Master Project Winter Term 2019/20

Causal Reasoning on Enterprise Data

Motivation

The questions that motivate most data analysis in an enterprise context are of causal nature. E.g., what are the causes and effects of events in manufacturing processes under observation? Nevertheless, the associational nature of state-of-the-art statistical methods often leads to misinterpretation and incorrect deduction. A recently developed mathematical theory of interventions enables causal reasoning on the basis of observational data but is limited due to performance constraints especially in the context of enterprise data characteristics.

This master project has the intention to extend an existing modular IT system that enables the application of machine learning techniques for causal inference in a real-world context. Therefore, you will deep dive into the core engine and improve existing approaches of constraint-based causal structure learning. In the context of enterprise data, complex and heterogeneous data-characteristics, high dimensionality, or large data volume may lead to long execution times of such a pipeline or even deficient results. Your goal is to drastically improve the performance of existing approaches, e.g., through parallelization techniques, and further incorporate methods for heterogeneous data, which are not yet available in the existing pipeline.

As part of the journey, you will have the opportunity to deepen your knowledge about tools for data science, improve machine learning skills, and influence the performance of an end-to-end pipeline for causal inference.

Project Goals

1. Deep dive into concepts of causal inference
2. Analyze & improve statistical methods for constraint-based causal structure learning
3. Optimize your implementation for underlying hardware, e.g., multi-core, GPUs
4. Extend an existing tool for data scientists, to allow for comparison of different experiments with regards to performance & quality

Technology & Skills

The core of the work will be based upon an existing pipeline for causal inference and previously developed extensions (e.g., using In-Memory Database Management Systems or GPUs) of the causal structure learning algorithms. Prior understanding of the fundamentals of machine learning techniques (e.g., having attended the lecture Causal Inference – Theory and Applications in Enterprise Computing or equivalent) is expected as well as knowledge in one of the following areas (C++, R, Python, openMP, CUDA).

Contact

You are welcome to visit us in the "Villa" on Campus 2 or reach out to one of the following contacts:

Dr. Matthias Uflacker, Johannes Hügle, Christopher Schmidt