Lecture Series “Data Science & Health”

Winter Semester 2019/2020

Scheduling:

Rooms:
KIT: Geb 07.07, Vincenz-Prießnitz-Straße 1, Seminarraum 222
Heidelberg: Konferenzraum, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen IWR Im Neuenheimer Feld 205 - Mathematikon 69120 Heidelberg

Date/Time	Location	Lecturer	Topic
22.10. 9:45-11:15	KIT	Ralf Mikut	Time Series Analysis
22.10. 11:30-13:00	KIT	Lennart Hilbert	Fluorescence microscopy and digital image processing in molecular cell biology
5.11. 9:45-11:15	Heidelberg	NN	
5.11. 11:30-13:00	Heidelberg	Klaus Maier-Hein	Radiologic Data Science
19.11. 9:45-11:15	KIT	Achim Streit	Introduction to Distributed and Parallel Computing
19.11 11:30-13:00	KIT	Peter Sanders	Parallel Algorithms for Dummies
3.12. 9:45-11:15	Heidelberg	Holger Fröning	Any Growth is Bounded - On the Future of Performance Scaling
3.12 11:30-13:00	Heidelberg	Lena Maier-Hein	Surgical Data Science
17.12 9:45-11:15	KIT	Franziska Mathis-Ullrich	Minimally-Invasive Robots for Medicine
17.12. 11:30-13:00	KIT	Tamim Asfour	Data-Driven Learning of Sensorimotor Skills for Robots and Exoskeletons
7.1. 9:45-11:15	Heidelberg	Michael Gertz	Introduction to Text Analysis

7.1. 11:30-13:00	Heidelberg	Robert Strzodka	Essential Performance Considerations in Programming
21.1. 9:45-11:15	KIT	Anne Koziolek	Requirements Engineering for Data-driven Solutions
21.1. 11:30-13:00	KIT	Ali Sunyaev	Introduction to Decentralized Data Management with Distributed Ledger Technology
4.2. 9:45-11:15	Heidelberg	Martin Frank	Mathematical Aspects of Uncertainty Quantification
4.2. 11:30-13:00	Heidelberg	Filip Sadlo	Introduction to Visual Data Science

Topics for 90 min Lectures in Data Science & Health

Methods

Ralf Mikut:
Time Series Analysis
The goal of this lecture is to present typical goals, methods and use cases for the analysis of time series in health. It includes preprocessing aspects (i.e., outlier detection, handling of missing data), basic analysis methods (i.e., clustering of time series, feature extraction from time series, integration of a priori knowledge and interpretability) and analysis tools. The use cases include data from movement analysis and prosthesis control.

Anne Koziolek:
Requirements Engineering for Data-driven Solutions
This lecture will discuss the state of the art of requirements engineering for data science applications and machine learning applications. How can we approach problems systematically? What RE techniques can be transferred from traditional SW development to data science applications and ML applications?

Martin Frank:
Mathematical Aspects of Uncertainty Quantification
We give an overview of mathematical methods to propagate uncertainties (e.g. unknown data, unknown initial or boundary conditions) through models (mostly differential

equations). Given uncertain input, how uncertain is the output? The uncertainties are modeled as random variables, and thus the solutions of the equations become random variables themselves. We summarize the necessary foundations of probability theory, with a focus on modeling correlated and uncorrelated random vectors, and present the following methods: Monte-Carlo, Quasi Monte-Carlo, Multilevel Monte-Carlo, generalized Polynomial Chaos, Sparse grids.

Ali Sunyaev:
Introduction to Decentralized Data Management with Distributed Ledger Technology
The goal of this lecture is to introduce distributed ledger technology (e.g., Blockchain) as a cutting edge method for the decentralized management of sensitive data. The lecture will cover the foundations of distributed ledger technology as well as salient design characteristics of distributed ledger systems. Moreover, usage scenarios for distributed ledger systems in health care data science applications will be discussed.

Robert Strzodka
Essential Performance Considerations in Programming
Operating on large data can be very time consuming. Design of parallel algorithms and implementations for high performance computing (HPC) is a wide research area and gaining additional performance can be very hard. However, in many cases slow code execution is due to disregard of simple performance considerations in programming. So before turning to HPC techniques it is essential to know some basic dos and don’ts with respect to code efficiency. We will look at common pitfalls and how to avoid them. Code examples in C++ will be shown, but most of the ideas apply also to other languages.

