From Predictive Models to Confidence Factor in Software Projects considering aspects of Product, Process and People

Ayşe Başar Bener
Experts Summit - Siemens
June 9, 2010 Munich
Problems in the industry

• How can we effectively manage testing phase?
• When to stop testing?
• How can we effectively allocate resources to a project?
Problems in the Industry

• Decision making under uncertainty and risk management
• Confidence in 3Ps
  – Product
  – Process
  – People
Solution

• Building intelligent oracles
  – Data driven approach
  – Algorithm driven approach

• Empirical research
### Intelligent Oracles

NPV ~ Revenue (→ cost est) – Cost (→ quality: defects/ reliability)

<table>
<thead>
<tr>
<th>Algorithmic Approach</th>
<th>Machine Learning</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/ Effort Estimation</td>
<td>Defect Prediction</td>
<td>Reliability Modeling</td>
</tr>
<tr>
<td>Regression, Neural networks, SVM, clustering</td>
<td>Naive Bayes, Noise detection, dimensionality reduction, sampling strategies, X-company, classifier ensembles</td>
<td>Bayesian Nets, EM algorithm,</td>
</tr>
</tbody>
</table>

**Data Driven Approach**

- Social Networks
- Cognitive Science Metrics

**Metrics**

- Productivity Model: Product, size and process
- Defects, Churn, social network, static code, confirmation bias
- Product, people, process, defects
Cost/ Effort Estimation
Cost/Effort Estimation

- Understanding data structure?
  - Cross- vs. within-application domain
- Better predictor?
  - Point estimation
  - Interval estimation
  - Cost classification: dynamic intervals, classification algorithms, point estimates
- Case-based reasoning systems
Effort Estimation in a Case Study

- Company: IT company of a Turkish sub of a European Bank
- Identification of Project Features
  - Process, product, resource aspects
- Data Collection and Analysis
  - Questionnaire design
  - 17 features
- Calibration with Local Data
  - Local data repository: 29 projects
- Effort Estimation Model
  - 78% accuracy with less than 25% error
- Phase Based Effort Estimation
  - Effort estimation for SDLC phases
Defect Prediction
A problem has been detected and Windows has been shut down to prevent damage to your computer.

The problem seems to be caused by the following file: SPCMDCON.SYS

PAGE_FAULT_IN_NONPAGED_AREA

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:

*** STOP: 0x00000050 (0xFD3094C2,0x00000001,0xFBFE7617,0x00000000)

*** SPCMDCON.SYS - Address FBFE7617 base at FBFE5000, DateStamp 3d6dd67c
Problem

A Testing Workbench

- Uncertainty
- Confidence
  - Team
  - Process
  - Product
Problem Definition

• Software development lifecycle:
  – Requirements
  – Design
  – Development
  – Test (Takes ~50% of overall time)

• Detect and correct defects before delivering software.

• Test strategies:
  – Expert judgment
  – Manual code reviews
  – Oracles/ Predictors as secondary tools
Predictive Models

• Learning based
  – Using ML algorithms
  – Past data
    • Static code attributes
  – Defect data
    • Matched at file, module, or function level
Evaluation

<table>
<thead>
<tr>
<th>Early in a contractor/client relationship</th>
<th>precision</th>
<th>pd</th>
<th>pf</th>
<th>effort#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk adverse (e.g. airport bomb detection, morning sickness)</td>
<td>hi</td>
<td>hi</td>
<td>hi</td>
<td></td>
</tr>
<tr>
<td>Cost adverse (e.g. budget conscious)</td>
<td>med</td>
<td>lo</td>
<td></td>
<td>&lt; pd</td>
</tr>
<tr>
<td>Arisholm and Briand [2006]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\#effort = LOC in the modules predicted to be faulty
Data Collection: Prest

- A tool developed by Softlab
- Parser
  - C, Java, C++, jsp, PL/SQL
- Static Code Metric Collection
- Data Analysis
- Automated Defect Prediction
- Code.google
Other Data Used in Defect Prediction

Increasing Information Content By New Metric Sets

- Churn Metrics
  - Number of changes in code repository

- Developer Related Metrics
  - Top Developers
  - Unique edits in code history

- Social Network Metrics
  - Social Network of testers, developers and bug reporters
  - Metrics from the undirected graph [centrality, betweenness and degree]...

