CSEP504:

Advanced topics in software systems

- Tonight
 - India trip report
 - Apologies and status
 - Evaluation approaches for software engineering research
 - Software engineering economics

David Notkin • Winter 2010 • CSEP504 Lecture 7

India trip report

- Microsoft Research India
- 3rd India Software Engineering Conference
- $\succ \mathsf{SEA} \Rightarrow \mathsf{CDG} \Rightarrow \mathsf{BLR}$
 - Microsoft Research India guest house
- > Car and driver Bangalore \Rightarrow Mysore
 - Infosys campus
- > Car and driver Mysore \Rightarrow Mysore Palace \Rightarrow BLR
- $\succ \mathsf{BLR} \Rightarrow \mathsf{CDG}$
 - Radisson Blu
- $\succ \mathsf{CDG} \Rightarrow \mathsf{SLC} \Rightarrow \mathsf{SEA}$

Microsoft Research India

- Recently celebrated its first 5 years
- Close connections with
 Indian Institute of Science
- Strong External Research Program
- Ties with Microsoft India Development Center (Hyderabad) and soon with Yahoo
- 50-70 technical staff
 - Double that in summer (interns)

- Algorithms Research Group
- Cryptography, Security, and Applied Mathematics
- Mobility, Networks, and Systems
- Multilingual Systems
- Rigorous Software Engineering
- Technology for Emerging Markets
- Vision, Graphics, and Visualization
- Advanced Development and Prototyping

Rigorous software engineering

- Akash Lal
- Aditya Nori
- Sriram Rajamani
- Kaushik Rajan
- Ganesan Ramalingam
- Venkatesh-Prasad Ranganath
- Kapil Vaswani

Recent publications (selected)

- Kaushik Rajan, Sriram Rajamani, and Shashank Yaduvanshi, GUESSTIMATE: A Programming Model for Collaborative Distributed Systems. PLDI 2010
- Prakash Prabhu, G Ramalingam, and Kapil Vaswani, Safe Programmable Speculative Parallelism. PLDI 2010
- Aditya V. Nori and Sriram K. Rajamani, An Empirical Study of Optimizations in Yogi, ICSE 2010
 - Yogi: a scalable software property checker that systematically combines static analysis with testing.
- Nels E. Beckman, Aditya V. Nori, Sriram K. Rajamani, Robert J. Simmons, Sai Deep Tetali, and Aditya V. Thakur, Proofs from Tests. IEEE TSE 2010

Recent publications (con't)

- Dawei Qu, Abhik Roychoudhury, Zengkai Lang, and Kapil Vaswani, Darwin: An Approach for Debugging Evolving Programs. ESEC/FSE 2009
- B. Ashok, Joseph Joy, Hongkang Liang, Sriram Rajamani, Gopal Srinivasa, and Vipindeep Vangala, DebugAdvisor: A Recommender System for Debugging, ESEC/FSE 2009
- Benjamin Livshits, Aditya V. Nori, Sriram K. Rajamani, and Anindya Banerjee, Merlin: Specification Inference for Explicit Information Flow Problems. PLDI 2009
- Trishul Chilimbi, Ben Liblit, Krishna Mehra, Aditya V. Nori, and Kapil Vaswani, Holmes: Effective Statistical Debugging via Efficient Path Profiling. ICSE 2009
 - Holmes: a statistical tool to find the most likely cause of test failures by collecting and analyzing fine-grained path coverage data and identified code paths that strongly correlate with failure

Other projects

- Mining API specifications quantified temporal rules
- Debug Advisor a search using fat multi-dimensional queries (KBs of structured and unstructured data describing the contextual information) to find similar bugs and related information– people related to it, relevant source and binary files, etc.
- Shadowed upgrades
- ...much more!



