The Role of Expert Judgment in Improving Software Estimation Processes

Ira A. Monarch
Dennis R. Goldenson
Robert W. Stoddard II

SEPG NA 2010
Savannah, Georgia – 24 March 2010

Software Engineering Institute | **Carnegie Mellon**

Software Engineering Institute | **Carnegie Mellon**

Today's Talk

What are we trying to accomplish?

How we map the dynamics of software engineering research areas

The case of software estimation: An initial mapping

A better database & a better mapping

Opportunities for collaboration

What are we trying to accomplish?

Narrow the gap between software estimation research & practice

Determine its extent & nature by
- Mining existing research & descriptions of practice
- Monitoring the impact of the research on practice
- Using the results to suggest further research

Use language data analysis methods to
- Scope & analyze a fairly voluminous literature
- Uncover different areas of research & practice within software estimation

Seek feedback on the viability of our approach

Explore opportunities for mutual collaboration

Approach

- Collect bibliographic records & apply text analysis
- Study automatic processing results & interpret them

Search Criteria

Estimat* & Software → INSPEC

Document Review → BESTweb

→ • **titles**
• **abstracts** →
• **dates**

Apply Text Analysis

INSPEC Output
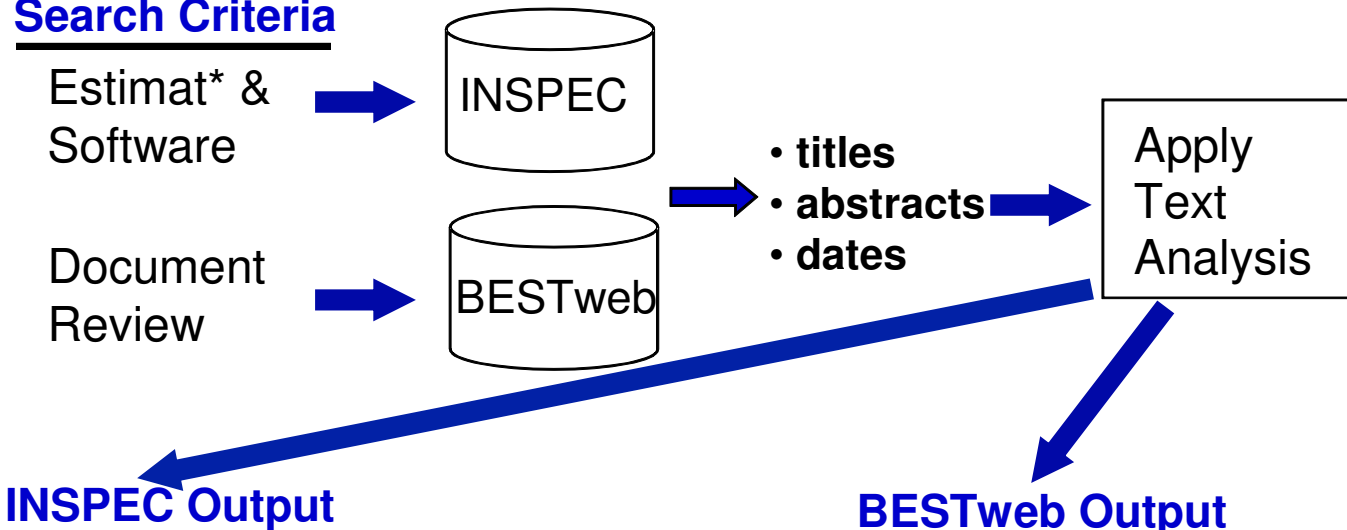
Time Periods	Themes & Categories
1966 -1996	• Expert System
1997 - 2004	• Expert Judgment
2005 - 2009	• Mix & Match

BESTweb Output

Time Periods	Categories, Themes & Concepts
1940 - 1995	• FEA, PM, RTA (judgments)
1996 - 2000	• FEA, PM (*Risk*), RTA (Forecasts)
2001 - 2009	• FEA, PM (*Reqmts*), EF, RTA (Sim)

