Large-Scale Patch Recommendation at Alibaba

Xindong Zhang¹, Chenguang Zhu², Yi Li³, Jianmei Guo¹, Lihua Liu¹, and Haobo Gu¹

1. Alibaba Group 2. University of Texas at Austin 3. Nanyang Technological University
Motivation

**50% time**
On average, 49.9% of software developers’ time has been spent in debugging.

**50% cost**
About half of the development costs are associated with debugging and patching.

Automated patch recommendation can significantly reduce developers’ debugging efforts and the overall development costs.
Challenges

Diverse Applications
Need a general approach

Insufficient test cases
Induce difficulty on patch validation

Lack patch labels
Accurate patch mining is difficult

Practical requirements
Highly responsible and low false positive
Solution

**Diverse Applications**

Patches are mined from internal codebase using generic features

**Insufficient test cases**

Independent of test cases and use developers’ feedback to validate and improve

**Lack patch labels**

Automatically mines bug and fix templates from historical changes

**Practical requirements**

Guarantee high responsiveness (scale of ms) and low false positive (22% and lower)
Commit fix#723 NPE check
author: Jack
+++++
- - - - -
+++++

- Commit message contains fix intentions
- 75% bug-fixing commits have such pattern:
  Delete bug snippet & Add patch snippet

PRECFIX

15 million commits
30 million files

Clustered Algorithm: DBSCAN
Clustered Strategy: Both defect & patch snippets
Optimization: Simhash-KDTree, API sequence
Similarity Comparison: Levenstein + Jaccard
Patch Category

API Modification

40%

Validation Check

26%

API Wrap

14%
**Results**

**EFFECTIVENESS**
False positive rate is 22% in patch discovery and it is supposed to be gradually reduced by feedback on discovered patch and contribution of new patch.

**EFFICIENCY**
Offline patch discovery costs 5 hours (extracting pairs, clustering, and extracting templates consumes 22, 270, and 5 min). Online patch recommendation is made within milliseconds.

**USER STUDY**
The majority (10/12) of the interviewed developers acknowledged the value of the patches, and all of them would like to see Prefix adopted in practice.

**DEPLOYMENT**
Prefix has been deployed in Alibaba for about one year so far. Every week, it recommends about 400 patches to developers on average, and receives about two to three false positive reports.