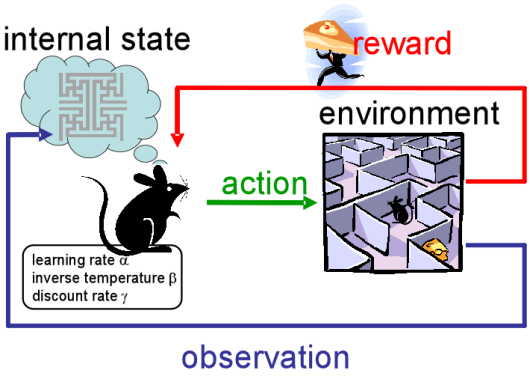
Master Project A: *Reinforcement Learning for riskbased Boardgames*

Apart from the classical supervised and unsupervised learning tasks, reinforcement learning (RL) is the third big paradigm in the area of machine learning or artificial intelligence. RL tasks usually deal with decision-making processes, where a so-called Agent will try to maximize a specified “reward” to solve a problem. In this case, we want to teach an RL-Agent to play the recent boardgame “The Quacks of Quedlinburg”[1]. This game combines deck-building and push-your-luck elements, meaning you will have to teach your algorithm to make good decisions and forward-looking thinking, while taking a risk at gambling might also work. You are tasked to come up with and develop possible strategies your Agent could use, define cost- and reward-functions and implement those using a given API and protocol to let your agent play a simplified version of the game on its one.



For this project, the following skills and characteristics are desired:

- Have fun playing advanced boardgames and planning out strategies
- Basic Reinforcement Learning Knowledge (but we will also give an short introduction)
- Basic knowledge about REST[2] and Flask[3]
- Advanced knowledge in Python

[1] https://www.schmidtspiele.de/files/Retail/72dpi_PNG/88220_Quacksalber_Quacks_Rules+Almanach_GB.pdf
[2] https://de.wikipedia.org/wiki/Representational_State_Transfer
[3] https://flask.palletsprojects.com/en/1.1.x/quickstart/#a-minimal-application

Master Project B: *Time Series Forecasting of Electric Engergy Consumption*

Time Series Prediction - sometimes also referred to as Time Series Forecasting - is an important statistical task and an interesting application for modern machine learning methods. In this project you will get to work on an exciting dataset concerning electricity demand. You will implement a project to deal with large amounts of multidimensional time series data and use a series of statistical and machine learning models to forecast future electricity demand.

Taking part in this project you will learn about and implement:

- Autoregressive and Generative Models
- Multiple types of time series prediction models including classical statistical and deep learning approaches:
  - Classical Timeseries Prediction, e.g. ARIMA
  - Recurrent Neural Networks (RNNs)
  - Long short-term memory models (LSTMs) & Gated Recurrent Units (GRUs)
  - 1d Convolutional Networks (CNNs, Wave-Net)
  - Seq2Seq Models

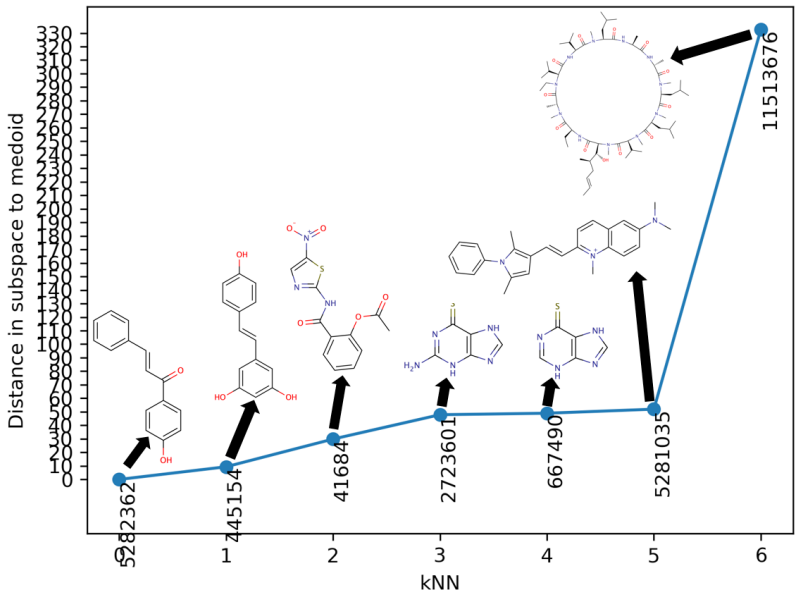
- Data processing for large amounts of timeseries data
- Training, Testing and using state of the art deep learning libraries

<u>Prerequisites:</u>

- Basic Knowledge of Machine Learning and Data science (Train-Test Split, Neural Networks, Deep Learning, Hyperparameter Optimization)
- Experience with Python and Git, especially working with datasets (numpy, pandas, etc)
- Best, but not required: Experience machine learning libraries like torch, TensorFlow or keras

Master Project C: *Clustering of Mol2Vec-Embeddings*

The current situation of the pandemic revealed that the importance of the so-called screening process for finding similar compounds with desirable properties as potent drugs against diseases is paramount. In this practical project, the objectives are to develop and implement different clusterings of molecular structures which have previously been embedded via Mol2Vec. With the resulting clusterings, it is further investigated how far they can be utilized to augment the screening procedure.



