

Research Seminar Data Science

Teachers: Elena Marchiori (coordinator) & Arjen de Vries

Objectives

- Learn how to evaluate research papers
- Learn what makes papers good
- Learn about how papers are refereed and published
- Obtain an overview of important recent developments in Data Science research
- Note: course inspired by the same course for Software Science at Imperial College, and some material then also heavily borrowed from that course <http://www.doc.ic.ac.uk/~dirk/Teaching/RS/>

Tasks

1. Interpret and analyze a seminal, "classic" paper
 - present your own selected paper
 - short summaries of those of others
2. Review a paper submitted to a top data science conference or journal
3. Evaluate a scientific presentation

1. Seminal paper

- Pick a seminal paper that had major impact in machine learning or information retrieval.
- Present the paper in class:
 - 15 minutes presentation to **initiate discussion** (i.e., no need to give a full-fledged summary),
 - 15 minutes discussion.
- Write a review on the paper presented (about 5-6 pages).
- Others: write short summaries (half a page) *before* the presentation, send queries to the speakers 24 hour before their presentations.

Presenting a paper

- Main goal: informative and lively talk that promotes discussion.
- Suggested outline:
 - Objective of the paper
 - Proposal of the paper
 - Evidence given
 - Shoulders of giants
 - Impact
 - Discussion points
- Same outline as above for written review + issues from discussion.

Short review

- Half a page, maximum one page
- Clearly-separated (use subheadings) sections covering
 - Summary (as briefly as you can – two or three sentences)
 - Evidence (what evidence is offered to support the claims?)
 - Strengths (what positive basis is there for publishing/reading *it/listening to it*?)
 - Weaknesses
 - Evaluation (if you were running the conference/journal where it was published, would you recommend acceptance/*would you invite the presenter for a talk?*)
 - Comments on the quality of the writing/presentation
 - Plus: Queries for discussion
- Deadline: 1 day before presentation/1 week after presentation

(in *italic*: for review on a presentation)

How to find a seminal paper

- Google scholar...
- For a recent, e.g., < 3-year old paper: you'd like to see at least 40 citations per year.
- For a not-so-recent paper: you'd like to see at least 200 citations in total.
- In case of doubt: send an email or make an appointment to discuss.

2. Review a (potentially) great paper

- Paper submitted to one of the main conferences in data science, such as IJCAI, ICML, UAI, SIGIR conferences:
- You will get assigned a "real" paper to review and a coach to help you. See for instance
 - <https://www.ijcai-18.org/>
 - <http://auai.org/uai2018/dates.php>
 - <https://icml.cc/>
 - <http://sigir.org/conferences/>
- Your output: a single report with
 - Your own review
 - Reflection on the review procedure as a whole

Info on reviewing

- Example reviewing criteria:
 - <http://www.auai.org/uai2015/reviewCriteria.shtml>
 - <http://www.informaqus.nl/sigir2017/review/guidelines-pc-fp/reviewing.html>
- How to review a technical paper:
 - https://www.cl.cam.ac.uk/teaching/0910/C00/L2/how_to_review.pdf
- The NIPS consistency experiment:
 - <http://blog.mrtz.org/2014/12/15/the-nips-experiment.html>

Confidentiality

- Paper reviewing for many conferences is double blind.
- You should *not* try to identify the authors.
- Nor share the paper with someone else.
- Same holds for the reviewer discussions / author feedback.

3. Evaluate a talk at a seminar/symposium

- ICTOpen
 - <http://www.ictopen.nl/>
 - March 19-20, Amersfoort
- Thalia symposium
 - <https://symposium.thalia.nu/>
 - This year's subject will be "Offensive Security".
 - February 23
- Data science weekly seminar
 - <http://www.ru.nl/datascience/research/seminars/>
- Only if you could not find a presentation to attend, consider video lecture like
 - http://videlectures.net/aaai2017_sanfrancisco/?q=2017
 - Contact the teachers prior to this last choice!

Preliminary planning – review+presentation seminal paper

- Presentations: on average every two weeks, max 3 papers
- Deadline review of your own presentation: 2 weeks after presentation

Preliminary planning - review conference paper

- End February: papers assigned for review (some may come one month later)
- April: review deadline
- End April: other reviews available
- June 18: final report (review + reflection on other reviews)

Output and assessment

- Review conference paper + reflection paper under review: deadline June 18, 40%
- Presentation + review seminal paper: deadline 2 weeks after presentation, 40%
- Short summaries "other" seminal papers + similar report on one presentation at a symposium or seminar: deadline 1 day before paper presentation / 1 week after presentation, 20%
- **Plagiarism:** If you use material written by someone else, make sure you acknowledge the source. Making effective use of sources is encouraged (provided copyright is respected).

What to do next?

- Check out the websites for Information Retrieval, Machine Learning conference proceedings of recent years.
- Fill in the form (see link posted in BlackBoard) by **February 13** the latest. We will try to make a schedule as soon as possible.
- Conference papers to review will be posted during the course by our teachers. Select a paper by sending a request to the teacher who posted that paper. Each paper can be reviewed by at most one student, so be alert and quick to choose.
- Find a seminar/symposium to attend.
- For any question contact Arjen and Elena.