

## Checklist for Revising a SIGKDD Data Mining Paper

© Eamonn Keogh

Dear Reader. I hope this checklist is useful to you. If you have any suggestions as how to improve it, please send them to me at [eamonn@cs.ucr.edu](mailto:eamonn@cs.ucr.edu) Best of luck with your research!

- Is the title clear and does it reflect the content and main findings?
- Does reading the abstract give the reader the main idea, and encourage them to read on?
- Is the abstract readable standalone (does it reflect the main story)?
- Do you have an explicit, falsifiable problem statement on the first page?
- Are key terms clear and familiar?
- Are the objectives clear and relevant to the audience?
- Are all variables, techniques and materials listed, explained and linked to existing knowledge.
- Are the results reproducible? Think *hard* about this. If I printed out your paper, handed it back to you, and then wiped your hard drive of all your code/data, could *you* reproduce your results?
- Are all results and comparisons relevant to the posed questions/objectives?
- Do the main conclusions reflect the posed questions?
- Is the text coherent, clear and focused on a specific problem/topic?
- Are proper tenses and voices used (active and passive)? (for the most part, active voice is better)
- Is the mathematical notation consistent and as simple as possible?
- Is the bibliography consistent and in the required format?
- Is the spelling of all proper names correct? (I do hate to see “Koegh” or “Keohe”)
- For every parameter in your method, do you should show, by logic, reason or experiment, that either there is a way to set a good value for the parameter, or the parameters value is relatively unimportant.
- Are all equations mathematically correct and explained in the text?
- AAAE? Are All Abbreviations Explained?
- Reconsider (avoid) using of subjective words “rather”, “very”, “better”, “may”, “appears”, “more”, “convinced”, “impression” in the text.
- Are all abbreviations, measurement units, variables and techniques internationally recognized?
- Are you using American spelling? (colour/color, gray/grey, programme/program etc.)
- Are you using American number formatting? Is  $\frac{1}{2} = 0.5$  or  $\frac{1}{2} = 0,5$ . Are you using American date formats?
- Are all figures/tables relevant and of good quality?
- Are all figures, tables, references and equations listed and mentioned in the text? Figures and tables should be mentioned *before* they occur.
- Are all references relevant, up to date, accessible and in the correct format?
- Have you referenced some papers from the journal/conference you are submitting to?
- If members of the editorial board/program committee of the venue you are submitting to have published in this area, have you referenced them?
- If you are using examples, are they culturally specific? Only Americans will get baseball references, only people from former English colonies will get cricket references etc.
- Do you have spurious precision? If Dr. Keogh is 75 inches tall, does that mean he is 190.500000194310 centimeters tall?
- Is there redundant text? Or to put it another way, is some text redundant? Tutorial text is redundant
- Do you have very long sentences that could be better written as two sentences?
- Are there words that could be removed without changing meaning of a sentence?
- Are there sentences that could be removed without changing meaning of a section?
- Are there sections that could be removed without changing meaning of the paper?
- If your work is funded by the NSF, DOD etc, have you acknowledged them?
- If someone gave you useful advice, code, data have you acknowledged them?

References to anything (to bibliography, to tables/figures) should be at the end of sentences

Text and figures should be your own. Reference the origin of all ideas, state clearly what ideas are your own.

- If the reviewing is double blind, are you sure you have not revealed your identity?
- Are you within stated page limits?
- Could a reviewer reasonably say “*I think/know/suspect that there is an easier way to get results this good*”
- Could a reviewer reasonably say “*nice paper, but I am not sure this is the venue for it*”.
- Could a reviewer reasonably say “*this is too similar to your last work*”.
- Did you acknowledge any limitations or weaknesses of your work?
- Could someone accuse you of stealing ideas/text/figures?
- Did you test print the paper (on-screen proof-reading is *not* good enough)
- Do you have many clichés? It is true that *good things come to those that wait, but on the other hand..*
- Do you needless words? Is it (absolutely) essential that these (actual) facts that help with (advance) planning of a (completely) unique event? Are **A** and **B** equal (to one another)?
- Are you assuming the reader will see a *color* version of your paper (“*Note that the red lines mean the algorithm..*”). You probably can only assume that for SIGGARPH and a handful of other conferences. Make sure your paper looks good in B/W.

• Have you confused...

Accept, Except, and Expect	Device and Devise	Last and Latter	Raise and Rise
Adverse and Averse	Discreet and Discrete	Later and Latter	Rational and Rationale
Advice and Advise	Disinterested and Uninterested	Lay and Lie	Recourse and Resource
Affect and Effect	Envelop and Envelope	Lead and Led	Respectively and Respectfully
All Together and Altogether	Explicit and Implicit	Leave and Let	Should and Would
Allude and Elude	Few (Fewer) and Little (Less)	Literally and Figuratively	Sometime, Some time, and Sometimes
Allusion and Illusion	Foreword and Forward	Many and Much	Stationary and Stationery
Allusive and Elusive	Formally and Formerly	Maybe and May Be	Their, There, and They're
A Lot (Much, Many)	Fortunate and Fortuitous	Media and Medium	To and Too
Amount and Number	Hanged and Hung	Farther and Further	Were, We're, and Where
Assure, Ensure, and Insure	Historic and Historical	Passed and Past	Which and Who
Choose, Chose, and Chosen	Imply and Infer	Precede and Proceed	Who and Whom
Cite and Site	Ingenious and Ingenuous	Prescribe and Proscribe	Whose and Who's
Complement and Compliment	Intense and Intent	Principal and Principle	Your and You're
Continual and Continuous	Its and It's	Quiet, and Quite	

### Basic Structure of a Paper<sup>1</sup>

- 1. Title
- 2. Introduce the topic and define (informally at this stage) terminology
- 3. Motivation: Emphasize why is the topic important
- 4. Relate to current knowledge: what's been done
- 5. Indicate the gap: what need's to be done?
- 6. Formally pose research questions.
- 7. Explain any necessary background material.
- 8. Introduce formal definitions.
- 9. Introduce your novel algorithm/representation/data structure etc.
- 10. Describe experimental set-up, explain what the experiments will show
- 11. Describe the datasets.
- 12. Summarize results with figures/tables.
- 13. Discuss results.
- 14. Explain conflicting results, unexpected findings and discrepancies with other research
- 15. State limitations of the study.
- 16. State importance of findings.
- 17. Announce directions for further research.
- 18. Acknowledgements.
- 19. References.

<sup>1</sup> Adapted from Hengl, T. and Gould, M., 2002. Rules of thumb for writing research articles.