- Conference developed primarily by Pankaj Jalote and Sriram Rajmani – build a stronger software engineering research community in India
 - Hyderabad, Pune, Mysore, Kerala, ...
- Three legs
 - Reviewed research papers, posters, etc.
 - Keynotes
 - ESEC/FSE and ICSE best paper presentations
- Mysore 2010: at Infosys training and education campus; about 200 attendees at ISEC



- Infosys founded 1981, now over 100K employees
 - Business and technology consulting, application services, systems integration, product engineering, custom software development, maintenance, re-engineering, independent testing and validation services, IT infrastructure services and business process outsourcing
- In 2007, received ~1.3M applications and hired ~3%
- NYSE INFY ADR: market cap of ~US\$34B; 2009 revenue about US\$4.7B, 11.7% growth

Infosys Mysore campus







UW CSE P504



Comments

- "Too many kinds of cookies in the same box."
- "Like Disneyland without the rides."

SE Research Center roundtable

- Should India have a software engineering research center something like the CMU Software Engineering Institute, Fraunhofer Institute, etc.?
- Most interesting point to me: why aren't more students in India interested in software engineering research?



- Me
- Kris Gopalakrishnan (CEO/MD Infosys)
- William Cook (UT Austin)





Best papers ESEC/FSE and ICSE

- Does Distributed Development Affect Software Quality? An Empirical Case Study of Windows Vista
 - Christian Bird, Nachiappan Nagappan,
 Premkumar Devanbu, Harald Gall, Brendan Murphy
- Asserting and Checking Determinism for Multithreaded Programs
 - Jacob Burnim, Koushik Sen
- DARWIN: An Approach for Debugging Evolving Programs
 - Dawei Qi, Abhik Roychoudhury, Zhenkai Liang, Kapil Vaswani

My perspective

- India's software engineering and software engineering research communities are vibrant
 - I heard some fascinating stories of start-ups leveraging the "cloud"
- There are educational and funding issues to address – real, but not insurmountable

Recap and status

- Lectures: tonight is the last one
- Grading: Sai has been on top of the structured reports; I have not been on top of the state-of-theresearch papers – this week's job
- Deadlines remain the same: I have some give on the March 14th deadlines, if needed, for the state-of-theresearch paper.
- Unassigned 10% of class grade
- Choppiest class I've ever taught due to travel, holidays, etc. Never again.

Evaluation of SE research

- You are in the field in industry
- You've read a number of SE research papers
- What convinces you?
 - Not necessarily to adopt a tool, but to consider an approach worthwhile enough to pursue in more detail
- Why?

Possible answers include

- Intuition
- Quantitative assessments
- Qualitative assessments
- Case studies
- ... other possible answers?

Which papers/ideas...

- ...have you found most compelling?
- Why those?

Brooks on evaluation

- The first user gives you infinite utility that is, you learn more from the first person who tries an approach than from every person thereafter
- In HCI, Brooks compared
 - "narrow truths proved convincingly by statistically sound experiments, and
 - broad 'truths', generally applicable, but supported only by possibly unrepresentative observations."
 - Grasping Reality Through Illusion -- Interactive Graphics Serving Science. Proc 1988 ACM SIGCHI

More on Brooks by Mary Shaw

- "Brooks proposes to relieve the tension through a certainty-shell structure – to recognize three nested classes of results,
 - Findings: well-established scientific truths, judged by truthfulness and rigor;
 - Observations: reports on actual phenomena, judged by interestingness;
 - Rules of thumb: generalizations, signed by their author but perhaps incompletely supported by data, judged by usefulness."
- What Makes Good Research in Software Engineering? International Journal of Software Tools for Technology Transfer, 2002

Shaw: research questions in SE

Type of question	Examples				
Method or means of	How can we do/create (or automate doing) X?				
development	What is a better way to do/create X?				
Method for analysis	How can I evaluate the quality/correctness of X?				
	How do I choose between X and Y?				
Design, evaluation, or	What is a (better) design or implementation for application X?				
analysis of a par-	What is property X of artifact/method Y?				
ticular instance	How does X compare to Y?				
	What is the current state of X / practice of Y?				
Generalization or	Given X, what will Y (necessarily) be?				
characterization	What, exactly, do we mean by X?				
	What are the important characteristics of X?				
	What is a good formal/empirical model for X?				
	What are the varieties of X, how are they related?				
Feasibility	Does X even exist, and if so what is it like?				
	Is it possible to accomplish X at all?				

