Artifact Evaluation Track

Wilhelm (Willi) Hasselbring
Kiel University, Germany
http://se.informatik.uni-kiel.de/

Petr Tuma
Charles University, Czech Republic
http://d3s.mff.cuni.cz/

April 12th, 2018
"Replication is the ultimate standard by which scientific claims are judged."
Viewpoint

The Real Software Crisis: Repeatability as a Core Value

Sharing experiences running artifact evaluation committees for five major conferences.

“Science advances faster when we can build on existing results, and when new ideas can easily be measured against the state of the art.”

Repeatability, not necessarily reproducibility

Several ACM SIGMOD, SIGPLAN, and SIGSOFT conferences have initiated artifact evaluation processes. Now ICPE, too.
Artifact Review and Badging:

A variety of research communities have embraced the goal of reproducibility in experimental science.
[more information]

**Artifacts Evaluated – Functional**
The artifacts associated with the research are found to be documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation.

**Artifacts Evaluated – Reusable**
The artifacts associated with the paper are of a quality that significantly exceeds minimal functionality.

**Artifacts Available**
Author-created artifacts relevant to this paper have been placed on a publically accessible archival repository.

**Results Replicated**
The main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.

**Results Reproduced**
The main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.
Some numbers for the first artifact evaluation at ICPE

• Some numbers for ICPE 2018
  – 59 submitted full research papers
  – 14 accepted full research papers
  – 6 submitted artifacts
  – 2 accepted artifacts, evaluated as functional
  – 0 accepted artifacts, evaluated as reusable

• Accepted Artifacts Available
  – At the SPEC Research Group Zenodo repository
  – At the ACM Digital Library
Adaptive Performance Optimization under Power Constraint in Multi-thread Applications with Diverse Scalability

Stefano Conoci
DIAG – Sapienza, University of Rome
Rome, Italy
conoci@diag.uniroma1.it
conoci@lockless.it

Bruno Ciciani
DIAG – Sapienza, University of Rome
Rome, Italy
ciciani@diag.uniroma1.it
ciciani@lockless.it

Pierangelo Di Sanzo
DIAG – Sapienza, University of Rome
Rome, Italy
disanso@diag.uniroma1.it
disanso@lockless.it

Francesco Quaglia
DiCII – University of Rome Tor Vergata
Rome, Italy
quaglia@uniroma2.it
quaglia@lockless.it

Characterizing the Microarchitectural Implications of a Convolutional Neural Network (CNN) Execution on GPUs

Shi Dong
Dept. of Electrical and Computer Engineering
Northeastern University
shidong@ece.neu.edu

Xiang Gong
Dept. of Electrical and Computer Engineering
Northeastern University
xgong@ece.neu.edu

Yifan Sun
Dept. of Electrical and Computer Engineering
Northeastern University
yifansun@ece.neu.edu

Trinayan Baruah
Dept. of Electrical and Computer Engineering
Northeastern University
tbaruah@ece.neu.edu

David Kaeli
Dept. of Electrical and Computer Engineering
Northeastern University
kaeli@ece.neu.edu
Thanks to the Evaluation Committee!

Thomas F. Düllmann, University of Stuttgart, Germany
Holger Eichelberger, University of Hildesheim, Germany
Vincenzo Ferme, University of Lugano, Switzerland
Nikolas Roman Herbst, University of Würzburg, Germany
Vojtěch Horký, Charles University, Czech Republic
Alexey S. Ilyushkin, TU Delft, The Netherlands
Holger Knoche, b+m Informatik AG, Germany
Haiyang Sun, University of Lugano, Switzerland
Michael Vierhauser, University of Notre Dame, United States of America
Jóakim von Kistowski, University of Würzburg, Germany
Felix Willnecker, fortiss GmbH, Germany
Is it worth making the effort?

“If I have seen further it is by standing on the shoulders of giants.”
Isaac Newton, 1676

“Science advances faster when we can build on existing results, and when new ideas can easily be measured against the state of the art.”

[Krishnamurthi & Vitek 2015]
Impact of Artifact Evaluation

Fig. 1. Average citation counts of AE and non-AE papers for conferences that used AE in 2013 to 2016 (conferences: VISSOFT, PPoPP, POPL, PLDI, PACT, OOPSLA, ISSTA, FSE, ECRTS, ECOOP, CGO, CAV).

Source: Bruce R. Childers, Panos K. Chrysanthis: „Artifact Evaluation: Is it a Real Incentive?”
In: Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE5.2), 2017

It seems that repeatability and reproducibility of performance research results brings specific challenges

• However, it is also of particular importance to our field
Recommendations for next edition

• Introduce a “rebuttal” process
  – To address replication hurdles after “kicking the tires”
  – Reviewers should see issues that other reviewers raised
  – One round should be enough
  – Could be handled via Easychair

• Authors should explicitly state hardware and system software requirement in the abstract
  – Such that reviewers can consider this in the bidding phase
  – With specific hardware and system software requirements, it is recommended to provide remote access to such systems

• Provide a checklist for reviewers to "standardize" what to look into, what not, when to stop/reject etc.
  – Could also be made available to authors

• Consider the artifact evaluation also for the
  – Results replicated badge
  – As, for instance, SIGMOD does

• Publication only at ACM Digital Library?
CFP: Open Computer Science

https://www.degruyter.com/view/j/itit