Achim Streit
Introduction to Distributed and Parallel Computing
This lecture gives an introduction to the basis of distributed and parallel computing, in particular to Grid and Cloud Computing as well as basic architectures for parallel computing systems i.e. HPC-systems. The content will cover history (ie. where everything started), principles and real-world examples.

Peter Sanders:
Parallel Algorithms for Dummys
Building on the lecture by Achim Streit we demonstrate basic approaches designing parallel algorithms. Preliminary list of topics:

- single program multiple data paradigm
- collective operations (reduction, broadcast, synchronization,...)
- load balancing techniques
- domain decomposition
- parallel sorting
- overview of available basic toolbox algorithms

Holger Fröning
Any Growth is Bounded - On the Future of Performance Scaling
The end of Moore's law has been predicted for many years, in combination with statements about exaggerations of its end. However, we are currently observing a phase of substantial slow-down in key metrics associated with silicon device manufacturing, and, as a result, huge

issues with sustained performance scaling. In this talk, we will review the most important recent and current trends in computer architecture, in particular in the light of diminishing returns from feature size scaling. Ultimately, we will see that energy efficiency is and will be key for a continuous performance scaling, and explore a couple of promising ideas in this context. The talk will conclude with a couple of anticipated directions, both for researchers but also practitioners.

Imaging

Klaus Maier-Hein
Radiologic Data Science
Applications and high-level overview of computational models and methods for problems appearing in medical image computing and radiologic data science.

Lennart Hilbert
Fluorescence microscopy and digital image processing in molecular cell biology
Fluorescence microscopy is a core technique of modern biological research, and comprises advanced optical hardware as well as computational control and processing. We will first discuss the modern overall workflow, going from a cell biological sample to digital data ("the microscope"), and from digital data to conclusions ("image processing"). We will discuss example challenges and solutions along the steps of this work-flow, including particular sample labeling, microscopy, data storage, and analysis approaches. To give hands-on experience, we will also explore a 3D microscopy data set using the Open Source software FIJI (https://imagej.net/Fiji/Downloads)

Surgery and Intervention 4.0

Franziska Mathis-Ullrich
Minimally-Invasive Robots in Medicine,
Minimally-invasive robots have the potential of revolutionising many aspects of medical interventions. Robot-assisted surgery has been associated with reduced risk of tissue damage, smaller incisions in the body, higher precision of surgery and direct feedback to the surgeon. This lecture introduces several minimally-invasive commercialised robotic systems and research platforms for medical interventions. We will discuss several robotic approaches to optimize several medical applications utilizing soft robots, continuum robots and micro- and nanorobot for specific.

Lena Maier-Hein:
Surgical data science
Surgical Data Science (SDS) is an emerging scientific field that focuses on the acquisition, modeling and analysis of data in order to improve the quality of interventional healthcare. It encompasses all clinical disciplines which involve interventions to manipulate anatomical structures with a diagnostic, therapeutic or prognostic goal. In this paradigm, data may relate to any step of the patient care, may concern the patient, caregivers, as well as

technology used to deliver care, and may be analyzed in the context of generic domain-specific knowledge. The unique scientific challenges related to the analysis of data from interventions include those related to speed, robustness as well as the heterogeneity and complexity of the procedures. This lecture will introduce basic concepts related to SDS and present state-of-the-art approaches to address core remaining research challenges. Particular focus will be put on methods for dealing with limited annotated data, uncertainty quantification and compensation, as well as meaningful performance assessment.

Tamim Asfour:
Data-Driven Learning of Sensorimotor Skills for Robots and Exoskeletons

Personalized Medicine

Michael Gertz
Introduction to Text Analysis
Topics will include basic text processing methods such as sentence splitting, part-of-speech tagging, and named entity recognition. On texts processed that way, we outline tasks such as information retrieval, topic detection and modeling, and information networks. The focus will be on medical and biomedical applications settings and texts, respectively. Some examples will be given in a Python programming framework.

Filip Sadlo
Introduction to Visual Data Science
This lecture introduces to the field of visual data science, with a focus on volume rendering and feature extraction from scalar, vector, and tensor fields. We elaborate on the interdependence between the analysis techniques and respective problems in medicine, and establish basic expertise regarding their effectiveness and interpretability.