Writing a paper

Volodya Vovk (with input from John Shawe-Taylor and using Arieh Iserles's material)

Centre for Machine Learning
Department of Computer Science
Royal Holloway, University of London

RHUL, Egham, Surrey
20 February, 2019

My plan

- Scientists must write
- How to write a paper (content and form)
  - finding the information
  - overall structure and specific parts of the paper
- How to publish a paper

A lot of lists; I know they are incomplete: please add.

Scientists must write

> *A naturalist's life would be a happy one if he had only to observe and never to write. (Charles Darwin)*

Scientists are different from other professions. Plumbers do not need to write about pipes. A scientific paper is the culmination of a scientific research.

Plan

How to write a paper

Standard scheme:

- think
- plan
- write
- revise (90%?)

Key Ingredient (1)

Unless you are trying to bluff your way in science, a paper needs a key ingredient.

Key Ingredient	Risk	Payoff
Experimental testing or application of a previously proposed technique	1	1
Review section of the literature	2	2
Generalization of previous result for another specialist area	2	2
New perspective on a result or results in one or more areas	3	4

Key Ingredient (2)

Key Ingredient	Risk	Payoff
Sideways translation of previous result(s) into a new model or application	3	4
New algorithm for solving a known problem, perhaps more efficiently, or a novel heuristic for solving a difficult problem	4	4
Resolution of an open problem in a particular area	5	5
Novel theory to analyse some theoretical or real-world problem	5	6

How to arrive at the key ingredient?

- Start with an easy problem to get something under your belt.
- This will also get you familiar with the particular subject matter (the most important aspect of research is building up an intuitive picture of the area).
- Don't underestimate what you have: most new perspectives/results seem trivial with hindsight (this fact may be obscured by, e.g., complicated notation).
- There is a natural progression from basic standard experimental research to novel ideas. Don't shy away from the mundane, but do some "venture capital investment" of time in trying to develop new ideas and perspectives.

Finding information (1)

Warning: you should not have too much respect for the printed word (especially if you are aiming at a high risk/payoff key ingredient).

> *It is that which we do know which is the greatest hindrance to our learning, not that which we do not know. (Claude Bernard)*

> *The advantage of a certain amount of ignorance is that it keeps you from knowing why what you have just observed could not have happened. (Frederick Hopkins)*

Finding information (2)

- Reading before you think may prevent you from finding entirely new approaches to the problem.
- But the final paper must cite and compare with relevant literature.

Sources of Information

- Internet
- Friends or supervisor
- Mailing lists
- Library
  - encyclopedias: general and specialized
  - special books and journals (`http://www.rhul.ac.uk/library/`, esp. `http://www.jstor.org/`)
  - Science Citation Index, `http://wok.mimas.ac.uk/` (follow up a paper, yours or somebody else's)
  - abstracting journals

Structure of the paper (1)

Most of scientific papers: IMRAD (see Day's book).

- Introduction
- Methods (and Materials)
- Results
- Discussion

Structure of the paper (2)

My usual scheme (suitable for theoretical papers): IRPAD.

- Introduction (why I have written this paper)
- Results and their interpretation
- Proofs
- Discussion

Other schemes?

- 2019: it might make sense to have literature review as a separate section (it is usually part of Introduction)

Specific parts of a paper (1)

Title: the most widely read part of the paper. The fewest possible words that adequately describe the contents of the paper.

Authors: problems with who to include and the order of the names.

Table of contents: very useful but not always allowed.

Abstract: the map of your paper. See “title”; why you have written this paper.

Specific parts of a paper (2)

Introduction: your purpose in writing this paper; background (such as existing literature); your principal results and conclusions. *Remember the referees will probably be drawn from the references you cite.*

"Discussion": you can use it for discussion of weaknesses of your approach; interesting directions of further research.

Bibliography: no problems if using BibTeX.

