

Advice to Machine Learning Authors

This editorial contains suggestions to authors of papers in the area of machine learning, although much of it applies to the broader field of artificial intelligence. I have distilled these comments from my five-year experience as an editor of *Machine Learning*, focusing on problems that tended to recur in different papers. Many comments are slanted toward papers that describe running systems, but others will be useful for different types of papers. Authors should focus on those suggestions relevant to their own research emphasis.

I have divided the suggestions into a number of categories, which should be self-explanatory. I expect most readers will agree with many of the points, but undoubtedly some will be more controversial. Despite this, I believe that listing them explicitly in this manner will at least encourage authors to think about the issues before drafting their papers, and thus reduce the need for revisions at later dates.

1. Content

- *State the goals of your research* and the criteria by which readers should evaluate your approach. Categorize the paper in terms of some familiar class; e.g., a theoretical analysis, a comparative experimental study, a description of some new learning algorithm, or a computational model of human learning.
- *Specify the performance and learning tasks* that are the focus of your research, clearly distinguishing between the two aspects. If there is no performance system, propose some other means of evaluating the learning behavior.
- *Describe the representation and organization* of your system's knowledge, along with the representation of training data. Include examples of each in the paper.

- *Explain both the performance and learning components* of your system in enough detail that readers can reimplement them. If possible, include both a pseudocode description and an extended example. Ideally, use some metaphor (like search through a hypothesis space, but others are possible) to describe its learning behavior, and state whether the system is incremental or nonincremental.
- *Evaluate your approach to learning*, avoiding unsubstantiated or rhetorical claims. It is fine to state that your approach is better than others, but make sure you give evidence or at least careful arguments to support your claims. For example, present experimental or theoretical evidence of performance improvement, show successful accounts of psychological phenomena, or give evidence of generality.
- *Relate your approach to other methods*, discussing similarities, differences, and advances over previous work. It is not enough to simply list references to relevant work. Place your method in historical context and clearly specify your intellectual debts.
- *State the limitations of your approach* and suggest directions for future research. Go beyond a list of problems to propose tentative solutions.

2. Titles and Summaries

- *Use an informative title*, but not one that is overly long. If possible, describe both the basic approach and the application domain, as in “An Incremental Hill-Climbing Approach to Chemical Discovery” or “Acquisition of Planning Expertise by Explanation-Based Learning”. Generally this can be accomplished with five or six content words and a few articles and pronouns. If you must say more, add a brief subtitle.
- *Keep your abstract brief*. If this is more than one paragraph (six or seven sentences), it is no longer an abstract. Also, do not repeat text from the abstract in your introduction. They should serve different purposes, with the former summarizing the text and the latter introducing the reader to the research.
- *Use informative section headings*. A lazy reader should be able to browse through your section headings and still have some idea what your paper is about. Headings like “Representation” or “Evaluation” are too generic. The ideal heading will contain three to five

content words, and will mention the name of your system or describe the general method, as in “HAL’s Organization of Memory” or “Experimental Studies of Induction over Explanations”. Never use pronouns (e.g., “it”) in headings.

- *Include introductory sentences* at the beginnings of sections and subsections to help readers make the transition. Also, never refer to a section heading as if it were part of the text.

3. Partitioning the Text

- *Use sections that are reasonable in length* and roughly the same length, except possibly for the introduction and conclusion. The same holds for subsections. Be consistent about whether you include an introductory paragraph before the first subsection.
- *Never include only one subsection* in a section. Subsections are intended to divide a section into components, and a single component serves no purpose. For the same reasons, avoid subsections with only one paragraph.
- *Use paragraphs that are reasonable in length*, not more than about six sentences or fewer than three sentences. Occasionally, it is all right to use a one- or two-sentence paragraph for emphasis, but this should be the exception. (Do not use this editorial as a role model for technical articles.)
- *Keep sentences to a reasonable length*, neither too short nor too long. Avoid run-ons and fragments. Do not try to say too much in a single sentence, but do not make them too choppy. (This may be obvious to many authors, but I have seen it all too often.)

4. Issues of Flow

- *Do not describe things before their time.* For example, do not talk about performance or learning while discussing representational issues. In some cases, you may need to motivate some aspect by forward reference, but keep this to a minimum.
- *Avoid overuse of parenthetical expressions.* Anything longer than a few words will upset the sentence’s flow and deserves to be a footnote instead. However, use complete sentences in footnotes, and keep them to a reasonable length (one to three sentences).

