READING RESEARCH PAPERS

CS Graduate Skills Seminar Fall 2017

Alex Gittens

The importance of reading

- Reading is essential for your success as a researcher
  - Identifies research directions and trends
  - Clarifies your research community
  - Prevents duplicated and wasted efforts
- Must identify *WHAT* papers to read
- Should know *WHY* to read them
- And *HOW* to read them effectively
- Feel no shame in asking questions

CLASSES of papers

- Technical Papers
  - Novel research contributions
  - Define and investigate problem, documents conclusions
  - Results can be theory, or systems/engineer-oriented, or both
  - Main type of paper you will be reading and producing in grad school
- Survey Papers
  - Summarize results and directions in a particular research field
  - Attempt to impose high-level structure on research literature
  - Can be useful when starting in a new research direction
- Vision Papers
  - Advocate for new directions on old problems or looking at new problems

TYPES of papers

- Seminal
  - Foundational results, opened new research direction
  - Can be relatively old (70s, 80s)
- Influential
  - Highly cited, novel perspective and directions in field
  - Could be seminal or not
  - Considered required reading in your field
- Expository
  - Papers your advisor recommends reading to understand your field
  - Excellent problem descriptions and motivation
  - Detailed related works sections
  - May or may not be influential
- Relevant
  - Related to your specific research

VENUES for papers

- Conference Papers
  - A primary publication method in many subfields of CS
  - Fast moving: less than a year between submission and publication
  - Relate to the topic of the conference, published in proceedings
  - Peer-reviewed and voted on for acceptance
  - Usual short, around 8 pages
  - Accompanied by talks and/or posters
- Workshop Papers
  - Preliminary versions of conference papers
  - Usually short
  - Also peer-reviewed, lower bar for acceptance
  - May or may not be published in workshop proceedings

VENUES for papers

- Journal Papers
  - Relate to the topic of the journal
  - Usually longer than conference papers (exception: Letters type journals)
  - Stronger peer-review process, multiple rounds of submission
  - Slower-moving: usually 6+ months to a couple years from submission to publication (exception: Letters type journals)
  - In CS, often extended versions of conference papers
  - Surprisingly, sometimes less selective than conferences

VENUES for papers

- Preprints and Technical Reports
  - Not peer-reviewed; be extra careful in taking at face value
  - Serve as way to disseminate results quickly
  - Posted on author's webpages, preprint servers (e.g. arXiv)
  - Preprint: pre-publication version of a peer-reviewed paper
  - TRs: may or may not be in process of peer-review

HOW to read effectively

- Reading is like eating
  - We need to do it to get sustenance (knowledge)
  - There are some shared commonalities, but everyone finds their own way to chew and digest
- Know WHEN and WHY to read a paper
- ORGANIZE yourself
- Read SMART
- DOCUMENT your take-aways

HOW to read effectively

- In a nutshell:
  1. **Skim** for the main ideas and results
  2. **Re-read** to get the gist of the arguments/proofs and experiments
  3. **Re-read** critically, challenging the claims
  4. **Summarize** to ensure you understand the contributions and main ideas

WHY to read a given paper

- You find it interesting
- It was recommended to you (advisor or colleague)
- To learn new tools or methods relevant to your research
- As background or to cite for your research
- As background for reading another paper
- To prepare for a conference or meeting
- To review for a conference or journal
- Assigned reading in a course, or reading group

WHEN to read a given paper

- When writing your own paper
  - Look for related results and relevant tools
  - Give credit where it is due (related works)
  - Position your paper and explain its contribution
  - Survey the field, as a service to the reader (and bump your paper cites up)
- Knowledge Maintenance
  - Know what is going on in your field (preprints, workshops, conferences)
  - Find interesting problems to work on
- Pre-conference
  - Plan which talks you will attend, and read those papers
  - Read before the talk, before the poster session, before author leaves

HOW to find papers to read

- Ask your advisor
- Related works section of relevant papers
- Check forward citations of relevant papers (Google Scholar, …)
- Follow preprint servers (arXiv, …)
- Follow journals' RSS feeds of recent articles
- Check conference proceedings
- Check researchers' webpages for preprints
- It helps to know your research community (conferences, individual researchers, relevant journals)– ask your advisor!

PLANNING to read a paper

- Time management is major:
  - Length of paper:
    - journal papers and TRs (a day or two)
    - conference papers (several hours)
  - Purpose in reading affects time spent:
    - Knowing what's in a paper (skim)
    - Understanding the main ideas (read notation and main results, experiments)
    - Understanding the details (read everything closely)
    - Checking the details (do the calculus)

CAVEATS in reading: peer review

- Conference reviewing is problematic
  - Single-blind vs double-blind leading to biases of all kinds
  - Very stochastic!
  - Quick turn arounds on reviewing
  - Sometimes poor selection of reviewers
  - Reviewers fatigued by multiple papers at once
- Journal reviewing also has issues
  - Scrutiny inverse with paper length
- Preprints are not peer-reviewed at all

CAVEATS in reading: credulity

- Be aware of your own biases
  - Belief that publication means correctness
  - Belief that authors know how to position their work
  - Belief that authors mention all related works
  - Trust that experiments are meaningful (choice of metrics, datasets, etc.)
  - Trust that theory is meaningful

- Challenge your assumptions and biases. Do not depend on peer review. *Research papers are not textbooks*.

CAVEATS in reading: writing quality

- Sometimes difficulty in understanding is not solely due to you
- Good researchers are not necessarily good writers
- More effort is spend on polishing some papers than others

- To ameliorate:
  - Ask your advisor
  - Identify quality expository papers in your field, start with them
  - Familiarize yourself with notation and conventions of your field, folkloric results
  - Contact the corresponding author

BEFORE reading a paper

- Check the publication details
  - Publication venue, date
  - Is there a journal version of this conference/workshop/TR?
  - Is this the authoritative version?
  - Who are the authors? Identify professors, post-docs, students, affiliations
  - Who is the corresponding author for questions?
  - Check citation count: how influential is this work?
- Read the abstract
  - What do the authors think is their contribution?
  - Does this still seem worth reading?

