Nirikshan: Process Mining Software Repositories to Identify Inefficiencies, Imperfections, and Enhance Existing Process Capabilities

Monika Gupta, Ashish Sureka, Srinivas Padmanabhuni
Indraprastha Institute of Information Technology - Delhi, India

**RESEARCH MOTIVATION AND AIM**

“If you cannot measure, you cannot improve”

- Process capability can be improved by timely obtaining model of ongoing process.
- Discovering problematic patterns is difficult.
- Process mine vast data of software repositories using event log
- Implement solution approach on real world data.

**DATA SOURCE**

APIs, XML-RPC, JSON-RPC, Web Scraping

**SURVEY AND FIELD STUDY**

Validate initial hypothesis, Verify in-practice process and policies, Identify research problems

**TECHNICAL CHALLENGES**

- Mining from multiple perspectives such as control flow, time, case, organizational
- Gathering data from heterogeneous sources
  - Data for one process spanning multiple IS
- Mining hidden tasks
  - Activities in lifecycle not clearly identified
- Data incompatibility with process mining tools
  - No event log stored directly, and missing data

**TECHNIQUE**

- Process data from single repository
- Process spanning multiple repository
- Social network analysis, Team compatibility analysis, Distinguish LTC and novice, generalist and specialist
- Identify the required process capabilities and their impact on performance
- Design framework to help solve well known problems by applying process mining

**CONTRIBUTION**

- **Process mining**
  - Case ID, activity, ts, actor
  - ProM:
    - 1. Discover process model
    - 2. Performance analysis
    - 3. Conformance verification
  - Disco:
    - Recommendation system for workload management, team’s task allocation

- **Organizational perspective**
  - Social network analysis, Team compatibility analysis, Distinguish LTC and novice, generalist and specialist

- **Extend capabilities of SE tools**
  - Social network analysis, Team compatibility analysis, Distinguish LTC and novice, generalist and specialist

- **Application to solve SE problems**
  - Social network analysis, Team compatibility analysis, Distinguish LTC and novice, generalist and specialist

**PROPOSED SOLUTION**

- Data Extraction And Integration
  - Single Repository
  - Multiple Repositories
- Preprocessing
  - Survey professional to verify activities
  - Handle missing data. Ex: imputation
  - Order overlapping activities
- Process Mining
  - CaseID, Timestamp, Activity, Actor

**EVALUATION**

- Real Data
- Approach
- VALIDATE Survey professionals for feedback & improve

**REFERENCES**


**ACKNOWLEDGEMENTS**

Presented research work is supported by Prime Minister’s Fellowship awarded to the author. I acknowledge SERB, CII and Infosys Limited for their support.