Centre for Data Analytics



Recomputation.org: Experience of Its First Year and Lessons Learned

Lars Kotthoff

lars.kotthoff@insight-centre.org

11 December 2014











Recomputation.org

- repository for computational experiments
- "If we can compute your experiment now, anyone can recompute it 20 years from now"
- based on principles laid out in the recomputation manifesto

Gent, Ian P. The Recomputation Manifesto, April 2013. http://arxiv.org/abs/1304.3674.

The Recomputation Manifesto

- 1. Computational experiments should be recomputable for all time.
- 2. Recomputation of recomputable experiments should be very easy.
- 3. It should be easier to make experiments recomputable than not to.
- 4. Tools and repositories can help recomputation become standard.
- 5. The *only* way to ensure recomputability is to provide virtual machines.
- 6. Runtime performance is a secondary issue.

Motivation – Shrdlu

- famous early AI program
- source code available
- but can't run it
- "The files are available for anyone who wants to use them (let me know if you get something running!)."



What does it mean to be "recomputable"?

- ability to rerun experiments
- ability to get the same results
- ability to get the same behaviour

Same behaviour

- often the "behaviour" is unspecified and only known through the results
- how the results are achieved is secondary
- in practice not necessary to replicate behaviour

Same Results

- experiments often stochastic/randomised
- results may require analysis/aggregation
- in practice often enough to get qualitatively similar results

Rerun

- non-standard equipment or lots of resources may be necessary
- if it runs, it doesn't mean that it will finish
- cannot guarantee that everything will run all the time
- How do I know it's right?

Our Approach

- work with authors of experiments
- they say how to "recompute"
- they can control/check results

Our Approach



- everything in virtual machines
- Vagrant to allow packaging and distribution
- experiments run automatically on start, results available in shared folder

Where are we now?

- 1 "demo" experiment, 14 experiments from published papers, and 3 other experiments available
- prototype web interface to support authors in making recomputable experiments
- stickers

Recomputation @ ECAI ×		
🔇 > 😋 🔯 burgs://kinimvie.cs.st-andrews.ac.uk:8443		☆ 🐵 🍇 ≡
Recomputation @ ECAI		
Experiments for Lars Kotthoff		
	Enter VM name Based on	
test		
stopped	Start VM Open Terminal	Box up

What have we done so far?

- tutorials at 2 major AI conferences
- worked with authors of accepted papers to make them recomputable
- summer school on empirical methods in CS







St Andrews, Scotland, 4-8 August 2014

Insight Centre for Data Analytics

11 December 2014

Slide 13

Experiences – Tutorials

- lots of interest in the AI community
- working with every author individually does not scale
- asking authors to do more work themselves leads to lower conversion rate

"So, you will need approximately 50.000.000 seconds [1.6 years] of computation on 2.27 GHz processors, memory limit of 3.9 GB (timeout is 2h)."

Experiences – Resources

"It requires a GPU/CUDA card of a certain version. I don't know if virtual machines work."

"[the script] delegates the theorem proving tasks to Geoff Sutcliffe's automated theorem proving servers in Miami."

Experiences – Resources

"My experiments are based on games that humans played with/against a computer-agent in lab conditions. In order to run these experiments, it's needed to have a research assistant in the lab that will take care on the process of the experiment [...]. My experiments were done only by students from specific departments and mean age."

Experiences – Licenses

- some experiments use proprietary software such as IBM CPLEX
- some authors use other people's tools that they don't own
- cannot include in virtual machine image and distribute
- solution: allow user to provide relevant software on host machine

Experiences – Summer School



Insight Centre for Data Analytics

Experiences – Summer School

- case studies on issues related to recomputation: ethics, parallel and distributed experiments, non-CS (computational) experiments
- paper written during the school was itself recomputable

 available on recomputation.org
- lots of issues related to recomputation that require further research!

Obstacles

- make it easier for researchers to make their experiments recomputable
- infrastructure
- tools

Outlook

- promising first 18 (+ a bit) months
- lots of interest from different communities
- now need to "mature"

