Reviewing Research Papers

ICS690 Seminar

Fall 2015

presentation by
Henri Casanova
As a graduate student, and later as a researcher/professor, one of your most important achievements is to publish high-quality research articles.

Two kinds:
- in conference proceedings
- in journals

“High-quality”:
- Peer-reviewed venues
- Prestigious venues
  - Not all venues are equal
  - Low acceptance rates (it’s a good idea to keep track)
Peer-review

- **Definition**: the evaluation of creative work or performance by other people in the same field in order to maintain or enhance the quality of the work or performance in that field.

- Each work is reviewed by a set of (anonymous) reviewers who are known experts in the field.

- Reviewers should be **independent**
  - Not close collaborators of authors (i.e., reviewers declare conflicts of interest).
  - Sometimes reviews are “double blind” (i.e., reviewers don’t know who the authors are).
Peer-review is Good

- Using multiple independent experts should make it possible to catch weaknesses in the work being reviewed.

- Experts should be able to provide constructive criticism that will lead to a re-submission of an improved version of the work, thereby improving the field.
Peer-review is Bad?

- Peer-review is slow
- Especially true for journals
  - e.g., several of my articles took 2 years from “submit” to “published”
  - several journals are trying to fix this and it’s getting better
  - but, at some lever, if the editor doesn’t do his/her job, a paper can get stuck
Peer-Review is Bad?

- Some argue that peer-review induces a bias against highly original work
  - Reviewers as a whole tend to be more tolerant of works that are consistent with their own views
- History is full of examples of great works that were ridiculed by peer-review because they contradicted the views of the day
  - Famous example: continental drift
    - Idea rejected from 1920’s until 1970’s in spite of some papers that we now can see were convincing
    - Then universally accepted “at once”
    - So we now have a bunch of incorrect peer-reviewed papers published pre-1970’s
- Galileo
Peer-Review is Bad?

- Peer-review can miss scientific errors
- The “publish or perish” syndrome leads to many paper submissions, therefore referees are often overloaded with reviews, and cannot spend enough time doing the reviews
  - Hence “farming out” reviews
- It is not rare for an author to find out later that one of his published papers had an error that was not found by the reviewer (hopefully not too fundamental)
Peer-Review is Bad?

- Peer-review can **miss scientific fraud**
  - Implicit trust that authors behave honestly
    - In some fields, it’s very hard or even impossible to reproduce published results in the time-frame needed for the review (and one would rather spend time doing one’s own research!)
    - A referee's judgment is a guess of plausibility, assuming that the author is honest and the description of the research adequate.

- Automatically Generated CS Research Papers:
  - Used to evaluate “peer-reviewed” conferences
  - Generated a paper that was accepted to the SCI 2005 conference, and recently one to CITE 2008 by the infamous “Herbert Schlangemann” (see Blog)
Peer-Review is Bad?

- Peer-review can be biased
  - I reject your work so that I can steal your ideas!
  - Implicit trust that reviewers behave honestly

- Hence the importance of technical reports!
Peer-Review is Bad?

- Peer-review is not consistent
  - A paper can be rejected at a conference and accepted at another of comparable quality
  - “Luck of the draw” due to the choice of the particular referees for a paper
    - “Tough” referees (justly or unjustly so)
    - “Lenient” referees
  - “Peer review evaluation decision” and “scientific quality” are not equal, just hopefully correlated
- Prepare yourselves for peer review decisions that you think are unfair. Get over it :)
Peer-Review is what we have

- In spite of all its flaws, peer-review is what we use, and it’s the best we can do at the moment

- Let’s try to understand the process a bit better, and then see what it takes to be a good referee
Outline

- Introduction
- Peer-review for Conferences
- Peer-review for Journals
- Being a Referee
Peer-Review for Conferences

- Talking here about conferences that publish full papers and the selection process is stringent
  - Like in Computer Science
  - Not like in, e.g., biology
- A conference has a Program Chair
- The Program Chair is responsible for ensuring the quality of the program, i.e., the set of articles in the conference’s proceedings
  - There may be Program Vice-Chairs, for different scientific areas
- The Program Chair forms a Program Committee (PC)
  - A set of experts in the field who have agreed to review papers
  - Being on PCs is good for an academic career
Peer-Review for Conferences

- PC members receive a set of papers to review
  - Some conferences insist that PC members review papers themselves
  - Some conferences allow the use of “outside referees”, while keeping PC members responsible for the reviews
- Papers are evaluated by filling in a review form
- Each paper is evaluated by 3, 4, or 5 PC members
- There is then a PC meeting
  - Face-to-face full-day meeting, on-line, or hybrid
  - Resolve “conflicting reviews” issues via discussion among PC members, editing reviews
  - Come up with list of accepted papers
  - Authors are notified
  - Possible phase of rebuttals
Peer-Review for Conferences

- Typical Structure of a review form:
  - Set of criteria to be rated on some scale
  - Overall recommendation
  - Comments for the authors
  - Comments for the PC, not shown to the authors
Peer-Review for Conferences

