KU LEUVEN

The Ins and Outs of Reviewing

— an abridged version of the ECMLPKDD 2019 tutorial

https://dtai.cs.kuleuven.be/events/ReviewingTutorial/materials.php

Hendrik Blockeel and Jesse Davis

KU Leuven, Department of Computer Science

Leuven.AI



Goals of Reviewing

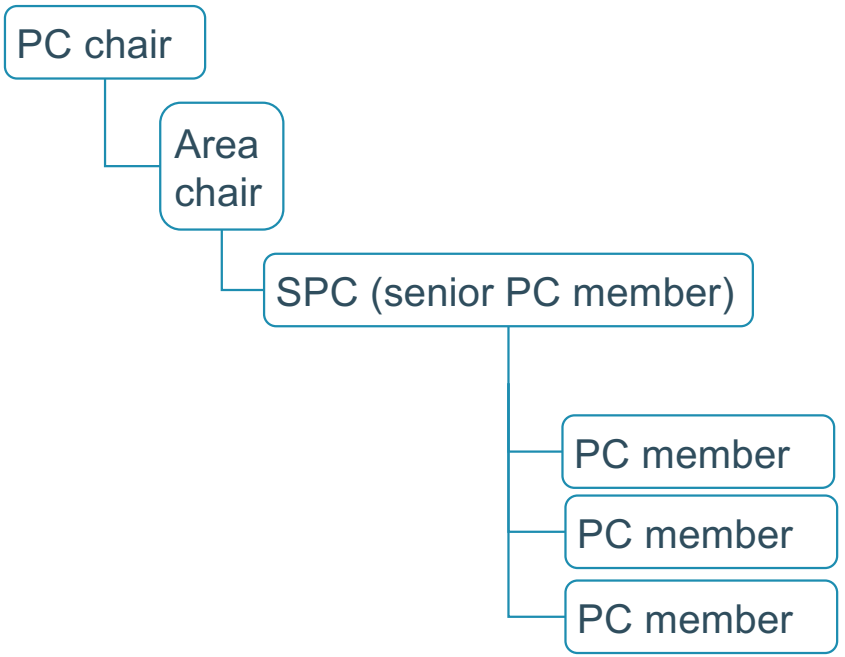
- **Workshops:** Incomplete, work-in-progress
  - Check for “kernel’’ of an idea
  - Try to help focus subsequent research
- **Top-level conferences:** Mature, stand-alone contribution
  - Check for novelty, significance, clarity and technical correctness
  - Given the format, there can be some loose ends
- **Journals:** Complete contribution or unifying several strands of work
  - Extended conference papers: 25% – 30% new work
  - All the details are present, no loose ends

CS Peer Review in Practice Attempts to Assess

- **Technical correctness:** Are there errors in the theory, algo, experiments?
- **Clarity:** Can the reader understand what you did?
- **Novelty:** Does the paper propose a new question, setting, or approach?
- **Significance:** How important or generally applicable are your findings? (i.e., 0.01% reduction in error on MNIST vs. XGBoost)
- **Reproducibility:** Can you or someone else reproduce, e.g., results?

Organization of conference reviewing

People involved in the reviewing of a single paper



Final decision, balances areas

Selects the best papers from his/her area, based on meta-reviews

Inspects 3 reviews, leads discussion, writes a motivated recommendation (“meta-review”) *a.k.a. meta-reviewer*

Writes a review

Writes a review *a.k.a. reviewers*

Writes a review

What is Your Job as a Reviewer?

- 1: Help SPC / chairs guarantee the quality of accepted papers
  - They should be interesting
  - They should be accessible
  - They should satisfy scientific standards
- 2: Help authors improve this paper (and/or write better papers in the future)
  - Help them understand why the paper did not get in
  - Provide hints on how to improve it
- The way you review should reflect these goals

Not Your Job as a Reviewer

- Evaluate the intelligence of the authors
- Change the paper into how *you* would have written it
- Evaluate how interesting the paper is to *you* (rather than to some audience)
- Try to make the authors solve a different problem
- Make authors feel bad about having submitted this paper
- Try to increase citations to your own work

A Good Review…

- Starts with a brief summary of what the paper is about (so the SPC and authors know *your view of what the paper is really about*)
  - They see what you consider the important message of the paper
  - It provides context for the remainder of the review
- Evaluates the paper on all important criteria (correctness, readability, impact, …)
  - Mentions strengths and weaknesses of the paper
  - Mentions how *important* these strengths and weaknesses are
  - *Motivates* these claims (why is this weakness sufficient reason for rejection?)
- Uses arguments that are precise enough to be refutable
- Provides constructive feedback to the authors

Criterion: Correctness

- Read the paper critically: do you believe what is being claimed? If you find it hard to believe, check extra carefully.
- Are design choices motivated? Are arguments provided? Are they solid?
- Are the experiments sound and convincing? Do they take away all doubt?
- Pay attention to formal correctness of definitions, theorems, …
- Check proofs! If you can't: *inform the SPC you were not able to.*
- Distinguish:
  - Small errors that are easily fixed and don't cause confusion 
  - Errors that cause misunderstandings, making it more difficult to understand the rest of the paper, or that jeopardize conclusions 