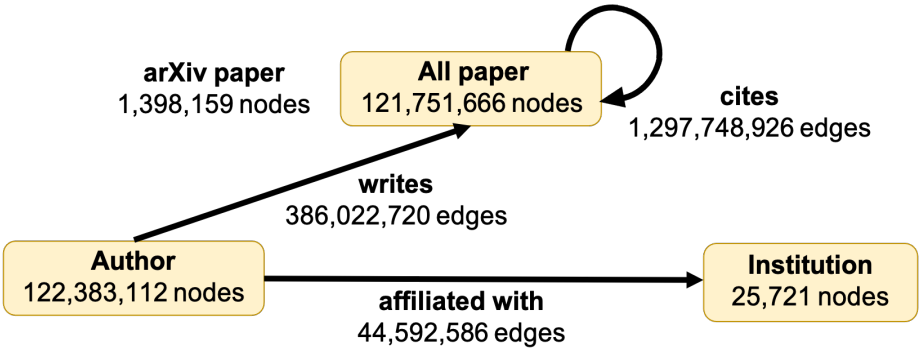
For this topic the following skills would be ideal:

- Knowledge of clustering (like from the KDD I lecture or BDMA)
- Experiences with Python are beneficial

Knowledge of chemistry, biochemistry or pharmacology is not required.
For questions, don’t hesitate to contact me at kazempour [at] dbs.ifi.lmu.de
Teach a Reinforcement Learning Agent to play a “deck-building” Boardgame

Master Project D: *Open Graph Benchmark – Large-Scale Challenge @ KDD Cup 2021*

For this project you will participate as a group in the OBG-LSC competition for the ACM conference on knowledge discovery from data (KDD) [1]. The goal is to effectively apply deep learning techniques, namely graph neural networks (GNNs), on large (billion-edge) graphs. In this case, the graph is a Microsoft academic graph (MAG), which is a heterogeneous graph (*i.e* with typed nodes and edges), where the task is to predict the subject area of a paper. This is a particularly relevant task, as the volume of scientific publication is roughly doubling every 12 years. Automating the process of labelling the subject area can alleviate the burden of manual labelling ans assist in automatic search and organization of research data.
The challenge is to develop the best possible model, and the winners get to present their approach on the KDD Cup workshop. More info on the competition and baseline can be found here: [1], [2].



Schema of the MAG240M-LSC graph. Figure taken from [1].

Taking part in this project you will learn about and/or implement:

- HINs: Heterogeneous Information Networks (networks with typed nodes and edges)
- GNNs: Graph Neural Networks
- Data handling for large-scale (billion-edge) graphs
- Training, Testing and using state of the art deep learning libraries

For this project the following skills are required:

- Basic Knowledge of Machine Learning (train/val/test split, neural networks, deep learning, hyperparameter optimization)
- Experience with Python datasets (numpy, pandas, etc)
- git-basics
- Best, but not required: Experience in DL libraries like PyTorch, TensorFlow, or keras
- Team spirit
- A competitive spirit ;)

[1] https://ogb.stanford.edu/kddcup2021/
[2] https://arxiv.org/abs/2103.09430

Master Project E: *Reinforcement Learning for Autonomous Fleet Navigation*

The goal of Autonomous Navigation became to one of the most game-changing but also most challenging tasks for the next years. Many approaches had been developed to automatize navigation actions for various kinds of vehicles but so far only assistance systems have made its way to the field. The latest and most promising approach to make progress in this area is the paradigm of reinforcement learning (RL). Therefore, traffic situations are captured as a current state and an RL agent learns reward-driven actions and strategies leading to the next and long-term best movement and mission goal state.
In this master project you will develop RL agents capable to autonomously navigate a fleet of vehicles. The agents need to be trained to be able to fulfill mission goals in the most efficient manner. For this task you will explore the required action and state space and develop and tune reward-based loss functions to enhance the intelligence of your agents step by step. Starting with a greedy prototype the fleet performance shall be compared with each new version. In a final presentation you will present the results of your data science project, which will not only cover its technical concept and implementation but also how you organized and structured your project, schedule and team to achieve these results.

In order to master this project it would be ideal (but not essentially required) if you have the following skills and characteristics:

- knowledge in Reinforcement Learning gained in courses like e.g. Intelligent Systems, etc.
- knowledge in Deep Learning
- advanced knowledge in Python
- familiarity with python libraries like pandas, numpy, torch/tensorflow, scikit-learn, etc.
- eager to learn and master challenging tasks as part of a team