Improving Collaboration Efficiency in Fork-based Development

Shurui Zhou (shuruiz@andrew.cmu.edu)
Carnegie Mellon University

Fork-based development is a lightweight mechanism that allows developers to collaborate with or without explicit coordination. When the number of forks grows, it becomes difficult to maintain an overview of what happens in the community, which would lead to additional inefficient practices.

**Problem**
- Lack of awareness
- Lost Contribution
- Redundant Development
- Fragmented Community

**Solution**
- Identifying Natural Interventions from Best Practices
- Social Fork
- Hard Fork

**Analysis/Evaluation**
- [FSE 2019]
- [under submission]

**Empirical Study**
- Tools - Enhancing Awareness and Communication
- INFOX: Identifying Features from Forks
- Effectiveness
- Usefulness
- [ICSE 2018]
- [SANER 2019]
- [Future work]

**Identifying Features**
- Dependency graph for code changes (static analysis)
- Clustering features (community detection)
- Labeling features (NLP)

**Identifying Redundancies**
- Manually analyze duplicate PRs
- Developing clues as indicators
- Operationalization
- ML predicting redundancies

**Result**: Better modularity and centralized management are associated with more contributions and a higher fraction of accepted Pull Requests.

**Result**: Achieved 90% accuracy on a set of known features. Also, INFOX can provide actionable insight for developers of forks.

**Result**: Achieved 57–83% precision for detecting redundancies; saved 1.9–3.0 commits of effort on average.

Hi, your PR#24 might be a duplicate of PR#17.

Hire me! @shuishui blue