Paper is not a story

About the early disclosure of your results (unlike O. Henry's endings):

> *Reading a scientific article isn't the same as reading a detective story. We want to know from the start that the butler did it.*

Be clear in your own mind what is the take home message or slogan of your paper. Put it in the title and abstract and discuss/motivate in the introduction.

You won't be read unless you are accessible.

Is the form important?

- If your paper is sloppy in form, 40% of your readers will not believe what you say! (Maybe subconsciously.)
- If the paper contains a lot of maths, your readers will not be able to check everything and will need to rely on clues.
- If somebody writes

  “pp. 2-7” instead of “pp. 2–7”

  you might think: he did not care to get his dots and dashes right; maybe the paper was written in a hurry and the argument is also flawed?
- Minor fault(s) $\rightarrow$ major fault(s) suspected.

Clarity

Important factors of clarity:

- plain English (see Gowers)
- “readability”, as measured by different readability formulas
- grammar (active vs. passive voice etc.)
- proper use of LaTeX

Readability formulas

Readability formulas used by Microsoft Office (Outlook, Word)

After Word (e.g.) completes a grammar check, readability statistics are displayed, for example:

Flesch Reading Ease: This index computes readability based on the average number of syllables per word and the average number of words per sentence. Scores range from 0 (zero) to 100. The average writing score is approximately 60 to 70. The higher the score, the greater the number of people who can readily understand the document.

Limited usefulness: it is more important how the sentences fit together (does the next sentence require a double take?).

Ten commandments (1)

In Day's book:

1. Each pronoun should agree with their antecedent.
2. Just between you and I, case is important.
3. A preposition is a poor word to end a sentence with.
4. Verbs has to agree with their subject.
5. Don't use no double negatives.

Ten commandments (2)

6. Remember to never split an infinitive.
7. When dangling, don't use participles.
8. Join clauses good, like a conjunction should.
9. Don't write a run-on sentence it is difficult when you got to punctuate it so it makes sense when the reader reads what you wrote.
10. About sentence fragments.

Don’ts (1)

Things to avoid:

- starting a new paragraph after every displayed equation;
- forgetting to use “~” or “\ ” after a dot that does not end a sentence;
- writing variable x as x;

Don’ts (2)

- forgetting to spell check the final version.

  Thou shalt commit adultery. (Bible, published in England in 1631, the 7th commandment)

  Know ye that the unrighteous shall inherit the Kingdom of God. (Bible, 1653)

Dos

Enhancing positive impression:

- paying attention to fonts;
- using different kinds of dashes for different purposes (-, –, —, −).

See almost any book on LaTeX.

Plan

Publication process

What to do after your paper is written?
Let's discuss:

- Choosing the right journal.
- Choose the right conference.

(In computer science, people often do both for the same paper.)

How to choose a journal

- Opinion of your friends, colleagues, supervisor.
- Check the publisher (is it familiar, respectable and trustworthy?).
- Check the Editorial Board (are the names familiar to you and do they include world-class people in your area?).
- “Objective” (easy to compute but measuring something else) measures, such as the impact factor or a similar metric (not comparable between different fields!).
- Unless you are persuaded that this is a good journal with good reputation, don’t submit and exercise great caution reading papers therein.
- All this is equally valid for electronic and traditional journals.

How to choose a conference

- It's more difficult.
- You can always ask more experienced colleagues (as for journals).
- Australian list:

  http://www.core.edu.au/conference-portal

  (They also rank journals.)

Machine learning: FOCS, STOC, IJCAI, UAI, ICML, COLT,. . . .
Again a risk/payoff problem: you need to make an unbiased judgement of your paper.

Acceptance/rejection (1)

Different reactions from the editor:

- acceptance (very rare)
- conditional acceptance (you should modify the paper)
- rejection

Acceptance/rejection (2)

Rejection letter from a Chinese economics journal:

> *We have read your manuscript with boundless delight. If we were to publish your paper, it would be impossible for us to publish any work of a lower standard. And as it is unthinkable that, in the next thousand years, we shall see its equal, we are, to our regret, compelled to return your divine composition, and to beg you a thousand times to overlook our short sight and timidity.*

How journals work (1)

Peer review system.