FEA: Formal Estimation Approach, PM: Process Management, RTA: Research Topics & Approaches, EF: Expert Forecasting

Text Analytics & Concept Maps

Automated text analysis tools used to identify recurring concepts & themes (clusters of concepts)

- Concepts include synonyms based on strongly related co-occurring terms
  — Constituted in automatically generated affinity lists
  — Named by most representative term in affinity list
- Themes are clusters of concepts with similar co-occurrence patterns
  — More strongly related to each other than to concepts in other clusters
  — Named by automatic selection of the concept most strongly related to other concepts in the cluster

Themes are represented graphically as Venn diagrams
- Concept names label dots that are in circles representing themes
- Dots can be linked by lines whose brightness represents frequency of co-occurrence
- Dots can appear in the overlap of two (or more) circles
- Circle size does not always indicate importance since circles can be sparsely populated

Example Concept Map from INSPEC Analysis
1997-2004
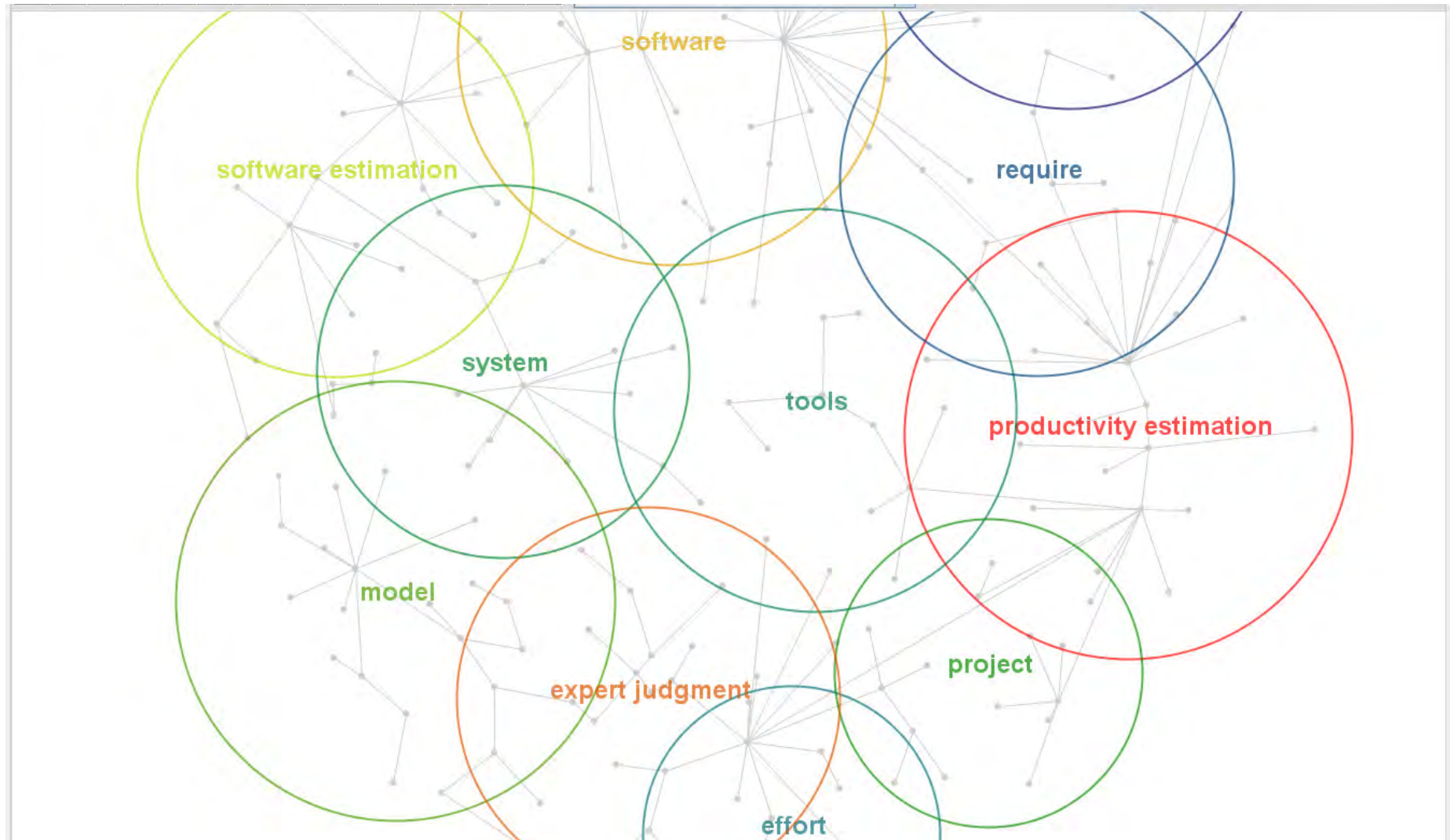
Text Analysis of INSPEC Estimation Literature

