UNDER REVIEW

RESCIENCE C

Editorial

ML Reproducibility Challenge 2020

Koustuv Sinha[1,2,4], Sasha Luccioni[2,3], Jesse Dodge[6], Jessica Zosa Forde[5], Robert Stojnic[4,7], and Joelle Pineau[1,2,4]

[1]School of Computer Science, McGill University, Montreal, Canada – [2]Montreal Institute of Learning Algorithms (Mila), Canada – [3]Université de Montréal, Canada – [4]Facebook AI Research, Montreal, Canada – [5]Brown University, USA – [6]Allen Institute for AI, USA – [7]PapersWithCode, USA

Edited by
(Editor)

Received
21 May 2021

Published
–

DOI
–

1 Introduction

Reproducibility is a key ingredient for an impactful scientific discovery, which allows future practitioners to build on the shoulders of published work. Reproducibility is also an important step to promote open and accessible research, allowing the scientific community to quickly integrate new findings and convert ideas to practice more seamlessly. In the spirit of promoting a culture of reproducible science in the Machine Learning community, we have hosted the fourth iteration of the ML Reproducibility Challenge in 2020. Unlike previous years, in this challenge we increased the scope to include a broad range of top Machine Learning conferences, including NeurIPS, ICML, ICLR, CVPR, ECCV, ACL and EMNLP. The goal of this challenge was to investigate the reproducibility of accepted papers published in these top conferences, and in-turn contribute to the better understanding of their central claims. In this special issue of ReScience C Journal, we present the peer-reviewed accepted papers of the 2020 ML Reproducibility Challenge.

2 Challenge

The goal of the challenge was to reproduce the central claims of papers published in top Machine Learning conferences of the year. Unlike the last iteration (NeurIPS 2019), in this year we focus on the central claim of the papers, and participants were open to choose to work on either all claims or partial claims depending on the complexity of the project. Participants were also free to reuse authors' code when available, while being encouraged to explore beyond simply running the code provided to verify reproducibility. The challenge involved a "Claim paper" step, where early on participants were encouraged to submit a claim on the paper they wished to work on using the OpenReview portal. The objective of the claiming process was to help participants narrow down their task by writing a short summary of items they wished to explore in reproducing the papers.

As in the last iteration, participants were free to claim multiple papers, and multiple teams could claim the same paper. In this year's iteration, a total of 244 claims were submitted, which is a 41% increase over last year. However, the total number of final submissions was slightly lower at 82 papers (vs 84 in previous year). We had participation from 48 institutions (47 universities and 1 industry organization). Top participating institutes consisted of University of Amsterdam, Netherlands, Indian Institute of Technology Gandhinagar, India, University of Waterloo, Canada, San Jose State University,

Correspondence should be addressed to Koustuv Sinha (koustuv.sinha@mail.mcgill.ca)
The authors have declared that no competing interests exist.

USA. In these cases (and several others), a high participation rate occurred when a professor at the university used this challenge as a final course project.
It is also worth noting that this iteration of the challenge witnessed a significant jump in the quality of the reproducibility reports. After extensive peer review, in this special issue we present the top 23 accepted reports, selected from 82 submissions, thus driving up the acceptance rate from 11% last year to 28% this year.

3 Reproducibility Summary and Template

The substantial increase in the quality of the submitted reports is attributed to two key decisions: reducing the scope of the challenge to cover central claims of the paper, and introducing the Reproducibility Summary Template to help authors to communicate their results and findings clearly and concisely, given that scientific communication is challenging. While there are many different types of papers, there are also common elements across ML, NLP, and vision. In a reproduction report, the main emphasis is on good reporting.
This year we introduced a first-page summary and optional template. This template:

- Has a place for reporting all items on the reproducibility checklist
- Acts as guide for researchers
- Shows that they understood the main claims, and the evidence that supports those claims.
- Allows readers to quickly look up what they're interested in – readers know what sections to check

4 Going beyond

Another crucial factor contributing to the increase in quality of the reports is that this year's edition saw many authors going above and beyond the original paper, running additional experiments and analyses and converting code between frameworks (e.g. Tensorflow → PyTorch). This was particularly encouraging since it is an indicator of the evolution of the Reproducibility Challenge from a simple replication of the initial results to a broader scope in terms of depth of engagement with reproducing research. We are impressed with the work of this year's authors and look forward to seeing further developments of the challenge!

5 Platforms

This challenge is conducted with the support of PapersWithCode and OpenReview. PapersWithCode is an open, collaborative platform to discover latest trending machine learning research papers with their codebases, which enables rapid re-usability and reproducibility of published works. PapersWithCode enabled us to reach a wide audience of students and researchers who participated in the competition. As last time, OpenReview provided crucial logistic support by providing an unique platform to claim and submit reproducibility reports. After submission, all reports went through a thorough peer review process consisting of hundreds of reviewers from the Machine Learning community, and OpenReview provided an easy-to-use platform for managing reviews and administrative processes. Finally, we used a public Github repository to perform the final editorial process of converting accepted papers into ReScience format, and thereby publish 23 high quality reports in this special issue.

