

D

Start

E
Preliminary Evaluation

• Hadoop Core
  – 3 years of development
  – 15 minor releases (13 KSLOC → 53 KSLOC)
  – 309 modification tasks

• Reverse engineer class diagrams from code, convert to ACN

• Assess quality of predictions against actual solution
Assessing Quality

• Precision = \frac{\# \text{ correct}}{\# \text{ recommended}}

• Recall = \frac{\# \text{ correct}}{\text{solution size} - \text{change source size}}

• \( F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \)
$F_1$ Measure Results
Future Work

• Improve selection of top candidates in ACN-based approach to improve recall
• Compare against other impact scope prediction approaches
• Evaluate on addition software systems
Conclusion

• When system is young, with limited history
  – Both ACN-based and Hybrid approaches outperforms data-mining approach

• As system matures
  – Purely ACN-based approach degrades
  – Hybrid approach still comparable to data-mining approach