

Dear area chairs,

the official discussion period starts this coming Tuesday (Feb 16)!

You are looking at three phases of work now: feedback to reviewers, discussion period, and recommendations to us. Our goal is to provide top notch reviewing to each and every submission, so we want to weed out careless reviews, and we want to make sure all decisions are well found.

1) Feedback to Reviewers

We wanted to reiterate our request to carefully read all reviews in your areas. We want each review to be a good review, and some reviews need a little coaxing from the area chairs. This is probably best done by personal emails. Here is a list of issues that often come up with reviews.

— Look in the reviews for justification of the overall score provided by the reviewer. When scores (either low or high) are not justified, ask the reviewers to extend their reviews and to explain their scores.

— Reviewers may mention that the reported results are not surprising or that the work has been done before, without providing references to prior work. Unless the prior work really is well known (such as part-of-speech tagging for English newswire), ask the reviewers to edit their review to provide discussion of specific prior work, or to omit the claim.

— For papers with ok/high scores, look at the textual reviews for clues about what the field will learn from these papers. If you don't get a sense of what aspects of the presented work are novel, useful or insightful, prompt the reviewers to justify their scores.

— Sometimes, reviewers dismiss a paper as being outside the scope of NAACL. This may indeed be the case for some papers, but for others it may simply mean that the paper is outside the interests of a given reviewer. If you feel a paper is clearly responsive to the call for papers (available here: <http://naacl.org/naacl-hlt-2016/cfp.html>), prompt the reviewer to edit his or her review in light of his or her misreading of the call.

— If a reviewer says he finds the paper hard to understand and gives a low score for this reason (and hopefully also a low confidence score), and the other two reviewers do not have this problem, you may wish to assign an additional reviewer.

2) Discussion Period

For the discussion period, you of course want the reviewers to discuss papers in which there is an outlier. Other issues that you should use the discussion period for:

— In addition, if a review has a problem as discussed above, and the reviewer has not been responsive, you can bring up the issue again in the discussion.

— If a reviewer has low confidence score because they cannot make up their mind, i.e., they see some problems in the work but also like some aspects and are torn on what to recommend, use the discussion period to elicit more opinions on the pros and cons of the paper.

3) Recommendations to us.

You will be recommending for each paper whether it should be accepted, rejected, or unsure (which should be a very small category). In general, we ask you to use your own judgement and to discuss all papers with your co-area chair, even those that seem obvious (since they may not

seem obvious to a second pair of eyes).

Here are some additional suggestions:

— In many cases, the discussion will not result in an agreement. Form your own opinion based on all available information from the reviews, the author response, the reviewer discussion, and the abstract of the papers. For a small number of papers, you may need to read the papers themselves to make the final judgement.

— If a paper is flagged as being outside the scope of NAACL, or of limited interest, consider whether maybe a slightly less typical paper would still be worthy of appearing at NAACL.

— Reviewers sometimes give low score while pointing out missing references to their own work. The references may be indeed appropriate but the reviewer may be too subjective in their evaluation. Try to balance their opinion with that of the other reviewers and with your own.

— When the reviewers have given a low confidence rating, get a sense where the lack of confidence come from. If they do not feel they understood the paper, it is ok to downplay their score when making decisions.

Above all, please resist the convenience of sorting by `overall_score` when you make your final recommendations.

Ani Nenkova and Owen Rambow