ACL 2021

Second Workshop on Insights from Negative Results in NLP

Proceedings of the Workshop

November 10, 2021
Online and Punta Cana, Dominican Republic

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-93-0

Introduction

Publication of negative results is difficult in most fields, and the current focus on benchmark-driven performance improvement exacerbates this situation and implicitly discourages hypothesis-driven research. As a result, the development of NLP models often devolves into a product of tinkering and tweaking, rather than science. Furthermore, it increases the time, effort, and carbon emissions spent on developing and tuning models, as the researchers have little opportunity to learn from what has already been tried and failed.

Historically, this tendency is hard to combat. ACL 2010 invited negative results as a special type of research paper submissions[1], but received too few submissions and did not continue with it. *The Journal for Interesting Negative Results in NLP and ML*[2] has only produced one issue in 2008.

However, the tide may be turning. Despite the pandemic, the second iteration of the *Workshop on Insights from Negative Results* attracted 39 submissions and 14 presentation requests for papers accepted to "Findings of EMNLP". NeurIPS 2021 also accepted the second iteration of *"I (Still) Can't Believe It's Not Better"*[3].

The workshop maintained roughly the same focus, welcoming many kinds of negative results with the hope that they could yield useful insights and provide a much-needed reality check on the successes of deep learning models in NLP. In particular, we solicited the following types of contributions:

- broadly applicable recommendations for training/fine-tuning, especially if X that didn't work is something that many practitioners would think reasonable to try, and if the demonstration of X's failure is accompanied by some explanation/hypothesis;
- ablation studies of components in previously proposed models, showing that their contributions are different from what was initially reported;
- datasets or probing tasks showing that previous approaches do not generalize to other domains or language phenomena;
- trivial baselines that work suspiciously well for a given task/dataset;
- cross-lingual studies showing that a technique X is only successful for a certain language or language family;
- experiments on (in)stability of the previously published results due to hardware, random initializations, preprocessing pipeline components, etc;
- theoretical arguments and/or proofs for why X should not be expected to work.

In terms of topics, 19 papers from our submission pool discussed "great ideas that didn't work", 11 dealt with the issues of generalizability, 3 were on the topic of "right for the wrong reasons", 2 papers focused on reproducibility issues, and 4 papers in other relevant topics. Some submissions fit in more than one category. We accepted 20 short papers (51.2% acceptance rate) and granted 4 presentation requests for Findings papers.

We hope the workshop will continue to contribute to the many reality-check discussions on progress in NLP. If we do not talk about things that do not work, it is harder to see what the biggest problems are and where the community effort is the most needed.

[1] https://mirror.aclweb.org/acl2010/papers.html

[2] http://jinr.site.uottawa.ca/

[3] https://i-cant-believe-its-not-better.github.io/neurips2021/

Organizers:

João Sedoc, New York University (USA)
Anna Rogers, University of Copenhagen (Denmark)
Anna Rumshisky, University of Massachusetts Lowell (USA)
Shabnam Tafreshi, University of Maryland: ARLIS (USA)

Program Committee:

Abeer Aldayel, King Saud University (Saudi Arabia)
Amittai Axelrod, DiDi Labs (USA)
Mark Anderson, Cardiff University (UK)
Nada Almarwani, Taibah University (Saudi Arabia)
Sawsan Alqahtani, Amazon (USA)
Wazir Ali, University of Electronic Science and Technology (China)
Aditya Bhargava, Torento (Canada)
Federico Bianchi, Bocconi University Milano (Italy)
Jeremy Barnes, Forsiden IFI (Germany)
Vasudha Bhatnagar, University of Delhi (India)
Yash Butala, Indian Institute of Technology Kharagpur (India)
Meghana Moorthy Bhat, Ohio State University (USA)
Shubham Chandel, Microsoft (USA)
Shubham Chatterjee, University of New Hampshire (USA)
Somnath Basu Roy Chowdhury, University of North Carolina at Chapel Hill (USA)
Yulong Chen, Zhejiang University, Westlake University (China)
Aleksandr Drozd, Tokyo Institute of Technology (Japan)
Alexandra DeLucia, John Hopkins University (USA)
Daria Dzendzik, ADAPT Centre (Ireland)
Lingjia Deng, Bloomberg (USA)
Luis Fernando D'Haro, University of Pennsylvania (USA)
Darren Edmonds, Donald Bren School of ICS (USA)
Antske Fokkens, VU (Amsterdam)
Elisabetta Fersini, University of Milano (Italy)
Haley Fong, City University of Hong Kong (Hong Kong)
Jason Fries, Stanford University (USA)
Catherine Finegan-Dollak, IBM (USA)
Salvatore Giorgi, University of Pennsylvania (USA)
Pedram Hosseini, George Washington University (USA)
Sardar Hamidian, Comcast (USA)
Christopher Klamm, (UKP) Lab - TU Darmstadt (Germany)
Huda Khayrallah, John Hopkins University (USA)
Neha Nayak Kennard, University of Massachusetts Amherst (USA)
Alka Khurana, University of Delhi (India)
Olha Kaminska, Ghent University (Belgium)
Siddharth Karamcheti, Stanford University (USA)
Constantine Lignos, Brandeis University (USA)
Shayne Longpre, Apple (USA)
Sotiris Lamprinidis, Copenhagen Business School (Copenhagen)
Tal Linzen, New York University (USA)