- *Do not overuse itemizations;* often a paragraph with the same material will communicate as well. Items can highlight important points or steps, but make sure they improve clarity rather than reduce it. Close off each list with some concluding sentence(s).
- *Itemize at the right level,* giving neither too much or too little detail. In general, items should be shorter than paragraphs but more than a few words. If you have more than four or five elements in a list, move to a higher level of abstraction.

5. Describing Your System

- *Give your system a name.* This will give your text more variety, and it will let other authors refer to something concrete in their review of your work. However, make sure you select a name that has never been used before, at least in your specific research area.
- *Do not overuse the name* of your system. Alternate between using the system name and an equivalent term, like “the system” or “it”. Especially avoid ending one sentence and starting the following with the system’s name. If the system name occurs more than three or four times in one paragraph, you are probably overdoing it.
- *Describe your system using paraphrases or pseudocode.* Such natural-language descriptions can significantly aid the reader, even if low-level routines remain unexpanded. If you must include a program listing, place it in an appendix.
- *Avoid using language-specific terms* and formalisms, as many readers will not know your implementation language. Reformat representations that involve list structure to make them more readable. For LISP, avoid special symbols like DEFUN, T, and NIL; for PROLOG and production systems, paraphrase specific rules in English.
- *Use mnemonic names* when referring to subroutines or parameters in your system. Avoid using internal system names; this makes your description read like a core dump.
- *Be wary of detailed program traces.* These can be helpful, but they are not a replacement for a careful system description. If you do include one, make sure you paraphrase it in English and include running commentary. Consider placing it in an appendix where it will not hurt the flow of the paper.

6. Terminology and Notation

- *Avoid abbreviations*, especially if you use them only a few times. Even if you use them more often, do not use more than a few abbreviations in a given paper. If you want to save keystrokes, define a macro or use global substitution. Never include an abbreviation in a title or heading.
- *Avoid needless jargon*. When possible, use terminology shared by other researchers in the field, rather than inventing your own. If you must coin new terms, explain their relation to existing concepts.
- *Avoid obscure terms* in favor of familiar ones. Borrowing unfamiliar terms from another field, even if legitimate in that field, can make a paper very difficult for readers unacquainted with that area.
- *Avoid terms that lend themselves to confusion*, especially when other words would serve equally well. Think carefully about the specialized terms that you employ.
- *Avoid unnecessary formalism* that does not occur in proofs or otherwise aid communication. If you do use formal notation, clarify its meaning in the text. Even readers with mathematical sophistication will appreciate the effort.

7. Figures and Tables

- *Include figures and tables* to help readers visualize your system and your results. Err on the side of too many rather than too few; they are easier to remove than to add. Anything that can be typeset should be labeled as a table; other material should be labeled as figures. Place figure captions below figures and place table headings above tables.
- *Refer to figures or tables in the text* and discuss them, at least briefly. Do not rely on figures to the exclusion of the text.
- *Include enough detail* in each figure caption or table heading to give readers some idea of the contents without reading the text. For instance, “Improvement in classification accuracy for three induction methods on the congressional voting domain” is better than “Learning curves on the voting domain” or “Behavior of three induction methods”. Avoid abbreviations here.

- *Label all distinct components* of figures and tables. Give a name for each axis of a graph, include a legend that briefly describes each curve, and specify the contents of each row and column in a table.
- *Do not refer to the location* of tables and figures, as in “the table below”. Their exact position may change during typesetting.

8. Issues of Style

- *Prefer active sentences* to passive ones whenever possible. This is much easier to achieve if you write in the first person (‘I’ or ‘we’) and if you use the system name for a sentence’s subject. For instance, “ID5 constructs decision trees incrementally . . .” is better than “Decision trees are constructed incrementally . . .”.
- *Avoid the use of contractions* in technical papers. Expressions like “I’ll” or “can’t” sound overly chatty.
- *Use ‘which’ and ‘that’ appropriately*. In general, use ‘which’ in relative clauses that could be removed without loss of meaning, as in “an incremental inductive algorithm, which is similar to HAL.” In contrast, use ‘that’ in relative clauses that are required as part of the sentence, as in “an inductive system that acquires concepts incrementally”. A good heuristic is that clauses beginning with ‘which’ are nearly always surrounded by commas.
- *Avoid long chains of adjectives*, such as “incremental instance-based learning algorithms”. Break such chains into manageable chunks, as in “incremental algorithms for instance-based learning”.
- *Do not interrupt the flow* of a sentence. For instance, the construction “For example, a method that uses . . .” reads better than “A method that, for example, uses . . .”.

In summary, remember that communication is central to the scientific process, and that few readers – including reviewers and editors – are willing to wade through difficult text. Spending an extra hour making your paper clear and easy to read can save tens of hours across the entire research community, as well as increase the paper’s chances of publication.