Analytics for Software Project Management –

Where Are We and Where Do We Go?

Guenther Ruhe

Joint paper with:
Maleknaz Nayebi
Roberta Cabral Mota Mujeeb Mufti
What do we mean by “Analytics for software project management?”
Types of Analytics (Kaisler et al. 2014)

- **Descriptive**: A set of techniques for reviewing and examining the data set(s) to understand the data and analyze business performance (what had happened?)
- **Diagnostic**: A set of techniques for determine what has happened and why (why did it happen?)
- **Predictive**: A set of techniques that analyze current and historical data to determine what is most likely to (not) happen (what will happen?)
- **Prescriptive**: A set of techniques for computationally developing and analyzing alternatives that can become courses of action – either tactical or strategic – that may discover the unexpected (what should happen?)
- **Decisive**: A set of techniques for visualizing information to facilitate human decision-making.
Project management

- Application of knowledge, skills, tools and techniques to project activities to meet the project requirements.
- Project management is accomplished through the application and integration of 47 logically grouped project management processes divided into five process groups: initiating, planning, executing, monitoring and controlling, and closing. (Duncan 2013).
Analytics for Software Project Management –

Where Are We and Where Do We Go?

- SM Study
- RQ’s
- Findings
- Discussion
Systematic mapping study selection process

- **Searching databases**: 15,406 papers
- **Apply inclusion criteria**: 7,306 papers
- **Apply exclusion criteria**:
  - Exclusion based on title & keywords: 320 papers
  - Exclusion based on abstract: 193 papers
  - Exclusion based on full-text: 115 papers

Action 15
Keywords used in electronic libraries

AND
{“Software Management”, “Project Management”, “Software Development”, “Software Project Management”}.

Inspec, Science Direct, Scopus, IEEE, ACM Dig. Library
Where the papers were coming from?
Research questions

RQ1 (Types of analytics):
What types of analytics has been used across the different software project management knowledge areas defined in the Software PMBOK?

RQ2 (Access to data):
To what extent was data used from open repositories or made publicly available?

RQ3 (Validation of results):
To what degree was validation done and if so, what was the percentage using real world data?

RQ4 (Reuse and replication):
How much are the retrieved papers (i) cross-referencing each other and (ii) using mutual datasets?
# Knowledge areas & analytical techniques

Distribution of papers across knowledge areas of SPM & types of analytical techniques

<table>
<thead>
<tr>
<th></th>
<th>Descriptive</th>
<th>Diagnostic</th>
<th>Predictive</th>
<th>Prescriptive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder management</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Procurement management</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Risk management</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Communication management</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Human resource management</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Quality management</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Cost management</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Time management</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Scope management</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Integration management</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>other</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14</strong></td>
<td><strong>11</strong></td>
<td><strong>56</strong></td>
<td><strong>34</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>
Pareto chart - publications across PMBOK knowledge areas

PMBOK knowledge areas

- Frequency
- Cumulative percentage
Availability of datasets

- Open data set
- Closed dataset
- Open data/total # of used datasets
Distribution of papers using validation with real vs. synthetic data sets
Cross-referencing

94 papers with no connection to each other

26 papers with connections

Classification of analytical questions

Buse, Zimmermann: Information needs for software development analytics. ICSE 2012: 987-996
## Comparative analysis

<table>
<thead>
<tr>
<th>Comparing</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance in practice</td>
<td>Descriptive</td>
<td>Diagnostic</td>
<td>Prescriptive</td>
<td>Predictive</td>
</tr>
<tr>
<td>Difficulty in practice</td>
<td>Predictive</td>
<td>Prescriptive</td>
<td>Descriptive</td>
<td>Diagnostic</td>
</tr>
<tr>
<td># of research papers</td>
<td>Predictive</td>
<td>Prescriptive</td>
<td>Descriptive</td>
<td>Diagnostic</td>
</tr>
</tbody>
</table>
Additional: Usage of hybrid techniques
Main findings - Discussion

- 93.9% of papers provide some form of validation.
- 37.3% made data openly accessible.
- Just 23% of the papers connected, replicated or reused previous models.
- Only 4% shared joined data.
- Open: Evaluation of industrial usefulness of results.
- Open: No trend from supporting developers towards also supporting managers.
References