Yiyuan Li, University of North Carolina, Chapel Hill (USA)
Andrei Mircea Romascanu, McGill University (Canada)
Ansel MacLaughlin, Amazon (USA)
Ashutosh Modi, CSE - IIT Kanpur (India)
Deepa Muralidhar, Georgia State University (USA)
Edison Marrese-Taylor, University of Tokyo (Japan)
Jay Mundra, IIT Kanpur (India)
Karo Moilanen, AIG (USA)
Kenton Murray, John Hopkins University (USA)
Enrico Meloni, Università degli Studi di Firenze (Italy)
Markus Müller, Amazon (USA)
Tristan Naumann, Microsoft Research (USA)
Anmol Nayak, Bosch (India)
Constantin Orăsan, University of Surrey (England)
Jessica Ouyang, The University of Texas at Dallas (USA)
John Ortega, JP Morgen (USA)
Adam Poliak, Barnard College (USA)
Ellie Pavlick, Brown University (USA)
Leibny Paola Garcia Perera, John Hopkins University (USA)
Michal Ptaszynski, Kitami Institute of Technology (Japan)
Jordan Rodu, University of Virginia (USA)
Neville Ryant, Linguistic Data Consortium (USA)
Ali Seyfi, George Washington University (USA)
Michael Saxon, UC Santa Barbara (USA)
Mihai Surdeanu, University of Arizona (USA)
Sashank Santhanam, UNC Charlotte (USA)
Vivek Srivastava, TCS Research (India)
Yow-Ting Shiue, University of Maryland (USA)
Silvia Terragni, University of Milano-Bicocca (Italy)
Jannis Vamvas, Zurich University (Switzerland)
Derry Tanti Wijaya, Boston University (USA)
Jin-Ge Yao, Microsoft Research (USA)
Mahsa Yarmohammadi, University of Pennsylvania (USA)
Xiang Zhou, University of North Carolina at Chapel Hill (USA)
Zhuosheng Zhang, Shanghai Jiao Tong University (China)

Invited Speakers:

Zachary Lipton, Carnegie Mellon University
Noah Smith, University of Washington / Allen Institute for AI
Rachael Tatman, Rasa
Bonnie Webber, University of Edinburgh

Table of Contents

Program

Wednesday, November 10, 2021

8:45–9:00	*Opening remarks*
9:00–10:00	*Invited talk: Noah Smith (University of Washington / Allen Institute for AI)* *What Makes a Result Negative?*
10:00–11:15	*Poster session 1*
11:15–11:30	*Social break / coffee time*
11:30–12:30	*Invited talk: Bonnie Webber (University of Edinburgh)* *The Reviewers and the Reviewed: Institutional Memory and Institutional Incentives*
12:30–13:00	*Oral presentation session 1*
13:00–14:00	*Lunch break*
14:00–15:00	*Invited talk: Zachary Lipton (Carnegie Mellon University)* *Some Results on Label Shift and Label Noise*
15:00–16:15	*Poster session 2*
16:15–16:30	*Social break / coffee time*
16:30–17:00	*Oral presentation session 2*
17:00–18:00	*Invited talk: Rachael Tatman (Rasa)* *Chatbots can be good: What we learn from unhappy users*
18:00–18:15	*Closing remarks*

The program is subject to change, please check the EMNLP 2021 conference website for the final program and schedule in different time zones. The program will also be available at `https://insights-workshop.github.io`. All times above are specified in Atlantic Standard Time (GMT-4).