ML Reproducibility Challenge 2022

Koustuv Sinha[1, ID], Maurits Bleeker[7, ID], Samarth Bhargav[7, ID], Jessica Zosa Forde[5, ID], Sharath Chandra Raparthy[1,3,4, ID], Jesse Dodge[6], Joelle Pineau[1,2,3, ID], and Robert Stojnic[1,8, ID]

[1]Meta AI, USA / Canada − [2]School of Computer Science, McGill University, Montreal, Canada − [3]Mila - Quebec AI Institute, Montreal, Canada − [4]Université de Montréal, Canada − [5]Brown University, USA − [6]Allen Institute for AI, USA − [7]University of Amsterdam, Netherlands − [8]PapersWithCode, USA

1 Introduction

The importance of reproducibility in science cannot be overstated. It is one of the key mechanisms in place to enforce the high standards of scientific discoveries, and a key ingredient for an impactful scientific discovery, allowing future practitioners to build on the shoulders of prior work. Reproducible science also promotes 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 sixth iteration of the ML Reproducibility Challenge in 2022. Following the trend of inclusivity and breadth, this iteration involves a challenge to reproduce papers published in nine top conferences in Machine Learning, including NeurIPS 2022, ICML 2022, ICLR 2022, ACL 2022, EMNLP 2022, CVPR 2022, ECCV 2022, AAAI 2022, IJCAI-ECAI 2022, ACM FAccT 2022, SIGIR 2022, and also for papers published in top ML journals in 2022, including JMLR, TACL and TMLR. An important objective of this challenge is to contribute toward improving the understanding of the central claims of the papers published in these top conferences, by inviting participants to run reproducibility study on them. In this special issue of ReScience C Journal, we are proud to present the peer-reviewed accepted papers of the 2022 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. Participants were invited 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.

As in the last iteration, participants were free to claim multiple papers, and multiple teams could claim the same paper. In this iteration, we observed a slight decline of reproducibility report submissions to 74, compared to 102 from last year. Reproducibility reports were spread across all top conferences, with most papers chosen from CVPR 2022, and the least from ACL 2022. A majority of the participants were students using the challenge as a part of their machine learning courses from various institutions around the world, including but not limited to: KTH (Royal Institute of Technology Stockholm, Sweden); Åbo Akademi University, Finland; University of Amsterdam, Netherlands;

University of Ljubljana, Faculty of Computer and Information Science; University of Michigan, USA; Carnegie Melon University, USA; and Vrije Universiteit Amsterdam. After in-depth peer review, in this special issue we present the top **45** accepted reports, selected from 74 submissions, showcasing a significant increase in paper acceptance numbers. This increase is largely due to significant improvements in the quality of the reports & their methodology, which is encouraging to see.

3 Best Paper Awards

Following the tradition set last iteration, we are presenting best paper awards to a few select reports to highlight the excellent quality all-round of their reproducibility work. The selection criteria consisted of votes from the Area Chairs, based on the reproducibility motivation, experimental depth, results beyond the original paper, ablation studies, and discussion/recommendations. Since the quality of these top papers are exceptionally high, we decided to change the "Best paper" award nomenclature to "Outstanding Paper" and "Outstanding Paper (Honorable Mentions)" to closely reflect the individual paper qualities of the best performing papers. We believe the community will appreciate the strong reproducibility efforts in each of these papers, which will improve the understanding of the original publications, and inspire authors to promote better science in their own work.

3.1 Outstanding Paper Award

- Kaiser Sun, Adina Williams, Dieuwke Hupkes; *A Replication Study of Compositional Generalization Works on Semantic Parsing*

- Seungjae Ryan Lee, Seungmin Lee; *[Re] Pure Noise to the Rescue of Insufficient Data*

3.2 Outstanding Paper Award (Honorable Mentions)

- Yannik Mahlau, Lukas Berg, Leo Kayser; *[Re] On Explainability of Graph Neural Networks via Subgraph Explorations*

- Alexander Shabalin, Ildus Sadrtdinov, Evgeniy Shabalin; *[Re] "Towards Understanding Grokking"*

- Skander Moalla, Manuel Madeira, Lorenzo Riccio, Joonhyung Lee; *[Re] Reproducibility Study of Behavior Transformers*

4 Platforms

This challenge is conducted with the support of PapersWithCode[1], OpenReview[2] and Kaggle[3]. PapersWithCode is an open, collaborative platform to discover latest trending machine learning research papers with their codebases, which enables rapid reusability and reproducibility of published works. PapersWithCode enabled the challenge organizers to reach a wide audience of students and researchers who participated in the competition. As was the case last year, 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. We used a

---

[1] https://paperswithcode.com
[2] https://openreview.net/group?id=ML_Reproducibility_Challenge/2022
[3] https://www.kaggle.com/

public Github repository[4] to perform the final editorial process of converting accepted papers into ReScience format, and thereby publish 45 high quality reports in this special issue.

4.1 Kaggle Awards

Kaggle deserves a special mention as they partnered with us in this iteration to provide awards to the best papers and reviewers. Kaggle has provided awards in the form of Google Cloud Compute (GCP) credits worth of 500k USD, which is extremely beneficial to conduct exploratory research leveraging high performance computing platform of Google. Kaggle has sponsored this award to outstanding papers and reviewers based on a final decision of the Kaggle awards committee[5]. We thank Kaggle for providing such generous award and enabling reproducible research in the Machine Learning community.

5 Acknowledgement

We thank the board and program committee of NeurIPS, ICML, ICLR, ACL, EMNLP, CVPR, ICCV, AAAI and IJCAI for partnering with us in this massive initiative and supporting the challenge. We thank the OpenReview team for their constant support in hosting and building the customized portal used in our challenge. We thank Nate Keating, D. Sculley and team from Kaggle for partnering with MLRC 2022 and providing generous awards to the challenge winners, participants and reviewers. We thank 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.

6 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 (D.Sculley, Samarth Bhargav, Maurits Bleeker, Jessica Zosa Forde, Sharath Chandra Raparthy and Koustuv Sinha) 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 who responded to our call for help to review some additional reports at the last minute. Following the trend from the last iteration, we also announce Outstanding Reviewer Award to select reviewers for their high quality and timely reviews for the challenge. The selection criteria involved votes from the Area Chairs after careful review of the reviews posted in the challenge. We thank the reviewers for their exceptional effort and hope they will continue to support us in future iterations.

6.1 Outstanding Reviewers

Divyat Mahajan

Furkan Kınlı

Fan Feng

Tobias Uelwer

Siba Smarak Panigrahi

Prateek Garg

Maxwell D Collins

Harman Singh

Pascal Lamblin

Taniya Seth

Gabriel Bénédict

Olivier Delalleau

Philipp Hager

Saketh Bachu

Sunnie S. Y. Kim

7 Conclusion

Reproducibility of central claims of papers published in Machine Learning conferences has been a center of considerable attention over the past several years. In recent 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 annual challenges and publishing high-quality peer-reviewed reproducibility reports will contribute more information about existing published papers, and help strengthen their core contributions in the process, while also promoting open, accessible and sound machine learning research.