Shaw: types of SE results

Type of result	Examples
Procedure or tech- nique	New or better way to do some task, such as design, implementation, measurement, evaluation, selection from alternatives, Includes operational techniques for implementation, representation, management, and analysis, but not advice or guidelines
Qualitative or descrip- tive model	Structure or taxonomy for a problem area; architectural style, frame- work, or design pattern; non-formal domain analysis Well-grounded checklists, well-argued informal generalizations, guidance for integrating other results,
Empirical model	Empirical predictive model based on observed data
Analytic model	Structural model precise enough to support formal analysis or auto- matic manipulation
Notation or tool	Formal language to support technique or model (should have a calcu- lus, semantics, or other basis for computing or inference) Implemented tool that embodies a technique
Specific solution	Solution to application problem that shows use of software engineer- ing principles – may be design, rather than implementation
	Careful analysis of a system or its development
	Running system that embodies a result; it may be the carrier of the result, or its implementation may illustrate a principle that can be applied elsewhere
Answer or judgment	Result of a specific analysis, evaluation, or comparison
Report	Interesting observations, rules of thumb

Shaw

 Types of validation

Type of validation	Examples				
Analysis	I have analyzed my result and find it satisfactory through ormal				
	analysis) rigorous derivation and proof				
	(empirical model) data on controlled use(controlled				
	carefully designed statistical				
	experiment) experiment				
Experience	My result has been used on real examples by someone other than				
	me, and the evidence of its correctness / usefulness / effectiveness				
	isalitative model) narrative(empirical model, data,				
	usually statistical, on practice				
	(notation, tool) comparison of this with similar results in				
	technique) actual use				
Example	Here's an example of how it works on				
	(toy example) a toy example, perhaps motivated				
	by reality				
	(slice of life)a system that I have been developing				
Evaluation	Given the stated criteria, my result				
	(descriptive model) adequately describes the phenomena				
	of interest				
	(qualitative model) accounts for the phenomena of interest				
	(empirical model) is able to predict because,				
	or gives results that fit real data				
	Includes feasibility studies, pilot projects				
Persuasion	I thought hard about this, and I believe				
	(technique) if you do it the following way,				
	(system) a system constructed like this would				
	(model) this model seems reasonable				
	Note that if the original question was about feasibility, a working				
	system, even without analysis, can be persuasive				
Blatant assertion	No serious attempt to evaluate result				

Tichy et al. on quantitative evaluation

- Experimental evaluation in computer science: A quantitative study. *Journal of Systems and Software* 1995
 - Tichy, Lukowicz, Prechelt & Heinz
- Abstract:

A survey of 400 recent research articles suggests that computer scientists publish relatively few papers with experimentally validated results. The survey includes complete volumes of several refereed computer science journals, a conference, and 50 titles drawn at random from all articles published by ACM in 1993. The journals of *Optical Engineering (OE)* and *Neural Computation (NC)* were used for comparison. .. (con't)

Con't

Of the papers in the random sample that would require experimental validation, 40% have none at all. In journals related to software engineering, this fraction is 50%. In comparison, the fraction of papers lacking quantitative evaluation in OE and NC is only 15% and 12%, respectively. Conversely, the fraction of papers that devote one fifth or more of their space to experimental validation is almost 70% for OE and NC, while it is a mere 30% for the computer science (CS) random sample and 20% for software engineering. The low ratio of validated results appears to be a serious weakness in computer science research. This weakness should be rectified for the long-term health of the field. The fundamental principle of science, the definition almost, is this: the sole test of the validity of any idea is experiment. —Richard P. Feynman. Beware of bugs in the above code; I have only proved it correct, not tried it. —Donald E. Knuth

Technology transfer: briefly

- Not a consumer problem
- Not a producer problem
- An ecosystem issue

A few examples

Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure (1995)

"Brooks-Sutherland" report

Computer Science and Telecommunications Board (CSTB)



FIGURE 1.2 Government-sponsored computing research and development stimulates creation of innovative ideas and industries. Dates apply to horizontal bars, but not to arrows showing transfer of ideas and people. Table 1.1 is a companion to this figure.