1966-1996: *Expert System* appears as a primary theme whereas **expert judgment** is a much less significant concept used to qualify *estimation models* like **COCOMO** and *expert systems.*

1997-2004: *Expert Judgment* appears as a primary theme, second only to *productivity estimation*, is again used to qualify various **formal estimation models**, whether they use **statistics, regression, fuzzy logic,** or **neural networks**, and to help estimate **function points, SLOC, size** and **effort**.
Its use is also affirmed in practical estimation contexts, and even claimed to be as accurate as, or even more accurate than, **formal estimation models.**

2005-2009: Neither **formal models/machine learning** nor **expert judgment** is central, perhaps continuing a mix and match approach started in 1997-2004. However, a new theme surfaces concerning *overruns* to acknowledge that in spite of the existence of these approaches, *overruns* of **schedule** and **cost** were not being prevented.

Need a Roadmap Beyond the Debate between Formal Estimation vs Expert Judgment

A debate between Magne Jorgensen and Barry Boehm moderated by Stan Rifkin published in IEEE Software in March/April of 2009
- dovetails very nicely with our text analysis of the INSPEC Database.

Boehm and COCOMO, not surprisingly, were representatives of what continues to be a dominant part of the software estimation paradigm
- though Boehm pointed out in 2000 that using a combination of methods and processes, including "expert judgment," was the best approach to be followed.

Jorgensen's view of mix & match differs from Boehm's
- "My main claim in this discussion is that organizations' process improvement work and research initiatives should focus on better judgment-based effort estimation processes, not on introducing or improving formal estimation models."

The debate is important because two important research stances begin to interact.

However, there are *other* important strands of work that need to be considered
- e.g., process & performance management.

Text Analysis of BESTweb Estimation Literature

Jorgensen's BESTweb system (www.simula.no/BESTweb)

- A web-based system supporting research on **software** cost & effort estimation using a database front-end client

- Gives access to information about journal & conference papers identified as relevant through the use of various categories

- For a text analysis of titles & abstracts

Most pertinent BESTweb categories

- Estimation Approaches

- Research Topics

- Research Approaches

Analysis leads to a rich analysis of evolving state of software estimation over the last 40 years

- Provides a better basis for aligning research with software estimation practice

Software Engineering Institute | **Carnegie Mellon**

The Role of Expert Judgment in Improving Software Estimation Processes
Monarch, Goldenson & Stoddard
24 March 2010

10

© 2010 Carnegie Mellon University

BESTweb 2001-2009

BESTweb Analysis Results Overall:
Trends and Directions[1]

