Measurement and Modeling in Software Engineering

Lionel C. Briand, P. Eng.

Canada Research Chair in Software Quality Engineering
Department of Systems and Computer Engineering
Carleton University, Ottawa, Canada

May 16th, 2005

General definitions

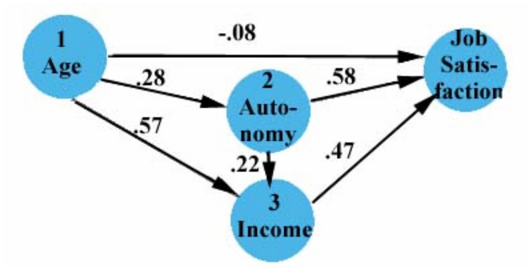
- *Measurement:* A logical rule for assigning numbers to observations to represent the quantity of a trait or characteristic possessed.
- *Modeling:* A theoretical construct that represents physical, biological or social processes, with a set of variables and a set of logical and/or quantitative relationships between them.
- Measurement of software product and process attributes and (Empirical) modeling

2

Models: A variety of purposes, forms, etc.

- Many purposes: assess, decide, predict, simulate
- Used throughout the natural and social/economic sciences (we can benefit from them)
- Models take a variety of forms …
  - Mathematical functions, relations (e.g., reliability)
  - Stochastic processes (e.g., Markov model of performance)
  - Rules of inference (e.g., fault-prone components)
  - Decision models, e.g., integration order
  - Cause-effect relations (e.g., pair programming decreases fault density)
  - Simulation model (e.g., cost-effectiveness)
- Quantitative or Qualitative
- Deterministic or statistical

Cause-Effect Modeling using Path Analysis



Examples of qualitative models

- *Psychology: Myers-Briggs personality type* is a representation of a person's preferences, using four scales. These scales can be combined in various ways to produce 16 personality types. Types are typically denoted by four letters--for example, INTJ (Introverted intuition with extraverted thinking).
- *Political science:* Political movements that take hold in one country are likely to spread to geographically neighboring ones.

Models: Structuring the debate

- An *argumentative framework* for applying logic and mathematics that can be independently evaluated (for example by testing) and that can be applied for reasoning in a range of situations.
- A *reusable tool* for discovering new facts, providing systematic logical arguments, as explicatory or pedagogical aids and for evaluating hypotheses
- In all domains, this often leads to hotly contested debates in the academic world

SE Applications of Measures and Models

- Assess "ilities"
  - Reliability
  - Testability
  - Maintainability

Decision models

- Predict
  - Dependent variables, e.g., maintainability
  - Explanatory variables, e.g., complexity

Prediction models

- Quantify variables of experiments
  - Diagnosability of a failure (i.e., ease of locating faults)

Causal models

Desirable properties

- A Measure must be
  - Based on explicit and justified assumptions
  - Objective
  - Repeatable (Reliable)
  - Precise
  - Unbiased (no systematic error)
- Depending on the objectives, a model must be
  - Based on explicit and justified assumptions
  - Work under all possible or well-defined conditions of application
  - Congruent with reality
  - Easy to use (e.g., no variable difficult to measure or estimate)
  - Easy to interpret (e.g., complex interaction expressions)
  - Accurate for predictive models

Repeatability implies an operational definition

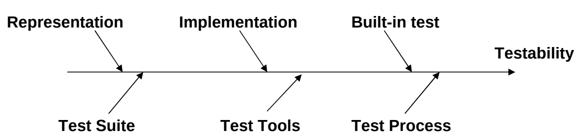
- An *operational definition* of a quantity is a specific process whereby it is measured.
  - How is effort data collected? (what activities does that include?)
  - How is polymorphism accounted in measuring coupling?
- Needed to ensure consistent measurement across studies
- Often missing in SE measurement papers!
- Difficult for qualitative measures …
  - How to we prepare the interviewee/respondents?
  - What information do we provide to them?

