Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

2016 © Microsoft Corporation. All rights reserved.
AUTONOMOUS AI-BASED MEASUREMENT SYSTEMS
SOFTWARE MEASUREMENT RESEARCH @ CHALMERS | UNIVERSITY OF GOTHENBURG

MIROSŁAW STARON & TEAM
COMPUTER SCIENCE AND ENGINEERING, CHALMERS | UNIVERSITY OF GOTHENBURG
Who am I?

- Professor in Software Engineering at Chalmers | University of Gothenburg

- Specialization in software measurement
  - Autonomous artificial intelligence based measurement
  - Measurement knowledge discovery
  - Simulation of outcome before decision formulation
  - Metrological foundations of measurement reference etalons

- Actively working with the standards
  - ISO/IEC 25000 (series) - Software Quality Requirements and Evaluation (SQuaRE)
  - ISO/IEC 14598 - Information Technology - Software Product Evaluation
  - ISO/IEC 26262 – Road Vehicles – Functional Safety
Our research team

Academia

- Miroslaw Staron
- Miroslaw Ochodek, PhD, Machine learning
- Peter Pickerill, Joshua Jungen (MSc)
- Sam Halali (MSc)

Industry

- Wilhelm Meding (Ericsson)
- Per Sundvall, Peter Eriksson, Jimmy Nilsson, Micael Caiman (Ericsson)
- Kent Niesel, Sajed Miremadi (Volvo Cars)
- Jonas Landgren, Christoffer Höglund (Saab EDS)
- Poupak Banishad, Anders Henriksson (Volvo Group Truck Technology)
- Ola Söder, Magnus Bäck (Axis Communications)
- Darko Durisic (Volvo Cars)
- Gert Frost, Brian Dalby (Grundfos)

PhD alumni:
- Niklas Mellegård (RISE)
- Rakesh Rana (Nordea)
- Vard Antinyan (Volvo Cars)
- Darko Durisic (Volvo Cars)
Software Center – a collaboration between 11 companies and 5 universities

- We work together to accelerate the adoption of novel approaches to software engineering
- Our mission with the Software Center is to contribute to maintaining – and strengthen – Sweden’s leading position in engineering industrial software-intensive products.
Ability to work in collaboration is a key success factor

Company-to-company-to-academia research environment

- Working on site at a company
  - Initiates new ideas and research challenges
  - Enables direct validation of ideas and results
  - Reduces research cycle times
  - Constant collaboration allows to reduce the effort for the companies

- Enabling industry to evolve measuring
  - Knowledge and competence
    - ISO 15939
    - Information quality
  - Tools and methods
    - Code stability (E) ->
    - Change Waves (SAAB) ->
    - Defects (GTT) ->
    - Dashboards (VCC) ->
    - Implicit architectural dependencies (E) ->
    - Test selection (Axis) ->
    - Identifying Risky Code fragments (E) ->
    - Identifying Risky Requirements (GTT) ->
    - ...
    - Performance profiling (Sony) ->
    - Stress test of measurement programs (Grundfos) ->
    - ...

[Diagram showing data flow and metrics]
Ability to work on site is a key success factor

Industrial impact

- > 40 000 measurement systems
  - Automated data collection, analysis and visualization
- Self-healing measurement infrastructure
  - Saves hours of effort per week
- Information Quality
  - Prevents wrong decisions
- Estimations of cost of standard implementation
  - Saves engineering hours
- Pinpointing low quality requirements
  - Saves engineering hours
Why I got into metrics

Modern usage of software metrics

• Autonomous artificial intelligence based measurement

• Measurement knowledge discovery

• Simulation of outcome before decision formulation

• Metrological foundations of measurement reference etalons
Our timeline

- Automated Information Quality
- Self-healing of measurement systems & Release readiness
- KPI Quality & 1 000 metrics in portfolio
- Software Analytics

2006:
1 company
1 university
1 manual measurement system

2008:
4 companies
2 universities
4 000 automated measurement systems

2010:
4 companies
2 universities

2012:
Robust measurement programs

2014:
> 40 000 automated measurement systems

2016:
7 companies
2 universities

2017:
7 companies
2 universities
AUTONOMOUS ARTIFICIAL INTELLIGENCE BASED MEASUREMENT SYSTEMS
Measurement infrastructure

