STAT 539 Writing research papers

Marina Meilă

Department of Statistics
University of Washington

April 4, 2013
Paper organization at a higher level

- **Abstract**: “should I read this paper?” (Is this paper relevant to me?)
- **Introduction**: informal problem description, places the problem in context, related work
- **Problem formulation**
- **Methodology/Proposed solution**
- **Discussion** (placing the work in perspective)
  - what is achieved and what is not
  - related work again
  - what else could be done with this method
- **Experimental evaluation**
  - demonstrate empirically that method works
  - verify provable properties
  - learn about unknown/conjectured properties
  - compare with other methods
  - demonstrate that it can solve real problems (better than competition)
  - discuss the results!
- **Conclusion**
  - a summary of achievements
  - more discussion
  - a final advertisement of your method
What is reproducible research?

- The purpose of research is not just to solve a problem for someone, but to also teach a larger group of people how to solve similar problem. Otherwise it’s called *development*.
- The research (results) you publish must be reproducible.
- A research paper must contain enough information to allow a reader to reproduce/verify its results.
- Verify that your assigned papers are reproducible, and how they achieve this.
The research (results) you publish must be reproducible!

- Theorems/methods have proofs

- Algorithms
  - describe in enough detail that they can be implemented by an informed reader
  - balance between what described and what referenced only

- Data
  - make data available (whenever possible)
  - use benchmark data sets (wide and indirect comparisons possible)
  - give sources for data sets not your own

- Parameters of algorithms
  - describe methods to set the parameters
  - for each experiment: give all parameters settings, for all algorithms used
Figures and experimental results

- Show only relevant results
- General principles for displaying information (figures vs. tables, etc), see Tufte
- Make clear annotations, as close to data as possible. Spare comments (e.g. “lower is better”)

**Rule** Figure/Table caption explains **what** is in the figure/table; comments are in the main text

- Discussing the experimental results
  - don’t repeat what’s in the tables/figures (but refer to them)
  - appeal to reader’s intuition
  - but be precise
Reproducible research one step further

(From Buckheit and Donoho)

we . . . distill Claerbout’s insight into a slogan

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.
Do reproducible research and increase your impact too

- Make code available (and easily usable by others)
- Make a “project webpage”
  - publications
  - code, data, instructions, figures
  - other resources that you find useful
Planning your project

- what the final writeup should contain
- what you will implement
- what extensions you will make to the assigned paper
- the schedule

This is a draft. Will be revised in the next 2-3 weeks.
Plan – first draft

- the paper summary
- what parts of the paper you will implement
- what programming language will you use for implementation. if you think there is a reason to use code not written by you, which package and why (it is understood that you will use graphics/plotting software, and optimization software like LP, QP solvers, LBFGS as needed)
- are there any theoretical/methodological parts that you want to extend/develop?
- are there any problems with the paper that you will try to fix?
- data sets you will use (from the paper or from elsewhere) (optional, for now) a plan of the experimental evaluation
- extensions
- a schedule of the implementation stages by week (including week 2)
- a list of other related papers (background, followup research, competing methods, applications) that are relevant to your project