Knowledge-Based Systems (KBS)

Machine Learning and Knowledge Engineering

Winter semester 2021/2022

Prof. Dr. Martin Atzmüller

Osnabrück University & DFKI
Semantic Information Systems Group
https://sis.cs.uos.de/

Prof. Dr. Martin Atzmüller – Background

- since 08/2020: ROSEN-Group-Endowed Chair of Semantic Information Systems (Full Professor, W3)
  Also – since 10/2021: Affiliated Professor at DFKI
- 2017-08/2020: Tilburg University, Department of Cognitive Science and Artificial Intelligence (Associate Professor)
- 2017-2019: Université Sorbonne Paris Nord - Computer Science Lab (Visiting Professor)
- 2010-2017: University of Kassel - Knowledge and Data Engeering Group (Postdoc, Habilitation, Privatdozent)
- 2003-2010: University of Würzburg: Chair for Artificial Intelligence and Applied Computer Science
  (Promotion, Postdoc)

Prof. Dr. Martin Atzmüller – Research

- The research of the ROSEN-Group-Endowed Chair of Semantic Information Systems and the according research group, headed by Prof. Dr. Martin Atzmueller, centers on Artificial Intelligence, Machine Learning and Data Science:
  - **Machine Learning on Complex Data**:
    One major focus is on machine learning and data mining on complex data such as graphs, complex networks and temporal data, also with a human-centered perspective.
  - **Artificial Intelligence, Human-Centered Data Science**
    $\rightarrow$ **Explainable and Interpretable Machine Learning**:
    Leveraging the massive amounts of Big Data by intelligent analytics and semantic interpretation, the research focuses on how to 'make sense' of complex information and knowledge processes. This is enabled using explainable and interpretable machine learning and semantic data mining, including semantic technologies.
- Projects: `https://sis.cs.uos.de`

Prof. Dr. Martin Atzmüller – Teaching

- Bachelor:
  - Database Systems
  - Seminars on Data Science and Machine Learning
- Master:
  - Knowledge-Based Systems:
    Machine Learning and Knowledge Engineering
  - Machine Learning on Complex Data: Graphs and Sequences (forthcoming)
  - Seminars on Data Science and Machine Learning
- ... according Bachelor-/Master thesis

Knowledge-Based Systems – Contents

Course Organization

1. Preliminaries & Prerequisites
2. Lecture & Practical Sessions
3. Practical Session Organization
4. Practical Exercises & Project
5. Literature, Further Reading

Preliminaries – This Winter Semester

- Please check all the rules and regulations of the university
- In particular: For lectures/practical session – book your seats in advance
- Proof of 3G status
- You are responsible for adhering to the rules & regulations
- Masks/lecture halls: IF (distance < 1.5m) THEN (wear mask)
- If you feel any symptoms of a cold etc. it is better to stay at home – then, get well soon!
- We will have "in presence" and online sessions - plan will be announced soon

Some preliminary remarks ...

Knowledge-Based Systems - Requirements

- Knowledge-Based Systems is a *Master Course*
- ... in English
- Therefore, basically all prerequisites for the *Computer Science Master Program* apply as *expected*.
- Furthermore, the following knowledge & skills are expected:
  - Lectures: Database Systems, Artificial Intelligence (see module handbook for the respective courses and content).
  - Most importantly, knowledge on:
    - Basics on data representation, modeling & management;
    - Basics on logics, probability theory/uncertainty, inference etc.
  - Programming experience (preferably in Python)

Structure: Lectures & Practical Sessions

"Knowledge-Based Systems" provides the following components:

- Lecture (4h / week): Presentation of the learning materials (Plenary/Video lectures, Q&A etc.)
- Practical Session (2h / week): Exercises, practical consolidation
- Self Study: **Studying the materials of the lectures/practical sessions**
- Examination: Final examination – Project ("Hausarbeit")

Important:

- From the module handbook MSc Computer Science: KBS = 9 LP $\rightarrow$ 270h total (!)
- Therefore: Plan, and allocate time accordingly

Format: Lectures & Practical Sessions

- Lectures (Monday, 16:00-18:00; Tuesday 10:00-12:00) as pre-recorded Video Lectures:
  $\rightsquigarrow$ lectures will be pre-recorded via OpenCast (and provided online via Stud.IP)
- Practical Session (Tuesday, 14:00-16:00): Practical sessions will be partially "in presence" (Präsenz), and partially online ("live, online" via BBB).
- There will be different (sometimes mixed) options for the practical sessions (assignment, inverted classroom/questions, tutorials)

Assignments

- For assignments, you have to upload your solution(s) via VIPS
- You can work on the assignment in groups (max. 2 persons)
- The assignments will be graded *pass/fail*
- Requirement for admission to the final examination (*Prüfungsvorleistung*): **PASS on 4 out of 5 assignments**
- Please note, that I will ask (some of you) to present (and discuss) your *sample solution(s)* in class
- Also, attending the practical sessions is strongly recommended – can also be related/relevant to/for the assignments