Typical process for a journal submission:

1. You send your manuscript to the editor.
2. The editor (who is typically not paid) send your manuscript to several (typically 2 or 3) referees (your peers). The referees are given deadlines, usually a few months.
3. Referees (who are often busy people) study your manuscript and send their reviews to the editor.

How journals work (2)

4. The editor makes his/her decision based on the referees' comments; typically "reject" or "modify". You are informed about the decision. *After 6 months it is sensible to remind the editor about your paper.* "Reject": you send the paper to a different journal or forget the whole thing. Suppose the decision is "modify".
5. You decide whether you want to go on with your submission. If yes, you make those of the changes suggested by the referees that you agree with and explain in the cover letter to the editor why you disagree with the other changes. Go to step 1. In step 2 the editor is likely to send the paper to the same referees (or to accept your paper immediately).

How journals work (3)

Sometimes you send a manuscript to the editor-in-chief, who then sends the manuscript to an associate editor. Papers are usually accepted by the associate editor and rejected by the editor-in-chief (the best possible approximation to anonymity).

Even if a referee's remark is unfair, it is likely to point out to a problem: you should try to improve your exposition; a typical reader will have even less time than the referee.

How a conference works (1)

Whether your paper is rejected or accepted is decided by a programme committee.

Different programme committees work differently.

How a conference works (2)

A possible scheme (role of the Programme Chair small):

1. All members of the committee read some papers (say, at least 7) and assign to every paper read 2 numbers:
   - grade (e.g., 0 horrible, 10 excellent)
   - confidence (e.g., 1 “not confident”, 3 “I am an expert in this field and this paper”)

   The numbers and comments of all members are circulated to all other members by e-mail.
2. A discussion (perhaps heated) follows. In view of the discussion, a new set of 2 numbers is assigned. The Programme Chair asks some members to review papers which are reviewed by too few people.

How a conference works (3)

3. The final discussion and the final numbers assigned.
4. Every submission gets a grade computed as the weighted average (the confidences serve as the weights) of the grades given to this submission.

The programme chair decides on the threshold; the papers above the threshold are accepted. Borderline papers: occasional exceptions.

Plan

Traditional model (1)

The traditional model of publishing:

- Journals are all printed in hard copy;
- Authors surrender their copyright to the publisher;
- If you wish to read a paper, either find a library subscribing to the journal, ask the author for a copy or photocopy from somebody else's hard copy.
- The costs of editing, copy-editing, typesetting, printing, marketing etc., plus (hopefully reasonable) profit of the publisher are covered by the (usually pretty steep) cost of subscription, which typically falls on university libraries.

Traditional model (2)

- This model served the scientific community well for close to 350 years.
- The oldest scientific journal, “Philosophical Transactions of the Royal Society”, was established in 1665.
- But it is now being increasingly challenged by alternative models of electronic publishing.

Model 1: Green Open Access

- The journal is printed and distributed in the usual way, except that...
- It is accompanied by an electronic copy on a dedicated website.
- Access to the website for the first n months is limited to subscribers (this usually means all staff and students of subscribing institutions).
- After n months (typically $n \in [6, 12]$) access to papers becomes unlimited; alternatively, after this cutoff time, authors have the right to post on their website and distribute the pdf file of their published (as distinct from the original draft) paper.

Model 2: Gold Open Access

- Fully electronic journal, freely available to everybody on the Internet;
- Authors are obliged to pay either page charges or article processing charges;
- Authors typically retain the copyright and are free to distribute the article in its final form, either as a hard copy or electronically.

