Ausgewählte Kapitel der Systemsoftware (AKSS)

Benchmarking Crimes (Gernot Heiser: http://gernot-heiser.org/benchmarking-crimes.html)

09. Juni 2021

Phillip Raffeck, Tim Rheinfels, Simon Schuster, Peter Wägemann

Lehrstuhl für Informatik 4
Friedrich-Alexander-Universität Erlangen-Nürnberg
BENCHMARKING CRIMES – A REALITY CHECK

HOW CAN WE REBUILD THE TRUST OF OUR CUSTOMERS? LET'S BRAINSTORM.

WE COULD STOP USING MISLEADING BENCHMARK TESTS TO SELL SHODDY PRODUCTS THAT HAVE HIDDEN COSTS.

I HEARD SOMEONE SAY "LIE." LET'S WRITE THAT ONE DOWN.
Three Rules for Summarizing Results

- Philip J. Fleming & John J. Wallace: *How Not To Lie With Statistics: The Correct Way To Summarize Benchmark Results*
- Three Rules
  1. Do Not Use the Arithmetic Mean to Average Normalized Numbers
  2. Use the Geometric Mean to Average Normalized Numbers
  3. Use the Arithmetic Mean to Average Raw Results

- Arithmetic mean: $x_{\text{arith}} = \frac{1}{n} \sum_{i=1}^{N} x_i$
- Geometric mean: $x_{\text{geom}} = \sqrt[n]{\prod_{i=1}^{N} x_i}$
1st Crime: Selective Benchmarking

1. Not evaluating potential performance degradation
   - Progressive criterion: actual improvement
   - Conservative criterion: no degradation elsewhere

2. Cherry picking without justification

3. Selective data set hiding deficiencies
2nd Crime: Micro-Benchmarks vs. Macro-Benchmarks

Pretend µ-Benchmarks Represent Overall Performance

- Macro-benchmarks $\rightsquigarrow$ realistic picture
- Examples exist for exception
3\textsuperscript{rd} Crime: Overhead follows Throughput

Throughput degraded by $x\% \Rightarrow$ overhead is $x\%$

- Throughput comparisons require accompanying comparisons of \textbf{complete CPU load}
- What determined throughput in baseline?
- I/O throughput: use \textbf{processing time per bit}
4th Crime: Downplaying Overheads

- 6% to 13% overhead $\not\Rightarrow$ 7% increase of overhead
- Percentage vs. percentage points
Disjoint workloads for calibration & evaluation
Predictions based on models
Raw averages misleading

All standard deviations **must be below 1%**

Doubts: use Student’s **t-test**¹

Fit lines: use regression coefficients

¹Student (William Sealy Gosset): The Probable Error of a Mean. Biometrika. 1908
7th Crime: Benchmarking of Simulated System

- Simulation == model
- Correctness of model?
- Best model is reality
8th Crime: Inappropriate & Misleading Benchmarks

- Reader lured with misleading benchmarks
- Usage of relevant benchmarks
- Example: CPU-bound workload for evaluation of network stack
Significance of results hidden

State denominator
10th Crime: No Proper Baseline

- Compare against state-of-the-art approach
- Existing implementations
- Theoretical optimal solution
11th Crime: Evaluate Against Yourself Only

- Compare against **accepted standard**
- Avoid using model to compare against
12th Crime: Unfair Benchmarking of Competitors

- Provide comparable common ground (e.g., configurations)
- Objectivity/fairness
- Direct evaluations against competitors must be performed extremely thoroughly
Arithmetic mean: $x_{arith} = \frac{1}{n} \sum_{i=1}^{N} x_i$

Geometric mean: $x_{geom} = \sqrt[n]{\prod_{i=1}^{N} x_i}$

- Normalized numbers $\Rightarrow$ geometric mean
- Absolute numbers $\Rightarrow$ arithmetic mean
References

- Benchmark Crimes: http://gernot-heiser.org/benchmarking-crimes.html
- Dilbert: dilbert.com/strip/2010-10-15
- Cherry Picking: https://commons.wikimedia.org/wiki/File:Cherry_picking_(7848350200).jpg
- Eiffel Tower: https://commons.wikimedia.org/wiki/Commons:Photo_challenge/2014_-_September-October_-_Big_and_small
- Percentage Points: http://imgs.xkcd.com/comics/percentage_points.png
- Disjoint Sets: https://commons.wikimedia.org/wiki/File:Disjunkte_Mengen.svg
- Standard deviation: https://upload.wikimedia.org/wikipedia/commons/0/05/Alex_Dodge_2012_left.jpg
- Simulation: https://commons.wikimedia.org/wiki/File:Fahr-Simulation.jpg
- Misleading: https://de.wikipedia.org/wiki/Rotk%C3%A4ppchen
- Relative Numbers: https://upload.wikimedia.org/wikipedia/commons/6/68/Extrema_example_original.svg
- Baseline: https://www.pexels.com/photo/field-sport-ball-game-54330/
- Unfair Competitors: https://i.ytimg.com/vi/lXRL4gZdRYQ/maxresdefault.jpg