

Vertiefungsthemen Data Science

Wintersemester 2024 / 2025

Dozenten: Prof. B. Mitschang, Prof. H. Schwarz, Dr. C. Stach
Ansprechpartner: Christoph Stach, christoph.stach@ipvs.uni-stuttgart.de
Sprache: Deutsch/German

Beschreibung

Die effiziente und effektive Verarbeitung und Analyse von großen Datenmengen stellt heutzutage einen zentralen Tätigkeitsbereich von Datenwissenschaftler*innen dar. Treiber wie web-basierter und mobiler Anwendungen (z.B. E-Mails, Blogs, soziale Medien oder E-Commerce) sowie die stark wachsende Menge an Sensoren im Kontext von Industrie 4.0 und dem Internet der Dinge sorgen dafür, dass fortlaufend neue Daten generiert werden. In diesem Kontext spricht man häufig von „Big Data“. Während die Sammlung und Speicherung dieser Daten mittlerweile durch Fortschritte in der Speichertechnologie weitgehend problemlos möglich ist, gestaltet sich gerade die Verarbeitung dieser Datenflut, d.h. die Gewinnung von Wissen aus diesen (Roh-)Daten, als eine sehr komplexe Aufgabe.



Datenwissenschaftler*innen operieren hierbei an der Schnittstelle zwischen spezifischer Domänenexpertise, technischer und mathematischer Expertise. Sie müssen sich daher sowohl bezüglich der Implementierungsdetails für ihre Datenanalysen Gedanken machen (z.B. wie können auch große Datenmengen effizient verarbeitet werden) und Kenntnisse über existierende Dateninfrastrukturen besitzen (z.B. wie lassen sich Datenströme verarbeiten oder wie lassen sich Daten sicher speichern), als auch Wissen über die Daten selbst aneignen (z.B. welche rechtlichen Rahmenbedingungen für die Verarbeitung dieser Daten gelten). Um dieser anspruchsvollen Aufgabe gewachsen zu sein, ist es daher unabdingbar, dass sich Datenwissenschaftler*innen intensiv mit verfügbaren Konzepten und Werkzeugen zur Verarbeitung von Big Data auseinandersetzen. In diesem Seminar sollen daher Techniken zur Analyse von Daten diskutiert werden. Die behandelten Themen sind unter anderem:

- Partitionierungsbasierte Clustering-Algorithmen (z.B. PAM)
- Hierarchiebasierte Clustering-Algorithmen (z.B. BIRCH)
- Fuzzy-Logic-basierte Clustering-Algorithmen (z.B. FCM)
- Dichtebasierte Clustering-Algorithmen (z.B. OPTICS)

Anmerkungen

- Die Seminarthemen werden in einer ersten Besprechung voraussichtlich in KW29 oder KW30 vorgestellt. Der Termin dieser obligatorischen Besprechung wird per E-Mail bekannt gegeben.
- Im Rahmen des Seminars bearbeitet jede(r) Studierende eines der angebotenen Themen. Einstiegsliteratur hierzu wird von den Betreuer*innen bereitgestellt. Die Ergebnisse der Arbeit werden in einer ca. 20-seitigen Ausarbeitung zusammengefasst und im Rahmen eines ca. 30-minütigen Vortrags präsentiert.

Voraussetzungen

Kenntnisse aus „Einführung Data Science“ und „Modellierung“ werden vorausgesetzt.