Topics for Seminars & Projects

Prof. Myra Spiliopoulou

myra@ovgu.de

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Schedule for seminars and projects (B/M)

- **REGISTRATION:**
  1. You apply for a topic till: Friday, Oct 19, 2018 – 13:00
  2. Topic assignments made until Monday, Oct 22, 2018
  3. You sign the form and submit to the KMD team by
       Thursday, Oct 25, 2018, 12:30 – 13:00 in person
- **MEETINGS for seminars**
  1. Monday, Nov 5, 9:15 – 12:00
  2. Monday, Nov 19, 9:15 – 12:00
  3. Monday, Dec 3, 9:15 – 12:00
  4. Monday, Dec 10, 11:00 – 12:30
  5. Monday, Jan 7, 9:15 – 12:00
  6. Backup: Jan 14, 9:15 – 12:00
  7. Monday, Jan 21 9:15 - 12:30 (final)

> Attendance is mandatory

- **REPORT for seminars:** Friday, Jan 25, 2018 – 13:00
- **PRESENTATIONS/REPORT for projects: Feb/Mar 2019**

For Master DKE students: Teamprojects for area "Methods I"

According to the statutes, it is not permissible to place a teamproject in the area "Fundamentals". This holds for all teamprojects.

Teamproject (Master degrees only)

Prerequisites for all projects of this type:
- A team of THREE students
- GOOD software engineering skills
- Background in data mining / machine learning

unless otherwise specified

STREAM-1: Toolbox of incremental classifiers

GOAL: Extend an existing software tool with a set of state of the art classification algorithms for opinionated texts

- SUBTASK 1: Select algorithms, make them incremental
- SUBTASK 2: Incorporate the algorithms into the tool
- SUBTASK 3: Implement a testbed to test the algorithms for different configurations (including different vectorizations and windowing methods) in a selection of text streams
- SUBTASK 4: Evaluate on classification quality, execution speed

PREREQUISITES:

1. Data Mining / Machine Learning
2. Software engineering in Python
3. Desirable: Stream Mining, Information Retrieval

STREAM-2: Toolbox of incremental regressors

GOAL: Extend an existing software tool with a set of state of the art regression algorithms
- SUBTASK 1: Select algorithms, make them incremental
- SUBTASK 2: Incorporate the algorithms into the tool
- SUBTASK 3: Implement a testbed to test the algorithms for different configurations (including different vectorizations and windowing methods) in a selection of text streams
- SUBTASK 4: Evaluate on prediction quality, execution speed

PREREQUISITES:
1. Data Mining / Machine Learning
2. Software engineering in Python
3. Desirable: Stream Mining, Information Retrieval

Seminar (Master level)

Prerequisites for all assignments of this type:

- Background in data mining / machine learning

Topics – Methods I unless otherwise specified

1. Scalable methods for data processing and counting on conventional streams – Fundamentals
2. Scalable methods for stream classification
3. Scalable methods for stream regression
4. Scalable methods for stream clustering
5. Scalable methods for live stream monitoring, esp. newsfeeds and textstreams
6. Visual analytics for live stream monitoring

It is possible that two students take the same topic.
It is not permitted that two students work on papers written by the same team of authors.

TASKS for each topic

1. Specify a keyword-based search scheme for literature collection
2. Collect a first paper list ; describe the exclusion/inclusion criteria for the papers
3. Select a shortlist of 4 papers *that do not build upon each other*; describe the procedure of shortlisting

4. Review the 4 papers (you will obtain a review template)
5. Rank them and justify the ranking, propose the top-1 paper among those 4 (self-nomination):
   All self-nominations enter a competition
6. Provide a justified nomination (with votes) for the competitors (you cannot vote for your paper)

Final report: compilation of the deliverables to all tasks

Timeline – Part 1: Collecting and choosing papers

Monday Oct 22: (Some of)You get a topic assignment and first instructions on how to collect papers on this topic,

Then, you have TWO WEEKS (a bit less) for following todos:

- You devise keywords to collect literature

- You use the seed paper (if you get one) to find similar literature

- You build a first collection of papers and you write down the criteria for making this list

- you get feedback and further instructions

- you refine the collection

- you build a shortlist of papers that do not build upon each other

Friday Nov.2, 13:00: **DELIVERABLE 1 (formatted document & slides):** description of the collection procedure, description of the shortlisting procedure, list of papers & shortlist, a few words on each paper

Monday Nov. 5, 9:15-11:30: **MEETING 1:** You present Deliverable 1 and get feedback; you are involved in the discussion of the papers on each topic

Your Deliverable 1 is either approved or not. If not, your absolute deadline is the date of Deliverable 2.