Comments?

UW CSE P504

Software engineering economics

- The phrase dates to around 1981, when Barry Boehm published his tome with the same title
- His 1976 IEEE Transactions on Computers article "Software Engineering" identified engineering economics as one "scientific principle" in which software engineering fell short of hardware engineering
- To the first order, the focus of his book was on how to better estimate effort, cost and schedule for large software projects – COCOMO (COnstructive COst MOdel)

COCOMO basics

- Algorithmic software cost estimation modeled with a regression formula that has parameters derived from historical project data and current project characteristics
- The basic COCOMO equations take the form
 - Effort Applied = $a(KLOC)^{b}$ (person-months)
 - Development Time = c(Effort Applied)^d (months)
 - People required = Effort Applied / Development Time (count)

	а	b	С	d
Organic	2.4	1.05	2.5	0.38
Semi-detached	3.0	1.12	2.5	0.35
Embedded	3.6	1.20	2.5	0.32

Regression parameters Basic COCOMO

- Based on waterfall-based 63 projects at TRW Aerospace
- Projects from 2KLOC to 100KLOC, languages from assembler to PL/I
- The Basic Model designed for rough order-ofmagnitude estimates, focused on small to mediumsized projects
 - Three sets of parameters: organic, semideteched and embedded

Intermediate COCOMO

- Uses more parameters (cost drivers) that account for additional differences estimates
- Product attributes: required software reliability, complexity of the product, ...
- Hardware attributes: run-time performance constraints, memory constraints, ...
- Personnel attributes: software engineering capability, applications experience, programming language experience, ...
- Project attributes: use of software tools, application of software engineering methods, ...

Intermediate COCOMO

- The 15 sub-attributes are each rated from "very low" to "extrahigh" with six discrete choices
- Effort multipliers are empirically derived and the EAF is the product of the multipliers
 Ratings

	Very			-	Very	Entra	
Cost Drivers	Low	Low	Nominal	High	High	High	
Product attributes							
Required software reliability	0.75		1.00				
Size of application database			1.00				
Complexity of the product	0.70	0.85	1.00	1.15	1.30	1.65	
Hardware attributes							
Run-time performance constraints				1.11			
Memory constraints				1.06		1.56	
Volatility of the virtual machine envionment			1.00				
Required turnabout time		0.87	1.00	1.07	1.15		
Personnel attributes							
Analyst capability			1.00				
Software engineer capability			1.00				
Applications experience	1.42	1.17	1.00	0.86	0.70		
Virtual machine experience			1.00				
Programming language experience	1.14	1.07	1.00	0.95			
Project attributes							
Use of software tools	1.24	1.10	1.00	0.91	0.82		
Application of software engineering methods	1.24	1.10	1.00	0.91	0.83		
Required development schedule	1.23	1.08	1.00	1.04	1.10		

Intermediate COCOMO

- $E=a(KLOC)^b \times EAF$
 - And similarly for development time and people counts
- There is a separate table for parameters a and b across organic, semi-detached, embedded for Intermediate COCOMO

Detailed COCOMO & COCOMO II

- Detailed COCOMO also accounts for the influence of individual project phases
- COCOMO II was developed and released in 1997, aimed at (then) modern software projects
 - Newly tuned parameters
 - Accounted for move from mainframes to desktops, from batch to interface computation, to code reuse, etc.
1981 Boehm book also discusses

- Multiple-goal decision analysis
 - Most optimization theory assumes that there is a single objective function to maximize
 - Models like this one account for multiple goals that must be balanced in a definable manner
- Risk analysis
 - Foundation for his later work in the spiral model
- And more...