Model 3: Mixed Open Access

- Typically both printed and published electronically;
- Subscribers pay, whether for hard copy or electronic access—so far, everything like in the traditional model;
- Authors may choose to publish in a traditional manner or...
- may opt to pay page charges—in that case the electronic copy is Open Access and the authors are free to distribute the pdf file with no restrictions.
- Royal Holloway, University of London, pays for open access (at least for academics)!

Model 4: Diamond Open Access

- Purely electronic journal;
- All articles are Open Access but. . .
- there are no page charges: everything is free.
- This might sound too good to be true.
- But in fact it is true for some of the best journals and quai-journals in machine learning (e.g., *Journal of Machine Learning Research* and *Proceedings of Machine Learning Research*).

Standard arguments for Open Access (1)

- The public funds academic research, hence members of the public deserve access to academic research: this freedom of information is crucial in an open science-driven society and economy.
  - Is it really applicable to maths and computer science?

Standard arguments for Open Access (2)

- Journal subscription is expensive, beyond reach of Third World universities and even of many universities in developed countries. Page charges are in reality cheaper than subscription costs and they shift the cost of publication to richer universities, from where most of publishable research originates.
  - This requires a viable and transparent mechanism for universities to cover page charge of their faculty and students. Sceptics are afraid that substantial savings from subscription costs, instead of being ploughed into the payment of page charges, will be diverted by university administrators to hire even more administrators.

Standard arguments for Open Access (3)

- Governments and funding agencies have already started to insist on Open Access of all publicly-funded research. Whether right or wrong, it is not practical to argue with them. This is already the case with UK Research Councils and with US National Institutes of Health.

Arguments against Open Access (1)

- Traditional journals have had many years to establish themselves. We, as a community, know which are the top journals, with strict editorial process, skyhigh acceptance criteria and top authors, and which journals are tree-wasters. What about electronic journals?
- An Open Access journal might have nothing to do with lofty academic goals and be just a commercial operation, publishing possibly useless or wrong papers in exchange for hard cash.
  - Traditional journals can also be of low or non-existing quality. . . .

Arguments against Open Access (2)

- Electronic journals might come and go: what will happen to my electronic paper once they are gone? Will an archive be maintained and will it remain accessible on the Internet, in an up-to-date format?
  - It is typically safe to trust journals associated with learned societies, major university presses, and few reputable publishers (such as Wiley or Springer) and even then read carefully their archival policies.

Off to Internet

- Even before you send the paper to a conference or journal, it is a good idea to share it on your (or your group's) website.
- You may also deposit it in one of subject-specific Internet archives, e.g., ArXiv.
- Preprints (author-formatted) are usually allowed. (Avoid publishers who do not allow it!)
- I keep maintaining arXiv reports even after the journal version is published.
- arXiv is owned by Cornell University, probably no need to worry about their archival policies.

Books (1)

Robert A. Day,
How to write and publish a scientific paper, 4th ed.,
Cambridge: Cambridge University Press, 1995.
Available free on the Internet.

Newest incarnation:

Robert A. Day and Barbara Gastel,
How to write and publish a scientific paper, 8th ed.,
Cambridge University Press, 2016.
7th ed. (2012) available free on the Internet.

Books (2)

- George M. Hall (editor),
  How to write a paper, 5th ed.,
  Wiley, 2012.
  Earlier editions available free on the Internet.
- Ernest Gowers and Rebecca Gowers,
  Plain words, newest edition,
  Penguin, 2015.
- Robert Barrass,
  Scientists must write, 2nd ed.,
  London: Routledge, 2002.

Books (3)

 Books on LaTeX.

Great book of style:

 The Chicago Manual of Style, 15th ed.,
Chicago and London: University of Chicago Press, 2003.
First edition: 1906; 14th: 1993 (808.02 UNI in Bedford library); 15th: 2003; 16th: 2010.

A great article

When writing a theory paper: see

Arieh Iserles,
How to write a paper,
Arieh's web site (DAMTP, Cambridge), June 2014.

+ the references therein