Measurement programs

- *Measurement programs* in industry are socio-technical systems where the technology interacts with stakeholders in order to support their goals
  - E.g. Collaboration with Ericsson resulted in the development of over 40,000 measurement systems; collaboration with Volvo Cars resulted in new dashboards, courses (movies)

- We cover base measures, derived measures and indicators in the areas of
  - Product (e.g. size and complexity)
  - Process (e.g. KPIs)
  - Organization (e.g. development speed)
  - Project (e.g. release readiness)

- Our work leads to
  - More informed, accurate decisions
  - More precise and deeper insight
  - Early warnings and problem avoidance
The basis of modern measurement program
Self-healing measurement systems

- **Measurement system** is a set of measuring elements assembled together to measure quantities of a specific kind
  - Our self-healing mechanisms provide robustness to the measurement systems
  - Robustness leads to proper balance of resources
    - Resources used for development (value-adding) of measurement program
    - Minimal resources spent on maintenance of the measurement systems
  - Robust measurement systems lead to trustworthy measurement programs
    - Which we complement with fully automated information quality assessment
  - Increase reliability (MTBF) from days to months
    - Decrease the cost of maintenance from hours/week to minutes/week
Measurement systems – examples
Reliable self-healing **Information Products**

**Information Quality**

- Information quality assessment provides the stakeholders with reliable information
  - Information quality is a foundation for data veracity
  - Our automated assessment allows the metrics team to daily ensure that all 40,000 measurement systems provide veracious results
    - When error occurs the stakeholder are notified and the metrics team gets details of what has gone wrong

![Diagram of information quality assessment process]
Automated assessment Information Quality

- We use *data flow* theories to assess quality of the measurement information.

- Monitoring of data flow enables data veracity.
Machine Learning map
AI-based measurement

- We study the use of machine learning to
  - Identify behavior of SW code → finding where the relevant code is
  - Classify which defects are important, based on their description, to save time for analysis
  - Identify bottlenecks in continuous integration, based on integration stop-patterns
  - Identify which KPIs should be removed because they do not provide any value
AUTOMATED MEASUREMENT INFRASTRUCTURE ENABLES EFFICIENT DEVELOPMENT OF HIGH IMPACT MEASURES AND INDICATORS
Assessing company-wide measurement programs

Robustness of measurement programs

- Modern robust measurement programs
  - Cover a broad spectrum of measures
  - Are supported by professional metric teams
  - Use solid infrastructure
  - Support decisions taken in the company
  - Contribute to organizational maturity

- Our robustness assessment method (MeSRAM) identifies weak spots in measurement programs
  - Pinpointing improvement areas

- 7 companies assessed their measurement program
  - One company increased its robustness by 300% within one year
Documenting the value of each measure

Metrics portfolio

- Cataloguing measures in software engineering
  - Literature
  - Empirical (industry)
  - New measures based on unaddressed information needs

- Currently ca 2000 measures

- Our portfolio provides us with the possibility to quickly answer questions like *if I am an architect, what should I measure?*
Robust infrastructure enables succinct indicators

Predicting events and scenarios instead of numbers

Predicting *defects per month* for the entire project
- : long-term perspective
- - : volatile to changes

Machine learning predicting *defects per week* for 3 weeks in advance
- : high accuracy (92%)
- - : short-term perspective

Machine learning predicting *defects trend* for 1 week in advance
- : predicting trends
- - : short-term perspective
Beyond Machine Learning
Predicting when the product is ready for release

- Release readiness predicts when we can ship/deploy the product
  - Our method enables us to assess how reactive an organization is
    - Short release readiness time leads to frequent releases
  - We provide the possibility to tune the organization towards customer value
    - Shorter release readiness increases operational performance

Indicator forecasts when the product is ready for release given the current development speed

$$RR = \left( \frac{\#\text{defects}}{\text{defect\_removal\_rate} - (\text{test\_execution\_rate} - \text{test\_pass\_rate})} \right)$$
Finding needle in a haystack
Using heatmaps for analyzing large code bases

• Problem:
  – Does the code development follow set practices?
  – Does the code base stabilize towards the end of the project?
  – Are there any areas which can potentially be risky w.r.t.
    • Too intense development over short time?
    • Develop-pause-develop – indicating defect fixes?
    • ...
  – Which code areas should be tested first?