Boehm Sullivan "Software Economics" roadmap (ICSE 2000)

"The core competency of software engineers is in making technical software product and process design decisions. Today, however, there is a 'disconnect' between the decision criteria that tend to guide software engineers and the value creation criteria of organizations in which software is developed. It is not that technical criteria, such as information hiding architecture, documentation standards, software reuse, and the need for mathematical precision, are wrong. On average, they are enormously better than no sound criteria.

Con't

"However, software engineers are usually not involved in or often do not understand enterprise-level value creation objectives. The connections between technical parameters and value creation are understood vaguely, if at all. There is rarely any real measurement or analysis of how software engineering investments contribute to value creation. And senior management often does not understand success criteria for software development or how investments at the technical level can contribute fundamentally to value creation. As a result, technical criteria tend to be applied in ways that in general are not connected to, and are thus usually not optimal for, value creation."

Thinking about value

- Decision theory (or utility theory) defines a framework for decisions under uncertainty, depending on the risk characteristics of decision makers
- This is closely related to (again) multi-objective decision-making
- Classical corporate finance uses net present value (NPV) as an investment decision criterion and computes it by discounted cash flow analysis (DCF) – can't make a business case without these

NPV example from Wikipedia

 A corporation must decide whether to introduce a new product line. The new product will have startup costs, operational costs, and incoming cash flows over six years. This project will have an immediate (t=0) cash outflow of \$100,000 (which might include machinery, and employee training costs). Other cash outflows for years 1-6 are expected to be \$5,000 per year. Cash inflows are expected to be \$30,000 each for years 1-6. All cash flows are after-tax, and there are no cash flows expected after year 6. The required rate of return is 10%.

Con't

- The table shows the present value (PV) for each year
- The NPV is the sum of the PVs
- In this case, it's \$8,881.52
- A positive NPV means it would be better to invest in the project than to do nothing – but there might be other opportunities with higher NPV

Year	Cashflow	Present Value
T=0	$\frac{-100,000}{(1+0.10)^0}$	-\$100,000
T=1	$\frac{30,000-5,000}{(1+0.10)^1}$	\$22,727
T=2	$\frac{30,000-5,000}{(1+0.10)^2}$	\$20,661
T=3	$\frac{30,000-5,000}{(1+0.10)^3}$	\$18,783
T=4	$\frac{30,000-5,000}{(1+0.10)^4}$	\$17,075
T=5	$\frac{30,000-5,000}{(1+0.10)^5}$	\$15,523
T=6	$\frac{30,000-5,000}{(1+0.10)^6}$	\$14,112

Real options

- DCF/NPV treats assets as passively held not actively managed
- But projects are (or can be ⁽ⁱ⁾) actively managed
 - Management usually has the flexibility to make changes to real investments in light of new information. (e.g., to abandon a project, enter a new market, etc.)
- The key idea of real options is to treat such flexibility as an option, and to (in some cases) price them using techniques related to those for financial options

Baldwin and Clark (2000)

- Baldwin and Clark view Parnas' information hiding modules as creating options
- They value these and develop a theory of how modularity in design influenced the evolution of the industry structure for computers over the last forty years
- Non-modular systems must be kept or replaced as a whole whole
- A system of independent modules can be kept or replaced (largely) individually based on judgments of improvement (or not)
- Modularity provides a portfolio of options vs. an option on a portfolio



DSMs: design structure matrices

- The parameters are A, B, and C
- The X in row B, column A means that good choice for B depends on the choice made for A.
- Parameters requiring mutual consistency are *interdependent*, *resulting in symmetric marks:* (B,C) and (C,B).
- When one parameter choice must precede another the parameters are said to be *hierarchically dependent:* (B,A).
- *Independent* parameters can be changed without coordination.

	Α	В	С
А			
в	Х		Х
с		Х	

Figure 1: DSM for a design of three parameters.

