Machine Learning Lab Course

Organizational Meeting

lecturer: Prof. Dr. Stephan Günnemann
Team

- Prof. Dr. Stephan Günnemann
- Marin Bilos
- Daniel Zügner

This is a practical course (Praktikum) for **Master** students!

*Name of module: Large-Scale Machine Learning (IN2106, IN4192)*

website: [ml-lab.in.tum.de](http://ml-lab.in.tum.de)
Machine learning is a steadily growing field

More ML papers are published

Source: aiindex.org
Industry is closely following academic growth

More startups are based on AI

Source: aiindex.org
Knowing ML is becoming an important skill set

More jobs require AI skills

Source: indeed.com
Why attend our Machine Learning lab course?

1. Get the chance to implement and apply state-of-the-art ML algorithms
2. Gain hands-on experience working on real-world data, solving real-world tasks by working on projects offered by our industry partners.
   – Successful projects might even qualify for a subsequent master thesis.
3. Work on large-scale problems with the support of state-of-the-art GPU computing resources.
Requirements

- Requirements for the lab course
  - **strong programming skills** (Java, Python, C++, Java, etc.)
  - strong knowledge in data mining/machine learning
  - you should have passed relevant courses (the more, the better)
    - Mining Massive Datasets
    - Machine Learning
    - Our seminars
  - self-motivation

- Additional selection criteria
  - other **relevant** experience (projects in companies, experience as a HiWi)
    - you can send an overview of your experience to us *(see end of slides)*
Organization

- Groups of 3 to 4 students
- Each team will work on a different project
- Groups are allowed to (should) collaborate!
  - exchange your experience with the other groups
  - how do the other groups tackle certain problems?
- Students get access to our GPU compute servers
  - Each of the servers has:
    - 4x NVIDIA GPU with 11GB RAM
    - 10-core CPU
    - 256 GB RAM
  - scale up your models and data!
Organization

- Weekly meetings (around 120 minutes)
  - each group should briefly report their progress, open problems, and next steps

- Regular documentation of your work
  - status reports and documentation (we have set up a wiki)
  - use of a central code repository
This semester's industry partners

SIEMENS  metris.io  BMW
Industry project: Metris

- In natural language processing the problem of ambiguity makes many tasks hard for algorithms to solve, while humans excel at them
  - E.g. finding what something refers to: Bayern as a state or as a football team
  - Or finding connections between different terms having the same meaning: Angela Merkel or The Chancellor
  - However, humans do not have capacity to process big amounts of data

- Task: Mining large text data to find entities and perform context disambiguation
  - Working on large text corpus of a Munich-based startup
Industry project: BMW

- Big multinational companies have workers using different languages, but it would be useful if all could have access to a shared knowledge base.
- Problem: Searching this knowledge base (hard even with one language)
  - Both the query and the result are written in natural language
  - Problem of finding relevant information from a written description
  - Problem of representing the current knowledge so that it is easy to search
  - Harder with multiple different languages
- Task: Information extraction and data representation in multilingual setting
  - Working with big data provided by BMW
Industry project

- Topic 1: Mining (massive) public transit data
- Topic 2: Representative present today: Mohamed Khalil
**Academic project: Graph neural network robustness**

*Graph neural networks* are successfully used to solve many tasks. Despite their success, there is still much to learn about their robustness w.r.t. corruptions in the data.

- The goal of this project is to deepen our understanding of graph neural networks.
- Therefore, you will build a test suite and compare different graph neural network architectures and their robustness w.r.t. noise and adversarial attacks.
- If successful, this is a great starting point for a publication and/or a master thesis.
The growing complexity of communication networks makes their analysis and maintenance increasingly challenging. Computer networks suffer from badly secured Internet of Things devices, hacker attacks or outdated software.

- Interdisciplinary project with the Chair of Communication Networks
- The goal is developing machine learning models for anomaly detection in dynamic communication networks
  - Using node embedding methods, probabilistic graphical models or community detection algorithms
More projects

- We’re **doubling the course size** next semester (12 -> 24 spots).
- We are still in discussions with **industry partners**, so we may add more industry projects before the semester starts.
- There will also be additional academic projects about the state of the art in **graph neural networks** and/or **robust temporal learning**.
Registration via the matching system!

*Module name: Large-Scale Machine Learning (IN2106, IN4192)*

+ fill out the application form (see next slide)
Your Experience

- Fill out our brief online form about your experience by July 24
  - provide us with a list of your experience in data mining/machine learning (courses, projects, etc.)
  - please send us a short overview (bullet list, not a complete CV)

- We will post a link to the form by tomorrow (July 16) at ml-lab.in.tum.de.
Also visit:

Poster session with this semester's Lab course students. Entrance of Interims, July 22 at 3:45pm

Join and ask questions!