Topics for Seminars & Projects

Prof. Myra Spiliopoulou

myra@ovgu.de

Timeline for seminars and projects (Bachelor / Master)

- REGISTRATION:

  - You decide for a topic, in agreement with the supervisor.

  - You sign the form and submit to the KMD team by

    Friday 19.10.17

- PRESENTATIONS for seminars: 22.01.2018, 09:00–12:00

  Attendance is mandatory!

- REPORT for seminars: 29.01.2018 till 13:00 s.t.

- PRESENTATIONS/REPORT for projects: February 2018

ALL MEETINGS FROM NOW ON: KMD LAB (R 021)

For Master DKE students: Teamprojects can be assigned to area "Methods I" only

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

IT–Softwareprojects (Bachelor degrees only)

Prerequisites for all projects of this type:
- A team of THREE students
- GOOD software engineering skills
- Background:
  - Data mining / ML: at least one member
  - Data mining / ML / statistics: at least one member

LEA-5: How do humans learn? Modeling cluster evolution

This IT-softwareproject extends earlier projects.

- LEA-1 built an interactive environment, with which the experimentator can observe the behaviour of experiment participants, as they form evolving groups during the experiment.

- LEA-2 extended this environment with visualization aids and with constraint-based clustering.

- LEA-3 built a testbed for the evaluation of the methods developed in LEA-2 under different parameter settings.

Goal of LEA-5 is to build an online clustering algorithm for the testbed data records used in LEA-3 and provide an appropriate visualization for this algorithm.

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

Privacy–preserving sequence mining

Goal of this project is to expand a privacy–preserving kNN classification algorithm, originally intended for flat data records to an algorithm for the classification of multidimensional sequences.

The teamproject encompasses

1. brief literature overview on the subject of privacy–preserving classification of flat data and on sequence data

2. design and implementation of the algorithm

3. specification of the evaluation criteria

4. design and implementation of a test environment, using public domain datasets

5. evaluation, discussion of the results and demonstration

MDKE area: "Methods I"

Framework for Time Series Imputation

Goal: implement a framework for imputing time series (univariate & multivariate, same length & different length) and evaluate the quality of the imputation in Java.

Focus: exchangeability of imputation algorithms

Prerequisites: 3 Students, Classification, UI-Design in Java

The framework encompasses 4 components:

1. Loading and preparing the datasets (from UCR repository)

2. Imputation algorithms (uni- and multivariate)

3. Evaluation through RMSE and TSC

4. Visualisation of TS and Results

Literature review of imputation algorithms, discussion of the results and demonstration

Christian.beyer@ovgu.de

MDKE area: „Methods 1"

Seminars (Master level)

Prerequisites for all seminars of this type:

- Background in data mining / machine learning

Seminar subjects

1. Predictive maintenance: the potential of neural networks

   MDKE area: "Fundamentals"

2. Predictive maintenance: the potential of bayesian networks

   MDKE area: "Fundamentals"

3. Kinetics in sports mining: recognizing movements with convolutional neural networks

   MDKE area: "Methods I"

For each seminar subject

TASK: Write a literature overview of the identified methods

SUBTASKS:

- Explain the learning task addressed by each method

1. Collect literature

2. Describe the formal problem solved *by each method*, focussing on commonalities among the methods.

3. Describe how each method works, focussing on differences among the methods.

4. Specify at least two criteria that allow you to compare methods.

5. For each criterion, explain which method is best and which methods are not performing well.

The specification of search criteria for literature collection, the decision about relevance/irrelevance of each article and the specification of the comparison criteria are mission-critical.

Thank you very much!

Questions ?