• Results
  – Using heatmaps based on code-churns to visualize code stability

• Impact
  – Saved time for test design – emphasizing important areas
  – Used to predict “forgotten” changes in the code

QUASAR@Car
Predicting the cost of standard change

• Problem: Efficiently manage the evolution of large software systems based on the evolution of domain-specific meta-models (AUTOSAR meta-model).

• Results: methods and tools for automated
  – analysis of the domain-specific meta-model changes for different roles,
  – estimation of cost and time to adopt the changes in the used modeling tools and
  – prediction of the impact of the changes to the existing models and requirements.

• Impact
  – Weeks of analysis -> minutes
  – Used at two companies
Requirements review: Rendex
Requirements quality model

- Problem: How to automatically assess which requirements need improvement?

- Results
  - Rendex quality model
  - Rendex tool for requirement quality measurement
  - Evaluation at five companies

- Impact
  - Used at three companies
  - Days of review work -> minutes
Machine learning

Finding design guidelines violations

- Problem: How to discover violations of design guidelines?

- Results
  - CCFlex tool for line counting

- Impact
  - Validated at two companies
  - 70% - 99% accuracy
  - Weeks of review -> minutes
NEW MEASURES
OPEN UP FOR MODERN VISUALIZATION TECHNIQUES
To have a good visualization we need to understand the organizational needs. How organizations talk measurement:

- Modern and future organizations leave the top-down dissemination towards bottom-up and horizontal patterns.
  - Our work captures these patterns into a dashboard selection model.
  - Our model provides efficient and effective selection of
    - the right visualization of
    - the right measure to
    - the right stakeholder.
Novel measurement requires novel visualization

Dashboard selection model

- Our model catalogues over 100 visualization models
  - Able to address over 100 different mechanisms to visualize the measurement data
  - Our selection model increases the speed of adoption of the visualization
    - Links the model to examples
    - Links the examples to the "Metric portfolio" through reference measures

<table>
<thead>
<tr>
<th>Type</th>
<th>Report</th>
<th>Dashboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data acquisition</td>
<td>Manual</td>
<td>Automated</td>
</tr>
<tr>
<td>Stakeholders</td>
<td>Individuals</td>
<td>Group</td>
</tr>
<tr>
<td>Delivery</td>
<td>Fetched</td>
<td>Delivered</td>
</tr>
<tr>
<td>Update</td>
<td>Periodically</td>
<td>Continuously</td>
</tr>
<tr>
<td>Aim</td>
<td>Information</td>
<td>Decision support</td>
</tr>
<tr>
<td>Data flow</td>
<td>Raw data</td>
<td>Indicators</td>
</tr>
</tbody>
</table>
Modern provisioning model
Measurement-as-a-Service (MaaS)

- MaaS is
  - a measurement licensing and delivery model in which
  - measures are licensed on a subscription basis, centrally hosted, collected and delivered on demand

- Reduces the cost of measurement by 50% through
  - centralizing the storage of data,
  - measurement competence, and
  - resources for measurement
Keeping up with modern dissemination technology

**MetricsCloud**

- Our metrics cloud disseminates the measurement results on all levels
  - Vertical inside one organization
  - Horizontal over similar roles

- Our solution leads to resilient measurement results
  - Local storage
  - Distributed computation
  - Global synchronization
WRAP-UP
Wrap-up

Modern usage of software metrics

- Autonomous artificial intelligence based measurement
- Measurement knowledge discovery
- Simulation of outcome before decision formulation
- Metrological foundations of measurement reference etalons
WRAP-UP
Machine learning

Finding design guidelines violations

- Problem: How to discover violations of design guidelines?
- Results
  - CCFlex tool for line counting
- Impact
  - Validated at two companies
  - 70% - 99% accuracy
  - Weeks of review -> minutes