119 responses

Summary

Are NLP conferences (ACL, NAACL, EACL, and so forth) the main venues where you present your work?

- Yes: 103 (87%)
- No: 14 (12%)

What do you see as the primary goal of reviewing?

- Enforced later [87]
- Other [12]
- Identify any and all flaws
- Ensure that all results

Other

When should reviewers suggest further work?

- When the presented experiments are too preliminary to draw any interesting conclusions [98] (82%)
- When there are obvious directions for further experimentation [6] (5%)
- When the paper did not compare results with work the reviewer is familiar with [13] (11%)

I am a

- Researcher (indust): 16 (13%)
- Professor: 42 (35%)
- Student: 46 (39%)
- Other: 15 (13%)
- Researcher (industry): 16 (13%)
Should the reviewer take into account the amount of work and effort that has gone into the experiments reported in the paper?
- Yes. Even if there are minor flaws and the results are not hugely superior to the state of the art, their effort should be considered. (48, 40%)
- No. The authors should have chosen more wisely where to invest their effort. (67, 56%)

Is it part of a reviewer’s job to make certain that portions of the work have not been previously published?
- Yes, reviewers should make every possible effort to prevent this from happening. (46, 39%)
- It is ok to point out the overlap if aware of it. (61, 51%)
- No, authors know they are solicited to submit unpublished work and it is their decision to judge if the overlap will hurt their reputation. (9, 8%)

What is a reasonable reviewing load for conference reviewing?
- 3—1 It is not good
- 9—6. This will ensure that the same standards are applied to a larger portion of the papers. Reviewers can always select secondary reviewers if they need help with this load.
- 5—4 This will give reviewers a good sense of the general quality of the submissions but each reviewer can do all of his/her assignments.
- 3—1 It is not good to burden reviewers with too much papers; the lighter their load the better.

Should we look for ways to identify and reward good reviewers?
- Yes, and good reviews should be identified by feedback from area chairs (65, 55%)
- Yes, and good reviews should be identified by feedback from the authors (25, 21%)
- No, this is too subjective a question (11, 9%)
- No, there are too many good reviewers to acknowledge (3, 3%)
- No, writing good reviews is expected behavior, there is no need for special acknowledgements (13, 11%)

To what extent should different ACL conferences coordinate the review of YOUR papers that were previously rejected?
- No coordination necessary; we will fix the issues that were mentioned. Having access to prior reviews will bias and confuse the reviewers. (53, 45%)
- No coordination necessary. We need a fresh set of opinions; mistakes do happen. (41, 34%)
I would like to get the same reviewers as in the previous conference.
I will certainly hope that I get different reviewers.

Should area chairs be expected to read all the papers in their area and write a meta-review summarizing all reviews and discussions that have led to the final decision on the paper?
Yes, this would ensure that the chairs for each area have read all papers and can mitigate problems with reviewers’ strictness or leniency.
No, this would be too much work.

How should reviewers be recruited?
Always provide area chairs with the complete list of reviewers from previous year. They may want to use it to balance geographic areas or expertise.
Recruit reviewers for the conference and use them across areas as needed, no need to commit to a particular area.

Should there be a minimum length of reviews?
Yes, the review should make it clear why a paper is good or bad, not just state the reviewer’s opinion.
No, some papers are so clearly good or bad that not much needs to be said.

Should reviews contain explicit suggestions about what needs to be done in order for the work to become acceptable for publication?
Yes, the purpose of reviewing is to ensure that all reasonable work is eventually published
No, it is the authors’ problem to figure that out
What overall distribution of recommendation scores (assuming a 1-10 scale) should reviewers aim for?
- Uniform distribution: 10 (8%)
- Symmetric bell curve: there are few truly terrible or amazing papers submitted, so reserve 1s and 10s for those papers that are truly special: 45 (38%)
- Sparse distribution: try to take a stand on each paper, either supporting or opposing its acceptance: 49 (41%)
- Positive skew: the modal score should be a reject: 8 (7%)
- Negative skew: the modal score should be an accept: 0 (0%)

When a reviewer leans strongly towards either acceptance or rejection, but also feels that the paper is a bit out of their central expertise, which of the following courses of action is better?
- Give a recommendation score based purely on their opinion, but give a low confidence score: 101 (85%)
- Give a low confidence score, and also pull their recommendation score toward the middle of the scale: 14 (12%)

How important is it for area chairs to easily be able to recruit reviewers outside of the general conference pool to handle specific papers based on expertise?
- Very important: 46 (39%)
- Somewhat important: 57 (48%)
- Not very important, there’s generally enough expertise to go around: 11 (9%)

Author feedback periods have become more popular in the past few years in ICML conferences. How useful do you feel that these are?
- Very useful and worth the time: I have seen opinions swing from accept to reject and/or the reverse on the basis of author feedback, and the decisions are better for it: 47 (39%)
- Somewhat useful, but not worth the time: 39 (33%)
- Not useful: I have seen nothing that affects accept/reject decisions, or I have but the decisions were not any better overall: 27 (23%)
Journal-style reviewing is different from conference-style reviewing in a number of respects, perhaps most notably the lack of a cap on the number of back-and-forth iterations between authors and reviewers. Compared with journal-style reviewing, how do you feel conference-style reviewing stacks up?

- Definitely inferior: 27 (33%)
- Somewhat inferior: 28 (35%)
- About equally good: 27 (33%)
- Somewhat better: 11 (11%)
- Much better: 2 (1%)

What type of work do you consider novel?

- Presents empirical evidence to support a conclusion not previously known to be true: 111 (95%)
- Introduces an algorithmically or mathematically more complicated formulation of a problem in prior work: 42 (36%)
- Introduces a task that no one has worked on before: 106 (91%)
- Outperforms the state of the art approach for a task: 87 (74%)

People may select more than one checkbox, so percentages may add up to more than 100%.

For submissions where empirical results are presented, how much value should be ascribed to those that include code and/or data, compared to submissions that do not:

- The highest possible value. It is essential that readers are able to replicate published empirical findings, and submissions without code and/or data should be rejected.
- Submissions with such resources should be heavily preferred to those without, but not to the extent of outright disqualifying submissions without them.
- Submissions with such resources are desirable, but whether or not a submission includes them should only be a deciding factor when it is not otherwise clear whether a paper sh...
- Submissions with such resources are nice, but should not have any impact on a paper's acceptance.
- Such resources are irrelevant, and should not be allowed to accompany submissions.