Metodologias Experimentais em Informática
2019/2020

Resumo do Artigo

“Experimental Computer Science:
The Need for a Cultural Change”

1. Introduction

- Experimental Science is:
    - Observation
    - Hypothesis testing
    - Reproducibility
- How and whether these apply to computer science?
- 1.1 Computer Science
    - Science
    - Engineering
    - Mathematics
- 1.2 Experimentation
    - “Using experimental feedback can be argued to be the dominant force underlying the progress of the whole of computer science.”
    - Important questions about Experimental Methodology:
        - Is it based on empirical evaluation and data?
        - Was the experiment designed correctly?
        - Is it based on a toy or a real situation?
        - Were the measurements used appropriate for the goals of the experiment?
        - Was the experiment run for a long enough time?
- Experimental Science components:
    - Observation
    - Hypothesis testing
    - Reproducibility

2. Observation

- In computer science this means the measurement of real systems.
- Forming a model of the world, essence to learn about it.
- The model can then be used to make predictions, which can then be tested
- 2.1 Challenges
- 2.2 Metric:

        - A special challenge in performing measurements is coming up with appropriate metrics.
        - Find relationships between units.
        - What metrics are the most effective is determined by experimentation
    - 2.3 Surprises:
        - Experiments are out to obtain new (and unexpected) knowledge.
        - Without experiments, nature doesn't have the opportunity to tell you anything new.
    - 2.4 Modeling:
        - Well-executed measurements provide us with data.
        - Modeling this data is the process that turns them into information and knowledge.
        - The risk of fitting the measurements to the theory, rather than the theory to the measurements
    - 2.5 Truth
        - The scientific method is based on the quest for truth by means of objective observations.
    - 2.6 Sharing
        - Getting data is hard. Getting good data is even harder. It is therefore imperative that data be shared, so that the most benefit possible will be gleaned from it.
        - sharing data enables two important things:
            - Exploration
            - Reproducibility

3. Hypothesis Testing

- Hypothesis testing is at the very core of the scientific method.
- Where experimentation comes in.
- See whether what you think you know is indeed true.
- Experimental science starts with observation.
- Based on the measured observations, one builds a model.
- The model is an abstraction of the world, and embodies a generalization of the results of the measurements; it is an expression of what you have learned from them.
- To justify itself, a model must be used.
    - Make predictions about the world, and in particular, about aspects that have not been measured yet.
    - Such predictions are actually hypotheses about what the outcome of the missing measurements will be
- To turn a hypothesis into bone-fide knowledge, it has to pass the test of experimentation.

        - A special test is designed, which will measure specifically whether the hypothesis makes the right prediction.
- Cycle:
    - A measurement led to a model,
    - The model to a hypothesis,
    - And now the hypothesis is used to guide another measurement.
    - This in turn may lead to a refinement of the model, and so on.
- 3.1 Emergent Hypotheses
    - The term “hypothesis” is used at two quite different levels:
        - Macro level
            - Macro level hypotheses are concerned with the shaping of a whole field.
            - As opposed to the micro level employed in individual research projects.
        - Micro level
            - Where a specific, concrete, atomic prediction is to be tested.
    - We accept an hypothesis because it has passed extensive tests: generations of computer scientists have tried to refute it and failed.
    - A basic hypothesis in performance analysis is that queueing networks provide an adequate model for making predictions.
    - Emergent hypotheses
        - They are not proposed and then tested systematically.
        - They emerge as a summarizing principle that unites a large body of work.
    - Micro-hypotheses (or *ad-hoc hypotheses*)
        - They are formulated for a specific need, and then tested to see that they fulfill this need.
- 3.2 Hypothesis-Driven Experiments
    - Micro level
        - Is concerned with individual experiments and measurements.
        - we need to explain the results of our measurements.
        - Coming up with a model of how the system behaves, and showing that the model agrees with the measured results.
        - It is also required that new measurements fit the model.
        - This hypothesis needs to be tested.
            - If it passes the test, and then another test, and another, we gain confidence that the model is indeed a faithful representation of the system’s innermost working.
        - The only way to find the right explanation is to check it experimentally.
- 3.3 Refutable Theory

        - Regarding possible explanations of system performance as mere hypotheses, and devising experiments to check their validity, are not only a mechanism for finding the right explanation.
        - Hypothesis testing and refutable theory are the fast lane to scientific progress.
        - Using explanations without thinking about and actually demonstrating proper experimental verification may lead us to false conclusions.
        - Why do limited resources promote good experimental research?
            - Because if you have limited resources, you need to think about how to best invest them.
            - What will yield the best returns on your investment.
        - The best return is obtained by carefully designed experiments, and specifically, those that can best distinguish between competing theories.
    - “But sir, what hypothesis does your experiment disprove?”
    - “But sir, what experiment could disprove your hypothesis?”

4. Reproducibility

- It is beneficial to allow others to repeat our work, both to verify it and to refine and extend it.
- 4.1 Mistakes
    - Mistakes are typically found by testing the software.
    - Testing has become a major part of development.
    - When multiple groups from different places using different methods got the same numbers, they were most probably right.
    - We can gain confidence if others repeat the work and obtain similar results.
    - To enable others to repeat a study, the work has to be reproducible. This has several important components:
        - Describe the work in sufficient detail.
        - Make software available in a usable form.
        - Make raw data available, especially input data.
        - Enable access to infrastructure.
- 4.2 Understanding
    - Improve our understanding of the measured system.
    - Reproducing a result means determining which details are important and which are inessential.
    - Finding out what are the important parameters that affect performance, the mechanisms by which they affect performance, and the degree to which they affect performance.
    - Verification is the means to achieve such a level of understanding.
    - To make repeatability easier, results should be accompanied by full details on how they were obtained.

        - The point of reproducibility is to reproduce the insights, not the numbers.
        - It is more qualitative than quantitative.
    - 4.3 Standardization
        - A canonization of a certain approach or process.
        - Designers forgo possible optimizations in the interest of interoperability.
        - Experimental computer science requires even more standardization.
        - Papers should include a methodology section that describes how the work was done in enough detail to allow it to be reproduced.
        - The “best practices”.
        - Improvements in methodology that do not immediately translate to significant impact are considered uninteresting.
    - 4.4 Progress
        - Replication is more about moving forward than about reviewing the past.
        - Replication fosters progress because it is hardly ever completely precise.
        - Each replication also introduces a small variation.
        - Such variations open the door to meta-studies.
        - Progress is built from a combination of breakthroughs and small steps.
            - Breakthroughs typically result from new insights, that are based on cumulative experience.
            - The small steps result from a choice between multiple candidates, just like evolution depends on the selection of the fittest among several variants.

5. What Next?

- One possible contribution would be to create a forum to foster the publication of experimental work: a Conference on Experimental Computer Science (CECS).
- A general conference dedicated to experimentation in computer science in general would be beneficial for bolstering the field as a whole, and for encouraging interactions and cross-fertilization between the subfields of computer science.