Criterion: Accessibility

- Is this paper accessible (readable, understandable) to a wide audience?
- Clearly distinguish:
  - Difficult to understand due to inherent difficulty of the subject matter 
    - Some papers are simply too specialized to be readable for all. We don't want to exclude all those papers.
  - Difficult to understand due to avoidable causes 
    - Ambiguity in writing
    - Unnecessary complexity
- Papers should be as simple as possible, but not simpler

Criterion: Scope

- When reviewing, be aware of what the scope of the conference / journal / track is
- E.g.: paper purely on convex optimization: does it fit in this machine learning conference?
  - Depends on level of detail, clarity, background expected from reader
  - For how many conference attendants will this be interesting?
  - How many will be able to follow this?
- E.g., research track vs. application track
  - “The paper does not introduce methodological novelty, it just solves an application problem” - well, depending on the track, that may be just right !

Criterion: Interestingness

- Not every paper needs to present a new method that outperforms all existing ones
- The real contribution lies in what we learn from the paper, and how interesting that is
  - Obviously, “new algorithm A outperforms the state of the art” is more interesting than “new algorithm A does not outperform the state of the art”
  - But there are other ways in which results can be interesting
  - Perhaps A is very simple and performs surprisingly well - shedding light on how (un)impressive the state of the art really is
  - Perhaps there are arguments for believing A *should* outperform the state of the art - so if it doesn’t, where’s the flaw in those arguments ?
- A paper is useful if we **gain new knowledge** that we can build on in practical applications, or in future research

Criterion: Interestingness

Interestingness increases monotonically with

- **Surprise** : no-one expected this to be true
- **Convincingness**: the results were more or less expected, but are now established beyond doubt, or in greater detail
- **Insight**: the results are interpreted and explained in a way that yields new insights
- **Applicability**: the more widely applicable these results are, the better

Criterion: Reproducibility

- A paper must have enough detail for the work to be reproducible, in principle
- Note: reproducible ≠ replicable!
  - It is not enough to be able to *replicate* the experimental results, e.g., by running a script that comes with the paper
  - It must be possible to *independently* write your own implementation of the proposed method / algorithm, run it on similar data, and get similar results (where “similar” is defined as “leading to qualitatively identical conclusions”)

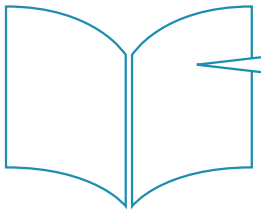
Criterion: Credibility

- What if :
  - a paper reports great results (“we beat all the competition”) with some very complex approach
  - you find it hard to believe, it seems too good to be true
  - yet, the results are there, and you don’t have concrete evidence that something is wrong
- What to do? “I have no choice but to accept these results, since I can’t prove them wrong”?
- No. It is the authors’ responsibility to *convince* you that the approach works, e.g., by
  - identifying the one thing in their approach that caused the big leap forward, and demonstrating that it indeed made the difference (e.g. present results with and without that key element)
  - providing enough details about the experimental procedure to rule out all imaginable methodological mistakes
- *Understanding why this works* is important. If a paper does not provide such an understanding, that may be a valid reason for rejection
- But: you cannot just say “I don’t believe the results”! You must explain why you don’t find them credible.

Not a Criterion: Complexity

- A paper is not an exam. Its purpose is not to “prove that the author is worthy of a PhD”.
- The paper does not have to look difficult. It does not have to be complicated. It does not have to present a solution that only clever people could think of. It only has to be a useful contribution (= informative, and of some interest)
- The basic rule: if reading the paper takes less time than finding and validating the solution yourself, it’s worth being published.

Example



"We propose a very simple method that is shown to outperform the existing state of the art as found in the literature."

(Literal quote from review)

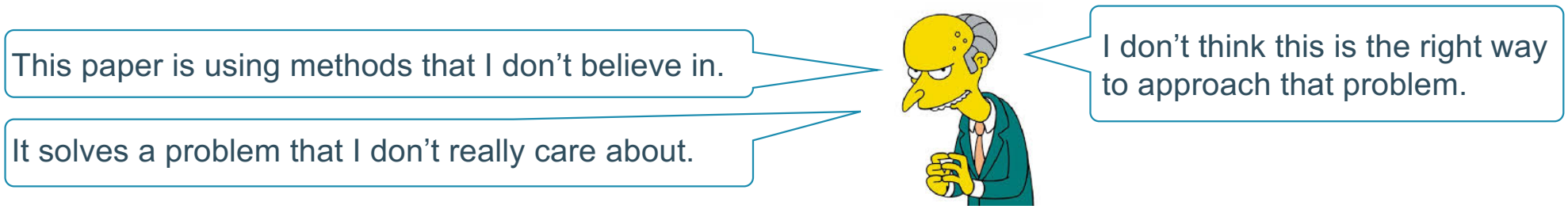
The proposed method does not have much novelty, and it may be already used in the field.



Not acceptable as a reason for rejection. Either you know for sure that this is already known / used, and you prove that (provide a reference), or you don't know for sure, and then you cannot use this argument.