Material from

Sullivan, Griswold, Cai, Hallen. The structure and value of modularity in software design. ESEC/FSE 2001

Splitting

- DSMs may not show largely independent designs
- In these cases, one approach is to apply splitting
- Break a dependence with a new parameter that constrains the values of the original parameters – this means, in part, that they depend on it
- Fix the value of the new parameter so that the original parameters to be changed independently as long as they are only changed in ways consistent with the new constraint
- For example, introduce a new interface (I, in the below example)



Figure 3: DSM for a modular design obtained by splitting.

Parnas KWIC



I - Alph Alg

K - Out Data

L - Out Alg

M - Master

х

х

Х

XXXXX

х

Х

Figure 5: DSM for strawman modularization

Figure 7: DSM for information hiding modularization

Х

х

Х

NOV (net option value)

- A module creates an opportunity
 - to invest in k experiments to create candidate replacements,
 - each at a cost related to the complexity of the module
 - if any of the results are better than the existing choice, to substitute in the best of them
 - at a cost that related to the visibility of the module to other modules in the system

KWIC NOV



Figure 13: Options Values for Information Hiding Design

- The option value of each module is the value at the peak
- Sum the module NOV's
 - 0.26 for the strawman design
 - 1.56 for the information-hiding



Figure 14: Options Values for Strawman Design

UW CSE P504



- The basic idea seems to make sense to many people
- One of the core problems is the notion of how to tune the model parameters
 - Financial markets set parameters based primarily on scads of historic data
 - COCOMO set parameters based on careful studies of a reasonably large set of reasonably similar software projects
 - Tuning parameters for modularity seems more complicated



Figure 1: Roadmap for research in software engineering economics.

Your turn

- In what ways does your organization link technical decision making with business-level decision making?
- And not?

McConnell's cone of uncertainty ICSE 2009 keynote



Governance of Software Development

- Clay Williams, IBM Research
- Slides directly taken from an NSF workshop
 presentation Governance @ IBM Future Directions

Governance of Software Development Strategic Initiative

 Goal: Develop the science and technology that enables the Rational software delivery platform to provide support for *governing the business* of software development.





Tempo - Overview

- Problem Statement
 - When project teams commit to a schedule, they are placing a bet. It would be extremely valuable for them to know the odds of winning.
- Approach
 - Capture "bottom-up" predictions regarding the time necessary to complete each task in a work breakdown.
 - Rather than discrete predictions, capture triangular distribution that reflects the fact predications are random variables.
 - Develop optimized scheduling approaches that rapidly reduce schedule risk in the project
 - Surface schedule risks to allow teams to better manage scheduling issues.
- What is hard?
 - Providing a tool that is easy to use and supportive of "what-if" risk mitigation analysis requires addressing subtle and difficult usability issues.
 - The variety of optimizations and analyses require significant mathematical skill.



Tempo in Rational Team Concert



Governance @ IBM

Architectural and Social Governance of Software Development

Research Goals

- Exploit / expand the role architectures play as "boundary objects" spanning multiple domains of discourse.
- Develop techniques for exploring key structural and behavioral properties of architectures (software, IT, and EA), the socio-technical dynamics of the teams producing and consuming them, and how these two areas can be aligned and engender communication beyond the technical domain.
- Develop / extend architectural approaches to support business decisions and value management.
- Understand the interplay across the value / architectural / socio-technical domains.

Collaborations

- CMU (Jim Herbsleb)
- Harvard Business School (Carliss Baldwin) pending
- Virginia (Kevin Sullivan)

Future Directions

Architecture and Business / Technical Alignment



My bottom line

- The long-term goal of software engineering economics is to help everybody make more sensible decisions
 - Technical decisions
 - Business decisions
 - Project management decisions
- Not one of these is primary with the others secondary but that is how we each seem to treat the others
- Better understanding the links among them is crucial; the models may give us opportunities to better understand these links
- I am always scared that quantification tends to lead to a focus on the quantities, and there is often a disconnect between the quantities we can measure and want we want to do

Questions?

- For tonight?
- For the quarter?

Course evaluations...

UW CSE P504