Timeline – Part 2: describing a paper *briefly*

After meeting 1 (Nov 5), you have TWO WEEKS (a bit less) for following todos:

- You perform a first reading of *all* 4 shortlisted papers

- You write down for each one: what exactly is the problem solved by the authors, why is this problem important, what are the keypoints of the solution

- You choose one of the papers to ask for feedback (concrete questions only!), you use the feedback to describe the other papers in a similar way

- You prepare a short presentation of each of the papers: it must be one slide only, at most two; you can use text in it, but not text from the paper

Friday Nov. 16, 13:00: **DELIVERABLE 2 (slides):** short presentations for the 4 shortlisted papers – you have ca. 2 minutes per paper

Monday Nov. 19, 9:15-11:00: **MEETING 2:** You present Deliverable 2, you get feedback; you engage in the discussion of the papers on the other topics; you may need to do changes in the deliverable

11:00 – 12:30: **MEETING 2 CONTINUED:** the review form is presented and explained (next slide)

Your Deliverable 2 is either approved or not. If not, your absolute deadline is the date of Deliverable 3.

Timeline – Part 3: writing your first review

The second part of Meeting 2 (Nov 19) is on the discussion of the review form (aka: review template).

Then, you have TWO WEEKS (a bit less) for following todos:

- You pick one of the papers

- You extend the description for this paper: with what other algorithms did the authors compare, what were the comparison criteria, how did they show that their approach is "good"

- You start filling the review template; you use it as guideline as you re-read the paper

- You get feedback (to concrete questions only!)

- You fill the review template; you can quote from the paper but only for the purposes of argumentation and to less than 25% of the whole review

Friday Nov. 30, 13:00: **DELIVERABLE 3 (filled review template for one paper):** review of at least one paper

Monday Dec. 3, 9:15 - 12:30: **MEETING 3:** You present the 1st review you wrote; you get feedback; you get involved in the discussion of the other reviews

Your Deliverable 3 is either approved or not. If not, your absolute deadline is the date of Meeting 4.

Timeline – Part 4: writing all reviews and one self-nomination

After Meeting 3 (Dec 3) you have ONE + FOUR WEEKS for following todos:

- You write reviews for all 4 papers on your topic

- You rank them and select the best one

- You write a "self-nomination" for this paper: you explain your ranking scheme and justify your decision for this paper

- You prepare a common repository for paper sharing (joint task, see next part)

After one week:

Monday Dec. 10, 11:00 – 12:30: **MEETING 4:** how to rank papers, how to write a self-nomination, how to nominate a paper on a topic worked by another student

…

... the four weeks for your todos

...

Monday Jan 7, 8:30: **DELIVERABLE 4 (filled review template for all papers, formatted document/slides):** reviews of all 4 papers, self-nomination document that compares the 4 papers and selects one for the competition

Your Deliverable 4 is either approved or not.

Timeline – Part 5: nominations for competition

Monday Jan. 7, 9:15- 12:30: **MEETING 5:** You present one of the reviews you wrote (which one is randomly chosen by the rest of the class); you present the self-nomination you wrote; you discuss the self-nominations

Tuesday Jan.8, 18:00: **DELIVERABLE 5 (a link):** self-nomination and all papers shared in a common repository

Then, you have A BIT LESS THAN TWO WEEKS for following todos:

- You read the self-nominations written by each other student for their top-1 paper and

- You write a "nomination" document, where you rank their top-1 papers, and cast votes

- You may decide to read the papers of the other students, but you are not expected to do so

Friday Jan 18, 13:00: **DELIVERABLE 6 (formatted document/slides):** nomination of papers from the other topics; the nomination is a ranked list

Monday Jan. 21, 9:15-12:30: **MEETING 6:** You present the nominations you wrote for the papers in the other topics; voting is done; awards are given

Friday Jan 25, 13;00: **FINAL DELIVERABLE (compiled report):** compilation of all documents of the previous deliverables into one document

Goal of the seminar

In this seminar, you learn to read papers in such a way that you become able to highlight the advantages and shortcomings of a paper

- With respect to the problem it solves

- In comparison to other papers that solve the same problem

- In comparison to other papers in the same research area

- In comparison to papers that are in another, remote research area (!!!)

This way of reading a paper is very demanding. But if you succeed, then you have acquired a deep understanding of what each paper does, and demonstrated that you are able to describe and evaluate a scientific work.

You also learn to search for papers and to set criteria on which papers to choose among hundreds of papers on the same subject.

These abilities are essential for the work you will need to do in a master thesis.

Evaluation criteria

1) Quality of the paper collection criteria

2) Quality of the short descriptions of the papers (counts the least)

3) Quality of the reviews (counts a lot)

4) Quality of the self-nomination (counts a lot)

5) Quality of the nomination for other papers (counts the most)

Where "Quality" encompasses

- technical quality (counts a lot)

- ability to identify advantages and disadvantages in a work (counts a lot)

- clarity of presentation (influences the other two)

Warning

You cannot pass this seminar if

- You do not understand the contents of the papers you read.

- You cannot describe what you read to others with your own words. Quoting from the paper is fine, but only if it helps you make a point formulated in your own words.

Warnings concerning grading:

The seminar is graded by 5.0 (FAIL) if

- one of the following cases appears for any of the papers you review:

  - The amount of quotations from the paper exceeds 25%.

  - The scores you give to the paper (the review template contains many scores) are not justified by the text in the review.

  - The review contains materials from papers that are not quoted.

- A deliverable is not submitted in time.

- A deliverable is not approved.

- Some deliverables are missing from the final report.

- Absence from the meetings

- No active participation in the meetings

BEWARE of the deadlines.

Thank you very much!

Questions ?