Not a Criterion: Personal Opinion



- Try to be objective. You may believe these things, but only use objective arguments in your review.
- It is OK to state a personal opinion, as long as you make clear it is an *opinion*, and your final recommendation *does not rely* on that opinion
  - E.g., you can tell the SPC "I don't really believe these results, my personal experience with this problem tells me they are highly implausible. There's likely a bug somewhere, I just can't put my finger on it"
  - If another reviewer does find a bug, the SPC can put two and two together. Reviewing is a collaborative effort.

Not a Criterion: Amount of Work Spent on This

There are not enough experiments. You should test on at least 10 datasets.



A classic !

- Sure, more extensive experimentation *can* be necessary …
  - If the current set of results is simply not convincing (win on 2 datasets, lose on 1 : that does not prove you'll win more often than lose, on average)
  - If it is not clear how the current datasets were chosen (cherry-picking?)
  - If the usefulness of this method depends on whether it outperforms state-of-the-art method X, and X is not compared to
  - …
- … but then, just state the *real* argument, instead of stating "not enough experiments"

What About the Current Experiments?



- How have the **datasets** been chosen?
  - How many suitable datasets are readily available? How many have been used? Is it argued why exactly these have been used? Cherry-picking?
- How has **parameter tuning** been done?
- Are **competitors** SOTA? Are all competitors chosen? If not, how have they been selected? Are the conclusions aligned with this selection?
- Is there information leakage (may be subtle) from test set to training phase?

- All of this information is relevant for assessing how convincing the experimental results are

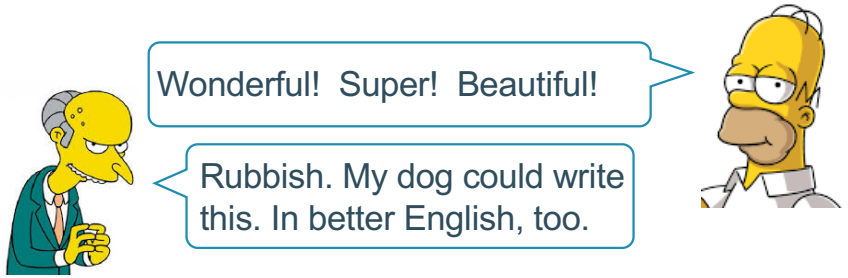
Ethical Aspects: Conflicts of Interest

- Whose paper can I review? When is there a conflict of interest?
- Use your own judgement. Whenever you would feel uncomfortable with a certain outcome for some paper, to the extent that it may affect your judgment, there is a COI.
  - Could be personal (your boyfriend wrote it)
  - but also: risk of being scooped, …
- **Perception also matters**! Even if you feel you can be objective, but others might perceive this as a COI, you may want to indicate a COI
- Standard sufficient conditions for COI: family relations, hierarchical relations, worked at same university / in same research group, recently collaborated, …
- COI is relative…
  - For an international conference, you may feel you know person X too well (e.g., because you've collaborated in a project on Y)
  - For a dedicated workshop on Y, you may know most of the authors as well as X, and declaring COI for all of them would effectively ban you from reviewing any paper

Ethical Aspects: Confidentiality

- Papers submitted for review are to be treated confidentially.
- Do not share them with others (unless for help with reviewing, and assuming they will respect confidentiality like you do)
- “What if reviewing a paper gave me ideas for follow-up work?”
  - You must wait until it’s published to start with that follow-up work (admittedly, it may be hard to postpone thinking about it)
  - It may be a good idea to contact the authors at that time, ask them if they’ve done any follow-up work, perhaps tell them about your idea and start a collaboration

Arguments are Essential



- If you claim that this is not new, it has been done before, add a *reference*.
- If you feel the authors ignore important related work, provide *references*.
- If you feel they should compare to X, say *why*.
- If you feel there are not enough experiments, say *why* more are needed. What would additional experiments contribute that's not in there already?
- If you feel the paper is needlessly complex, language is poor, …: show examples
- References and arguments are not only useful to the authors, they are **essential** because they make your claim verifiable / falsifiable.

E.g., "This was done before, by (ref)": Authors can look up the reference and accept the criticism, or point out in their rebuttal that there are important differences

Take a Stance

- Many reviewers tend to score papers near “borderline”
- Such recommendations are not very helpful to the SPC, even with arguments
- In your scores, make very clear what you think of each scored aspect
  - If you always score “writing” 4-6 on a scale of 1-10, you’re probably too cautious.
  - If you feel a paper contains mistakes that undermine its main results, make it a clear reject - don’t soften that result because the writing is excellent
  - The overall score need not be the average of subscores. (In fact, the minimum of those subscores may be more appropriate!)

Reviewing is a Collaborative Effort

Many conferences have a discussion phase. Take part in the discussion!

- Read author rebuttal, read other reviewers' reviews
- If you disagree with another reviewer, state that, and say why
- If another reviewer points out things that you missed, re-evaluate the paper with this in mind
- Don't be afraid to admit you missed some important problem, or to change your mind! (If reviewers were perfect, we'd need only 1 per paper)