Issues in Soft. Eng. Measurement

- Many attributes are not easy to quantify
  - Internal product quality attributes, e.g., cohesion
  - External product quality attributes, e.g., testability
  - Human perception, e.g., too usability
- Qualitative measures for many important attributes
- Multi-dimensional attributes, e.g., testability
- Validity of measures is context-dependent
  - Development process and artifacts
  - Application domain
- No guidelines for writing precise assumptions

Testability

- At a high level of abstraction, software testability is thought as the result of six factors (sub-attributes)



Binder, 1996

There are many ways a measure can be wrong

- Poor construct validity, e.g., strange properties in terms of measurement theory
- Low reliability (for qualitative measures)
- Lack of precision
- Data is too expensive or impossible to collect
- Difficult to use (e.g., no benchmark for decision making)
- Not a good predictor

Limits of measurement

- Can crucial decisions in SE be made *only* or *primarily* on the basis of what can be quantitatively measured?
- Not everything that counts can be counted. Some matters can only be judged, that is to say they can only be assessed in a qualitative way.
- Role of qualitative measurement?
- "The root of the difficulty is one of perception. Quantitative measurement appears to be objective and value free. Qualitative assessment appears to be subjective and value laden. In fact, quantitative measures – whether in the form of a funding formula or of a performance indicator – contain and conceal important value judgments. " (public policy)

Carleton UNIVERSITY
Canada's Capital University

squall

13

Issues in Soft. Eng. Modeling

- Causal model
  - Cannot prove causality
  - Theory + correlational evidence
  - Experimental approach (not always feasible)
- Assessment
  - Assessment is often relative to a benchmark
  - Benchmarks are not always easy to devise
- Decision making
  - Multi-criteria
  - Evaluation/Fitness function
  - Optimization problem (e.g., evolutionary techniques)
- Prediction
  - Many explanatory variables
  - Interactions among explanatory variables
  - Complex relationships

Carleton UNIVERSITY
Canada's Capital University

squall

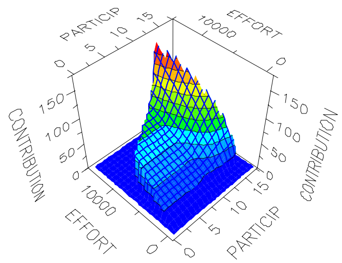
14

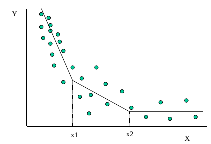
Modeling techniques

- A wide variety of techniques with different purposes, strengths and weaknesses ….
  - Statistics (regression)
  - Machine learning (decision trees)
  - Evolutionary algorithms (Genetic algorithms)
  - Simulation (Monte-Carlo, System dynamics)
  - Bayesian networks, Neural networks
  - Analytical Hierarchical Process
  - MARS: Multiple Adaptive Regression Splines
- Selection factors
  - Purpose: classification, continuous prediction, decision, optimization …
  - Variables: type, number, interactions, relationships
  - Data: amount, missing data, …

15

Examples: MARS to model inspection effectiveness



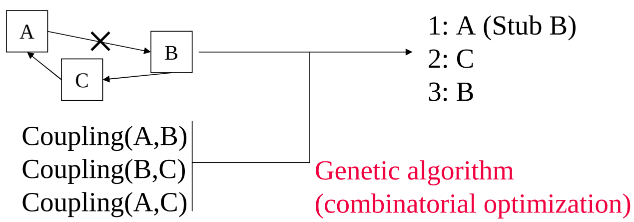


Briand, Freimut, Vollei, 2004

16

Genetic Algorithms and Coupling measures

- Many classes, dependencies, and cycles
- Devise optimal integration order on OO systems in order to minimize stubbing effort (Decision model)
- Coupling measures as surrogate cost measures for stubbing (Briand, Feng, Labiche, 2003)