Final Examination – Project ("Hausarbeit")

Excerpt – taken from the APO:

(3) [1]Eine Hausarbeit ist die selbstständige Bearbeitung und angemessene Dokumentation einer fachspezifischen Aufgabenstellung im Rahmen eines festgelegten Zeitraums. [2]In geeigneten Fällen können Hausarbeiten auch in Form von Gruppenarbeiten erbracht werden; die Eignung des Themas stellt die Prüfende oder der Prüfende fest. [4]Der Beitrag der einzelnen Verfasserin oder des einzelnen Verfassers muss die Anforderungen nach Satz 1 erfüllen und als individuelle Prüfungsleistung auf Grund der Angabe von Abschnitten, Seitenzahlen oder anderen objektiven Kriterien deutlich abgrenzbar und für sich bewertbar sein. [5]Der oder die Prüfende kann die Bearbeitungszeit einer Hausarbeit in begründeten Ausnahmefällen einmalig bis um die Hälfte der vorgegebenen Zeit verlängern. [6]Dem Prüfling kann Gelegenheit gegeben werden, für die Aufgabenstellung Vorschläge einzureichen.

Final Examination – Project ("Hausarbeit")

(Preliminary) Schedule:

- 2021-12-21 Presentation/discussion of possible project topics/directions (example topics, also cf. APO)
- 2022-01-11 Submission of topic proposals + discussion
- 2022-01-18 Discussion of elaborated project proposal
- 2022-02-28 Deadline for submission of project – short paper (English)
- Short "closing meeting" (5-10 minutes) for the project.

Final Examination – Project ("Hausarbeit")

- Important: You should think about and write a proposal for your (!) project (which we will of course discuss)
- It is possible to conduct the project in a group (max. 2 persons).
  - However, then, you need to include a paragraph in the paper who did what (and to which extent), e.g., algorithmic development, programming, preprocessing, writing (which section) etc., and which tasks where done jointly (then: approximate percentages).
  - This needs to be clearly attributable
  - Obviously, two persons can conduct a larger project (scope) than a single person

Final Examination – Project ("Hausarbeit")

- What could the project be about?
  - Implementation of a (adequately scaled) KBS ...
  - Implementation of a learning algorithm ...
  - Extension of an existing algorithm/method
  - Analysis/evaluation of an algorithm/method on a new dataset (or: in a new context)
  - ...
  - (I will provide some examples later ...)
- You need to write a short paper on this (introduction, related work, method, results, discussion) – max. 5 pages (including references; for a team/2-person project: max. 8 pages) in a two-column (ACM-like) template (will be provided).
  N.B. You should use LaTeX to write the paper. The paper should be written in English.

Literature

Main ressource:

- **Stud.IP: Slides/Course material ($\rightarrow$ Files, Tab: Dateien)**

Materials on course topics and (general) further reading:

- Atzmueller, Martin (2015) *Subgroup discovery*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 5(1):35-49
- Beierle, Christoph, and Gabriele Kern-Isberner (2019) *Methoden wissensbasierter Systeme: Grundlagen, Algorithmen, Anwendungen*. Springer
- Flach, Peter. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.
- Goodfellow, Ian, Bengio, Yoshua, and Courville Aaron (2016) *Deep Learning*. MIT Press
- Han, Jiawei, and Kamber, Micheline (2011) *Data Mining: Concepts and Techniques*. 3rd Edition, Morgan Kaufmann, Burlington.
- Puppe, Frank (2012) *Systematic introduction to expert systems: Knowledge representations and problem-solving methods*. Springer
- Russell, Stuart J., and Norvig, Peter (2010) *Artificial Intelligence: A Modern Approach*. Upper Saddle River, N.J. :Prentice Hall

Programming et al.

Some general remarks and recommendations

- In this course, nearly all techniques and methods will be introduced and defined in an algorithmic manner.
- Also, many will be introduced programmatically ...
- So, you are highly encouraged to test and play with them.
- We are also going to have a look at several exercises/assignments.
- For those, we are going to use Python 3 ...

Onto learning Python ...

- Getting the latest version for your OS:
  https://www.python.org/downloads/
- You will probably want to use a Python IDE (Integrated Development Environment)
- We will use Python (Jupyter) Notebooks in the exercises
- $\rightsquigarrow$ Miniconda:
  https://docs.conda.io/en/latest/miniconda.html
- Also possible:
  - `https://vscodium.com/`
  - `https://www.jetbrains.com/pycharm`
- For Python, there are a lot of good (video) tutorials, e.g.:
  `https://www.youtube.com/watch?v=rfscVS0vtbw`