

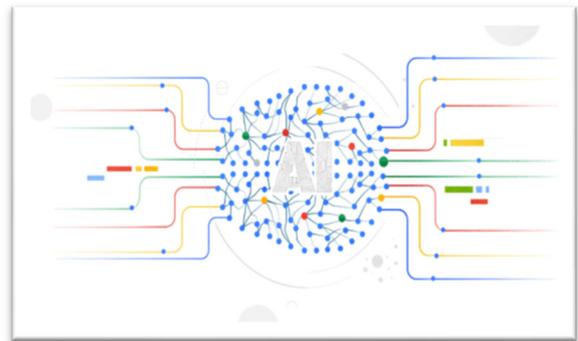
Data Management for End-to-End Machine Learning

Winter Term 2021/2022

Lecturers: Prof. Dr. B. Mitschang, Dr. P. Reimann, Dr. H. Schwarz
Contact: Holger Schwarz, holger.schwarz@ipvs.uni-stuttgart.de
Language: English

Description

The comprehensive analysis of large volumes of data is crucial in modern information systems. Some of the main driving forces are current trends such as Industrie 4.0 and the Internet of Things (IoT). In particular, applying machine learning (ML) to data of real-world scenarios imposes various data management challenges in several steps of an end-to-end ML application. Very often, the characteristics of real-world data impose challenges themselves. For instance, data may change over time (data set shifts in non-stationary environments). Here, major problems are how to detect and characterize these changes (e.g., different concept drifts), as well as how to consider them in adaptive machine learning strategies. Furthermore, organizations typically train a multitude of machine learning models from their data. These ML models and their lifecycle have to be managed in a comprehensive way. Afterwards, the most suitable models have to be selected for a concrete application, e.g., via approaches to model discovery or AutoML. Moreover, many machine learning models provide some kind of prediction. Here, it is mandatory for the users to understand why a certain prediction was provided by the model. Finding methods to offer such explanations and thus to help people interpret the models correctly, instead of treating them just as black boxes, is another important challenge. In this seminar, we address current technologies, concepts, algorithms and system infrastructures for above-mentioned challenges for data management in end-to-end machine learning, for example:



- Machine learning with complex data characteristics, e.g., in non-stationary environments
- Comprehensive management of machine learning models
- Automatic selection of machine learning models (e.g., AutoML, Meta-Learning)
- Explainable artificial intelligence / interpretable machine learning
- ...

Remarks

- In this advanced seminar, each student will work on one specific topic. Basic literature will be provided by the advisors. Each student will summarize results in a document of about 20 pages and a 30 minutes presentation.
- **Seminar topics will be assigned to the participants in a first meeting at the beginning of the winter term. Registered participants will be informed about the date by email.**

Prerequisites

Basic knowledge on database systems and information systems, e.g., from the lectures "Modellierung" or „Data Warehouse, Data Mining and OLAP“, is mandatory.