- Confirmation Bias
Defect Prediction: Case Study
About Software System

• The largest telecom operator in Turkey
• Two major components of the software system
• Predictions for all releases in six months
• Release period
  – One or two new functionalities and the rest is modifications/ upgrades for the current system
  – Every two weeks
  – 10 to 15 work packages
  – More than 400 GUIs of these components
• Testing process
  – 10 to 15 work packages with 1000-1500 test cases
  – 20 testers
  – 5 days on average
About Testing Process

• Ideal case:
  – Each tester needs to run 10 to 15 test cases per day (8 hours) → 80% of the functionality in total.
  – Manual inspections to ensure 100% test coverage

• In reality:
  – Only 3 days left for testing
  – Each tester can run 30 to 45 test cases
  – All test cases executed by 20 testers in 3 days → 48% of the functionality
## Benefits

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of executed test cases (per day)</td>
<td>10-15</td>
<td>9-13</td>
</tr>
<tr>
<td>Pre-release defects caught (%)</td>
<td>?</td>
<td>87%</td>
</tr>
<tr>
<td>The amount of functionality inspected (%)</td>
<td>48%</td>
<td>71%</td>
</tr>
<tr>
<td>Inspection effort (man-hours)</td>
<td>1.25</td>
<td>1.1</td>
</tr>
<tr>
<td>Post-release code defects (%)</td>
<td>59</td>
<td>32</td>
</tr>
</tbody>
</table>

Decrease by 11.2%

Decrease by 44%
Benefits

Trend of post-release code defects

Model calibration
Reliability Prediction
Reliability Prediction

- When to stop testing? When to release the software product?
  - Predict the final reliability, i.e. defect proneness, of a software product

- Build sub-networks for each software process
  - Requirements, design, development, test
  - Identify features affecting the overall process quality
  - Causal relationships

- Bayesian Network
  - Inference algorithms
Predictions: Case Study in Telecom

- We build the network and test on three consecutive releases
  - Output: Number of medium and high production defects for each subnet
    - Discretized as intervals (8 < CF < 15: Level 2)

<table>
<thead>
<tr>
<th>Releases</th>
<th>Testing (Predicted/Actual)</th>
<th>Requirements (Predicted/Actual)</th>
<th>Development (Predicted/Actual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.59</td>
<td>(3&lt;defects&lt;7) 3</td>
<td>(&lt;8) 5</td>
<td>(&lt;29) 28</td>
</tr>
<tr>
<td>2.60</td>
<td>(3&lt;defects&lt;7) 4</td>
<td>(&lt;8) 4</td>
<td>(&lt;29) 15</td>
</tr>
<tr>
<td>2.61</td>
<td>(7&lt;defects&lt;11) 6</td>
<td>(&gt;15) 34</td>
<td>(29&lt;defects&lt;40) 32</td>
</tr>
<tr>
<td>2.62</td>
<td>(7&lt;defects&lt;11) 8</td>
<td>(&gt;15) 17</td>
<td>(29&lt;defects&lt;40) 31</td>
</tr>
<tr>
<td>2.63</td>
<td>(7&lt;defects&lt;11) 9</td>
<td>(&gt;15) 16</td>
<td>(&lt;21) 21</td>
</tr>
<tr>
<td>2.64</td>
<td>(&gt;11) 8</td>
<td>(8&lt;defects&lt;15) 15</td>
<td>(&gt;40) 49</td>
</tr>
<tr>
<td>2.65</td>
<td>(&gt;11) 14</td>
<td>(&lt;8) 7</td>
<td>(29&lt;defects&lt;40) 30</td>
</tr>
<tr>
<td>2.66</td>
<td>(&gt;11) 6</td>
<td>(8&lt;defects&lt;15) 15</td>
<td>(&gt;40) 62</td>
</tr>
<tr>
<td>2.67</td>
<td>(&lt;3) 2</td>
<td>(&lt;8) 7</td>
<td>(&gt;40) 50</td>
</tr>
<tr>
<td>2.68</td>
<td>(&lt;3) 1</td>
<td>(8&lt;defects&lt;15) 10</td>
<td>(&lt;29) 28</td>
</tr>
</tbody>
</table>
Going Forward...

Building an integrated product offer
Transition from research to industry: scale
The Team

- Ayşe Başar Bener, PhD
- Turgay Aytaç, MSc
- Burak Turhan, PhD
- Ayşe Tosun, MSc
- Gül Çalıkılı, MSc
- Bora Çağlayan, MSc
- Ekrem Kocagüneli, MSc