Insights 2020

First Workshop on Insights from Negative Results in NLP

Proceedings of the Workshop

November 19, 2020
Online

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 978-1-952148-66-8

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. The first iteration of the *Workshop on Insights from Negative Results* attracted 35 submissions and 11 presentation requests for papers accepted to "Findings of EMNLP". Moreover, we are not alone: an independent workshop *"I can't believe it's not better!"* is held at NeurIPS 2020[3].

We invited submissions with 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, 15 papers from our submission pool discussed "great ideas that didn't work", 12 dealt with the issues of generalizability, 5 were on the topic of "right for the wrong reasons", and 2 more papers focused on reproducibility issues. We accepted 18 short papers (51.4% acceptance rate) and granted 5 presentation requests for Findings papers.

We hope that this event will be the first of 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/

Organizers:

Anna Rogers, Univeristy Copenhagen (Denmark)
João Sedoc, Johns Hopkins Univeristy (USA)
Anna Rumshisky, University of Massachusetts Lowell (USA)

Program Committee:

Emily Alsentzer, MIT (USA)
Amittai Axelrod, DiDi Labs (USA)
William Boag, MIT (USA)
Anneke Buffone, Facebook (USA)
Aleksandr Drozd, RIKEN (Japan)
Allyson Ettinger, University of Chicago (USA)
Stefan Evert, Friedrich-Alexander-Universitäat Erlangen-Nürnberg (Germany)
Jason Alan Fries, Stanford (USA)
Leibny Paola Garcia, Johns Hopkins University (USA)
Matt Gardner, Allen AI (USA)
Sharath Chandra Guntuku, University of Pennsylvania (USA)
Constantine Lignos, Brandeis University (USA)
Tal Linzen, Johns Hopkins University (USA)
Kyle Lo, Allen Institute for Artificial Intelligence (USA)
Ana Marasović, Allen Institute for Artificial Intelligence (USA)
Matthew B. A. McDermott, MIT (USA)
Neha Nayak, University of Massachusetts Amherst (USA)
Mark Neumann, Allen Institute for Artificial Intelligence (USA)
Denis Paperno, Université de Lorraine (France)
Ellie Pavlick, Brown University (USA)
Masoud Rouhizadeh, Johns Hopkins University (USA)
Jordan Rodu, University of Virginia (USA)
Neville Ryant, University of Pennsylvania (USA)
Djamé Seddah, Université Paris-Sorbonne (France)
Andy Schwatz, Stony Brook University (USA)
Emma Strubell, University of Massachusetts Amherst (USA)
Ekaterina Vylomova, University of Melbourne (Australia)
Chris Welty, Google Research (USA)
Matthijs Westera, Universitat Pompeu Fabra (Spain)
Mark Yatskar, Allen AI (USA)

Invited Speakers:

Rada Mihalcea, University of Michigan (USA)
Byron C. Wallace, Northeastern University (USA)

Table of Contents

Program

Thursday, November 19, 2020

7:00–7:15	*Opening remarks*
7:15–8:15	*Invited talk: Rada Mihalcea (University of Michigan)* *The ups and downs of word embeddings*
8:45–9:15	*Thematic session: representation learning*
8:15–8:45	*Thematic session: dialogue*
9:15–10:00	*Social break / meal time*
10:00–11:00	*Invited talk: Byron C. Wallace (Northeastern University)* *Negative results yield interesting questions, or: a bunch of stuff that didn't work*
11:00–11:30	*Thematic session: question answering*
11:30–12:00	*Thematic session: natural language inference*
12:00–12:30	*Thematic session: lessons learned the hard way*
12:30–13:00	*Social break / meal time*
13:00–14:00	*Interactive orals*
14:00–14:45	*Panel discussion*
14:45–15:00	*Breakout*
15:00–15:15	*Closing remarks*
15:15–16:00	*Happy hour*

The program is subject to change, please check the EMNLP 2020 virtual 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 PST.