International Workshop on BIG Data Software Engineering (BIGDSE’15)

“Software Analytics to Software Practice: A Systematic Literature Review”

Tamer Mohamed Abdellatif, Luiz Fernando Capretz, Danny Ho

Tamer Mohamed
PhD Candidate, Software Engineering
Agenda

- Overview
- Research Contribution
- Research Methodology
- Review Conduction and Results
- Observations and Recommendations
Overview

- Software analytics (SA) is a new branch of big data analytics (2011) [1].
- Differs from direct software analysis.
- Obtains valuable insights by analyzing and linking many software artifacts (not only source code).
- Covers different phases of software development process.
- Supports all stakeholders in decision-making.
- SA should provide visualization and useful interpretation.

Research Contribution

- Identify research state-of-art related to SA.
- Identify hot topics which can be considered for researchers’ future work.
- Conduct gap analysis which will guide researchers to elaborate and close gaps.
Research Methodology

- Review planning and protocol definition.
- Search period January 2000 to December 2014.

Research Methodology (Cont’d)

Planning
- Define SLR goal
- Define RQs
- Search strategy
- Study selection criteria:
  - Inclusion/Exclusion
  - Define Quality Assessment (QA) checklist
- Data extraction form
- Data synthesis methods

Review Conduction
- Primary studies selection
- Data extraction
- Data synthesis

SLR Reporting
- Results
- Limitations
- Observations and Recommendations

Review Protocol
Extracted Data
SLR Report
### Research Questions

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Quality Assessment Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1: Which software practitioners does the available SA research target?</td>
<td>QA1: Study contribution is clearly stated.</td>
</tr>
<tr>
<td>RQ2: Which domains are covered by SA studies?</td>
<td>QA2: Software artifacts that are used are clearly explained.</td>
</tr>
<tr>
<td>RQ3: Which software artifacts are extracted?</td>
<td>QA3: SA characteristics are clear, different from those of direct statistics and provide advanced insights.</td>
</tr>
<tr>
<td>RQ4: If different artifacts are used, are they linked together?</td>
<td>QA4: The results and application(s) are described in detail.</td>
</tr>
</tbody>
</table>

6/10/2015
### Research Methodology (Cont’d)

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA concepts were applied to extract insights from software project artifacts.</td>
<td>Studies that were irrelevant to software analytics.</td>
</tr>
<tr>
<td>Research was relevant to software project lifecycle phases.</td>
<td>Studies that were irrelevant to software projects.</td>
</tr>
<tr>
<td>Research was directly related to software industry and stakeholders.</td>
<td>Studies that were relevant to generic data analytics and were not directly relevant to SA or software artifacts.</td>
</tr>
<tr>
<td>For duplicate publications of the same study, the newest and most complete one was selected.</td>
<td></td>
</tr>
</tbody>
</table>
Review Conduction

Search for:
“Software analytics”
OR “Software analytic”
OR “Software development analytics”
OR “Software development analytic”

IEEE Xplore Library
- 41 Studies

ACM Digital Library
- 102 Studies

1st Filtration Phase (Inclusion/Exclusion)
- 41 Studies

2nd Filtration Phase (Apply quality assessment >= 50%)
- 19 Selected Studies
Review Results

*RQ1: Which software practitioners does the available SA research target?*
RQ2: Which domains are covered by SA studies?

- 58% Software Analytics Platform
- 21% Incident Management and Defect Prediction
- 11% Dashboard and Collaboration
- 5% Maintainability and Reverse Engineering
- 5% Software Effort Estimation
Review Results (Cont’d)

**RQ2: Which domains are covered by SA studies?**

1) Maintainability and Reverse Engineering

- Maleku: extracts information from software repositories.
- SQuAVisiT: visual software analytics tool, maintainability assessment and metrics.
- SAMAO: maintainability and visualization software analytics tool for mobile apps.

Observation:
- Work is similar to direct software analysis and literature metrics.
Review Results (Cont’d)

RQ2: Which domains are covered by SA studies?

2) Team Collaboration and Dashboard

- Baysal et al.:
  - Qualitative dashboard as a complement to quantitative dashboard.
  - Qualitative dashboard with metrics such as highlighting new comments.

- Observation:
  - Promising work but current features are very direct.
Review Results (Cont’d)

**RQ2:** Which domains are covered by SA studies?

3) Incident Management and Defect Prediction

- **Service Analysis Studio (SAS):**
  - Analyzes and links multiple data sources – such as performance counters and operating system logs.
  - Mining technique to find the suspicious execution patterns.
Review Results (Cont’d)

RQ2: Which domains are covered by SA studies?

4) SA Platform

- CODEMINE:
  - Software analytics common platform.
  - Common analytics framework for multiple client SA applications at Microsoft.
  - Provides data from different software artifacts such as source code and project schedule.
RQ2: Which domains are covered by SA studies?

5) Software Effort Estimation

G. Robles et al.:
- Effort estimation of OpenStack open-source project.
- Guesses effort roughly based on management system activities such as the time between two commits.
- Calibrates rough estimates based on other estimates collected by survey.
- Tackles open source projects estimation issue due to the distributive and collaborative nature.
RQ3: Which software artifacts are extracted?
Review Results (Cont’d)

RQ3: Which software artifacts are extracted?

- 9 (47%) studies analyzed only 1 artifact
  - 4 studies analyzed only source code.
  - The others analyzed artifacts such as code repository, issue tracker and call stack.
- The other studies analyzed artifacts that included emails, wikis, operating system logs and service transaction logs.
Review Results (Cont’d)

**RQ4: If different artifacts are used, are they linked together?**

- Valid for only 10 studies, where more than 1 artifact are analyzed.
- Only 8 studies scored 100% in QA3 and complied with SA concept.
Observations and Recommendations

- There is **scarcity** in available work.
- SA research work is still **embryonic**.
- Suggested open research areas:
  - Analyzing **multiple** artifacts.
  - Targeting **higher-level** business decision making like portfolio management, marketing strategy, and sales directions.
Future Work

- Extend our review by covering more digital libraries such as INSPEC, Google Scholar, and other digital libraries.
- The extended version will be submitted to ICSE’16.
Thank you