6 Conclusion

Reproducibility of central claims of papers published in Machine Learning conferences has been a center of considerable attention over the past several years. Conferences such as NeurIPS, ICLR, AAAI, ICML, EMNLP have routinely included reproducibility workshops and challenges to cultivate the culture of reproducible science in the community. Several conferences have also introduced code submission policies and Reproducibility Checklists to further advance the cause and build momentum of reproducible science. We hope our continued endeavour of hosting regular reproducibility challenges and publishing high-quality peer-reviewed reproducibility reports will contribute more information about the existing published papers, and help strengthen their core contributions in the process, while also promoting open, accessible and sound machine learning research.

7 Acknowledgements

We thank the board and program committee of NeurIPS, ICML, ICLR, ACL, EMNLP, CVPR and ECCV for partnering with us in this massive initiative and supporting the challenge. We thank the OpenReview team (in particular Andrew McCallum, Parag Pachpute, Melisa Bok, Celeste Martinez Gomez, Pam Mandler and Mohit Uniyal) for their constant support in hosting and building the customized portal used in our challenge. We thank Robert Stojnic, Ross Taylor and Elvis Saravia from PapersWithCode for hosting and supporting the challenge along with its logistics. We thank the ReScience board (in particular Nicolas Rougier, Konrad Hinsen, Olivia Guest and Benoît Girard) for presenting the accepted reports in their esteemed journal. Finally, we thank all of our participants who dedicated time and effort to verify results that were not their own, to help strengthen our understanding of the concepts presented in the papers. A special thank you to Ana Lucic (University of Amsterdam), who instructed and supported several of the student teams whose reports are featured in this issue.

8 Reviewers

Our reviewers need a special section dedicated to thank them for their tireless efforts in screening and providing valuable feedback to the Area Chairs to select the best papers. We were fortunate enough to attract a large pool of reviewers, who spent their precious time to critically review the reports. We would like to specifically acknowledge our Emergency reviewers (marked in *) who responded to our call for help to review some additional reports at the last minute. We hope that our reviewer base will keep supporting us in this endeavour in future.

Abhinav Agarwalla
Akshita Gupta
Ali Hürriyetoğlu
Andreas Ruttor
Andrew Drozdov
Anis Zahedifard
Arna Ghosh
Azin Shamshirgaran
Bharathi Srinivasan *
Chao Qin
Charbel Sakr
Chuan Li
Clement Laroche
David Arbour
Di He
Dmitriy Serdyuk
Donghyeon Cho
Dylan Hadfield-Menell
Emmanuel Bengio
Ernest K. Ryu
Fan Feng
Fatemeh Koochaki
Fernando Martínez-Plumed
Gagana B
Georgios Leontidis
Haitian Sun
Hanna Suominen
Hao He
Heng Fang
Huaibo Huang

Huseyin Coskun
Ishani Vyas
Jiakai Zhang
Jiangwen Sun
Jie Fu
Jitong Chen
John Frederick Wieting
Kanika Madan
Katherine Lee
Kaushy Kularatnam
Koustuv Sinha *
Leo M Lahti
Leonid Kholkine
Levent Sagun
Li cheng
Lijun Wu
Linh Tran *
Lluis Castrejon
Mahzad Khoshlessan
Maneesh Kumar Singh
Mani A,Marija Stanojevic
Maria Maistro
Marija Stanojevic
Marin Misur
Massimiliano Mancini
Matthew Kyle Schlegel
Matthew Ryan Krause
Maxime Wabartha
Maxwell D Collins
Md Imbesat Hassan Rizvi
Melanie F. Pradier
Michal Drozdzal
Mingrui Liu
Monjoy Saha
Neal Fultz
Nikolaos Vasiloglou
Olga Isupova *
Olivier Delalleau
Opeyemi Osakuade
Otasowie Owolafe *
Pablo Robles-Granda
Pascal Lamblin
Patrick Philipp
Paul Tylkin
Peter Henderson
Prasad Sudhakara Murthy
Praveen Narayanan
Radha Chitta
Rajanie Prabha
Sadid A. Hasan
Samira Shaikh
Sandhya Prabhakaran
Satya Prakash Dash
Seohyun Kim
Sepehr Janghorbani
Shuai Kyle Zheng
Siwei Wang
Steffen Udluft
Sunnie S. Y. Kim
Swetha Sirnam *
Tammo Rukat
Taniya Seth
Tobias Uelwer
Ujjwal Verma
Vibha Belavadi
Víctor Campos
Wenbin Zhang
Wenhao Yu
Xavier Bouthillier
Xavier Sumba
Xiang Zhang
Xiao Zhang
Xin Guo
Yufei Han
Yuntian Deng
Zhangjie Cao