



FAKULTÄT FÜR
INFORMATIK

Topics for Seminars & Projects

Prof. Myra Spiliopoulou

myra@ovgu.de



Schedule for seminars and projects (B/M)

- **REGISTRATION:**

1. You apply for a topic till: **Friday, Apr 5 – 13:00, 2019**
2. Topic assignments made until **Monday, Apr 8, 2019**
3. You sign the form and submit by
Thursday, Apr 11, 12:30 – 13:00 in person

- **MEETINGS for seminars**

1st meeting: Monday, Apr 15, 13:00
then weekly (schedule follows)

Attendance is mandatory

- **REPORT for seminars:** **Wednesday, Jul 3, 15:00**
- **PRESENTATIONS/REPORT for projects:** **July/Aug 2019**

For Master DKE students:

M-DKE new: teamprojects belong to “Applied Data Science” **only**

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

SIGNAL-1: Time series management tool

This teamproject builds a tool that stores multidimensional time series of arbitrary length into a repository and supports segment matching queries

- SUBTASK 1: Schema management utility
- SUBTASK 2: Data storage utility
- SUBTASK 3: Segment matching utility
- SUBTASK 4: Demonstrator that collects time series from public domain repositories, using the segment matching utility

PREREQUISITES:

1. Data management
2. Software engineering in Python
3. Desirable: Stream Mining

SIGNAL-2: Time series recorder & collector

This teamproject builds a module that records multi-dimensional time series and stores them in a data repository

- SUBTASK 1: Specification of the sensory inputs

Input channel 1: keystrokes (one finger only)

Input channel 2: stress – device is already available

- SUBTASK 2: Construction of the device that collects the sensory inputs
- SUBTASK 3: Data acquisition and storage (cf. SIGNAL-1)
- SUBTASK 4: Design and realization of an experiment for the collection of inputs by $N \geq 20$ participants

PREREQUISITES:

1. Hardware engineering
2. Software engineering in Python
3. Desirable: Stream Mining

Subtasks 1, 2: two MDigiEng students

Seminar (Master level)

Goal & Prerequisites

Goal 1: Learn to assess the scientific quality of an article

Goal 2: Learn to compare scientific articles

Goal 3: Learn to evaluate scientific documents

Prerequisites:

- Familiarity with data mining / machine learning
- Eagerness to explore advanced scientific topics
- Willingness to do intensive work
- Availability for fast track work during June 2019

This seminar is for students close to the master thesis.

TASKS for the 1st round

1. Read two papers
2. Prepare a short presentation for each one
3. Prepare a review for each one (template will be given)
4. Present in class – as proponent for the one paper and as opponent for the other
5. Engage in a discussion for each of the paper you have not read
6. For each paper presented and discussed in class, contribute to the decision on whether the paper should be retained in the 2nd round.

Timeline:

- Tasks 1-3 till ca. Apr 20
- Tasks 4, 5, 6: multiple meetings (last one: April 29 or May 6)

TASKS for the 2nd round

1. Read one-two more papers
2. Prepare a review for each one
3. Rank the papers (criteria will be discussed in class)
4. For each paper presented and discussed in class, contribute to the decision on the final ranking position of this paper (4 ranks total)

Timeline:

- Tasks 1-3 till ca. May 15
- Task 4: multiple meetings (last one: May 27)

TASKS for the 3rd round

1. Read one PhD thesis (from a list of 3-5 theses)
2. Prepare a presentation and a review
3. Engage in the in-class discussion to rank all manuscripts
4. Write a final discussion and justification for the thesis you have read

Timeline:

- Assignment: June 3 (might be June 4)
- Task 2 - 1st draft: June 17; final: June 20
- Task 3 – meetings: June 17, June 24, July 1
- Task 4: July 3

TASKS for the 3rd round

1. Read one PhD thesis (from a list of 3-5 theses)
2. Prepare a presentation and a review
3. Engage in the in-class discussion to rank all manuscripts
4. Write a final discussion and justification for the thesis you have read

Timeline:

- Assignment: June 3 (might be June 4)
- Task 2 - 1st draft: June 17; final: June 20
- Task 3 – meetings: June 17, June 24, July 1
- Task 4: July 3

ACM SIGKDD 2019
Best PhD Award

Goal of the seminar

In this seminar, you learn to read scientific work in such a way that you become able to judge its merits

- With respect to the problem it solves
- In comparison to other works that solve the same problem (1st round)
- In comparison to other works in the same research area (2nd round)
- In comparison to works in another, remote research area (3rd round)

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 short descriptions of the papers (counts the least)
- 2) Quality of the reviews for the papers (counts more)
- 3) Quality of the arguments as proponent and opponent (counts more)
- 4) Quality of the 3rd round review (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

1. You do not understand the contents of the papers you read.
2. You cannot describe what you read to others with your own words.
3. You cannot argue for (or against) a paper with 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 documents 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

Applying for the seminar

When you apply for the seminar, choose two papers from those in the next slides:

- All papers from KDD 2018, research track
- Choose from two groups – at least from one small group
- Explain *briefly* why you chose from these groups.

Papers for 1st and 2nd round

Group 1:

- Efficient Attribute Recommendation with Probabilistic Guarantee
- Ranking Distillation: Learning Compact Ranking Models With High Performance for Recommender System
- xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems
- Local Latent Space Models for Top-N Recommendation
- Learning from History and Present: Next-item Recommendation via Discriminatively Exploiting User Behavior
- Local Latent Space Models for Top-N Recommendation

Papers for 1st and 2nd round

Group 2:

- Data Diff: Interpretable, Executable Summaries of Changes in Distributions for Data Wrangling
- Learning and Interpreting Complex Distributions in Empirical Data
- Active Feature Acquisition with Supervised Matrix Completion
- Stable Prediction across Unknown Environments
- FAHES: A Robust Disguised Missing Values Detector
- SUSTain: Scalable Unsupervised Scoring for Tensors and its Application to Phenotyping
- Efficient Mining of the Most Significant Patterns with Permutation Testing

Papers for 1st and 2nd round

Group 3:

- Multi-label Learning with Highly Incomplete Data via Collaborative Embedding
- Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts
- Variable Selection and Task Grouping for Multi-Task Learning
- Towards Mitigating the Class-Imbalance Problem for Partial Label Learning
- Stabilizing Reinforcement Learning in Dynamic Environment with Application to Online Recommendation
- New Incremental Learning Algorithm for Semi-Supervised Support Vector Machine

Papers for 1st and 2nd round

Group 4:

- A Dual Markov Chain Topic Model for Dynamic Environments
- Decoupled Learning for Factorial Marked Temporal Point Processes

Papers for 1st and 2nd round

Group 5:

- Model-based Clustering of Short Text Streams
- TaxoGen: Constructing Topical Concept Taxonomy by Adaptive Term Embedding and Clustering
- When Sentiment Analysis Meets Social Network: A Holistic User Behavior Modeling in Opinionated Data
- Dynamic Embeddings for User Profiling in Twitter
- Active Opinion Maximization in Social Networks

Thank you very much!

Questions ?