- Typical criteria that are to be rated on a scale (often from 0 to 5)
  - Originality and contribution
  - Relevance to the conference
  - Quality of presentation
  - Reference to previous work
  - Confidence of the reviewer in his/her assessment
  - Technical strength/soundness
Peer-Review for Conferences

- Typical overall recommendation options
  - Strong Accept
  - Accept
  - Weak Accept
  - Borderline / Accept only if room
  - Weak Reject
  - Reject
  - Strong Reject
  - Award quality (e.g., best paper)
Peer-Review for Conferences

- Comments for authors
  - This is where the reviewer argues his rating of the paper and provides constructive criticism
  - Often starts with a brief summary of the paper
    - Which helps PC chair and other PC members understand the context of the review during the PC meeting
  - Is often split in two parts
    - high-level comments
    - detailed comments
  - Should not be succinct in general
    - Nothing worse than the “1-line review”, especially if the ratings are not good
Peer-Review for Conferences

- Comments for the PC
  - This is where the reviewer can make comments that should not be seen by the authors
  - Examples:
    - There is a clear fraud here
    - I rated this a “weak reject” but I was really close to a “reject”
    - There is one element of the paper that I am not able to evaluate
    - I am really not sure that this paper is relevant to the conference
    - This paper is exactly on the same topic as submission #XX, which I also reviewed
The process for journals is a bit different
- Coordination through an editor who chooses reviewers
- Somewhat different criteria/questions

 Biggest difference is overall recommendation:
- Accept without changes
- Accept with minor revision
  - Trust that authors will make the minor revisions
- Accept with major revision
  - Authors resubmit paper within 2-3 months, with a document that gives specific answers to the reviewers’ comments
  - Reviewers make a decision based on these answers
- Reject
Outline

- Introduction
- Peer-review for Conferences
- Peer-review for Journals
- Being a Referee
Being a Referee

- Being a referee is something people learn to do ad-hoc
  - Your advisor gives you a paper to review
  - You have no idea what to do
  - Hopefully he/she helps you and critiques your review
  - You keep doing reviews, until you know how

- An article that focuses on refereeing papers in computer science

- How to become a referee?
  - Writing papers
  - Participate in societies
  - Be recommended by somebody
  - Offer to help

- Let’s go through some of the main points and fill in some blanks
The Task of the Referee

Key questions:

- What is the purpose of the paper?
  - Can you even tell?
- Is the paper appropriate? (for the venue)
- Is the goal significant?
  - Real problem? do we care about solving it?
- Is the method of approach valid?
  - Bad assumptions, bad models, bad scientific method, ...
- Is the actual execution of the research correct?
  - Bad math? Bad proofs? Bad experimental design? ..
- Are the correct conclusions drawn from the results?
- Is the presentation satisfactory?
- What did you learn?
- Is it novel?
Classes of papers

- [1] talks about 7 categories:

  1. Major results; very significant (<1% of papers)
  2. Good, solid, interesting work; a definite contribution (<10% of papers)
  3. Minor, but positive, contribution to knowledge (10-20%)
  4. Elegant and technically correct but useless. This category includes sophisticated analyses of flying pigs
  5. Neither elegant nor useful, but not actually wrong
  6. Wrong and misleading
  7. So badly written that technical evaluation is impossible
Standards

- A difficult thing is to make a recommendation appropriate for the venue
  - a “best paper award” paper at some workshop could very well be rejected from a better conference
    - This just happened to me!

- Know how competitive the venue is
  - Ask your peers
  - Ask conference organizers
  - Look at acceptance rate
  - Impact factors?
Novelty?

- Evaluating novelty is always difficult
  - Hopefully you’re aware of at least a few related published papers
  - Referees don’t have time to do lit reviews

- The paper should explicitly state why it’s novel and as a reviewer you can be more or less convinced

- It’s a good idea to do quick lit searches to see whether relevant work’s missing
  - 5 minutes on Google Scholar can help
  - As a student, you were given the paper to review because it should be relevant for you, and so doing a long lit searches will help you in the long run
Cite Work Carefully

- It’s not acceptable to say “this is already done” without giving precise references to published work
  - Worse: “this is so obvious it clearly was published before”
- If you reject a paper for novelty issues, you really have to be ultra precise about your claims
Detecting Fraud

- Plagiarism
  - A good idea to use Google to look for full sentences of the abstract/intro, and the title
  - It takes only a few minutes

- Self-plagiarism, unfortunately, is rampant
  - Authors going after the “minimum publishable unit” and pushing it too far
  - There are rules about self-plagiarism, especially between conference and journal publications
  - An author that does not reference his/her own previously published work, it’s highly suspicious
You’re not an Expert

- Often you will have to review papers on topics that you don’t feel you’re an expert on
  - This may even be by design
- You have to be upfront about it
  - In comments to the PC/editor
  - By rating your “confidence” as low
- You should still be able to do a partial evaluation, but for novelty
- Or, you can try to find an appropriate referee for the paper
  - If allowed by the venue
Preserving Anonymity