1. In all time periods overarching concerns divide into

    a. *Formal Estimation Approaches*

    b. *Process & Project Management*

    c. *Research Topics & Approaches*

2. In last time period a new approach, *Expert Forecasting*, emerges from *Research Topics & Approaches*

3. In all time periods, **COCOMO** and **Function Point Analysis** are the primary exemplars of *Formal Estimation Approaches*, but in the 2nd & 3rd time periods many other approaches and techniques surface, e.g.,

    a. 1996-2000: **Mark II**, **LOC**, **ANGEL**, **IFPUG**, **neural**, **regression**, **COCOMO II**

    b. 2001-2009: **CBR**, **Bayesian**, **ISBSG**, **MMRE**, **AQUA**, **ERP**, **MRE**, **UML**, **JAVA**, **CART**, **Genetic Algorithm**, **IFPUG**, **Estimation by Analogy**, **neural**, **regression**

    c. In 2001-2009, relations are also more rich: e.g., **data sets**--**prediction**, **regression**--**accuracy**, **analysis**--**models**, **ERP**--**measures**

BESTweb Analysis Results Overall:
Trends and Directions$_2$

4. For ***Process & Project Management***

    a. There are no exemplars in the 1st & 3rd time periods & only **MBASE** appears in the 2nd

    b. Relations in the 1st two time periods are sparse, although **risk** & **uncertainty** begin to appear in the 2nd time period

    c. However, relations in the 3rd time period are much denser & notable, e.g., **Maintenance--cost**; **early--requirements**, **development--experiment**, **quality--schedule**, **project--information**, **project--data & manage--risk-- uncertainty**

5. For ***Research Topics & Approaches***

    a. Exemplars in the first two time periods are specific but not proper names, e.g., **subjective probability distributions** & **forecasting performance**

    b. However, exemplars in the 3rd time period contain specific & proper name examples, i.e., **Simulation** & **Monte Carlo**

    c. Relations in the 1st time period focus on **expert judgment** & **decision making**. The focus in the 2nd time period is on **learning**, **accuracy** & **confidence** as well as **expert judgment** . However things change in the 3rd time period.

BESTweb Analysis Results Overall:
Trends and Directions$_3$

6. In the 3rd time period: A new approach emerges from **_Research,_** with some remaining overlap, & a new research focus begins

    a. The new approach is **_Expert Forecasting_**. There are no exemplars. Relations include **Judgmental--ambiguity**, **experts--knowledge**, **reliability--knowledge**, **experts--series**, **series--intervals**, **standard--intervals**, **better--statistical--standard, experts--adjustment** & **experts--rules**.

    b. As pointed out, there are research exemplars, **Simulation** & **Monte Carlo**, and they are new.

    Relations include **Large--simulation**, **group--differences**, **Monte Carlo--value**, **value--simulation** & **performance--findings**.

14

Implications & Opportunities for Collaboration[1]

Action research is needed to

- Better understand isolated approaches & research areas for potential synergy
  — Formal estimation approaches, expert judgment & project management
- & translate identified improvements more widely into practice

Some ideas for collaboration –

- Voice of the customer (VoC) has been used very successfully, but there is a lingering question about its generalizability
- Explore combining automated language data analysis, VoC techniques and people, e.g.,
  — KJ, Kano, QFD & wide-band Delphi
  — And semi-automated text analysis & collaborative software tools to include *much* more information from *many* more stakeholders
  — Estimation, subject matter experts *& other key stakeholders* from organizations doing estimation (including management & customers)
  — Expert representatives of the various approaches described here

Implications & Opportunities for Collaboration$_2$

Use text analytic findings to seek synergy among estimation approaches

- Parametric models
- Discrete event, system dynamic and agent based simulation
- Monte Carlo scenarios
- Bayesian belief networks
- Causal modeling of prediction intervals & risk as a function of controllable and non controllable x-factors
- Better calibrating the models
  - In principle any expert judgment can be an explicit model parameter

Use decision support environments to augment model predictions iteratively for

- Risk mitigation in the presence of uncertainty
- Taking corrective actions based on management & engineering judgment
  - For which re-estimation cannot be justified
  - Although a good model interface should make what-if analyses more feasible!

What else? Let's continue the dialogue.

Thank You for Your Attention!

Ira Monarch   Dennis Goldenson   Bob Stoddard
iam@sei.cmu.edu   dg@sei.cmu.edu   rws@sei.cmu.edu

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Software Engineering Institute | Carnegie Mellon