BEFORE reading a paper

- Skim:
  - Get very broad outline of paper contents
  - Understand how it relates to your research interests

- Decide whether to continue reading
  - Is this relevant to you?
  - Is the quality of the paper up to par?
  - Do you need to read other background material first?

Multi-pass Reading

- Recall purpose in reading:
  - Knowing what's in a paper
  - Understanding the main ideas
  - Understanding the details
  - Checking the details
- First, skim:
  - Understand positioning
  - Understand main results
  - Understand meaning of experiments
  - Formulate your take-aways from the paper

Multi-pass Reading

- Second pass, go deeper:
  - Identify the main tools and ideas used. Which are new?
  - Any flaws or omissions in methodology or theory?
  - Look for simple implications of complex or difficult to parse claims: are they reasonable? (e.g. non-obvious exponential dependencies)
- Third pass, challenge:
  - Are the technical details correct?
  - Can the results be obtained more simply?
  - If code is available, were the experiments done as described?

Anatomy of a Research Paper

- Introduction
- Related Works
- Notation
- Main Results and/or Algorithms
- Experimental Results
- Conclusions
- Bibliography
- Appendices : Theory, Supplemental Experiments

Reading the Introduction and Related Works

- Purpose of the Introduction:
  - Describe the problem being addressed
  - Motivate interest in this problem
  - Position the paper's results in the broader area of research
  - Explain the importance of the results
- Purpose of Related Works:
  - Give fair comparison to similar work
  - Provide reader with context to judge results
- You judge:
  - How interesting/important is this problem?
  - How novel and reasonable are the paper's results?
  - What related works would I benefit from reading?
  - Keep in mind the authors' claims as you read the rest of the paper

Reading the Main Results

- Describes solution to the problems raised in the introduction
  - Algorithms
  - Software, Hardware
  - Novel theoretical understanding
- You judge:
  - In what sense is the problem solved? Partially? Completely?
  - Is the solution fully and unambiguously described?
  - How efficient is the solution?
  - How meaningful are these results?
  - Do the results match the claims made in the introduction and abstract?

Reading the Main Results

- Decide *ahead of time* on your criteria for measuring quality of the solution
  - What would a reasonable solution look like or guarantee?
  - Scalability?
  - Robustness?
- If theory, look at simplified models (e.g. restate tensor results as matrix results)

Reading the Experimental Results

- Provides experimental validation of results
  - Describes experimental design and setup
- You judge:
  - Are the relevant questions answered? (depends partly on authors' claims)
  - Are the baselines appropriate (extant?), and strong enough?
  - Are the metrics meaningful and sufficient for the problem?
  - Are datasets reasonably challenging, representative, and illustrative? Are the results statistically meaningful?
  - Are these experiments in keeping with standard practice?
  - Do the results support the claims made in the introduction and abstract?

Reading the Experimental Results

- Decide *ahead of time* on your criteria for judging quality of the experiments
  - Avoids bias towards agreeing with authors' choices
- Example considerations (depends on your area):
  - Hyperparameter selection
  - Hidden costs (model selection)
  - Bias in data selection
  - Red flag: no drawbacks at all, the new method is always the best
  - Accuracy-vs-time tradeoff

Reading the Appendices

- Expands on body of the paper with further details
  - Usually in theory papers, contains the technical proof details
  - In empirical papers, contains further experimental validation
- You judge:
  - Given why I'm reading this paper, is this relevant material?
  - How does this reflect or supplement the main claims of the paper?
- For theory papers (read the appendices!):
  - Understand how the parts hang together first before reading in detail
  - Identify the crucial insightful results vs the mechanical lemmata
  - Work through the proofs yourself, try to reorganize and simplify

Reading the Conclusion

- Summarizes the main points of the paper
- Looks forward to future directions
- You judge:
  - Do the conclusions match/repeat the claims made earlier?
  - Do you think the research directions are worth pursuing?
  - Are there more valuable contributions that you feel should be listed?

AFTER reading

- Are you confident in your understanding?
  - Read it again
  - See related works for different perspectives
  - Talk to your advisor
- Correctness of the paper
  - Talk to your advisor
  - Contact the corresponding author
- Big picture
  - What was the value of this paper for your research?
  - What new tools or approaches did you learn?
  - Where will you go from here: more reading? Work on this problem?

ORGANIZING reading

- Keep track of papers you want to read, *and why*
  - Bibliography tools: BibTeX, …
  - Paper managers: Mendeley, Papers, …
  - Annotated pdf readers
- Keep track of papers you read, and your thoughts
  - Judgements and questions
  - Ideas for related research directions
  - Related works to follow up on

ORGANIZING reading

- Time management:
  - Schedule regular times for reading
  - Don’t linger on tough spots, skip and revisit

- Reading order:
  - First, read easily digestible papers
  - Next, influential papers
  - Next, relevant papers
  - Skim: seminal papers. Especially if they are old

Some closing thoughts

- Read as if the authors are friends asking for honest feedback
- Active reading w/ pen and paper (or tablet and stylus)
- Challenge everything: do I know how to do this better, make this more concise, transparent?
- Post-mortem: can I teach the ideas in this paper to someone else w/o jumping into equations?

Even more closing thoughts

- Try to position paper techniques and results in your personal knowledge graph
  - Tie in with what you already know
  - Make it more approachable and see as something you could produce
  - *Own the knowledge in what you read*

Questions?