- Be careful not to reveal your identity as a reviewer
  - e.g., “You should reference the following seminal papers [Casanova 1997], [Casanova 2001], [Casanova 2002], [Casanova 2004], [Casanova 2008], [Casanova 2009]”

- Once you start being on many PCs, you may get the same paper twice or even more
  - If you feel uncomfortable about this, just declare a conflict of interest or inform the PC Chair/editor that you may be biased
Constructive Criticism

- It is important to always be civil and not be offensive:
  - “The authors are stupid...”
  - “This was a waste of ink and paper...”
  - “Go back to grade school...”
  - “I hope the authors were drunk...”
  - (These are all real)

- This serves no purpose whatsoever, and could be dangerous in case there is a mishap and reviews are made public and non-anonymous!
  - I’ve seen this happen as well

- Let’s now look at a few excerpts of reviews I’ve written
  - With pieces removed to ensure anonymity
  - I’m not saying I’m the best referee around!
This is an interesting paper that makes a clear contribution. The experimental evaluation is solid and convincing, and the technique is based on well-argued principles. The authors' justification of the fact that XXX measures can be added to an existing application is reasonable, and I would even argue that applications that cannot measure XX are simply ill-suited to production workloads in the current technology landscape.

The English, while understandable, needs quite a bit of work. While the authors are non-native speakers hopefully they can find a way to improve it.
This paper targets the important topic of XXX. Unfortunately there are many problems:

- The original scientific contribution is difficult to identify. The XXX approach is from another paper, and in fact if this paper gave more details about it it wouldn't hurt.

- The content of the paper up to Section 4.2 makes no novel contribution, as most of the ideas are described in rather vague and high-level terms and are well-known concepts (i.e., job priorities, job replication, using a ring for fault-tolerance).

- Part of this reviewer's inability to understand the research contribution certainly stems from the presentation of the paper, and in particular the English. The authors, like many, are not native English speakers, but unfortunately the English in the paper makes it very difficult to follow. So there may be interesting contributions therein, but they are not clearly identified and/or described.
This paper tackles an important topic and I applaud the authors for performing experiments on a real-world platform, which is always labor-intensive. Unfortunately there are many problems with the paper:

- Part of the experimental results are not clearly presented
- The algorithms to which the authors compare their work are not very convincing as they suffer from obvious weaknesses
- Although the authors run real-world experiments, which is great, they use a very naive workload
- The discussion of related work is rather confusing and could use some restructuring
A discussion/evaluation of XXX in these terms would have made the paper much more interesting. I understand that this would require more space, but some of the other content could easily be reduced.

The main problem is the lack of experimental evaluations.

The paper is not sufficiently self-contained and gives too few details regarding models, assumptions, and algorithms.

The paper is not self-contained and extremely incremental. All scheduling algorithms have been published by the authors or by others. Furthermore, in publications by these authors, these algorithms have been evaluated for some of the metrics used in this work. So, this paper is essentially an addendum to a body of work.

The related work section is much too long. The entire left column on page 2 describes extremely well-known algorithms that can be found in all undergraduate OS textbooks.
Review Samples

- Overall, I found the main idea somewhat interesting but not particularly exciting or original, the experiments poorly described, and the presentation overly confusing. With a lot of work, this paper could be worthwhile.

- Another major problem with this work is that there is no comparison with previous approaches, and in particular with the work in [22]. While the authors do reference that work, they do not provide convincing qualitative or quantitative comparison with it.

- Any guess as why XXX is larger than XXX? This seems counter-intuitive and warrants an explanation.
Review Samples

- It turns out that an extended abstract of this (already very poor) paper was submitted to the XXX’08 workshop under the title "XXX" (content from Section I, II, and II). That extended abstract promises to deliver what the content of this submission is! This is a rather outrageous case of double submission.
- One of the authors is from a fictitious institution!
- “performance” is misspelled as “perfromance”
I rated this paper as a "weak reject" because although the novel contribution is not very clear, the authors have built two useful systems that they compare to existing standard technology. Publishing this paper in the proceedings wouldn't be embarrassing I think. It could be put in the “accept if room” category.

In spite of the paper's weaknesses, it is an interesting problem and there are great opportunities here, hence my overall rating of "Maybe". I could easily downgrade to a "Weak Reject".

This paper is just a proposal and a future work section. I gave it a “borderline” because in case more papers are needed, the authors here could perhaps given an interesting talk on this topic in spite of the paper not having a clear research contribution.

I am not 100% familiar with the inner workings of XXX but I found the content convincing.

I was really between a weak accept and an accept and could go either way.
Conclusion

- When writing a paper, think as a reviewer
  - You should be “upfront” about the weaknesses of your work!
- A few key points
  - When writing a review, the key aspect is the novel contributions. Are they highlighted well in your paper?
    - A list in the introduction is increasingly common
  - Parameter value choosing for experiments
    - Avoid “voodoo” parameters that raise eyebrows
    - How does everything you do relate to real-world problems?
  - Related/Previous work clearly identified
    - Along with their limitations
  - Significance of the contribution clearly stated
  - Remember to tell a story that’s well-articulated
    - Easier said than done