Scientific Method

M. Anton Ertl

anton@mips.complang.tuwien.ac.at

http://www.complang.tuwien.ac.at/anton/

Institut für Computersprachen

Technische Universität Wien

Goals in Science

General public, readers New, relevant Knowledge

Author Publish papers, increase reputation

Publication/reviewer ensure high quality of papers

Evaluation criteria

- Relevance
- Appropriateness for the journal/conference
- Originality
- Quality of presentation
- Technical quality

Methods

Physics

data→theory (hypothesis)→prediction
→experiment→data
reproducability

Social sciences

Surveys, Interviews, Experiments
statistics

Mathematics, theoretical computer science

Proof

Systems/Engineering

Problem→possible solution
→implementation/simulation→benchmarks

Problems of benchmarks

- Only a few cases out of many possible ones
- Too many interfering factors in runs on real machines
- Simulations take a long time
  realism vs. interference

Methods of theoretical computer science

Statements about all cases, but:

- Often qualitatively, not quantitatively
- If quantitatively, what is the distribution of actual data?
- If an assumption is made, does it correspond to reality?
- Simplifying assumptions or restricted problem statements
- Optimal methods are often NP-hard or worse
- Usually no consideration of interaction with other components

Other methods

Model based on statistical data

- Behaviour in computer science is often non-continuous or not continuously differentiable
- Functions involving max when parallelism is in play.
- Statistical data comes from benchmarks (with their problems)
- Very narrow applicability, see also synthetic benchmarks.

Important principles (from Johnson)

- Perform Newsworthy Experiments
- Tie your Paper to the Literature
- Use Instance Testbeds that Can Support General Conclusions
- Use Efficient and Effective Experimental Designs
- Use Reasonably Efficient Implementations
- Ensure Reproducability
- Ensure Comparability
- Report the Full Story

- Draw Well-Justified Conclusions and Look for Explanations
- Present Your Data in Informative Ways

Frequent mistakes

- Non-representative benchmarks
- No rationale for the choice of benchmarks
- synthetic benchmarks
- microbenchmarks
- tuning for the benchmark
- Compare the best case of method A to the typical case of method B (cannot show the superiority of A)
- Missing explanations for anomalies

- Conclusions without support
- One-dimensional comparison of techniques that compete on multiple dimensions
- Comparisons of techniques without keeping the rest the same
- Missing the big picture
- Indirect Metrics
- Comparison of relative numbers with different bases
- Missing description of the measurement setup
- Inefficient implementations

- Wrong average
- Loss of code and/or data
- Using running time as a stopping criterion (for heuristic algorithms)

Presentation

- Confusing content
- Percentages with unclear meaning (is 150% factor 1.5 or 2.5?)
- Percentages for which the base is unclear
- Overloaded charts
- 3D charts
- Line chart between unordered points (use bar chart)
- Hard-to-discern lines/bars etc.
- Missing axis title

- Missing units
- Missing/bad reporting of aggregation
- Different scaling for comparable graphs
- Different orders between lines and legend