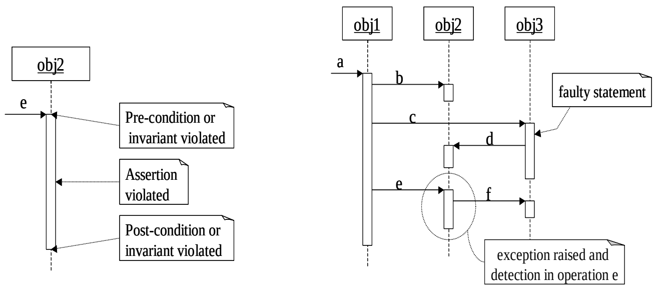
17

Context assumptions must be explicit

- Reliability models: fault correction? fault introduction?
- Diagnosability: representative fault diagnosis process?
- Example: diagnosability as distance measure on sequence diagrams (Briand, Labiche, Sun, 2003)
- Context assumptions are necessary to
  - Select important dimensions for an attribute, and therefore appropriate measures
  - Decide of proper ways to measure in context
  - Decide of proper model techniques and structures
  - Ensure the validity of a measure or model

18

Diagnosability measure



Exception raised	Methods analyzed	Measure
at e's entry	a, b, c	3
during e, before the call to f	e, a, b, c	4
during e, after the call to f	e, f, a, b, c	5

Dealing with "Missing" Model Variables

- Missing model variables: cannot be measured or easily estimated
- Compromise between cost of data collection and accuracy: Simplifying assumptions
- Example: Cost-effectiveness model of fault-proneness models to drive inspections (Briand & Wuest, 2002)
- Alternative: expert-based estimates
- Example: Cost-effectiveness model of inspections (Briand , Freimut, Vollei, 2000)

State of the Art

- Many measures for internal attributes of software code and design, but it is still difficult to determine what to use when attempting to measure an external quality attribute, e.g., testability.
- Many data analysis and modeling techniques, but we are probably not fully making use of what is available.
- Reliability models have been widely studied, but other quality aspects are rather neglected, from a measurement standpoint.
- Still no standard way to report on SE empirical, measurement-based work.
- But the level of maturity / sophistication of the research work is definitely increasing.
- What about the practice of software measurement?

Carleton UNIVERSITY Canada's Capital University

squall

Lessons learned

- Defining a measure / model is not simply a mathematical exercise: context and simplifying assumptions
- A measure / model must be clearly justified based on process, product information in context
- The validity & usability of a measure / model often varies with context
- Expect wide variability across environments, but as a science we must be able to generalize to develop a body of evidence
- Assumptions must be explicit and investigated empirically
- Measures must be defined to be useful in terms of specific objective (assessment, decision making, prediction). There are many interesting but useless measures (e.g., not applicable). What model will your measure be part of?

Carleton UNIVERSITY Canada's Capital University

squall

Models in other domains

- "Models, at best, are flawed maps of an uncertain terrain—abstractions of reality."
- "Find the model that best approximates your world...so the more models the better. Be aware of when your model is wrong and the corrections that need to be made. Use variables that are intuitive and easily understood. Make sure the models are calibrated properly."
- "How can we know when a model is wrong? How can we make it consistent? What are the criteria for judging models? Too much attention to complicated models is a dangerous waste of time."
- Application domain?

Carleton UNIVERSITY
Canada's Capital University

squall

23

Vic's Work on Measurement & Modeling

- He emphasized the scientific method, and therefore measurement, as a means to achieve quality and productivity improvement in software engineering (*Quality Improvement Paradigm*).
- He suggested a way to link corporate measurement programs to context assumptions, measurement objectives (*Goal Question Metric Paradigm*).
- He proposed an organizational model to implement measurement programs in a typical software development setting (*Experience Factory*).
- A large body of measurement and experimental works …
- He created a *school of thought*, that has helped spread the application of the scientific method throughout software engineering research.

Carleton UNIVERSITY
Canada's Capital University

squall

24

Conclusions

- Measurement and modeling in Software Engineering: A growing community, increasingly part of main stream.
- Much progress has been made in 20 years (quality of research, body of knowledge).
- Still difficult to generalize and converge towards usable bodies of evidence.
- CS graduates (and researchers) still mostly ignorant of the scientific method, measurement and modeling issues.
- We are not *fully* comparable to any other social or natural science with respect to measurement & modeling. We can learn a lot from them, but we need to devise our own methodologies and standards.

Carleton University
Canada's Capital University

squall

25